Alibaba Cloud
Apsara Stack Enterprise

User Guide

Version: 1901

Issue: 20190528

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.
5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified,

reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.

6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

Style	Description	Example
	This warning information indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 **Danger:** Resetting will result in the loss of user configuration data.
	This warning information indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 **Warning:** Restarting will cause business interruption. About 10 minutes are required to restore business.
	This indicates warning information, supplementary instructions, and other content that the user must understand.	 **Note:** Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructio ns, best practices, tips, and other contents.	 **Note:** You can use **Ctrl** + **A** to select all files.
>	Multi-level menu cascade.	**Settings** > **Network** > **Set network type**
Bold	It is used for buttons, menus, page names, and other UI elements.	Click **OK**.
`Courier font`	It is used for commands.	Run the `cd /d C:/windows` command to enter the Windows system folder.
`Italics`	It is used for parameters and variables.	`bae log list --instanceid` *`Instance_ID`*
[] or [a|b]	It indicates that it is a optional value, and only one item can be selected.	`ipconfig` *`[-all|-t]`*
{} or {a|b}	It indicates that it is a required value, and only one item can be selected.	`switch` *`{stand | slave}`*

Contents

Legal disclaimer I
Generic conventions I
1 What is the Apsara Stack console? 1
2 Log on to the Apsara Stack console 3
3 Web page introduction 4
4 Configuration of system initialization 6
4.1 Configuration instruction 6
4.2 Configuration process 8
5 Resource management 10
5.1 Quota management 10
5.1.1 Quota parameters 10
5.1.2 Create cloud resource quotas 12
5.1.3 View total and used quotas 12
5.1.4 Change quotas 13
5.1.5 Delete quotas 13
5.2 View and export project resource overview 14
5.3 Configuration of resource notifications 15
5.3.1 Configure resource notification objects 15
5.3.2 View resource notification objects 16
5.3.3 Delete a resource notification object 16
6 Alert management 17
6.1 Overview 17
6.2 Alert contacts 17
6.2.1 Create an alert contact 17
6.2.2 Add an alert contact to alert groups 18
6.2.3 Query an alert contact 19
6.2.4 Change alert contact information 19
6.2.5 Delete alert contacts 19
6.3 Alert groups 20
6.3.1 Create an alert group 20
6.3.2 Change alert notification methods 21
6.4 Alert rules 22
6.4.1 Create an alert rule 22
6.4.2 Create multiple alert rules 24
6.4.3 View alert rules 25
6.4.4 View alert history 26
6.4.5 Change an alert rule 26
6.4.6 Pause alert rules 27
6.4.7 Start alert rules 27

6.4.8 View alert notification objects..... 28
6.4.9 Delete alert rules.....28
6.5 Configuration of alert notification..... 29
6.5.1 Configure email alert notification..... 29
6.5.2 Configure DingTalk alert notification.....29
6.5.3 Configure SMS alert notification..... 30
6.6 View alert information..... 31
7 Monitoring management.....33
7.1 Overview.....33
7.2 View dashboard.....33
7.3 CloudMonitor..... 35
7.3.1 Overview..... 35
7.3.2 View CloudMonitor overview.....35
7.3.3 Cloud monitoring metrics..... 36
7.3.4 View monitoring charts..... 42
7.4 System reports..... 42
7.4.1 Create a report download task..... 42
7.4.2 Change the report name.....43
7.4.3 Preview and download a report.....44
7.4.4 Delete a report download task.....44
7.5 Task center.....45
7.5.1 View running tasks.....45
7.5.2 View previous tasks..... 45
7.6 Operation logs..... 45
7.6.1 View logs.....45
7.6.2 Delete logs..... 48
8 RAM management..... 49
8.1 Overview.....49
8.2 RAM roles.....50
8.2.1 View a role policy..... 50
8.2.2 Create a RAM role.....51
8.2.3 View role details.....52
8.3 RAM users.....53
8.3.1 Create a RAM user.....53
8.3.2 View RAM user details..... 54
8.3.3 Change the description of a RAM user..... 54
8.3.4 Grant permissions to a RAM user..... 55
8.4 RAM authorization policies.....56
8.4.1 Create a RAM authorization policy.....56
8.4.2 View RAM authorization policy details.....59
8.4.3 Delete a RAM authorization policy..... 59
9 System maintenance.....61
9.1 Department management.....61

9.1.1 Create a department....61
9.1.2 Change the department name....61
9.1.3 View projects of a department....62
9.1.4 Obtain the AccessKey of a department....62
9.1.5 Delete a department....62
9.2 Project management....63
9.2.1 Create a project....63
9.2.2 Add a project member....63
9.2.3 Change the project name....64
9.2.4 View project details....65
9.2.5 View project members....65
9.2.6 View resource information of a project....65
9.2.7 Release resources....66
9.2.8 Delete a project....66
9.3 Role management....67
9.3.1 Default roles....67
9.3.2 Add a custom role....67
9.3.3 View role details....69
9.3.4 Change a custom role....69
9.3.5 Delete a custom role....70
9.4 User management....70
9.4.1 Create a user....70
9.4.2 View basic information of a user....72
9.4.3 Change user information....72
9.4.4 Change the logon policy of a user....72
9.4.5 Change user roles....73
9.4.6 Authorize third-party access....73
9.4.7 Reset logon password....74
9.4.8 Export initial password....74
9.4.9 Enable and disable a user....75
9.4.10 Delete a user....75
9.4.11 Restore a user....76
9.5 Logon policy management....76
9.5.1 Create a logon policy....76
9.5.2 View a logon policy....78
9.5.3 Bind a logon policy to multiple users....78
9.6 Configure storage path for attachments....79
9.7 Configure ECS startup....80
9.8 Configuration of system style....80
9.8.1 Configure the theme....80
10 Personal information management....81
10.1 Change personal information....81
10.2 View AccessKey of your personal account....81
10.3 View third-party AccessKey....82

10.4 Change your avatar..... 82
10.5 Change your logon password..... 82
11 Elastic Compute Service (ECS)..... 84
11.1 What is ECS..... 84
11.1.1 Overview..... 84
11.1.2 Instance types..... 85
11.1.3 Instance lifecycle..... 98
11.2 Instructions before use..... 99
11.2.1 Overview..... 99
11.2.2 Prohibitions..... 99
11.2.3 Suggestions..... 100
11.2.4 Restrictions..... 100
11.2.5 Precautions for using ECS instances in Windows..... 102
11.2.6 Precautions for using ECS instances in Linux..... 102
11.2.7 DDoS protection..... 103
11.3 Quick start..... 103
11.3.1 Overview..... 103
11.3.2 Log on to the ECS console..... 103
11.3.3 Create a security group..... 104
11.3.4 Create an instance..... 105
11.3.5 Connect to an instance..... 108
11.3.5.1 Overview..... 108
11.3.5.2 Connect to a Linux instance using the SSH command in Linux or Mac OS X..... 109
11.3.5.3 Connect to a Linux instance using a remote connection tool in Windows... 109
11.3.5.4 Connect to a Windows instance using the remote desktop connection function in Windows..... 110
11.3.5.5 Connect to an ECS instance by using the Management Terminal..... 113
11.4 Instances..... 115
11.4.1 Overview..... 115
11.4.2 View an instance..... 115
11.4.3 Edit an instance..... 116
11.4.4 Start, stop, or reboot an instance..... 117
11.4.5 Delete an instance..... 117
11.4.6 Modify configurations..... 118
11.4.7 Change ownership..... 118
11.4.8 Change the ECS instance logon password..... 119
11.4.9 Change the VNC password..... 120
11.4.10 Join a security group..... 121
11.4.11 Customize instance data..... 122
11.4.12 Change private IP..... 126
11.4.13 Install a certificate..... 126
11.4.14 Install the CUDA and GPU drivers for a Linux instance..... 139
11.4.15 Install the CUDA and GPU drivers for a Windows instance..... 143

11.5 Disks....146
11.5.1 Overview....146
11.5.2 Create disks....147
11.5.3 View disks....149
11.5.4 Roll back a disk....150
11.5.5 Edit disk attributes....150
11.5.6 Attach a disk....151
11.5.6.1 Overview....151
11.5.6.2 Attach a disk on the Instance Details page....151
11.5.6.3 Attach a disk on the Disk List page....152
11.5.7 Partition and format disks....153
11.5.7.1 Overview....153
11.5.7.2 Partition, format, and attach data disks in Linux....153
11.5.7.3 Partition and format data disks in Windows....157
11.5.8 Resize a system disk....162
11.5.8.1 Overview....162
11.5.8.2 Create a snapshot for a system disk....163
11.5.8.3 Create an image from a snapshot....164
11.5.8.4 Change a system disk....165
11.5.8.5 Set a snapshot policy for a system disk....167
11.5.9 Detaching a disk....167
11.6 Images....168
11.6.1 Overview....168
11.6.2 Select a suitable image....169
11.6.3 Create a custom image....169
11.6.3.1 Overview....169
11.6.3.2 Create custom images from snapshots....169
11.6.3.3 Create a custom image from an instance....170
11.6.4 View images....170
11.6.5 Copy images....171
11.6.6 Share images....171
11.6.7 Import images....172
11.6.7.1 Overview....172
11.6.7.2 Notes for importing images....172
11.6.7.3 Convert image file format....176
11.6.8 Export images....181
11.6.9 Delete images....182
11.7 Snapshots....182
11.7.1 Overview....182
11.7.2 Create a snapshot....183
11.7.3 View snapshots....184
11.7.4 Delete snapshots....184
11.7.5 Application scenarios....185
11.8 Automatic snapshot policies....186

11.8.1 Overview..... 186
11.8.2 Create an automatic snapshot policy..... 186
11.8.3 View automatic snapshot policies.....187
11.8.4 Edit an automatic snapshot policy.....187
11.8.5 Configure an automatic snapshot policy.....188
11.8.6 Configure an automatic snapshot policy for multiple disks.....188
11.8.7 Delete an automatic snapshot policy.....189
11.9 Security groups..... 189
11.9.1 Overview..... 189
11.9.2 View security groups.....190
11.9.3 Add security group rules.....190
11.9.4 Remove an instance from a security group.....192
11.9.5 Delete a security group.....192
11.10 Elastic network interfaces..... 193
11.10.1 Overview..... 193
11.10.2 Create an ENI.....193
11.10.3 View ENIs..... 194
11.10.4 Edit an ENI.....195
11.10.5 Attach an ENI to an instance.....195
11.10.6 Detach an ENI from an instance..... 196
11.10.7 Delete an ENI..... 196
11.11 Deployment sets.....197
11.11.1 Overview..... 197
11.11.2 Create a deployment set..... 197
11.11.3 View a deployment set..... 198
11.11.4 Edit a deployment set.....198
11.11.5 Delete a deployment set.....198
11.12 Install FTP software..... 199
11.12.1 Overview..... 199
11.12.2 Install VSFTP in CentOS.....199
11.12.3 Install VSFTP in Ubuntu and Debian.....200
11.12.4 Configure FTP through IIS in Windows 2003.....201
11.12.5 Install and configure FTP in Windows 2008.....202
11.12.6 Install and configure IIS and FTP in Windows 2012.....205
12 Auto Scaling (ESS).....211
12.1 What is ESS.....211
12.2 Usage..... 212
12.2.1 Overview..... 212
12.2.2 Precautions..... 212
12.2.3 Manual intervention.....213
12.2.4 Quantity limits.....214
12.2.5 Scaling group statuses..... 215
12.2.6 Scaling activity process..... 215
12.2.7 Removal of unhealthy ECS instances..... 217

12.2.8 Instance rollback after a scaling activity failure..... 217
12.2.9 Instance life cycle management..... 218
12.3 Quick start..... 219
12.3.1 Overview..... 219
12.3.2 Log on to the ESS console..... 219
12.3.3 Create a scaling group..... 220
12.3.4 Create a scaling configuration..... 223
12.3.5 Enable a scaling group..... 224
12.3.6 Create a scaling rule..... 225
12.3.7 Create a scheduled task..... 226
12.4 Scaling group..... 227
12.4.1 Overview..... 227
12.4.2 Query a scaling group..... 227
12.4.3 Modify scaling group information..... 228
12.4.4 Disable a scaling group..... 228
12.4.5 Delete a scaling group..... 229
12.4.6 Query ECS instances..... 229
12.5 Scaling configuration..... 230
12.5.1 Overview..... 230
12.5.2 Query a scaling configuration..... 230
12.6 Scaling rule..... 231
12.6.1 Overview..... 231
12.6.2 Query a scaling rule..... 231
12.6.3 Edit a scaling rule..... 231
12.6.4 Delete a scaling rule..... 232
12.7 Trigger tasks..... 232
12.7.1 Overview..... 232
12.7.2 Manually execute a scaling rule..... 232
12.7.3 Add an ECS instance..... 233
12.7.4 Remove an ECS instance..... 234
12.8 Scheduled tasks..... 235
12.8.1 Overview..... 235
12.8.2 Query a scheduled task..... 235
12.8.3 Edit a scheduled task..... 235
12.8.4 Disable a scheduled task..... 236
12.8.5 Enable a scheduled task..... 237
12.8.6 Delete a scheduled task..... 237
13 Object Storage Service (OSS)..... 238
13.1 What is OSS..... 238
13.2 Instructions..... 238
13.3 Quick start..... 239
13.3.1 Log on to the OSS console..... 239
13.3.2 Create a bucket..... 240
13.3.3 Upload objects..... 242

13.3.4 Obtain an object URL....243
13.4 Bucket....243
13.4.1 View a bucket....243
13.4.2 Delete a bucket....244
13.4.3 Change the capacity....244
13.4.4 Change the ownership....244
13.4.5 Change ACL settings....245
13.4.6 Configure static website hosting....246
13.4.7 Enable the logging function....247
13.4.8 Configure hotlinking protection....248
13.4.9 Configure CORS....249
13.4.10 Manage lifecycle rules....250
13.4.11 Configure cross-cloud data synchronization....251
13.5 Object....253
13.5.1 Search for objects....253
13.5.2 Delete objects....254
13.5.3 Configure ACL of an object....254
13.5.4 Create a folder....255
13.6 Image service....256
13.6.1 Create a style....256
13.6.2 Enable source image protection....257
13.7 Create single tunnels....258
14 Table Store....260
14.1 What is Table Store....260
14.2 Instructions....260
14.3 Quick start....262
14.3.1 Log on to the Table Store console....262
14.3.2 Create an instance....262
14.3.3 Create a table....263
14.4 Manage instances....265
14.4.1 View an instance....265
14.4.2 Release an instance....265
14.5 Manage tables....266
14.5.1 View table details....266
14.5.2 Update a table....266
14.5.3 Delete a table....267
14.6 Bind VPC instances....267
15 Network Attached Storage (NAS)....269
15.1 What is NAS?....269
15.2 Instructions....269
15.3 Quick start....270
15.3.1 Log on to the NAS console....270
15.3.2 Create NAS instances....271

15.3.3 Create permission groups....272
15.3.4 Create permission group rules....273
15.3.5 Add mount points....274
15.3.6 Mount NAS instances....275
15.4 NAS instance....277
15.4.1 View the NAS instance details....277
15.4.2 Delete NAS instances....278
15.5 Mount point....279
15.5.1 View the mount point list....279
15.5.2 Enable or disable mount points....279
15.5.3 Delete mount points....280
15.5.4 Modify the permission group of a mount point....281
15.6 Permission group....281
15.6.1 View the permission group list....281
15.6.2 Delete permission groups....282
15.6.3 Manage permission group rules....283
15.7 Migrate data....283
15.7.1 Data migration tool for Windows....283
15.7.2 Migrate local files or files stored in OSS to NAS instances....291
15.8 Directory-level ACL....298
16 Distributed File System (DFS)....300
16.1 What is DFS....300
16.2 Limits....300
16.3 Quick start....301
16.3.1 Log on to the DFS console....301
16.3.2 Create file systems....302
16.3.3 Create permission groups....303
16.3.4 Create permission group rules....304
16.3.5 Add mount points....306
16.3.6 Mount file systems....308
16.4 File systems....310
16.4.1 View file system details....310
16.4.2 Delete file systems....311
16.4.3 Change file system information....312
16.5 Mount points....312
16.5.1 View the mount point list....312
16.5.2 Manage mount points....314
16.6 Permission groups....314
16.6.1 View the permission group list....314
16.6.2 Change permission group information....315
16.6.3 Delete permission groups....316
16.6.4 Manage permission group rules....316
17 ApsaraDB for RDS.... 318

17.1 What is ApsaraDB for RDS? 318
17.2 Limits 320
17.2.1 Usage limits of ApsaraDB RDS for MySQL 320
17.2.2 Usage limits of ApsaraDB RDS for PostgreSQL 321
17.2.3 Usage limits of ApsaraDB RDS for PPAS 322
17.3 Quick start 323
17.3.1 Quick start 323
17.3.2 Log on to the RDS console 324
17.3.3 Create an instance 325
17.3.4 Initiate the configuration 327
17.3.4.1 RDS for MySQL 327
17.3.4.1.1 Configure a whitelist 327
17.3.4.1.2 Create a premier account 330
17.3.4.1.3 Create a standard account 333
17.3.4.1.4 Create a database 335
17.3.4.2 RDS for PostgreSQL 337
17.3.4.2.1 Configure a whitelist 337
17.3.4.2.2 Create a database and an account 340
17.3.4.3 RDS for PPAS 342
17.3.4.3.1 Configure a whitelist 342
17.3.4.3.2 Create a database and an account 345
17.3.5 Connect to an instance 347
17.3.5.1 Log on to an instance through DMS 347
17.3.5.2 Connect to a MySQL instance from a client 348
17.3.5.3 Connect to a PostgreSQL instance from a client 351
17.3.5.4 Connect to a PPAS instance from a client 354
17.4 Instances 357
17.4.1 Create an instance 357
17.4.2 View details 359
17.4.3 Restart an instance 359
17.4.4 Modify configurations 360
17.4.5 Release an instance 360
17.4.6 Configure parameters 361
17.4.7 Change ownership 362
17.4.8 Modify an instance name 362
17.4.9 Typical parameter configuration 363
17.4.9.1 Modifiable MySQL instance parameters 363
17.4.9.2 Best practices for MySQL instance parameter optimization 390
17.4.9.2.1 Overview 390
17.4.9.2.2 Unmodifiable MySQL instance parameters 390
17.4.9.2.3 Modifiable MySQL instance parameters 390
17.4.9.2.4 How to configure parameters 391
17.4.9.2.5 New MySQL parameters 394
17.5 Accounts 395

17.5.1 Create an account..... 395
17.5.2 Reset your password..... 398
17.5.3 Modify account permissions.....399
17.5.4 Delete an account.....399
17.5.5 Modify descriptions..... 400
17.6 Databases.....400
17.6.1 Create a database..... 400
17.6.2 Modify database description..... 402
17.6.3 Delete a database.....403
17.7 Access mode.....403
17.8 Backup and recovery..... 404
17.8.1 RDS data backup.....404
17.8.1.1 Automatic backup.....404
17.8.1.2 Manual backup.....405
17.8.2 RDS data recovery..... 406
17.8.2.1 Clone an instance.....406
17.8.3 Binary log (binlog).....408
17.9 Security.....408
17.9.1 Configure a whitelist..... 408
17.9.2 Audit logs..... 411
17.9.3 Configure SSL encryption.....411
17.9.4 Download SSL CA certificates.....413
17.9.5 Configure transparent data encryption..... 414
17.10 Read-only instances.....415
17.10.1 Overview..... 415
17.10.2 Create a read-only instance..... 416
17.10.3 View read-only instance details.....418
17.10.3.1 View instance details through a read-only instance.....418
17.10.3.2 View instance details through the primary instance.....418
17.11 Read/write splitting.....419
17.11.1 Overview..... 419
17.11.2 Enable read/write splitting.....421
17.11.3 Modify the latency threshold and weights of read requests.....424
17.11.4 Disable read/write splitting.....425
17.11.5 Monitor read/write splitting performance.....426
17.11.6 Rules of system weight distribution..... 426
17.12 Performance optimization.....428
17.12.1 Slow SQL statistics.....428
17.12.2 Missing index..... 429
17.13 Monitor system resources.....429
17.14 Data migration from the on-premises database to RDS.....431
17.14.1 Compress data.....431
17.14.2 Migrate MySQL data.....432
17.14.2.1 Use mysqldump to migrate MySQL data.....432

17.15 Typical applications....435
17.15.1 Store multi-structure data....435
18 KVStore for Redis....437
18.1 What is KVStore for Redis....437
18.2 Quick start....437
18.2.1 Get started with KVStore for Redis....437
18.2.2 Log on to the KVStore for Redis console....439
18.2.3 Create an instance....440
18.2.4 Configure a whitelist....442
18.2.5 Connect to an instance....443
18.2.5.1 Connect to KVStore for Redis instances from a Redis client....443
18.2.5.1.1 Overview....443
18.2.5.1.2 Jedis client....443
18.2.5.1.3 phpredis client....446
18.2.5.1.4 redis-py client....446
18.2.5.1.5 C or C++ client....447
18.2.5.1.6 .net client....449
18.2.5.1.7 node-redis client....451
18.2.5.2 Connect to KVStore for Redis through redis-cli....451
18.3 Manage instances....452
18.3.1 Create an instance....452
18.3.2 View instance details....454
18.3.3 Change the instance name....455
18.3.4 Change instance configurations....455
18.3.5 Configure a whitelist....456
18.3.6 Set the O&M time....456
18.3.7 Enable data transmission encryption....457
18.3.8 Clear instance data....457
18.3.9 Reset a password....458
18.3.10 Release an instance....458
18.3.11 Set parameters....459
18.4 Backup and recovery....459
18.4.1 Configure automatic backup policies....460
18.4.2 Manual backup....460
18.4.3 Archive backup data....461
18.4.4 Restore data....462
18.5 Import data....463
18.6 Commands supported by KVStore for Redis....464
19 ApsaraDB for MongoDB....468
19.1 What is ApsaraDB for MongoDB....468
19.2 Instructions....468
19.3 Quick start....469
19.3.1 Procedure....469

19.3.2 Log on to the ApsaraDB for MongoDB console..... 469
19.3.3 Create an instance.....470
19.3.4 Set a whitelist.....472
19.3.5 Obtain the seven elements required to connect to an instance.....473
19.3.6 Use Mongo shell to connect to an instance..... 474
19.4 Manage instances..... 475
19.4.1 Create an instance.....475
19.4.2 View instance details..... 477
19.4.3 Restart an instance.....477
19.4.4 Change specifications.....478
19.4.5 Switch to VPC.....478
19.4.6 Change an instance name.....479
19.4.7 Reset a password.....480
19.4.8 Release an instance..... 480
19.5 Security.....481
19.5.1 Set a whitelist.....481
19.5.2 Audit logs..... 481
19.6 Monitoring information.....482
19.7 Backup and recovery..... 485
19.7.1 Automatic backup.....485
19.7.2 Back up an instance manually.....486
19.7.3 Search for backups.....486
19.7.4 Restore data..... 487
19.7.5 Backup download.....487
19.7.6 Create an instance from a backup point..... 488
20 KVStore for Memcache.....489
20.1 What is KVStore for Memcache..... 489
20.2 Limits..... 489
20.3 Quick start..... 490
20.3.1 Start to use KVStore for Memcache.....490
20.3.2 Log on to the KVStore for Memcache console.....491
20.3.3 Create an instance.....492
20.3.4 Set a whitelist.....494
20.3.5 Connect to an instance from a client.....495
20.3.5.1 Overview..... 495
20.3.5.2 Java: Spymemcache.....495
20.3.5.3 PHP: memcached..... 499
20.3.5.4 Python..... 507
20.3.5.5 C#/. NET: EnyimMemcached.....507
20.3.5.6 C++.....508
20.4 Instances..... 514
20.4.1 Create an instance.....514
20.4.2 View instance details..... 517
20.4.3 Change an instance name.....518

20.4.4 Change the instance specification....518
20.4.5 Set a whitelist....518
20.4.6 Configure a maintenance time period....519
20.4.7 Clear the instance data....519
20.4.8 Reset a password....520
20.4.9 Set a data eviction policy....520
20.5 Backup and recovery....521
20.5.1 Automatic backup....521
20.5.2 Manual backup....521
20.5.3 Data restore....522
20.6 Supported protocols and commands....522
21 Data Management Service (DMS)....525
21.1 What is DMS?....525
21.2 Log on to an instance through DMS....527
21.3 SQL operations....529
21.3.1 Use the Command Window....529
21.3.2 Use the SQL window....532
21.3.2.1 Open an empty SQL window....532
21.3.2.2 Restore a saved SQL window....540
21.3.2.3 Manage commonly used SQL commands....541
21.3.2.4 Use the SQL template....542
21.3.3 Table operations (based on the Table directory tree)....542
21.3.3.1 Open a table-based SQL window....542
21.3.3.2 Edit table data....542
21.4 Database development....543
21.4.1 Overview....543
21.4.2 Table....543
21.4.2.1 Add a table....543
21.4.2.2 Edit a table....544
21.4.2.3 Delete a table....545
21.4.2.4 Create a similar table....545
21.4.2.5 Generate SQL statement templates....545
21.4.2.6 Query table information....546
21.4.2.7 Clear data....546
21.4.2.8 Perform table operations in batches....546
21.4.2.9 Maintain a table....547
21.4.3 Manage indexes....547
21.4.4 Manage foreign keys....548
21.4.5 Create partitions....549
21.4.6 Create a stored procedure....549
21.4.7 Create a function....550
21.4.8 Create a view....551
21.4.9 Create a trigger....552
21.4.10 Create event....553

21.5 Data processing..... 554
21.5.1 Import Data..... 554
21.5.2 Export data..... 555
21.5.2.1 Export a database..... 555
21.5.2.2 Export an SQL result set..... 556
21.6 Performance..... 557
21.6.1 Overview..... 557
21.6.2 Lock wait..... 557
21.6.2.1 View lock wait..... 557
21.6.2.2 Release lock wait..... 558
21.6.3 Sessions..... 558
21.6.3.1 View sessions..... 558
21.6.3.2 End a session..... 559
21.6.3.3 Optimize a session..... 559
21.6.4 View real-time performance..... 560
21.7 Extended tools..... 560
21.7.1 Table data volume statistics..... 560
21.7.2 E-R diagram..... 561
21.8 DMS for Redis..... 561
21.8.1 Function overview..... 561
21.8.2 Data management..... 563
21.8.2.1 Create a key..... 563
21.8.2.2 Edit a key..... 564
21.8.2.3 Set key timeout..... 566
21.8.2.4 Delete a key..... 567
21.8.2.5 Rename a key..... 567
21.8.3 Performance monitoring..... 568
21.8.3.1 View the homepage..... 568
21.8.3.2 View real-time performance..... 568
21.9 DMS for MongoDB..... 569
21.9.1 Function overview..... 569
21.9.2 Structure management..... 570
21.9.2.1 Create a collection..... 570
21.9.2.2 Create a database..... 571
21.9.2.3 Create an index..... 571
21.9.2.4 Edit an index..... 572
21.9.2.5 Delete a collection..... 572
21.9.2.6 Delete a database..... 572
21.9.2.7 Delete an index..... 572
21.9.3 User management..... 573
21.9.3.1 Create a user..... 573
21.9.3.2 Edit a user..... 573
21.9.3.3 Delete a user..... 573
21.9.4 Data management..... 574

21.9.4.1 Create a document....574
21.9.4.2 Edit a document....574
21.9.4.3 Query a document....576
21.9.4.4 Delete a document....577
21.9.5 View the homepage....577
22 Server Load Balancer (SLB)....579
22.1 What is Server Load Balancer?....579
22.2 Planning and preparation....580
22.3 Quick start....581
22.3.1 Overview....581
22.3.2 Log on to the Server Load Balancer console....582
22.3.3 Create an SLB instance....582
22.3.4 Add listeners....583
22.3.5 Add backend servers....584
22.4 SLB instances....584
22.4.1 SLB instance overview....584
22.4.2 Create an SLB instance....585
22.4.3 Start or stop an instance....586
22.4.4 View instance details....586
22.4.5 Modify attributes of an SLB instance....586
22.4.6 Modify the ownership of an SLB instance....587
22.4.7 Delete an SLB instance....587
22.5 Listeners....587
22.5.1 Overview....587
22.5.2 Configure a Layer-4 listener....588
22.5.3 Configure a Layer-7 listener....591
22.5.4 Configure forwarding rules....594
22.5.5 Configure access control....597
22.5.6 Stop a listener....598
22.5.7 Start a listener....598
22.5.8 Edit listener settings....599
22.5.9 Delete a listener....599
22.6 Backend servers....600
22.6.1 Backend server overview....600
22.6.2 Add backend servers....600
22.6.3 Modify the weight of an ECS instance....601
22.6.4 Remove a backend ECS instance....601
22.7 VServer groups....602
22.7.1 Add a VServer group....602
22.7.2 View a VServer group....603
22.7.3 Edit a VServer group....603
22.7.4 Delete a VServer group....604
22.8 Certificates....604
22.8.1 Certificate overview....604

22.8.2 Certificate format....605
22.8.3 Generate a CA certificate....606
22.8.4 Generate a client certificate....608
22.8.5 Upload a certificate....609
22.8.6 Convert the format of a certificate....611
22.8.7 Replace a certificate....612
23 Virtual Private Cloud (VPC)....613
23.1 What is VPC....613
23.2 Quick start....614
23.2.1 Tutorial overview....614
23.2.2 Log on to the VPC console....615
23.2.3 Create a VPC and a VSwitch....615
23.2.4 Create a security group....617
23.2.5 Create an ECS instance....618
23.3 VPC....619
23.3.1 Plan a CIDR block....619
23.3.2 Create a VPC....620
23.3.3 View a VPC....621
23.3.4 Modify VPC information....621
23.3.5 Delete a VPC....621
23.4 VSwitch....622
23.4.1 Create a VSwitch....622
23.4.2 View VSwitches....623
23.4.3 Edit VSwitch information....624
23.4.4 Delete a VSwitch....624
23.5 VRouter and route table....624
23.5.1 Overview....624
23.5.2 View VRouters and routing tables....625
23.5.3 Add routing entries....625
24 Log Service....627
24.1 What is Log Service?....627
24.2 Quick start....627
24.2.1 Procedure....627
24.2.2 Log on to the Log Service console....629
24.2.3 View the key pair....630
24.2.4 Create a Project....630
24.2.5 Create a Logstore....632
24.2.6 Configure an index....634
24.2.7 Set an alarm....637
24.2.8 Log consumption....640
24.3 Project....641
24.3.1 Project....642
24.3.2 Import projects....642

24.3.3 Change the project ownership....643
24.3.4 Modify a project comment....644
24.3.5 Delete a project....644
24.4 Logstore....645
24.4.1 Logstore....646
24.4.2 Modify Logstore configurations....646
24.4.3 Delete a Logstore....647
24.5 Shard....647
24.5.1 Shard management....647
24.5.2 Split a shard....650
24.5.3 Merge shards....651
24.6 Data collection....651
24.6.1 Data collection....651
24.6.2 Collect Nginx access logs....651
24.7 Collection by Logtail....663
24.7.1 Overview....663
24.7.1.1 Logtail overview....663
24.7.1.2 How Logtail collection works....668
24.7.2 Installation....671
24.7.2.1 Install Logtail (for Linux)....671
24.7.2.2 Configure startup parameters....672
24.7.3 Data sources....676
24.7.3.1 Text logs....676
24.7.3.1.1 Collect text logs....676
24.7.3.1.2 Configure a time format....681
24.7.3.1.3 Generate a topic....684
24.7.3.1.4 Import historical log files....686
24.7.3.2 Collect syslogs....688
24.7.3.3 Syslog collection reference....692
24.7.4 Machine group....697
24.7.4.1 Machine groups....698
24.7.4.2 Create an IP address-based machine group....699
24.7.4.3 Create a machine group with a user-defined identifier....700
24.7.4.4 View the machine group list....703
24.7.4.5 Modify a machine group....704
24.7.4.6 View the machine group status....705
24.7.4.7 Manage machine group configurations....705
24.7.4.8 Delete a machine group....706
24.7.5 Troubleshooting....707
24.7.5.1 View the local log collection status....707
24.7.5.2 Query error diagnostics....721
24.7.5.3 Troubleshoot log collection errors....726
24.7.6 Limits....728
24.8 Other collection methods....732

24.8.1 Use LogStash to collect logs....732
24.8.1.1 Logstash overview....732
24.8.1.2 Quick installation....733
24.8.1.3 Custom installation....733
24.8.1.4 Set LogStash to a Windows service....735
24.8.1.5 Create a LogStash collection configuration....737
24.8.1.6 Advanced functions....739
24.8.1.7 LogStash error handling....740
24.8.2 SDK collection....740
24.8.2.1 Producer Library....740
24.8.2.2 Log4j Appender....741
24.8.2.3 C Producer Library....742
24.8.3 Common log formats....742
24.8.3.1 Overview....742
24.8.3.2 Apache logs....742
24.8.3.3 Nginx logs....745
24.8.3.4 Python log....747
24.8.3.5 Log4j log....749
24.8.3.6 Node.js log....751
24.8.3.7 WordPress log....752
24.8.3.8 Delimiter log....753
24.8.3.9 JSON logs....756
24.8.3.10 ThinkPHP log....758
24.8.3.11 Use LogStash to collect IIS logs....759
24.8.3.12 Use LogStash to collect CSV logs....760
24.8.3.13 Use LogStash to collect other logs....762
24.9 Query and analysis....763
24.9.1 Indexing and query....763
24.9.2 Real-time analysis....765
24.9.3 Disable an index....768
24.9.4 Index types....768
24.9.4.1 Overview....768
24.9.4.2 Text type....771
24.9.4.3 Value type....773
24.9.4.4 JSON type....773
24.9.5 Query syntax and functions....774
24.9.5.1 Query syntax....774
24.9.5.2 Context query....780
24.9.5.3 Other functions....782
24.9.5.4 Quick analysis....786
24.9.5.5 Saved search....790
24.9.6 Analysis syntax and functions....792
24.9.6.1 General aggregate functions....792
24.9.6.2 Map functions....793

24.9.6.3 Estimating functions....795
24.9.6.4 Mathematical statistical functions....795
24.9.6.5 Mathematical functions....796
24.9.6.6 String functions....798
24.9.6.7 Date and time functions....799
24.9.6.8 URL functions....803
24.9.6.9 Regular expression functions....804
24.9.6.10 JSON functions....805
24.9.6.11 Type conversion functions....806
24.9.6.12 GROUP BY syntax....806
24.9.6.13 Window functions....808
24.9.6.14 HAVING syntax....810
24.9.6.15 ORDER BY syntax....811
24.9.6.16 LIMIT syntax....811
24.9.6.17 CASE WHEN syntax....812
24.9.6.18 Nested subquery....813
24.9.6.19 Arrays....813
24.9.6.20 Binary string functions....816
24.9.6.21 Bit operation....817
24.9.6.22 Comparison functions and operators....817
24.9.6.23 Lambda function....820
24.9.6.24 Logical function....823
24.9.6.25 Column alias....824
24.9.6.26 Geospatial functions....825
24.9.6.27 JOIN syntax....829
24.9.7 Advanced analytics....830
24.9.7.1 Excellent analysis cases....830
24.9.7.2 Optimize a query....831
24.9.8 Log analysis through JDBC....833
24.10 Alarms....837
24.10.1 Overview....837
24.10.2 Set an alarm....839
24.10.3 Notification methods....843
24.11 Log consumption....844
24.11.1 Preview logs....844
24.11.2 Consumption by consumer groups....844
24.11.2.1 Consumption by consumer groups....844
24.11.2.2 View consumer group status....847
24.11.3 Use Flink to consume logs....850
24.11.4 Storm consumption....857
24.11.5 Spark Streaming consumption....860
24.11.6 Consumption by StreamCompute....860
25 Domain Name System (DNS)....862
25.1 What is Apsara Stack DNS?....862

25.2 Log on to the DNS console....862
25.3 Global internal domains....863
25.3.1 Overview....863
25.3.2 View internal domain names....863
25.3.3 Add a domain name....863
25.3.4 Add remarks for a domain name....864
25.3.5 Delete a domain name....864
25.3.6 Delete multiple domain names....864
25.3.7 Manage DNS records....865
25.4 Global forwarding domains....865
25.4.1 Overview....865
25.4.2 View a forwarding domain....866
25.4.3 Add a domain name....866
25.4.4 Add remarks for a domain name....866
25.4.5 Change forwarding settings....867
25.4.6 Delete a domain name....867
25.4.7 Delete multiple domain names....867
25.5 Global default forwarding....868
25.5.1 Enable default forwarding....868
25.5.2 Change default forwarding settings....868
25.5.3 Disable default forwarding....868
26 API Gateway....870
26.1 Product overview....870
26.2 Quick start for consumers....870
26.2.1 Overview....870
26.2.2 Step 1: Obtain the API document....870
26.2.3 Step 2: Create an application....871
26.2.4 Step 3: Obtain authorization....872
26.2.5 Step 4: Call the API....873
26.3 Quick start for providers....874
26.3.1 Overview....874
26.3.2 Create a group....875
26.3.3 Create an API....875
26.3.4 Publish an API....879
26.3.5 Authorize applications to call APIs....880
26.4 Call an API....882
26.4.1 Manage applications....882
26.4.1.1 Create an application....882
26.4.1.2 View application details....883
26.4.1.3 Change an application....883
26.4.1.4 Delete an application....883
26.4.2 View existing APIs....883
26.4.3 Authorization....884
26.4.4 Encrypt the signature....884

26.4.5 Request signature instructions....884
26.4.6 API call example....888
26.5 APIs....889
26.5.1 Usage limits....889
26.5.2 Manage groups....890
26.5.2.1 Create a group....890
26.5.2.2 Environment management....890
26.5.2.3 Delete a group....892
26.5.3 Create an API....892
26.5.3.1 Overview....892
26.5.3.2 Create an API....892
26.5.3.3 Support HTTPS....897
26.5.3.4 Mock an API....898
26.5.4 API management....899
26.5.4.1 View and modify an API....899
26.5.4.2 Publish an API....899
26.5.4.3 Authorize applications to call APIs....900
26.5.4.4 Delete an API....902
26.5.4.5 Unpublish an API....902
26.5.4.6 View the version history of an API....902
26.5.4.7 Change the version of an API....903
26.5.5 Throttling policies....903
26.5.5.1 Create a throttling policy....903
26.5.5.2 Bind a throttling policy to APIs....904
26.5.5.3 Delete a throttling policy....904
27 Apsara Stack Security....905
27.1 What is Apsara Stack Security....905
27.2 Restrictions....906
27.3 Quick start....906
27.3.1 User permissions....906
27.3.2 Log on to Apsara Stack Security Center....907
27.3.3 Grant permissions in RAM....908
27.3.4 Switch regions....911
27.3.5 User interface of Apsara Stack Security Center....912
27.4 Threat Detection Service....913
27.4.1 Threat Detection Service Overview....913
27.4.2 Overview....914
27.4.2.1 View security overview information....914
27.4.2.2 View network traffic information....916
27.4.2.3 View visit analysis results....916
27.4.2.4 View information on visualization screens....917
27.4.3 Event analysis....920
27.4.3.1 View emergencies....920
27.4.4 Threat analysis....922

27.4.4.1 View threat analysis results....922
27.4.4.2 View threat attack information....924
27.4.5 Security reports....926
27.4.5.1 Add a report task....926
27.4.5.2 Manage a report task....929
27.4.6 Vulnerability scan....929
27.4.6.1 Manage application vulnerabilities....929
27.4.6.2 View server vulnerabilities....931
27.4.6.3 View weak password information....931
27.4.6.4 Add custom weak passwords....932
27.4.6.5 View configuration risks....934
27.5 Network security....935
27.5.1 Enable network security blocking....935
27.6 Application security....935
27.6.1 WAF overview....935
27.6.2 Connect domain names to WAF....936
27.6.2.1 Before you start....936
27.6.2.2 Add a protected domain name....937
27.6.2.3 Upload HTTPS certificates and private keys for domain names of HTTPS websites....940
27.6.2.4 Allow the access from a server IP address in the WAF cluster....942
27.6.2.5 Verify the WAF connection configuration for a domain name locally....943
27.6.2.6 Modify DNS resolution settings....944
27.6.3 Configure protection functions....945
27.6.3.1 Configure Web application protection....945
27.6.3.2 Configure malicious IP blocking....946
27.6.3.3 Configure HTTP flood protection....947
27.6.3.4 Configure precise access control....950
27.6.3.5 Configure blocked areas....953
27.6.4 View security reports....955
27.6.4.1 View security overview....955
27.6.4.2 View security reports....956
27.6.4.3 View business analysis results....958
27.7 Server security....959
27.7.1 Server security overview....959
27.7.2 Server list....961
27.7.2.1 Manage the server list....961
27.7.2.2 Manage groups....963
27.7.3 Threat protection....964
27.7.3.1 Vulnerability management....964
27.7.3.1.1 Manage Linux software vulnerabilities....964
27.7.3.1.2 Manage Windows vulnerabilities....965
27.7.3.1.3 Manage WCMS vulnerabilities....966
27.7.3.1.4 Manage other vulnerabilities....967

27.7.3.1.5 Configure vulnerability management....968
27.7.3.2 Baseline check....971
27.7.3.2.1 Baseline check overview....971
27.7.3.2.2 Configure baseline check....972
27.7.3.2.3 Set a baseline check policy....973
27.7.4 Intrusion detection....973
27.7.4.1 Unusual logons....973
27.7.4.1.1 How unusual logon detection works....973
27.7.4.1.2 Check unusual logon alerts....974
27.7.4.1.3 Configure logon security....974
27.7.4.2 Webshells....975
27.7.4.2.1 Manage webshell files....975
27.7.4.3 Suspicious servers....976
27.7.4.3.1 Manage server exceptions....976
27.7.5 Server fingerprints....977
27.7.5.1 Manage listening ports....977
27.7.5.2 Manage processes....977
27.7.5.3 Manage account information....978
27.7.5.4 Manage software versions....978
27.7.5.5 Set the server fingerprint refresh frequency....979
27.7.6 Log retrieval....979
27.7.6.1 Log retrieval overview....979
27.7.6.2 Search for logs....980
27.7.6.3 Supported log sources and fields....981
27.7.6.4 Inference rules and logical operators....986
27.7.7 Settings....987
27.7.7.1 Manage security settings....987
27.7.7.2 Install the Server Guard agent....988
27.7.7.3 Uninstall the Server Guard agent from a server....990
27.8 Physical machine security....990
27.8.1 View and handle file tampering events....990
27.8.2 View and handle suspicious processes....991
27.8.3 View and handle suspicious network connections....992
27.8.4 View and handle suspicious port listening events....993
27.9 Asset overview....993
27.9.1 Overview....993
27.9.2 Manage groups....994
27.9.2.1 Add a group....995
27.9.2.2 Delete a group....996
27.9.2.3 Sort groups....997
27.9.3 Asset information....997
27.9.3.1 Manage server assets....997
27.9.3.2 Manage NAT assets....998
27.9.3.3 Modify attributes for multiple assets....1000

27.10 Security audits....1001
27.10.1 Overview.... 1001
27.10.2 View audit overview....1002
27.10.3 Query audit events....1003
27.10.4 View raw logs....1003
27.10.5 Policy settings.... 1004
27.10.5.1 Add an audit policy.... 1004
27.10.5.2 Manage action types....1006
27.10.5.3 Set an alert recipient....1008
27.10.5.4 Manage event log archives....1009
27.10.5.5 Manage export tasks.... 1010
27.10.5.6 Modify system settings....1011
27.11 System management.... 1012
27.11.1 Manage your Alibaba Cloud account....1012
27.11.2 Rule database synchronization....1014
27.11.2.1 Synchronization overview.... 1014
27.11.2.2 Specify the upgrade mode for rule databases....1015
27.11.2.3 Refresh the cloud synchronization list.... 1016
27.11.2.4 Manually upgrade a rule database.... 1016
27.11.2.5 Import an offline upgrade package....1017
27.11.2.6 Roll back a rule database....1017
27.11.2.7 View version history of a rule database....1018
27.11.3 Alert settings.... 1018
27.11.3.1 Set alert recipients....1018
27.11.3.2 Set alert information....1019
27.11.4 Global settings.... 1019
27.11.4.1 Set CIDR blocks for traffic monitoring.... 1019
27.11.4.1.1 Add a CIDR block for traffic monitoring....1020
27.11.4.1.2 Manage CIDR blocks for traffic monitoring....1021
27.11.4.2 Region settings.... 1021
27.11.4.2.1 Add a CIDR block for a region.... 1021
27.11.4.2.2 Manage CIDR blocks for region detection....1022
27.11.4.3 Configure whitelists....1023
27.12 Optional security products....1024
27.12.1 Anti-DDoS settings....1024
27.12.1.1 Overview.... 1024
27.12.1.2 View DDoS events....1025
27.12.1.3 Anti-DDoS rules.... 1025
27.12.1.3.1 Set alert thresholds....1025
27.12.1.3.2 Manage anti-DDoS rules....1027
28 Key Management Service (KMS).... 1028
28.1 What is KMS.... 1028
28.2 Log on to the KMS console....1029
28.3 Create a CMK.... 1030

28.4 View CMK details.....1031
28.5 Enable a CMK.....1031
28.6 Disable a CMK.....1031
28.7 Schedule a CMK to be deleted.....1032

1 What is the Apsara Stack console?

The Apsara Stack console is a service capability platform that is based on the Alibaba Cloud Apsara Stack platform and is customized for government and enterprise customers. This platform focuses on improving IT management and solving operation problems for you, and is dedicated to provide a service capability platform of industrial cloud computing. It provides large-scale and cost-effective one-stop cloud computing and big data services for customers in many industries, such as government, education, healthcare, finance, and enterprise.

Overview

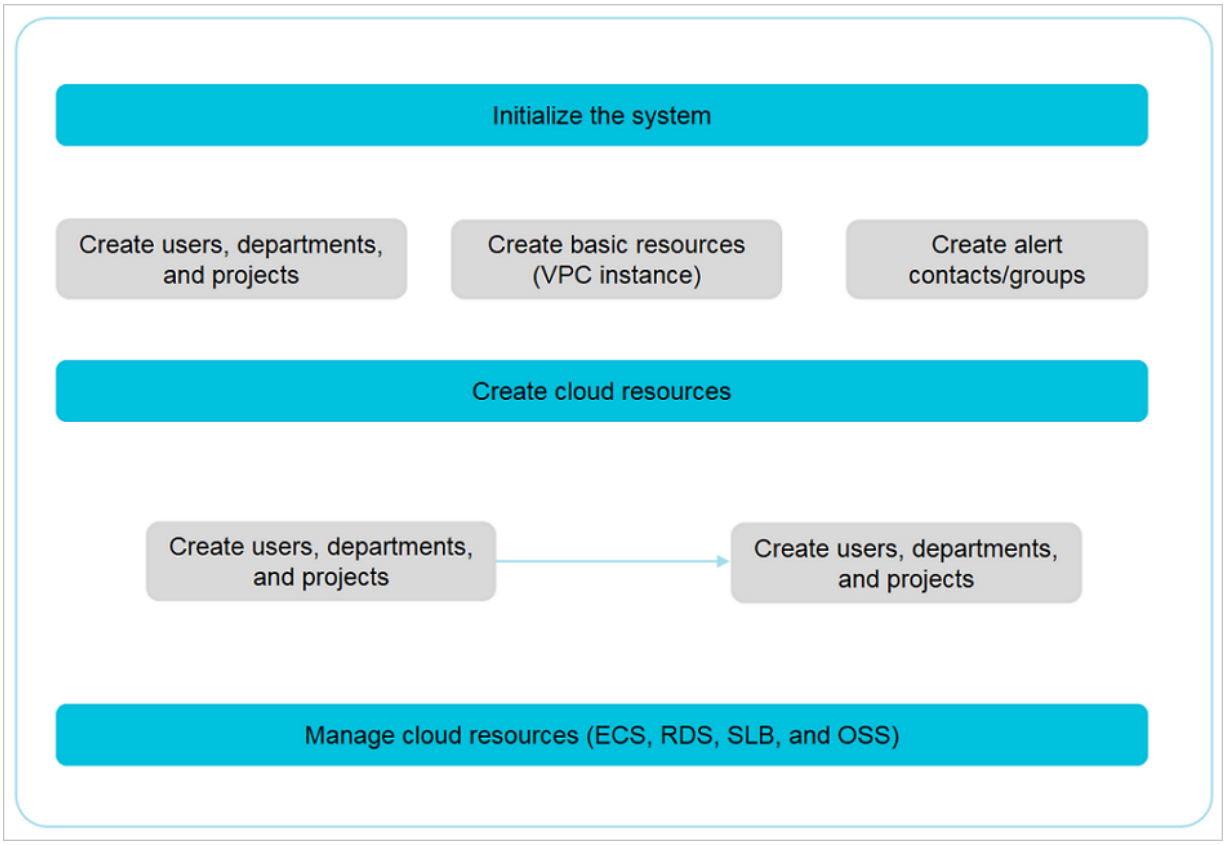
The Apsara Stack console builds the government and enterprise Apsara Stack platform that supports different business types, simplifies management and deployment of physical and virtual resources, and helps you easily and rapidly establish your own business system with higher resource utilization and lower Operation and Maintenance (O&M) costs. It shifts your attention from operation and O&M to business, brings the Internet economic model to government and enterprise customers, and builds a brand new ecological chain that is based on cloud computing.

Procedure

The main operations available in the Apsara Stack console are as follows:

- Initialize the system: Complete the basic system configurations, such as creating departments , projects, users, basic resources (Virtual Private Cloud (VPC) instances), alert contacts, and contact groups.
- Create cloud resources: The administrator directly creates resources as required.
- Manage cloud resources: Manage resources, such as starting, using, and releasing resources, and changing resource configurations and resource quotas.

Figure 1-1: Procedure of the Apsara Stack console



2 Log on to the Apsara Stack console

This topic describes how to log on to the Apsara Stack console as cloud product users.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

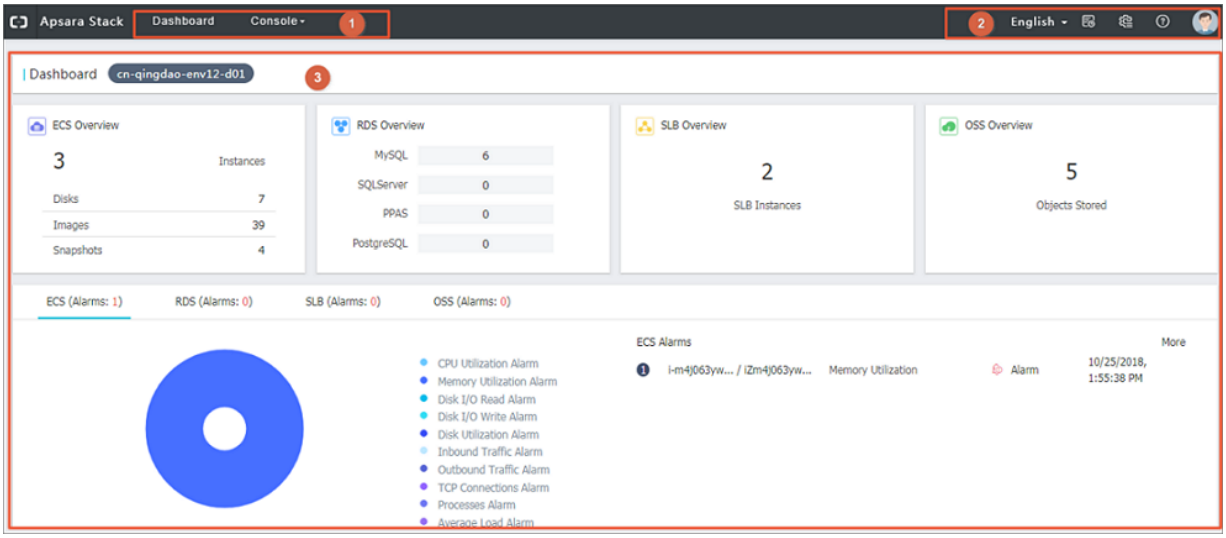
Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.

3 Web page introduction

The Web page of the Apsara Stack console consists of three areas: main menu bar, information area of the current logon user, and operation area.

Figure 3-1: Apsara Stack console



For more information about the functional areas of the Web page, see *Table 3-1: Functional areas of the Web page*.

Table 3-1: Functional areas of the Web page

Area		Description
1	Main menu bar	• Dashboard: displays the resource overview and monitoring status of each service in the Apsara Stack console. • Console: manages the overall system and all resources. It contains the following modules: — Compute, Storage & Networking: manages all types of basic cloud products and resources. — Database: manages all types of database products and resources. — Big Data: manages all types of big data products and resources. — Administration: manages the CloudMonitor, System Reports, Operation Log, and Task Center of the system. — Operations Center: manages resource allocation of the system. — User Center: manages the departments, projects, roles, users, and logon policies of the system.

Area		Description
		Note: The menu bar varies with different roles. See your menu bar for relevant functions.
2	Information area of the current logon user	•  : click this to select language. •  : click this to display the history and most frequently accessed menu items. •  : click this to go to the **System Configuration** page. •  : click this to go to the **Help** page. • Click your avatar and select **Personal Information** to go to the **Personal Information** page or select **Log Off** to log off of the console. On the **Personal Information** page, you can: — View your basic information. — Change your information. — Change your avatar. — Change your logon password. — View the AccessKey. — View the third-party AccessKey.
3	Operation area	Displays the function configuration page of the selected menu item.

You can go to any pages of cloud products and the system displays the **Customize Menu** in the left-side navigation pane. Click the **Customize Menu** and configure your common menu items.

Click  to expand and collapse the left-side navigation pane.

4 Configuration of system initialization

4.1 Configuration instruction

Before using the Apsara Stack console, the administrator must complete a series of basic configurations, such as creating departments, roles, projects, users, and initializing resources according to the configuration process.

The Apsara Stack console follows service principles to perform centralized management for users , roles, departments, and projects related to cloud data centers, which allows you to grant different resource access permissions to users.

- Department

  After the Apsara Stack console is deployed, the system creates a root department by default. You can create departments under the root department.

  The departments are displayed hierarchically and you can create sub-departments under each department.

- Role

  A collection of access permissions. When creating users, you must assign different roles to users to meet their access control requirements on the system.

- Project

  A container where resources are stored. All resources must be created under a corresponding project.

For the relationships among departments, users, projects, roles, and cloud resources, see *Relationships among departments, users, projects, roles, and cloud resources*.

Figure 4-1: Relationships among departments, users, projects, roles, and cloud resources

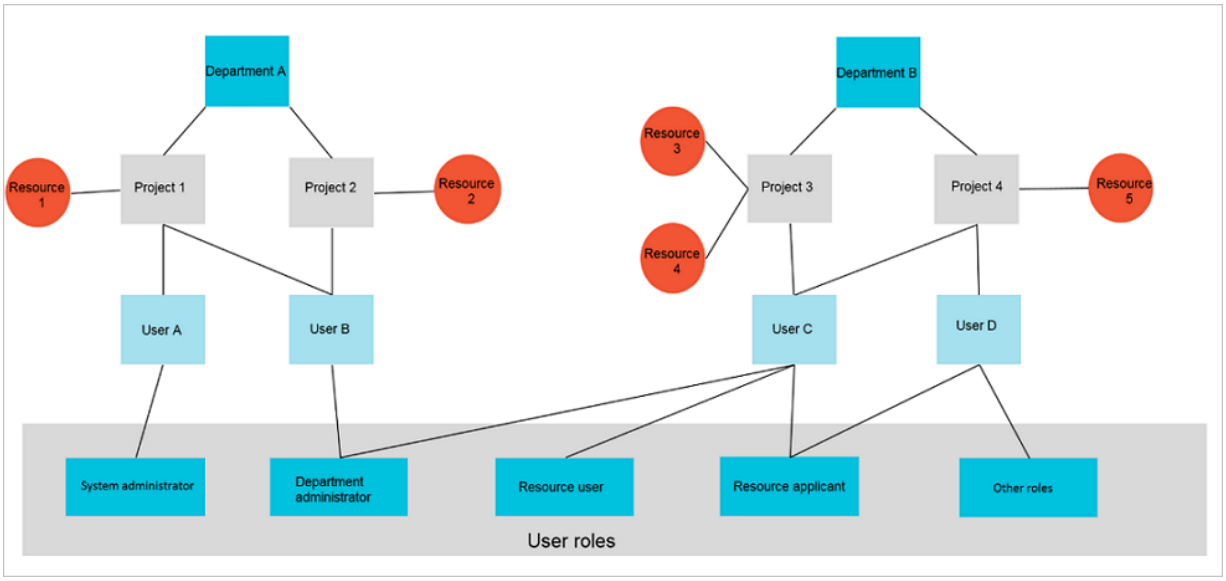


Table 4-1: Relationship table

Relationship	Relationship type	Description
Department and project	One to many	A department can have multiple projects, but each project can only belong to one department.
Department and user	One to many	A department can have multiple users, but each user can only belong to one department.
Project and user	Many to many	A user can have multiple projects, and a project can be assigned to multiple users who belong to the same level-1 department.
User and role	One to many	A user can have multiple roles, and a role can be assigned to multiple users.
Project and resource	One to many	A project can have multiple resources, but each cloud resource can only belong to one project.

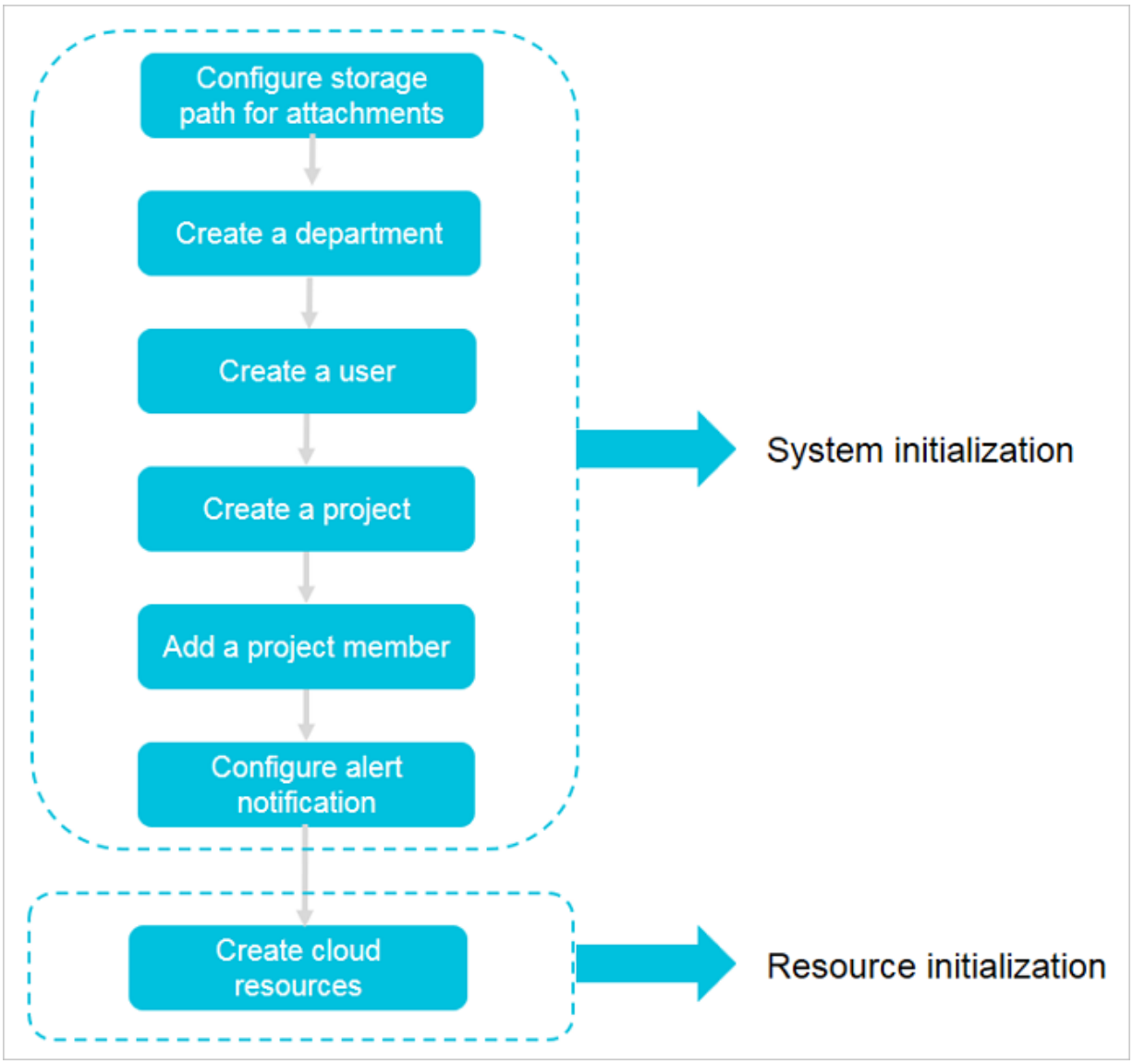
After initializing the system, the administrator initializes resources, creating cloud resources for project members to use.

4.2 Configuration process

This topic describes the initial configuration process of the system.

Before using the Apsara Stack console, the administrator must complete the initial configuration of the system as shown in *Figure 4-2: Initial system configuration process*.

Figure 4-2: Initial system configuration process



1. *Configure storage path for attachments*

   Configure the storage path for uploaded attachments.

2. *Create a department*

   Create a department to store projects and the resources in the projects.

3. *Create a user*

The administrator can create users and assign roles to users to meet their access control requirements on the system.

4. *Create a project*

   Create a project before applying for resources.

5. *Add a project member*

   Add a user to a project.

6. *Configuration of alert notification*

   Configure the alert notification to allow alert contacts to receive alert notifications by email, SMS, and DingTalk when alerts are triggered in real time.

7. Create cloud resources

   The administrator can create cloud product instances in the console of each cloud product according to the project requirements. For more information, see the detailed introduction of all cloud products.

5 Resource management

5.1 Quota management

5.1.1 Quota parameters

This topic introduces the quota parameters of products.

OSS

Parameter	Description
Total OSS Instances	Total number of buckets that you can configure for Object Storage Service (OSS)
Total OSS Capacity (GB)	Total size of buckets that that you can configure for OSS

ECS

Parameter	Description
Total CPU Quota (Cores)	Total number of CPU cores that you can configure for Elastic Compute Service (ECS) and the used cores
Total Memory Quota (GB)	Total memory size that you can configure for ECS
Total Disk Quota (GB)	Total number of cloud disks that you can configure for an ECS instance

RDS (including primary instances and read-only instances)

Parameter	Description
Total CPU Quota (Cores)	Total number of CPU cores that you can configure for ApsaraDB for Relational Database Service (RDS) (MySQL/PPAS/ PostgreSQL) and the used cores
Total Memory Quota (GB)	Total memory size that you can configure for ApsaraDB for RDS (MySQL/PPAS/PostgreSQL)
Total Storage Quota (GB)	Total storage size that you can configure for ApsaraDB for RDS (MySQL/PPAS/PostgreSQL)

SLB

Parameter	Description
Total External IP Addresses Quota	Total number of external IP addresses that you can configure for Server Load Balancer (SLB)
Total Internal IP Addresses Quota	Total number of internal IP addresses that you can configure for SLB

MongoDB

Parameter	Description
Total CPU Quota (Cores)	Total number of CPU cores that you can configure for ApsaraDB for MongoDB and the used cores
Total Memory Quota (GB)	Total memory size that you can be configure for ApsaraDB for MongoDB
Total Storage Quota (GB)	Total storage size that you can configure for ApsaraDB for MongoDB

VPC

Parameter	Description
VPCs	Total number of Virtual Private Cloud (VPC) instances that you can configure for VPC

NAS

Parameter	Description
Storage Quota	Total space size that you can configure for Network Attached Storage (NAS)

Redis

Parameter	Description
Total Memory Quota (GB)	Total memory size that you can configure for KVStore for Redis

SLS

Parameter	Description
Storage Quota	Total space size that you can configure for Log Service

5.1.2 Create cloud resource quotas

The Apsara Stack console supports configuring quotas to allocate resources reasonably among departments.

Context

You can create quotas. If the parent department (except a level-1 parent department) has a quota , the result that the quota of the parent department minus the quotas of other sub-departments is the maximum quota that can be configured for the sub-department. The result cannot be smaller than the amount of resources that you have created for the sub-department.

This topic takes the ECS quota as an example to describe how to create quotas. You can create quotas for other cloud resources in a similar way.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Operations Center** > **Quota Management**.
3. Click the **Quota** tab.
4. Select the **Department** and **Region**.
5. Click **ECS** as the product for which you want to create quotas.
6. In the upper-right corner of the ECS section, click the  icon.
7. Configure the total quotas and then click the  icon.

   For more information about the quota parameters, see *Quota parameters*.

5.1.3 View total and used quotas

The administrator can view the total and used quotas of cloud resources for different departments and regions.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.

2. In the top navigation bar, choose **Console** > **Operations Center** > **Quota Management**.
3. Click the **Overview** tab.
4. Select the region, department, and product. Then, click **Search**.
5. In the search result, view the total and used quotas of the product.

5.1.4 Change quotas

The administrator can change cloud resource quotas based on the department requirements.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Operations Center** > **Quota Management**.
3. Click the **Quota** tab.
4. Select the department and region.
5. Select the product whose quotas you want to change.

   The system displays the product quota section.
6. In the upper-right corner of the product quota section, click the  icon.

   Note:
   For ApsaraDB for Relational Database Service (RDS), select a product type first.
7. Enter the quotas and then click the  icon.

5.1.5 Delete quotas

The administrator can delete quotas as required.

Prerequisites

Before deleting quotas, make sure that all sub-departments of the selected department do not have any quotas.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Operations Center** > **Quota Management**.
3. Click the **Quota** tab.

4. Select the department and region.
5. Select the product whose quotas you want to delete.

   The system displays the product quota section.
6. In the upper-right corner of the product quota section, click the  icon.

   Values in the product quota section are cleared.

> **Note:**
> For ApsaraDB for Relational Database Service (RDS), select a product type first.

5.2 View and export project resource overview

The cloud resource overview displays the numbers of resources and alerts in each department and project in the Apsara Stack console in the format of lists.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or a user.
2. In the left-side navigation pane, click **Overview**.
3. Click the **Resource Overview** tab.

   On the **Resource Overview** tab, you can quickly view the numbers of resources and alerts in each department and project to learn about the distribution and running conditions of resources.

> **Note:**
> Click a number on the page and then you are redirected to the corresponding resource page where you can view the detailed resource information.

4. Click **Export** on the **Resource Overview** tab to add the resource overview report to the **Download Center** of reports.

What's next

After adding, choose **Console** > **Administration** > **System Reports**. Click **Download Center** to go to the report list page and download the resource overview report.

5.3 Configuration of resource notifications

5.3.1 Configure resource notification objects

The administrator can configure the resource notification objects to receive emails and SMS notifications from the Apsara Stack console when resources are created or deleted.

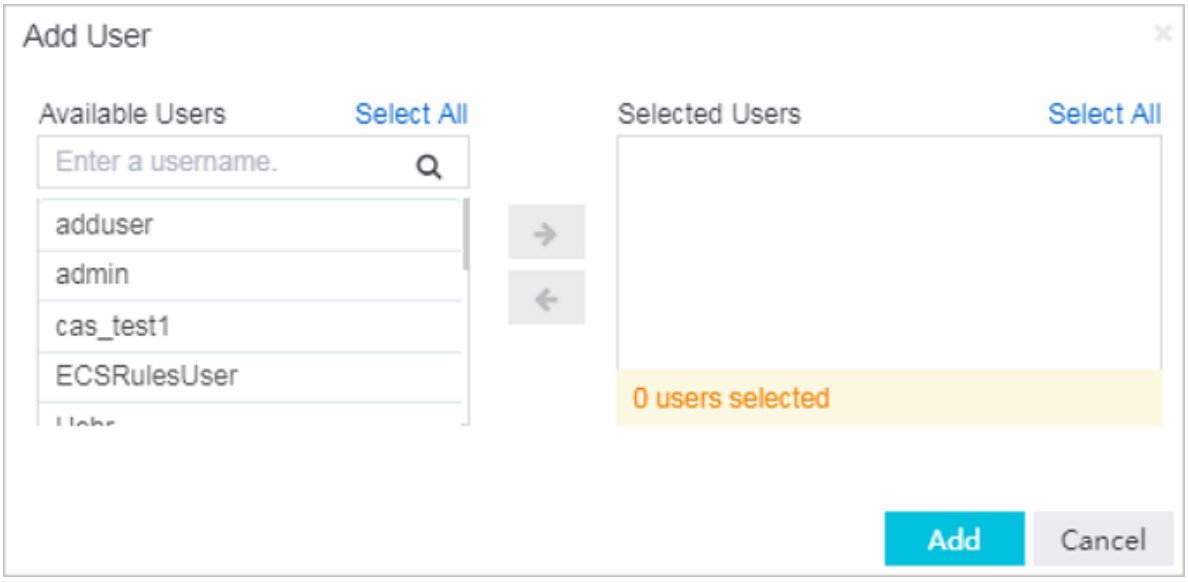
Context

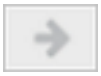
You can add users as notification objects by configuring the resource notifications. If resources are created or deleted in the Apsara Stack console, the system sends emails and SMS notificati ons to the configured notification objects. Users who are added as notification objects can keep up with the resource usage.

Note:
The Apsara Stack console can send resource notifications for Elastic Compute Service (ECS), Object Storage Service (OSS), ApsaraDB for Relational Database Service (RDS), and Server Load Balancer (SLB) resources.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the icon to go to the **System Configuration** page.
3. Click the **Resource Notification Configuration** tab.
4. On the **Resource Notification Configuration** tab, click **Add**.



5. In the displayed **Add User** dialog box, select users in the **Available Users** field and click the  icon to add them to **Selected Users**. Click **Add** to complete the configurations.

5.3.2 View resource notification objects

The administrator can view the information of resource notification objects.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon to go to the **System Configuration** page.
3. Click the **Resource Notification Configuration** tab.
4. On the **Resource Notification Configuration** tab, view the resource notification objects in the Apsara Stack console.
5. Optional: You can click the username of a resource notification object to view the detailed information of this notification object.

5.3.3 Delete a resource notification object

The administrator can remove users who are no longer required to be notified of new changes from resource notification objects because of business changes or other reasons.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon to go to the **System Configuration** page.
3. Click the **Resource Notification Configuration** tab.
4. Find the user to be deleted. Click the  icon in the **Actions** column and select **Delete**.
5. In the displayed dialog box, click **OK**.

6 Alert management

6.1 Overview

CloudMonitor provides real-time monitoring, alert, and notification services of resources to protect your products and business.

Currently, CloudMonitor can monitor metrics of Elastic Compute Service (ECS), Server Load Balancer (SLB), ApsaraDB for Relational Database Service (RDS), and Object Storage Service (OSS).

You can use the monitoring metrics of cloud products to set alert rules and notification polices to keep up with the running conditions and performance of instance resources for product services. Consider a scale-up in time after receiving an insufficient resource alert.

CloudMonitor has the following functions:

- Automatic monitoring: The monitoring is automatically started based on your created ECS resources or auto scaling groups. You are not required to start it manually or install any plug -ins. You can view the monitoring data of specific instances on the monitoring page after applying for resources.
- Flexible alert: You can configure alerts flexibly, such as setting alerts and thresholds for monitoring metrics, pausing and enabling alerts.
- Real-time notification: You can set the alert notification to receive notifications by SMS or email in real time. If the status of an alert rule changes, such as alerts are triggered, data is insufficie nt, or alerts are cleared, the system informs you by sending SMS or email.

6.2 Alert contacts

6.2.1 Create an alert contact

You can create an alert contact to receive alert notifications.

Context

An alert contact is a person who receives alert notifications. Alert notifications can be sent by SMS or email. When monitoring data meets the conditions specified in alert rules, the system sends alert notifications to the corresponding alert contacts.

Procedure

1. *Log on to the Apsara Stack console*.

2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. On the **Alert Contact** tab, click the **Alert Contact** sub-tab.
4. Click **Create Contact**.
5. Configure the alert contact. For more information, see *Alert contact configurations*.

   Table 6-1: Alert contact configurations

Configuration	Description
Username	The username of the alert contact
Department	The department to which the alert contact belongs **Note:** If you select **All**, the alert contact is a global one.
Project	The project to which the alert contact belongs
Cell Phone Number	The mobile phone number of the alert contact. It is used to send alert notifications to the alert contact by SMS. Make sure that the entered mobile phone number is correct. If the number is changed, update it in time on the platform.
Email	The email address of the alert contact. It is used to send alert notificati ons to the alert contact by email. Make sure that the entered email address is correct. If the email address is changed, update it in time on the platform.
DingTalk ID	The DingTalk ID of the alert contact

6. Click **OK**.

6.2.2 Add an alert contact to alert groups

You can add a created alert contact to alert groups for better management.

Prerequisites

- An alert contact is created. For more information, see *Create an alert contact*.
- An alert group is created. For more information, see *Create an alert group*.

Context

An alert contact can be added to multiple alert contact groups.

Procedure

1. *Log on to the Apsara Stack console*.

2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. On the **Alert Contact** tab, click the **Alert Contact** sub-tab.
4. Select the alert contact that you want to add to alert groups and then click **Add to Alert Group**.
5. In the displayed **Change Alert Group** dialog box, select alert groups and click **OK**.

6.2.3 Query an alert contact

You can query the information and alert groups of alert contacts on the **Alert Contact** page.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. On the **Alert Contact** tab, click the **Alert Contact** sub-tab.
4. Select the query condition based on name, cell phone number, email, or DingTalk ID. Enter the keyword in the search bar and then click **Search** to query the information and alert groups of an alert contact.

6.2.4 Change alert contact information

If the information of an alert contact is changed, you can change the alert contact information on the **Alert Contact** page.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. On the **Alert Contact** tab, click the **Alert Contact** sub-tab.
4. Find the alert contact whose information you want to change. Click the  icon in the **Actions** column and select **Change**.
5. In the displayed dialog box, change the contact information of the alert contact, namely the cell phone number, email, and DingTalk ID.
6. Click **OK**.

6.2.5 Delete alert contacts

You can delete one or more alert contacts that are no longer in use based on the business requirements.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. On the **Alert Contact** tab, click the **Alert Contact** sub-tab.
4. Perform the following operations:
   - Delete an alert contact

     Find the alert contact to be deleted. Click the  icon in the **Actions** column and select **Delete**.
   - Delete multiple alert contacts

     Select multiple alert contacts to be deleted and click **Delete Alert Contacts** in the upper-right corner.
5. In the displayed dialog box, click **OK**.

6.3 Alert groups

6.3.1 Create an alert group

You can create an alert group to classify alert contacts.

Context

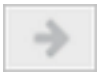
An alert group is a group of alert contacts. It contains one or more alert contacts. If an alert is triggered, all the alert contacts in the alert group can receive the alert notification.

When setting an alert rule, you must select an alert group to receive the alert notifications. For each monitoring metric, if the alert threshold is reached, the system sends alert notifications to the members in the alert group according to the configured notification methods.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. On the **Alert Contact** tab, click the **Alert Group** sub-tab.
4. Click **Create Contact Group**.
5. Configure the alert group.

Configuration	Description
Group Name	The name of the alert contact group, which must be 2 to 20 characters in length and contain letters, numbers, and underscores (_)

Configuration	Description
Department	The department to which the alert contact to be added belongs **Note:** If you select **All**, the alert group is a global one.
Project	The project to which the alert contact to be added belongs
Remarks	The description of the alert contact group, which must be 0 to 256 characters in length and can contain letters, numbers, hyphens (-), or underscores (_)
Choose Contacts	Add contacts to the alert contact group as follows: Select contacts in the **Existing Contacts** field and click the  icon to add them to the **Selected Contacts**. To remove a selected contact, select the contact and then click the  icon. **Note:** If the contact is not created, create an empty alert group first. Then, *create an alert contact* and *add the alert contact to the alert group*.

6. Click **OK**.
7. Optional: To remove an alert contact from the alert group, go to the **Alert Group** page and click **Delete** at the right of the alert contact.

6.3.2 Change alert notification methods

Phone notifications, email notifications, and DingTalk notifications are enabled by default. You can disable the unnecessary notification methods as required.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. On the **Alert Contact** tab, click the **Alert Group** sub-tab.
4. Find the alert contact whose alert notification methods you want to change. Enable or disable the phone notifications, email notifications, and DingTalk notifications by turning on or off the switches.

6.4 Alert rules

6.4.1 Create an alert rule

You can create an alert rule for an instance to monitor this instance.

Prerequisites

For Elastic Compute Service (ECS) instances, you must install the monitoring plug-in to collect the metric data at the operating system level.

Complete the following steps to install the plug-in:

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Monitoring** tab.
4. In the ECS instance list, find the instance to be monitored. Click the  icon in the **Actions** column and select **Install Plugin**.

> **Note:**
> The monitoring chart displays monitoring data 5 to 10 minutes after the monitoring plug-in is installed.

Context

We recommend that you create an alert group before setting an alert rule. You can also create an alert group when you are setting an alert rule. For more information about how to create an alert group, see *Create an alert group*.

Alert rules configured in CloudMonitor are used to monitor the server performance. In this way, you can detect and resolve server problems in time, which guarantees a secure, stable, and effective operation of servers.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Monitoring** tab.
4. Click a cloud product sub-tab.

5. Find the corresponding instance or bucket. Click the  icon in the **Actions** column and select **Alert Rules** to go to the **Alert Item** page.

> **Note:**
> You can also use the search function to find a specific instance or bucket and create an alert rule for the instance or bucket.

6. Click **Create Alert Rule**.
7. Configure the alert rule.

Configuration	Description
Monitor Metric	Select a monitor metirc from the drop-down list. For more information about monitor metrics, see *Cloud monitoring metrics*.
Reference Period	Select a reference period from the drop-down list. The reference period is the interval at which data statistics are generated.
Calculation Method	Select a calculation method from the drop-down list. The following calculation methods are available: • **Average**: If the average value of all monitoring data collected in a reference period exceeds the threshold, an alert is triggered. • **Maximum**: If the maximum value of the monitoring data collected in a reference period exceeds the threshold, an alert is triggered. • **Minimum**: If the minimum value of the monitoring data collected in a reference period exceeds the threshold, an alert is triggered.

8. Click **Next**.
9. Configure the notification object.

   A notification object is an alert contact. For more information about how to configure an alert contact, see *Create an alert contact*.

Configuration	Description
Alert Retries	Select the number of retries before an alert is triggered from the drop-down list If the value exceeds the threshold for consecutive reference period, an alert is triggered. The system notifies alert contacts only after the number of retries is exceeded.

Configuration	Description
Contact Notification Group	Select a contact notification group After you set an alert rule for a monitor metirc, the system sends an alert notification to the alert contacts if the monitoring data meets conditions configured in the alert rule.
Notification Time	Select the notification time, which is a time range during which the system sends alert notifications

10. Click **OK**.

6.4.2 Create multiple alert rules

You can create the same alert rule for multiple instances or buckets to monitor these instances or buckets.

Prerequisites

For Elastic Compute Service (ECS) instances, you must install the monitoring plug-in to collect the metric data at the operating system level.

Complete the following steps to install the plug-in:

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Monitoring** tab.
4. In the ECS instance list, find the instance to be monitored. Click the icon in the **Actions** column and select **Install Plugin**.

Note:
The monitoring chart displays monitoring data 5 to 10 minutes after the monitoring plug-in is installed.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Monitoring** tab.
4. Click a cloud product sub-tab and then select multiple instances or buckets.
5. Click **Create Alert Rules** in the upper-right corner.
6. Configure the alert rule.

Configuration	Description
Monitor Metric	Select a monitor metirc from the drop-down list. For more information about monitor metrics, see *Cloud monitoring metrics*.
Reference Period	Select a reference period from the drop-down list. The reference period is the interval at which data statistics are generated.
Calculation Method	Select a calculation method from the drop-down list. The following calculation methods are available: • **Average**: If the average value of all monitoring data collected in a reference period exceeds the threshold, an alert is triggered. • **Maximum**: If the maximum value of the monitoring data collected in a reference period exceeds the threshold, an alert is triggered. • **Minimum**: If the minimum value of the monitoring data collected in a reference period exceeds the threshold, an alert is triggered.

7. Click **Next**.
8. Configure the notification object.

   A notification object is an alert contact. For more information about how to configure an alert contact, see *Create an alert contact*.

Configuration	Description
Alert Retries	Select the number of retries before an alert is triggered from the drop-down list If the value exceeds the threshold for consecutive reference period, an alert is triggered. The system notifies alert contacts only after the number of retries is exceeded.
Contact Notification Group	Select a contact notification group After you set an alert rule for a monitor metirc, the system sends an alert notification to the alert contacts if the monitoring data meets conditions configured in the alert rule.
Notification Time	Select the notification time, which is a time range during which the system sends alert notifications

9. Click **OK**.

6.4.3 View alert rules

You can view your alert rules on the **Alert Rules** page after creating an alert rule.

Context

Alert rules are used to display monitoring metrics in the alert rules of CloudMonitor. In this way , you can quickly view monitoring metrics, which guarantees a secure, stable, and effective operation of servers.

Currently, the system provides the alert rules for Elastic Compute Service (ECS), ApsaraDB for Relational Database Service (RDS), Server Load Balancer (SLB), Object Storage Service (OSS), and KVStore for Redis. The operations used to manage alert rules are similar to each other. Therefore, take the ECS alert rules as an example in this topic.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Alert Rules** tab.
4. Click the tab of a cloud product, such as **ECS**. Then the ECS alert rules appear.
5. Enter the ID and name of a monitored resource, select a region, monitoring metric, alert status, and enabled status, and then click **Search** to query alert rules.

6.4.4 View alert history

After alerts are triggered, you can view the alert history on the **Alert Rules** page.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**. Then, click the **Alert Rules** tab.
3. Click the tab of a cloud product.
4. Find the alert rule whose alert history you want to view. Click the  icon in the **Actions** column and select **Alert History**.
5. Optional: In the displayed **History** dialog box, select the start date and end date of the alert and then click **Search**.
6. In the **History** dialog box, view the alert history.

6.4.5 Change an alert rule

You can change the alert rule on the **Alert Rules** page of your cloud product.

Procedure

1. *Log on to the Apsara Stack console*.

2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Alert Rules** tab.
4. Click the tab of a cloud product.
5. Find the alert rule to be changed. Click the  icon in the **Actions** column and select **Change** to change the alert rule.

   For more information about how to change alert rules, see *Create an alert rule* to configure alert rules.

6.4.6 Pause alert rules

You can pause one or more alert rules as required.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Alert Rules** tab.
4. Click the tab of a cloud product.
5. Perform the following operations:
   - Pause an alert rule.

     Find the alert rule to be paused. Click the  icon in the **Actions** column and select **Pause**.
   - Pause multiple alert rules.

     Select multiple alert rules to be paused and click **Pause** in the upper-right corner.
6. In the displayed dialog box, click **OK**.

   After you pause the alert rules, the system stops to send alert notifications to the corresponding alert contacts.

6.4.7 Start alert rules

You can start one or more paused alert rules as required.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Alert Rules** tab.

4. Click the tab of a cloud product.
5. Perform the following operations:
   - Start an alert rule.

     Find the alert rule to be started. Click the  icon in the **Actions** column and then select **Enable**.
   - Start multiple alert rules.

     Select multiple alert rules to be started and click **Start** in the upper-right corner.

6.4.8 View alert notification objects

After creating alert rules, you can view the notification object of each alert rule on the **Alert Item** page.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Alert Rules** tab.
4. Click the tab of a cloud product.
5. Click the alert contact or alert group in the **Alert Contact** column.

   The system displays the detailed information of alert contacts in the appeared dialog box.

6.4.9 Delete alert rules

You can delete one or more alert rules that are no longer in use.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Alert Rules** tab.
4. Click the tab of a cloud product.
5. Perform the following operations:
   - Delete an alert rule.

     Find the alert rule to be deleted. Click the  icon in the **Actions** column and select **Delete**.

- Delete multiple alert rules.

  Select multiple alert rules to be deleted and click **Delete** in the upper-right corner.

6. In the displayed dialog box, click **OK**.

6.5 Configuration of alert notification

6.5.1 Configure email alert notification

The administrator can configure the email alert notification to allow alert contacts to receive alert notifications by email when alerts are triggered.

Prerequisites

Make sure that the SMTP server URL is obtained before you configure the email notification.

To obtain the SMTP server URL and port, view the official description of the mailbox system to be configured. Generally, the SMTP server URL is in the format of smtp.xxxx.com. For example, the SMTP server URL of the 163 mailbox is smtp.163.com.

The system sends email notifications by using the configured email address and email password.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon.
3. On the **System Configuration** page, click the **Alert Notification Configuration** tab.
4. In the **Email Alert Notification Settings** section, click **Configure**.

   The **Email Alert Notification Settings** dialog box appears.
5. Enter the SMTP server URL, email address, and email password, and then select the SMTP server port.
6. Click **OK**.

   To change the configurations, click **Clear Settings** and reconfigure the settings.

6.5.2 Configure DingTalk alert notification

The administrator can configure the DingTalk alert notification to allow alert contacts to receive alert notifications by DingTalk when alerts are triggered.

Context

To send alert notifications by using DingTalk, you must obtain the CorpID, CorpSecret, and AgentID.

Procedure

1. Obtain the AgentID.
   a) Log on to oa.dingtalk.com as a DingTalk administrator.
   b) Click **Applications** and find the **Application Base** section.
   c) Click the  icon on an application and then select **Set**.
   d) In the displayed dialog box, obtain the AgentID.
2. Obtain the CorpID and CorpSecret.
   a) Log on to oa.dingtalk.com as a DingTalk administrator.
   b) Click **Applications** and find the **Create your app** section.
   c) Click **Open Application** to go to the DingTalk developer platform.
   d) In the left-side navigation pane, click **Account Management**. In the **Account Information** section, obtain the CorpID and CorpSecret.
3. *Log on to the Apsara Stack console* as a system administrator.
4. In the upper-right corner, click the  icon.
5. On the **System Configuration** page, click the **Alert Notification Configuration** tab.
6. In the **DingTalk Alert Notification Settings** section, click **Configure**.

   The **DingTalk Notification Settings** dialog box appears.
7. Enter the CorpID, CorpSecret, and AgentID.
8. Click **OK**.

   To change the configurations, click **Clear Settings** and reconfigure the settings.

6.5.3 Configure SMS alert notification

The administrator can configure the SMS alert notification to allow alert contacts to receive alert notifications by SMS when alerts are triggered.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon.
3. On the **System Configuration** page, click the **Alert Notification Configuration** tab.

4. In the **SMS Alert Notification Settings** section, click **Configure**.

   The **SMS Notification Settings** dialog box appears.

5. Enter the Notification URL, AccessKey ID, and AccessKey Secret.
6. Click **OK**.

   To change the configurations, click **Clear Settings** and reconfigure the settings.

6.6 View alert information

You can view alert information to have a clear grasp of the running conditions of Elastic Compute Service (ECS), ApsaraDB for Relational Database Service (RDS), Server Load Balancer (SLB), and Object Storage Service (OSS) and obtain the exception information in time.

Context

Alert information displays the information of alert rules that do not meet the requirements.

> **Note:**
> This topic takes the ECS alert information as an example to describe how to view the alert information. You can view the alert information of other cloud resources in a similar way.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Alert Information** tab.
4. You can filter alert information based on the region, monitored resource ID, monitored resource name, monitor metric, start date, and end date. See the following table for the field descriptions of alert information.

   Table 6-2: Field descriptions

Field	Description
Region	Region in which the monitored object resides
Monitored Resource ID/Name	Instance ID or name of the monitored object
Monitor Metric	Monitor metirc of the monitored object
Description	Detailed description of the alert information

Field	Description
Trigger Status	Alert trigger status, including **Alerts** and **Insufficient Data**
Threshold	Threshold of the monitor metric
Alert Value	Value of the monitor metirc when the alert is triggered
Start Time	Time when the alert is started
End Time	Time when the alert is ended

5. Optional: Click **Export** to export the current alert information to your local computer as an XLS file.

   The exported file is named *`Alert.xls`* and stored in *`C:\Users\Username\Downloads`* by default.

7 Monitoring management

7.1 Overview

CloudMonitor provides real-time monitoring, alert, and notification services for resources to protect your products and business.

Currently, CloudMonitor can monitor metrics of Elastic Compute Service (ECS), Server Load Balancer (SLB), ApsaraDB for Relational Database Service (RDS), and Object Storage Service (OSS).

You can use the monitoring metrics of cloud products to set alert rules and notification polices to keep up with the running conditions and performance of instance resources for product services. Consider a scale-up in time after receiving an insufficient resource alert.

CloudMonitor has the following functions:

- Automatic monitoring: The monitoring is automatically started based on your created ECS resources or auto scaling groups. You are not required to start it manually or install any plug -ins. You can view the monitoring data of specific instances on the monitoring page after applying for resources.
- Flexible alert: You can configure alerts flexibly, such as setting alerts and thresholds for monitoring metrics, pausing and enabling alerts.
- Real-time notification: You can set the alert notification to receive notifications by SMS or email in real time. If the status of an alert rule changes, such as alerts are triggered, data is insufficie nt, or alerts are cleared, the system informs you by SMS or email.

7.2 View dashboard

The Apsara Stack console uses charts to display the usage and monitoring conditions of existing system resources in all regions, which helps you learn about the usage of current resources.

Context

Note:
Resource types vary with region types. For resource types available to you, see your **Dashboard** page.

Procedure

1. *Log on to the Apsara Stack console*.

2. In the top navigation bar, click **Dashboard**.
3. Click the region to view the usage overview and alert information of each cloud resource in each region.
   - Usage overview of cloud resources

     In each cloud resource overview, click the total number of resources and then you are redirected to the corresponding resource page. You can view the detailed resource information on the resource page.

     For more information about the fields in the resource overview, see *Resource overview fields*.

     Table 7-1: Resource overview fields

Cloud product	Content	Description
Elastic Compute Service (ECS)	Number of ECS instances	Total number of ECS instances of the current user
	Number of disks	Total number of disks of the current user
	Number of images	Total number of images of the current user
	Number of snapshots	Total number of snapshots of the current user
Server Load Balancer (SLB)	Number of SLB instances	Total number of SLB instances of the current user
Object Storage Service (OSS)	Number of objects stored	Total number of OSS instances of the current user
ApsaraDB for Relational Database Service (RDS)	Number of ApsaraDB RDS for MySQL/PPAS/PostgreSQL instances	Total number of ApsaraDB RDS for MySQL/PPAS/PostgreSQL instances of the current user

   - Alert information of cloud resources

     Click the corresponding tab of each cloud resource to view their alert information. On each cloud resource tab, click **More** on the right and then you are redirected to the corresponding alert information page.

For more information about the alert rules of each cloud resource, see *Cloud monitoring metrics*.

7.3 CloudMonitor

7.3.1 Overview

CloudMonitor provides real-time monitoring, alert, and notification services for resources to protect your products and business.

You can use the monitoring metrics of cloud products to set alert rules and notification polices to keep up with the running conditions and performance of product instances. Consider a scale-up in time after receiving an insufficient resource alert.

CloudMonitor has the following functions:

- Automatic monitoring: The monitoring is automatically started based on your created Elastic Compute Service (ECS) resources or auto scaling groups. You are not required to start it manually or install any plug-ins. You can view the monitoring data of specific instances on the monitoring page after applying for resources.
- Flexible alert: You can configure alerts flexibly, such as setting alerts and thresholds for monitoring metrics, pausing and enabling alerts.
- Real-time notification: You can set the alert notification to receive notifications by SMS or email in real time. If the status of an alert rule changes, such as alerts are triggered, data is insufficie nt, or alerts are cleared, the system informs you by SMS or email.

7.3.2 View CloudMonitor overview

On the **CloudMonitor** page, click the **Overview** tab. You can view the number of instances, number of alert rules, number of alerts, and alert status of all cloud products.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. On the **Overview** tab, view the number of instances, number of alert rules, number of alerts, and alert status of all cloud products.

7.3.3 Cloud monitoring metrics

This topic describes the monitoring metrics of each product and the corresponding description.

CloudMonitor tests the service availability based on the monitoring metrics of cloud resources. You can configure alert rules and notification polices for these monitoring metrics to keep up with the running conditions and performance of product instances.

CloudMonitor can monitor resources of Elastic Compute Service (ECS), ApsaraDB for Relational Database Service (RDS), Server Load Balancer (SLB), Object Storage Service (OSS), and KVStore for Redis. Monitoring metrics supported by each service are described as follows.

Table 7-2: ECS monitoring metrics

Metric name	Metric description	Measured object	Calculation formula	Remarks
CPU Utilization	Used to measure the CPU utilization (%) of a measured object	ECS instance	CPU utilization of the ECS instance/ Total CPU cores of the ECS instance	None
Memory Utilization	Used to measure the memory utilization (%) of a measured object	ECS instance	Memory utilization of the ECS instance /Total memory of the ECS instance	The memory utilization calculated by CloudMonitor does not include cache utilization. Therefore, when you run the **`free`** or **`top`** command to query the memory utilization of a Linux server, the result may be inconsistent with the memory utilization displayed in the Apsara Stack console.
Disk I/O Read	Used to measure the volume of data read from a measured object per second (KB/s)	ECS instance	Total bytes read from the ECS instance disk /Statistical peirod	For Linux hosts, the disk I/O monitoring data is obtained by using the iostat tool. If your Linux host has no disk I/O data, check if iostat is installed on your machine . If not, Redhat or CentOS users can use yum to install the tool, and Ubuntu or Debian users can use apt-get to install the tool.
Disk I/O Write	Used to measure the	ECS instance	Total bytes written to	None

Metric name	Metric description	Measured object	Calculation formula	Remarks
	volume of data written to a measured object per second (KB/s)		the ECS instance disk /Statistical peirod	
Disk Utilization	Used to measure the disk utilizatio n (%) of a measured object	ECS instance	Used capacity of the ECS instance disk/Total capacity of the ECS instance disk	None
Inbound Traffic	Used to measure the inbound network traffic of a measured object per second (Kbit/s)	ECS instance	None	None
Outbound Traffic	Used to measure the outbound network traffic of a measured object per second (Kbit/s)	ECS instance	None	If the purchased bandwidth is used up, access fails or requests slow down. On the monitoring chart, eth0 indicates the intranet NIC of the server, and eth1 indicates the Internet NIC of the server.
TCP Connection s	Total number of TCP connections set up by the server	ECS instance	None	None
Processes	After you set an alert rule with this monitoring metric, the specified running processes are	ECS instance	None	To monitor the running conditions of processes on the server, set an alert rule with this monitoring metric to trigger the alert when the number of processes is unequal to the actual number of processes.

Metric name	Metric description	Measured object	Calculation formula	Remarks
	counted and the system displays the total number of these processes.			
Average Load	A concept in Linux, the average load value of the server	ECS instance		The average load value cannot be greater than 1. If your server has a multi-core processor, the average load value must be divided by the number of CPU cores and the result must be smaller than 1. Generally, if the average load value is greater than 1, processes are queued up and the server slows down.

Note:

For ECS instances, you must install the monitoring plug-in to collect the metric data at the operating system level.

Installation method: On the **CloudMonitor** page, click the **Monitoring** tab. In the ECS instance list, locate the instance to be monitored. Click the icon in the **Actions** column and select **Install Plugin**.

The monitoring chart displays monitoring data 5-10 minutes after the monitoring plug-in is installed.

Table 7-3: RDS monitoring metrics

Metric name	Metric description	Measured object	Calculation formula
CPU Utilization	Used to measure the CPU utilization (%) of a measured object	ApsaraDB for RDS instance	CPU utilization of the ApsaraDB for RDS instance /Total CPU cores of the ApsaraDB for RDS instance

Metric name	Metric description	Measured object	Calculation formula
Memory Utilization	Used to measure the memory utilization (%) of a measured object	ApsaraDB for RDS instance	Memory utilization of the ApsaraDB for RDS instance/ Total memory of the ApsaraDB for RDS instance
Disk Utilization	Used to measure the disk utilization (%) of a measured object	ApsaraDB for RDS instance	None
IOPS Utilization	Used to measure the number of I/O requests of a measured object per second. Unit: %	ApsaraDB for RDS instance	Number of I/O requests of the ApsaraDB for RDS instance/ Statistical period
Connection Utilization	Used to measure the number of connections between the applicatio n and the measured object per second. Unit: %	ApsaraDB for RDS instance	Number of connections between the application and the ApsaraDB for RDS instance per second/Statistical period

Table 7-4: SLB monitoring metrics

Metric name	Metric description	Measured object	Remarks
Outbound Packets per Second	Number of packets sent by SLB per second	SLB instance	None
Inbound Packets per Second	Number of request packets received by SLB per second	SLB instance	None
Inbound Data	Traffic consumed to access SLB from the external (Kbit/s)	SLB instance	None
Outbound Data	Traffic consumed by SLB to access the external (Kbit /s)	SLB instance	None
Active Port	Number of all connection s in the ESTABLISHED status	SLB instance	It can be interpreted as, but cannot be equivalent to, the concurrent connections. This is because a persistent

Metric name	Metric description	Measured object	Remarks
Connections			connection transmits multiple file requests simultaneously.
Inactive Port Connections	Number of all TCP connections except connections in the ESTABLISHED status	SLB instance	None.
New Port Connections	Number of times the first SYN_SENT status occurs in a TCP three-way handshake during a statistical period	SLB instance	Active Port Connections, Inactive Port Connections, and New Port Connections are all used to measure the number of requests for connecting a client to an SLB instance.

Table 7-5: OSS monitoring metrics

Metric name	Metric description	Measured object
Reads	Used to measure the number of reads of a measured object	OSS bucket
Internal Server Errors	Used to measure the number of errors of a measured object	OSS bucket
Public Network Inbound Traffic	Used to measure the inbound Internet network traffic (bytes) of a measured object per second	OSS bucket
Public Network Outbound Traffic	Used to measure the outbound Internet network traffic (bytes) of a measured object per second	OSS bucket
Classic Network Inbound Traffic	Used to measure the inbound intranet network traffic (bytes) of a measured object per second	OSS bucket
Classic Network Outbound Traffic	Used to measure the outbound intranet network traffic (bytes) of a measured object per second	OSS bucket

Metric name	Metric description	Measured object
Writes	Used to measure the number of writes of a measured object	OSS bucket
Storage Space Used	Used to measure the used storage space (bytes) of a measured object	OSS bucket

Table 7-6: KVStore for Redis monitoring metrics

Metric name	Metric description	Measured object	Unit
CPU Utilization	Used to measure the CPU utilization of a measured object	KVStore for Redis instance	%
Memory Utilization	Used to measure the proportion of current used memory to the total memory of a measured object	KVStore for Redis instance	%
Memory Used	Used to measure the used memory of a measured object	KVStore for Redis instance	Bytes
Connections Used	Used to measure the total number of connections of client	KVStore for Redis instance	-
The Percentage of Connections Used	Used to measure the proportion of current established connections to the total connections of the measured object	KVStore for Redis instance	%
Write Network Bandwidth	Used to measure the network traffic currently writen per second of a measured object	KVStore for Redis instance	Bytes/s
Read Network Bandwidth	Used to measure the network traffic currently read per second of a measured object	KVStore for Redis instance	Bytes/s
Failed Operations	Used to measure the times of failure of operating a measured object	KVStore for Redis instance	times/s
Write Bandwidth Usage	Used to measure the proportion of currently writen bandwidth to the total bandwidth of a measured object	KVStore for Redis instance	%

Metric name	Metric description	Measured object	Unit
Read Bandwidth Usage	Used to measure the proportion of currently read bandwidth to the total bandwidth of a measured object	KVStore for Redis instance	%
QPS Usage	Used to measure the currently used number of QPS of a measured object	KVStore for Redis instance	times/s

7.3.4 View monitoring charts

You can view the monitoring chart to learn about the running conditions of each instance or bucket.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the top navigation bar, choose **Console** > **Administration** > **CloudMonitor**.
3. Click the **Monitoring** tab.
4. Click the tab of a cloud product.
5. Find the instance or bucket whose monitoring chart you want to view. Click the  icon in the **Actions** column and select **Monitoring Chart**.

   You can view the monitoring data of each monitoring metric on the displayed page.

7.4 System reports

7.4.1 Create a report download task

You need to create a report download task on the **System Reports** page before previewing or downloading reports.

Context

You can create a download task for the following reports:

- Resource report

  A resource report summarizes the current number of instances for cloud products in the Apsara Stack console and the details of each instance, including the region, department, project, and status of the instance.
- Alert report

An alert report summarizes the alert information of cloud products.

- Resource usage evaluation report

  A resource usage evaluation report summarizes the usage of resources for cloud products. You can view resource usage evaluation reports to learn about the usage of each resource and prevent waste or overload use of resources.

- Resource monitoring report

  A resource monitoring report summarizes the monitoring information of cloud products.

- Quota report

  A quota report summarizes the quota information of cloud products in each department.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Administration** > **System Reports**.
3. On the **System Reports** page, click a report tab.
4. Configure the filter conditions or evaluation rules based on business requirements and click **Create Report Download Task**.
5. In the displayed **Create Report Download Task** dialog box, enter a report name and select a product. Then click **Create**.

   After the report download task is created, you can click **Download Center** to go to the download center page to view the status of this task.

7.4.2 Change the report name

The administrator can change the report name on the **Download Center** page after a download task is created.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Administration** > **System Reports**.
3. Click the **Download Center** tab.
4. In the task list, find the task whose report name you want to change. Click the  icon in the **Actions** column and select **Change Report Name**.

> **Note:**
> You can also query the download tasks based on the status or created date of the task to change report names.

5. In the displayed dialog box, enter the report name and click **OK**.

7.4.3 Preview and download a report

The administrator can preview and download a report based on the report name and type.

Prerequisites

You can only preview or download a report whose task status is **Complete**.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Administration** > **System Reports**.
3. Click the **Download Center** tab.
4. Find the report to be previewed based on the report name and type. Click the icon in the **Actions** column and select **Preview**.
5. Find the report to be downloaded based on the report name and type. Click the icon in the **Actions** column and select **Download**.
6. In the displayed dialog box, click **OK**.

   The downloaded file is stored in `C:\Users\Username\Downloads` by default.

7.4.4 Delete a report download task

The administrator can delete a report download task that is no longer in use.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Administration** > **System Reports**.
3. Click the **Download Center** tab.
4. In the task list, find the download task to be deleted. Click the icon in the **Actions** column and select **Delete**.
5. In the displayed dialog box, click **OK**.

7.5 Task center

7.5.1 View running tasks

Before a task is finished, the administrator can view the task details in **Execute Task**.

Context

You can query tasks quickly by department, task ID, task name, task type, and created time.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Administration** > **Task Center**.
3. Click the **Execute Task** tab.
4. Configure the query conditions and then click **Search**.

   In the search results, view the task details.

7.5.2 View previous tasks

The administrator can view the details of finished tasks in **Previous Tasks**.

Context

A previous task is a finished task.

You can query tasks quickly by department, task ID, task name, task type, and created time.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Administration** > **Task Center**.
3. Click the **Previous Tasks** tab.
4. Configure the query conditions and then click **Search**.
5. Optional: If a task fails, click **Error** in the **Status** column.

   View the failure details of the task.

7.6 Operation logs

7.6.1 View logs

The administrator can view logs to learn about the usage conditions of resources, such as Elastic Compute Service (ECS), ApsaraDB for Relational Database Service (RDS), and Server Load

Balancer (SLB). You can also learn about the running conditions of all function modules on the platform in real time.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Administration** > **Operation Log**.
3. On the **Operation Log** page, you can query operation logs by username, module, level, instance ID, start date, and end date.

   For more information about the fields in the search results, see *Field descriptions*.

   Table 7-7: Field descriptions

Field	Description
Time	Operation time
Username	Name of the operator
Module	• ECS: records all actions related to ECS instances, including creating, changing, deleting, and querying ECS instances. • RDS: records all actions related to ApsaraDB for RDS instances, including creating, changing, deleting, and querying ApsaraDB for RDS instances. • OSS: records all actions related to Object Storage Service (OSS) instances , including creating, changing, deleting, and querying OSS instances. • OTS: records all actions related to Table Store instances, including actions of Table Store instances and tables. • SLB: records all actions related to SLB instances, including creating, changing, deleting, and querying SLB instances. • VPC: records all actions related to Virtual Private Cloud (VPC) instances, including creating, changing, deleting, and querying VPC instances, and managing VSwitches and VRouters. • REDIS: records all actions related to KVStore for Redis instances, including creating, querying, and deleting KVStore for Redis instances. • ESS: records all actions related to Storage Service (ESS) instances, including creating, changing, deleting, querying, and enabling or pausing ESS instances. • MAINTENANCE: records actions related to the operations side. • MEMCACHE: records all actions related to KVStore for Memcache instances, including creating, changing, querying, and releasing KVStore for Memcache instances. • NAS: records all actions related to Network Attached Storage (NAS), including creating and deleting file systems.

Field	Description
	• KMS: records all actions related to Key Management Service (KMS), including creating, deleting, enabling, and disabling master keys of users. • DNS: records all actions related to Domain Name System (DNS), including creating, querying, and deleting domain names. • MONGODB: records all actions related to ApsaraDB for MongoDB instances, including creating, querying, and deleting ApsaraDB for MongoDB instances. • RAM: records all actions related to Resource Access Management (RAM), including creating, querying, and deleting server roles. • DFS: records all actions related to Distributed File System (DFS) instances, including creating, changing, deleting, and querying DFS instances. • CLOUD_API: records all actions related to instances in four modules (grouping, API, application, and plug-in) of API Gateway, including creating , changing, deleting, querying, binding, unbinding, releasing, authorizing instances, and bringing instances offline respectively. • AUTH: records all actions related to user roles, including adding and deleting user roles. • USER: records user activities, including logon time and logoff time. • PROJECT: records all actions related to projects, including creating, updating, querying, and deleting projects, and adding and deleting project members. • DEPARTMENT: records all actions related to departments, including creating, modifying, and deleting departments. • LOGINPOLICY: records all actions related to logon polices, including creating, changing, and deleting logon policies. • VISITCONFIG: records all actions related to access control, including changing access control. • REGION: records all actions related to regions. • INVENTORY: records actions related to inventory management on the operations side. • LOG: records actions related to operation logs, including obtaining operation log list. • ROLE: records all actions related to roles, including creating, deleting, and authorizing roles. • MAINTENANCEALARM: records actions related to monitoring alarms on the operations side. • TASKCENTER: records all actions related to the task center. • REPORT: records all actions related to reports, including previewing and downloading reports.

Field	Description
	• ALARM: records all actions related to monitoring alerts, including obtaining monitoring data and alert contact information. • BCSG: records all actions related to block storage gateway, including creating, deleting, and querying cloud resources, volumes, and storage groups. • IMAGE: records all actions related to images, including creating, changing, deleting, querying, and sharing images.
Region	The region in which the operation object resides
Level	The operation level, including Information, Notification, Warnings, Error, Important, Emergency, Alerts, and Debug
Actions	The operation type, including logon, logoff, and display
View Details	Brief introduction to the operation objective

4. Optional: Click **Export** to export the current logs to your local computer as an .xls file.

   The exported file is named *log.xls* and stored in *C:\Users\Username\Downloads*.

7.6.2 Delete logs

The administrator can delete logs within a specific time period if they are no longer in use.

Context

Note:
Logs cannot be recovered after being deleted, so proceed with caution.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Administration** > **Operation Log**.
3. On the **Operation Log** page, configure the query conditions and then click **Search**.
4. Click **Delete Log** to delete logs within a specific time period.

8 RAM management

8.1 Overview

Resource Access Management (RAM) is a resource access control service provided by Alibaba Cloud Apsara Stack.

RAM allows you to manage users (such as employees, systems, and applications) in a centralized way and control permissions to allow users to access specific resources under your name.

RAM has the following two functions:

- RAM role

  You must create a corresponding RAM role to authorize cloud services in a level-1 department to use or view other resources of the current department. This role contains the operations that cloud services can perform on resources.

  Only the system administrator and level-1 department manager can create RAM roles.

- RAM user

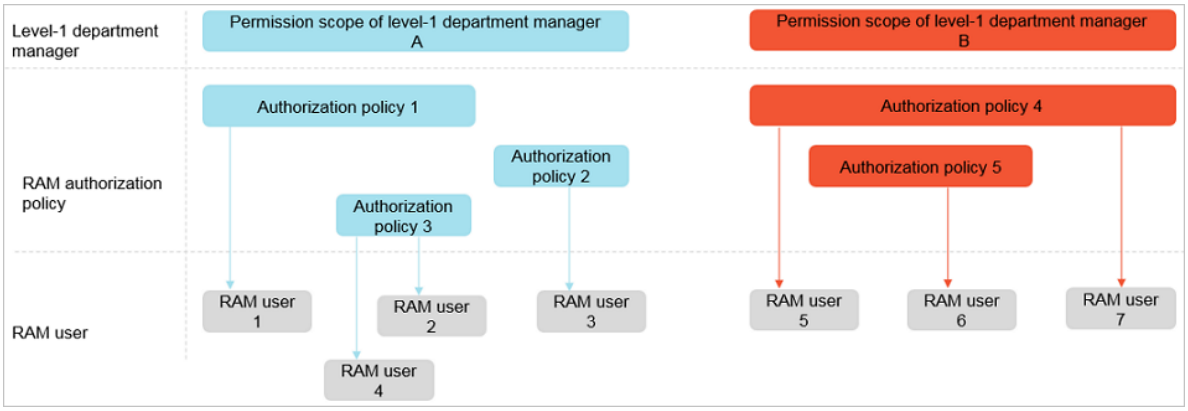
  To allow multiple users to use cloud resources in the same department, you can create multiple RAM users (users with the developer sub-account role) in the department to allow multiple RAM users to have different permissions to the same cloud resource.

  You can create RAM authorization policies to grant different permissions to different RAM users.

  RAM users can be considered as sub-accounts of the creator. A RAM user comes from the creator account based on the authorization policies, and the permission scope of a RAM user is smaller than or equal to that of the creator account.

  The RAM users are created by the system administrator or level-1 department manager.

Figure 8-1: RAM users and RAM authorization policies



8.2 RAM roles

8.2.1 View a role policy

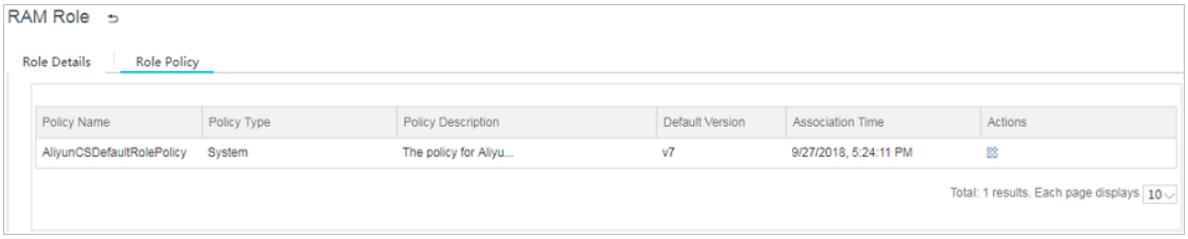
The administrator can view a role policy to learn about the detailed authorization of a role.

Prerequisites

A RAM role is created. For more information, see *Create a RAM role*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Resource Access Management**.
3. On the **RAM Role** tab, find the role whose policy you want to view. Click the icon in the **Actions** column and select **View Details**.
4. On the **RAM Role** page, click the **Role Policy** tab.



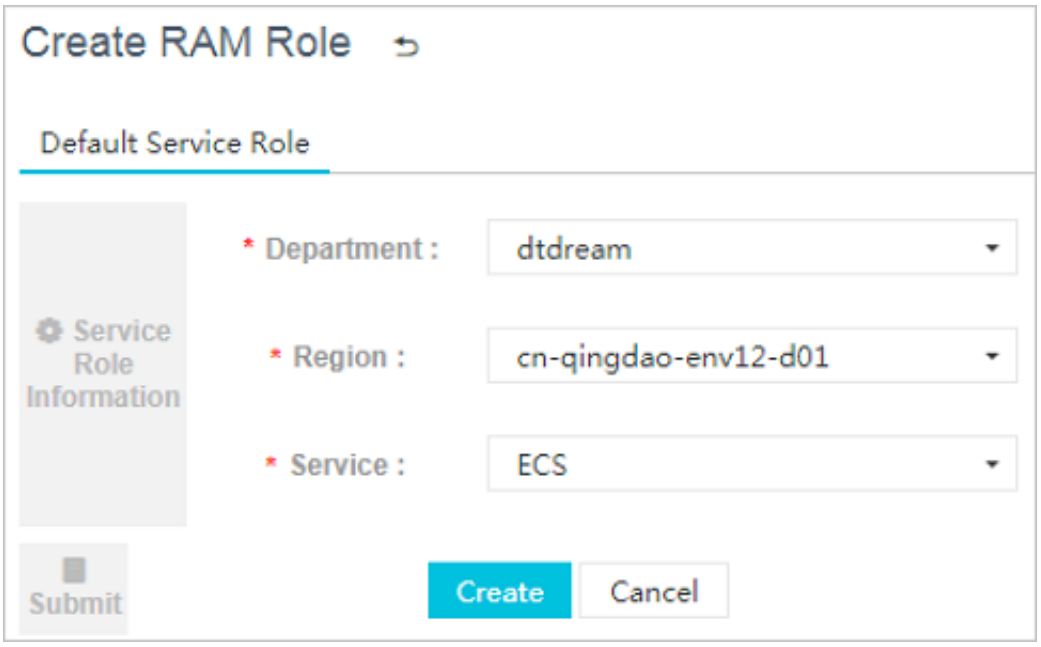
5. In the policy list, find the policy that you want to view. Click the icon in the **Actions** column and select **View Details** to view the policy details, namely the name, description, type, and contents of the policy.

8.2.2 Create a RAM role

To allow a cloud service to access other cloud resources, you must create the corresponding RAM role of this cloud service in the level-1 department.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator or level-1 department manager.
2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Resource Access Management**.
3. On the **RAM Role** tab, click **Create RAM Role**.



On the **Create RAM Role** page, select the level-1 department, region, and cloud service to be authorized. Then, click **Create** to complete the authorization.

4. The created RAM role appears in the RAM role list.

For more information about the RAM roles that can be created and their role names, see *Mapping between RAM roles and services*.

Table 8-1: Mapping between RAM roles and services

Role name	Service	Role description
AliyunECSImageExport DefaultRole	Elastic Compute Service	Used to grant ECS to use this role to export images
AliyunECSImageImport DefaultRole		Used to grant ECS to use this role to import images

Role name	Service	Role description
AliyunESSDefaultRole	Auto Scaling Service	Used to grant ESS to use this role to access resources of other cloud products

For example, select A as the **Region**, B as the **Department**, and **ECS** as the **Service** to create a RAM role. After the RAM role is created, ECS in region A, department B and its sub-departments can use the created RAM role to access resources of other cloud products in region A, department B and their sub-departments.

8.2.3 View role details

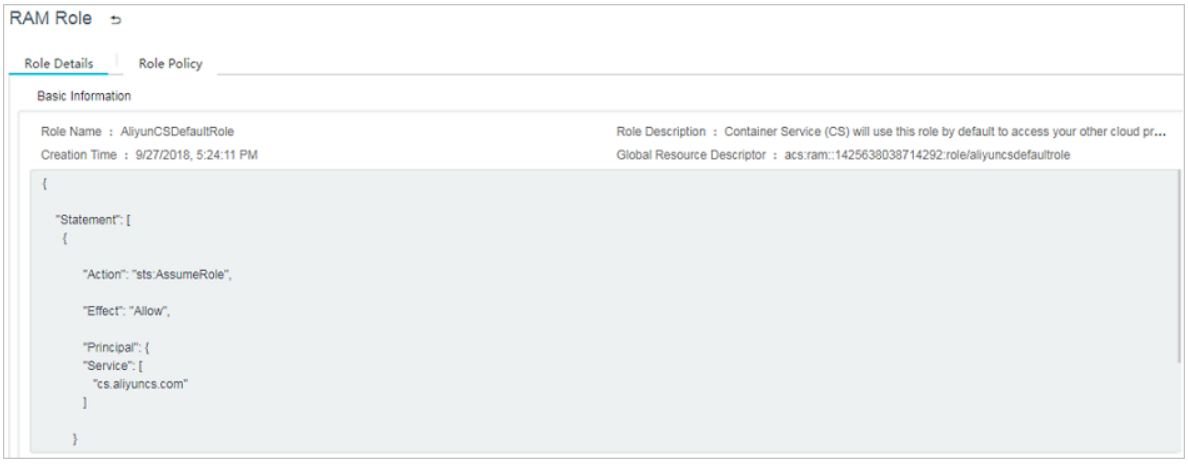
The administrator can view the role details to learn about the name, description, created time, and global resource descriptor of the role.

Prerequisites

A RAM role is created. For more information, see *Create a RAM role*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Resource Access Management**.
3. On the **RAM Role** tab, find the role that you want to view. Click the icon in the **Actions** column and select **View Details**.



On the **Role Details** tab, view the role details, namely the name, description, created time, and global resource descriptor of the role.

8.3 RAM users

8.3.1 Create a RAM user

To allow multiple users to use cloud resources in the same department, you can create multiple RAM users in the department.

Prerequisites

Before creating a RAM user, make sure that a department is created. For more information, see *Create a department*.

Context

RAM users are the operation objects who have specific access permissions to cloud resources. Currently, this function only applies to Object Storage Service (OSS) and Table Store.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Click **Add**. The **Add User** dialog box appears.
5. Configure the user information. Parts of the configurations are as shown in *Table 8-2: RAM user configurations*.

Table 8-2: RAM user configurations

Configuration	Description
Username	The name must be 3 to 30 characters in length and can contain letters, numbers, hyphens (-), underscores (_), and at signs (@). It must start with a letter or a number.
Display Name	The name must be 2 to 30 characters in length and only contain letters.
Role	Select **Developer Subaccount**. If you select other roles, the RAM user cannot be queried.
Department	Select the department to which the user belongs.
Logon Policy	The logon policy restricts the time period and IP addresses for the user to log on. By default

Configuration	Description
	, the default policy is automatically bound to newly created users. **Note:** The default policy does not restrict the time period and IP addresses for users to log on. To restrict the logon time and IP addresses of a user, you can change the user's logon policy or create a logon policy for the user. For more information, see *Create a logon policy*.

6. Click **OK** to create a RAM user.

8.3.2 View RAM user details

The administrator can view the details of a RAM user that they have created.

Prerequisites

A RAM user is created. For more information, see *Create a RAM user*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Resource Access Management**.
3. On the **Resource Access Management** page, click the **RAM User** tab.
4. Optional: Select the department where the RAM user resides from the **Department** drop-down list and enter the username in the **Username** field. Click **Search** to query the RAM user.
5. Find the RAM user that you want to view. Click the icon in the **Actions** column and select **View Details**.

   On the **User Details** page, view the detailed information of the RAM user, including the username, UID, contact information, and created time.

8.3.3 Change the description of a RAM user

The administrator can change the description of a created RAM user for better management.

Prerequisites

A RAM user is created. For more information, see *Create a RAM user*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Resource Access Management**.
3. On the **Resource Access Management** page, click the **RAM User** tab.
4. Optional: Select the department where the RAM user resides from the **Department** drop-down list and enter the username in the **Username** field. Click **Search** to query the RAM user.
5. Find the RAM user whose description you want to change. Click the  icon in the **Actions** column and select **Change**.
6. In the displayed dialog box, enter the description and then click **OK**.

8.3.4 Grant permissions to a RAM user

The administrator can bind created authorization policies to a RAM user based on the business needs.

Prerequisites

- A RAM user is created. For more information, see *Create a RAM user*.
- A RAM authorization policy is created. For more information, see *Create a RAM authorization policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Resource Access Management**.
3. On the **Resource Access Management** page, click the **RAM User** tab.
4. Optional: Select the department where the RAM user resides from the **Department** drop-down list and enter the username in the **Username** field. Click **Search** to query the RAM user.
5. Find the RAM user that you want to grant permissions. Click the  icon in the **Actions** column and select **Authorize**.
6. In the displayed dialog box, select the authorization policies from the **Available Authorization Policies** and then click  to add them to the **Authorization Policies Selected** .

7. Click **OK**.

8.4 RAM authorization policies

8.4.1 Create a RAM authorization policy

The administrator can create authorization policies as required to grant these authorization policies to RAM users.

Context

RAM authorization policies are the implementations of RAM user permissions. RAM users obtain the permissions by binding RAM authorization policies. Currently, this function only applies to Object Storage Service (OSS) and Table Store.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Resource Access Management**.
3. On the **Resource Access Management** page, click the **RAM Authorization Policy** tab.
4. Click **Create Authorization Policy**. The **Set Basic Information** step appears.
5. Configure the basic information of the authorization policy.

   For more information about the configurations, see *Authorization policy configurations*.

   Table 8-3: Authorization policy configurations

Configuration	Description
Policy Name	The RAM authorization policy name, which must be 1 to 128 characters in length and can contain letters, numbers, and hyphens (-), but must not start with dtrole
Region	The region to which the RAM authorization policy belongs
Department	The department to which the RAM authorization policy belongs
Project	The project to which the RAM authorization policy belongs
Product	The product of the RAM authorization policy. Currently, only OSS and Table Store are supported.
Policy Description	The description of the RAM authorization policy

6. Click **Next** to go to the **Add Rules and Conditions** page.

7. Configure the contents of the authorization policy.

   For more information about the configurations, see *Content configurations*.

Table 8-4: Content configurations

Configuration	Description
Effect	The authorization effectiveness includes **Allow** and **Deny**.
Action	The operations performed on specific resources For example, the access policy allows user A to perform the GetBucket operation on the resource SampleBucket. The operation is GetBucket.
Resource	The resource is the specific object that is authorized. For example, the access policy allows user A to perform the GetBucket operation on the resource SampleBucket. The resource is SampleBucket.
Condition	Composed of one or more condition clauses. A condition clause is composed of the keyword, operation type, and value.
Keywords	The keyword of the condition. For more information, see *Keywords*.
Operation Type	The operation type of the condition The following operation types are supported: • String • Date and time • Boolean • IP address For the methods that each operation type supports, see *Methods supported by operation types*.
Value	The value of the condition

Table 8-5: Keywords

Keyword	Type	Description
acs:CurrentTime	Date and time	The time when the Web server receives the request, which is in the format of ISO 8601. For example, 2012-11-11T23:59:59Z.

Keyword	Type	Description
acs:SecureTransport	Boolean	Whether the secure channel (for example, HTTPS) is used for sending the request
acs:SourceIp	IP address	The client IP address when the request is sent
oss:Delimiter	String	The delimiter used by OSS to divide object names into groups
oss:Prefix	String	The prefix of the OSS object name

Table 8-6: Methods supported by operation types

Operation type	Method
String	• StringEquals • StringNotEquals • StringEqualsIgnoreCase • StringNotEqualsIgnoreCase • StringLike • StringNotLike
Current time	• DateEquals • DateNotEquals • DateLessThan • DateLessThanEquals • DateGreaterThan • DateGreaterThanEquals
Date and time	• DateEquals • DateNotEquals • DateLessThan • DateLessThanEquals • DateGreaterThan • DateGreaterThanEquals
Boolean	Bool
IP address	• IpAddress • NotIpAddress

8. Click **Add Condition** and then click **Add Rules**.

   The authorization policy details are automatically generated in the **Authorization Policy** field.

9. Click **OK** to create the RAM authorization policy.

8.4.2 View RAM authorization policy details

The administrator can view the RAM authorization policy details to learn about the name, department, region, and created time of that authorization policy.

Prerequisites

A RAM authorization policy is created. For more information, see *Create a RAM authorization policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Resource Access Management**.
3. On the **Resource Access Management** page, click the **RAM Authorization Policy** tab.
4. Optional: Select the department and region where the RAM authorization policy resides from the **Department** and **Region** drop-down lists and enter the authorization policy name in the **Policy Name** field. Click **Search** to query the RAM authorization policy.
5. Find the RAM authorization policy that you want to view. Click the  icon in the **Actions** column and select **View Details**.

   On the **Policy Details** page, view the authorization policy details, namely the name, type, version number, the number of times that the authorization policy is referenced, created time, and description.

8.4.3 Delete a RAM authorization policy

The administrator can delete a RAM authorization policy that is no longer in use.

Prerequisites

A RAM authorization policy is created. For more information, see *Create a RAM authorization policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.

2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Resource Access Management**.
3. On the **Resource Access Management** page, click the **RAM Authorization Policy** tab.
4. Optional: Select the department and region where the RAM authorization policy resides from the **Department** and **Region** drop-down lists and enter the authorization policy name in the **Policy Name** field. Click **Search** to query the RAM authorization policy.
5. Find the RAM authorization policy to be deleted. Click the  icon in the **Actions** column and select **Delete**.
6. In the displayed dialog box, click **OK**.

9 System maintenance

9.1 Department management

9.1.1 Create a department

The administrator can create a department to store projects and resources in the projects.

Context

After the Apsara Stack console is deployed, a root department is created by default. You can create departments under the root department. The departments appear hierarchically and you can add sub-departments under each level of the department.

Departments added under the root department are level-1 departments and departments added under the level-1 departments are level-2 departments. Other departments are added in a similar way. In the Apsara Stack console, the sub-departments of a department refer to departments of all levels under the department.

Departments reflect the tree structure of an enterprise or business unit. A user can only belong to one department.

You can create a department under an existing department. The created department is a sub-department of the existing department.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Department Management**.
3. Select a department and click **Add Department**.
4. In the displayed **Add Department** dialog box, enter the department name.

   The name must be 2 to 50 characters in length and can contain letters and numbers.
5. Click **OK**.

9.1.2 Change the department name

If the department information is changed, the administrator can change the department name.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Department Management**.

3. Select the department whose name you want to change and click **Change Department**.
4. In the displayed dialog box, change the department name and click **OK**.

9.1.3 View projects of a department

The administrator can view projects of a department to learn about the project information in the department.

Context

Departments reflect the tree structure of an enterprise or business unit. A department can have multiple projects.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Department Management**.
3. Select the department that you want to view.

   Projects of this department appear in the right list.

9.1.4 Obtain the AccessKey of a department

The administrator can obtain the department AccessKey.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Department Management**.
3. Select a department and click **Get AccessKey** to obtain the account name, AccessKey, and PrimaryKey of the department.

   Note:
   The system automatically allocate the Apsara Stack account name, AccessKey, and PrimaryKey to the level-1 department. The sub-departments use the same account name, AccessKey, and PrimaryKey as their level-1 department.

9.1.5 Delete a department

The administrator can delete a department that is no longer in use.

Prerequisites

Note:

Make sure that the department to be deleted does not contain any users, projects, or sub-departments. Otherwise, the department cannot be deleted.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Department Management**.
3. Select the department to be deleted and click **Delete Department**.

9.2 Project management

9.2.1 Create a project

You must create a project before applying for resources.

Prerequisites

Make sure that a department is created before creating a project. For more information, see *Create a department*.

Context

You can create at most 20 projects under each level-1 department.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Project Management**.
3. On the **Project Management** page, click **Add Projects**.
4. In the displayed **Add Projects** dialog box, enter the project name and select a department to which the project belongs.
5. Click **OK**.

9.2.2 Add a project member

You can add a member to a project to allow the member to use the resources of the project.

Prerequisites

Before adding a project member, make sure that:

- A project is created. For more information, see *Create a project*.
- A user is created. For more information, see *Create a user*.

Context

The members of a project have the permissions to use resources of the project.

Deleting resources from a project does not affect the members of the project. Similarly, deleting members from a project does not affect the resources of the project.

You can delete the project members that are no longer in use. A deleted project member cannot access the resources of the project.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Project Management**.
3. Find the project that you want to add members. Click the  icon in the **Actions** column and select **View Details**.
4. Click the **Project Members** tab.
5. Click **Add Members**.
6. In the displayed **Add Project Members** dialog box, select a department and the corresponding project members.
7. Click **OK**.

   To remove one or more members from the project, complete the following steps:

   a. Select one or more members and click **Delete Members**.
   b. In the displayed **Delete Members** dialog box, select **Yes**.
   c. Click **OK**.

Result

The project member is added. You can view information about this member in the project member list.

9.2.3 Change the project name

If the project information is changed, the administrator can change the project name.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Project Management**.
3. Find the project that you want to change the name. Click the  icon in the **Actions** column and select **Change Project Name**.

4. In the displayed dialog box, change the project name and click **OK**.

9.2.4 View project details

The administrator can view project details to learn about the basic information of a project, including the name, ID, department, created time, and headcount.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Project Management**.
3. Find the project that you want to view. Click the  icon in the **Actions** column and select **View Details**.
4. On the **Project Details** page, view the project details.

9.2.5 View project members

To use resources of a project, you must be a member of the project. Check if you are in the member list of the project.

Context

The members of a project have the permissions to use resources of the project.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Project Management**.
3. Find the project. Click the  icon in the **Actions** column and select **View Details**.
4. On the **Project Details** page, click the **Project Members** tab. You can view all the members of the project and their contact information.

9.2.6 View resource information of a project

If you want to use certain cloud resources, you can view the resource information of a project in the project resource list.

Context

The project resource list displays all cloud resources of the project.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.

2. In the top navigation bar, choose **Console** > **User Center** > **Project Management**.
3. Find the project. Click the  icon in the **Actions** column and select **View Details**.
4. On the **Project Details** page, click the **Project Resources** tab.
5. On the **Project Resources** tab, view all cloud resources of the project.
6. Click the tab of a cloud product.
7. Find the resource. Click the  icon in the **Actions** column and select **View Details** to view the resource details.

9.2.7 Release resources

The administrator can release the resources that are no longer in use in a project.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Project Management**.
3. Find the project. Click the  icon in the **Actions** column and select **View Details**.
4. Select the **Project Resources** tab and then:
   - Release a single resource.

     Click the tab of a cloud product. Find the resource to be released. Click the  icon in the **Actions** column and select **Release Resources**. In the displayed dialog box, click **OK**.
   - Release multiple resources.

     Click the tab of a cloud product. Select multiple resources to be released, and then click **Delete** in the upper-right corner.

9.2.8 Delete a project

If a project is finished or changed, the administrator can delete the project that is no longer in use.

Prerequisites

 **Note:**

Make sure that the project to be deleted does not contain any resources or project members.

Context

Note:
If a project contains resources or project members, it cannot be deleted.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Project Management**.
3. Find the project to be deleted. Click the icon in the **Actions** column and select **Delete Project**.
4. Click **OK**.

9.3 Role management

9.3.1 Default roles

The system provides six default roles.

Role name	Description
Super Administrator	Initializes system and deploys cloud
System Administrator	Has the (read and write) permission to manage all resources on the cloud platform and on-site
Global Resource Inspector	Has the (read-only) permission to view all resources on the cloud platform
Department Manager	Has the permission to add, delete, and modify users in the department that the role belongs to
Resource Inspector	Has the (read-only) permission to view resources in the department that the role belongs to on the cloud platform
Resource User	Has the (read and write) permission to manage resources that belong to the role on the cloud platform

9.3.2 Add a custom role

The administrator can add custom roles in the Apsara Stack console to better assign permissions to users.

Context

A role is a collection of access permissions. Each role corresponds to a range of permissions. A user can have multiple roles, which means that this user has all the permissions defined in these roles. You can use a role to grant the same permissions to a group of users.

The system has six roles by default on the user side. The super administrator initializes system information and creates system administrators. Both system administrators and department managers are administrators, and the rest of default roles are normal users.

Before adding a custom role, note that:

- To add or change users, you must have the permissions to manage users and projects.
- To create Virtual Private Cloud (VPC)-related resources, you must have the permissions to view VPC instances, users, and projects, and permissions to manage users and projects.
- To create alert rules, you must have the permissions to manage the CloudMonitor Center.
- The total number of custom and default roles cannot exceed 20.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Role Management**.
3. Click **Add Role**. The **Add Roles** dialog box appears.

   Complete the configurations. For more information, see *Role configurations*.

   Table 9-1: Role configurations

Configuration	Description
Role Name	The name of the role, which must be 1 to 15 characters in length and contain letters or numbers
Description	(Optional) The description of the role, which must be 0 to 100 characters in length and contain letters, numbers, commas (,), semicolons (;), periods (.), underscores (_), and spaces
Permission Scope	• **Department** The permissions apply to all departments of the corresponding modules. • **Current Department/Subdepartments** The permissions apply to the department to which the user belongs and its subdepartments. • **Project**

Configuration	Description
	The permissions apply to the projects to which the user is added.
Select Permissions	Specify the operation permissions to cloud products To select all the permissions of a specific module, double-click the module. To select all the permissions of all modules, click **Import**.

4. Click **OK**.

9.3.3 View role details

The administrator can view permissions of a role on the **Role Management** page.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Role Management**.
3. Find the role whose permissions you want to view. Click the  icon in the **View Permissions** column. View the permissions of this role in the displayed dialog box.

9.3.4 Change a custom role

The administrator can change the description and permissions of a custom role.

Context

Note:
System default roles cannot be changed.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Role Management**.
3. Find the role to be changed. Click the  icon in the **Actions** column and select **Change**.
4. In the displayed dialog box, change the description, permission scope, and permission list of the role.
5. Click **OK**.

9.3.5 Delete a custom role

The administrator can delete a custom role that is no longer in use to manage roles better.

Context

> **Note:**
> System default roles cannot be deleted.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Role Management**.
3. Find the role to be deleted. Click the icon in the **Actions** column and select **Delete**.

   The **Confirm Deletion** dialog box appears.
4. Click **OK**.

9.4 User management

9.4.1 Create a user

The administrator can create users and assign roles to users to meet their access control requirements on the system.

Prerequisites

Before creating a user, make sure that:

- A department is created. For more information, see *Create a department*.
- A custom role is created if you want to customize the role. For more information, see *Add a custom role*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Click **Add**. The **Add User** dialog box appears.
5. Configure the user information.

Configuration	Description
Username	The cloud platform account name of the user. The name must be 3 to 30 characters in length and can contain letters, numbers, hyphens (-), underscores (_), and at signs (@). It must start with a letter or a number.
Display Name	The name must be 2 to 30 characters in length and can contain letters, numbers, hyphens (-), underscores (_), and at signs (@).
Role	Select a role for the user.
Department	Select the department to which the user belongs.
Logon Policy	Select a logon policy for the user. It restricts the time period and IP addresses for the user to log on. By default, the default policy is automatically bound to newly created users. **Note:** The default policy does not restrict the time period and IP addresses for users to log on. To restrict the logon time and IP addresses of a user, you can change the user's logon policy or create a logon policy for the user. For more information, see *Create a logon policy*.
Cell Phone Number	The mobile phone number of the user. It is used to notify the user of resource applications and usage by SMS. Make sure that the entered mobile phone number is correct. **Note:** If the number is changed, update it in time on the platform.
Landline	(Optional) The landline number of the user. It must be 4 to 20 characters in length and can contain numbers (0 to 9) and hyphens (-).
Email	The email address of the user. It is used to notify the user of resource applications and usage by email. Make sure that the entered email address is correct. **Note:** If the email address is changed, update it in time on the platform.

For the relationships among departments, users, and roles, see *Configuration of system initialization*.

6. Click **OK**.

9.4.2 View basic information of a user

The administrator can view the basic information of a user to learn about the department, role, and contact information of the user.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Find the user whose basic information you want to view. Click the  icon in the **Actions** column and select **User Information** to view the basic information of the user.

9.4.3 Change user information

If the user information is changed, the administrator can change the display name and contact information of the user.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Find the user to be changed. Click the  icon in the **Actions** column and select **Change**.
5. In the displayed **Change User** dialog box, change the display name and contact information of the user.

   For more information about how to change the personal information, see *Change personal information*.

9.4.4 Change the logon policy of a user

For better management, the administrator can change the logon policy of a user to change the permitted logon time and IP addresses for the user.

Prerequisites

A logon policy is created. For more information, see *Create a logon policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.

3. Click the **Users** tab.
4. Find the user. Click the  icon in the **Actions** column and select **Assign Logon Policy**.
5. In the displayed **Assign Logon Policy** dialog box, select the logon policy.
6. Click **OK**.

   After the logon policy of the user is changed, the user is limited by the new policy.

   If the user does not want to be limited by the bound logon policy, the user must submit an application to the administrator. After approving the application, the administrator binds a logon policy that meets the user's requirements to the user.

9.4.5 Change user roles

The administrator can change user roles by adding, changing, or deleting roles for a user.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Find the user to be changed. Click the  icon in the **Actions** column and select **Authorize**.
5. In the **Role** field, add, change, or delete roles for the user as required.
6. Click **OK**.

9.4.6 Authorize third-party access

To call APIs of the Apsara Stack console, the administrator must authorize third-party access to obtain the third-party AccessKey.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Find the user. Click the  icon in the **Actions** column and select **Authorize Third-Party Access**.
5. In the displayed dialog box, click **Authorize**.

Note:

Authorize Third-Party Access is enabled by default. You can click **Recreate an AccessKey** or **Remove the AccessKey** in the displayed dialog box.

To view the third-party AccessKey, see *View third-party AccessKey*.

9.4.7 Reset logon password

The administrator can reset the logon passwords for users if they forget their logon passwords.

Prerequisites

Only users who have the write permission to user management and project management can reset the logon passwords.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Find the user. Click the icon in the **Actions** column and select **User Information**.
5. On the **User Information** page, click **Reset Password**. The system automatically generates a new password and sends the new password to the user by SMS.
6. In the displayed **Reset Password** dialog box,
   - Click **Reset and Download**. The user password is reset to the initial password and locally downloaded in the TXT format.
   - Click **Reset Only**. The user password is reset to the initial password.

9.4.8 Export initial password

If **Reset Only** is selected when the logon password is reset, the administrator can export the initial user password and notify the user of the corresponding logon password orally.

Prerequisites

The password is reset. For more information, see *Reset logon password*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.

3. Click the **Users** tab.
4. Select the user whose initial password you want to export and click **Export Initial User Password**.

   The password file `UserInitPassword.txt` is generated.

9.4.9 Enable and disable a user

To prevent a user from logging on to the Apsara Stack console, the administrator can disable the user. A disabled user must be enabled before logging on to the Apsara Stack console.

Context

A user is activated by default after being created.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Perform the following operations:
   - Find an **Enabled** user. Click the  icon in the **Actions** column and select **Disable** to disable this user.
   - Find a **Disabled** user. Click the  icon in the **Actions** column and select **Enable** to enable this user.

9.4.10 Delete a user

The administrator can delete a user based on business requirements.

Prerequisites

The user has been removed from all projects. For more information, see *Add a project member*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Find the user. Click the  icon in the **Actions** column and select **Delete**.
5. In the displayed dialog box, click **OK**.

The deleted user still exists in the database, but does not belong to any department or have any role, and cannot log on to the Apsara Stack console.

9.4.11 Restore a user

After a user is deleted, the administrator can locate and restore the user in the **Deleted Users** list.

Context

Except for the department and role, the other basic information and the logon password of a recovered user are the same as those before the user was deleted.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Deleted Users** tab.
4. Find the user to be restored. Click the  icon in the **Actions** column and select **Restore**.

   The **Recover User** dialog box appears.
5. Select a department and a role and then click **OK**.

9.5 Logon policy management

9.5.1 Create a logon policy

The administrator can configure logon polices to control the logon address and time of users.

Context

A logon policy is used to control the logon time and address of users. After binding a logon policy to a user, the user can only log on to the Apsara Stack console within the time period and IP addresses configured in the logon policy.

A default logon policy is automatically generated when the Apsara Stack console provides services. This policy does not have any limits on the logon time and address, and cannot be deleted.

With the logon policies configured, users can access the Apsara Stack console at the permitted time and from permitted IP addresses. This improves the security of the console.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.

2. In the top navigation bar, choose **Console** > **User Center** > **Logon Policy Management**.
3. Click **Create Policy**.
4. In the displayed **Configure Policy** dialog box, enter the policy name, permitted logon time, and permitted logon address.

   Table 9-2: Logon policy configurations

Configuration	Description
Policy Name	The name must be 1 to 15 characters in length and contain letters or numbers, but must not be the same as an existing policy name. The name of the default policy cannot be changed.
Logon/Logoff Time	The permitted logon time is a time period. After being configured, users can only log on to the Apsara Stack console during the specified time period.
Client IP Address	The permitted logon address is an IP address Classless Inter-Domain Routing (CIDR) block. After being configured, users can only log on to the Apsara Stack console from the IP addresses within the specified CIDR block.

5. Click **OK**.
6. Optional: Bind a logon policy to a user. For more information, see *Change the logon policy of a user*.

   Note:

   - After binding a logon policy to a user, this user can only log on to the Apsara Stack console at the permitted time and from permitted IP addresses configured in the policy.
   - If the user does not want to be limited by the bound logon policy, the user must submit an application to the administrator. After approving the application, the administrator binds a logon policy that meets the user's requirements to the user.

What's next

After creating a logon policy, you can change or delete the existing logon policy.

- Find the logon policy to be changed. Click the icon in the **Actions** column and select **Change** to change the policy.

- Find the logon policy to be deleted. Click the  icon in the **Actions** column and select **Delete** to delete the policy.

  Note:
  You cannot delete the default logon policy.

9.5.2 View a logon policy

The administrator can view a logon policy to learn about the permitted logon time and IP addresses of a user.

Context

A default logon policy is automatically generated when the Apsara Stack console provides services. This policy does not have any limits on the logon time and IP addresses.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **Logon Policy Management**.
3. Optional: Enter the policy name in the search bar and click **Search**.

   The search result appears.
4. View the logon policy, namely the permitted logon time and IP addresses of users.

9.5.3 Bind a logon policy to multiple users

You can bind the same logon policy to different users. Users are limited by the logon policy when logging on to the Apsara Stack console.

Prerequisites

- Users are created. For more information about how to create users, see *Create a user*.
- A logon policy is created. For more information about how to create a logon policy, see *Create a logon policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the top navigation bar, choose **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Select multiple users and click **Assign Logon Policy** to bind a logon policy to multiple users.

9.6 Configure storage path for attachments

You can specify the storage path for uploaded attachments by configuring the system OSS.

Prerequisites

Before configuring the system OSS, select an OSS bucket as the system OSS and obtain the AccessKey ID and AccessKey Secret. AccessKey ID and AccessKey Secret are used to identify a visitor. The system uses AccessKey ID and AccessKey Secret to access OSS.

Obtain AccessKey ID and AccessKey Secret as follows:

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Object Storage Service**. In the bucket list, view the region and department to which the bucket belongs.
3. In the top navigation bar, choose **Console** > **User Center** > **Department Management**. In the department tree, find the region and department of the bucket. Select the department and then click **Get AccessKey**.

Context

By default, the storage path for attachments is not configured in the Apsara Stack console, and no attachment upload function is available. Configure the system OSS to specify the storage path for attachments to implement the high-reliability storage of large numbers of attachments.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon to the **System Configuration** page.
3. Click the **Storage Configuration** tab.
4. Set **Storage Type** to **OSS**.
5. Configure the system OSS.

   For more information about the configurations, see *System OSS configurations*.

   Table 9-3: System OSS configurations

Configuration	Description
OSS Endpoint	The endpoint address of OSS. Obtain the endpoint by viewing the bucket details.

Configuration	Description
Bucket Name	The name of the bucket.
AK ID and **AK Secret**	The keys used to access OSS. AccessKey ID is used to identify a user, and AccessKey Secret is a key used to authenticate a user.

6. Click **Save**.
7. Click **Test Connection**.

   To change the OSS configuration, click **Reset** and configure the system OSS again.

9.7 Configure ECS startup

On the **Resource Notification Configuration** tab, configure whether to automatically start the ECS instance after it is created.

Procedure

1. *Log on to the Apsara Stack console* as a system administrator.
2. In the upper-right corner, click the  icon.
3. On the **System Configuration** page, click the **Resource Notification Configuration** tab.
4. In the **ECS Startup Configuration** section, select the **Automatically start the ECS instance after it is created** check box.
5. Click **Save**.

   A system prompt appears, indicating the instance has been configured.

9.8 Configuration of system style

9.8.1 Configure the theme

You can change the theme of system as required.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the upper-right corner, click the  icon to go to the **System Configuration** page.
3. Click **Theme Configuration** tab.
4. Select the theme type as required, and then click **Save**.

10 Personal information management

10.1 Change personal information

If your personal information is changed, you can change the basic information of the personal information.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the upper-right corner, click your avatar and select **Personal Information**.
3. Click **Change** at the right of the item that you can change.
4. Change the information.
5. Then, click **Save**.

10.2 View AccessKey of your personal account

To guarantee the security of cloud resources, the system must verify the identity of the visitor to make sure the visitor has the related permissions. To access the cloud resources, you must obtain the AccessKey ID and AccessKey Secret of your personal account to authorize your logon.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the upper-right corner, click your avatar and select **Personal Information**.
3. In the **Alibaba Cloud AccessKey** section, view the AccessKey information of your personal account.

Note:

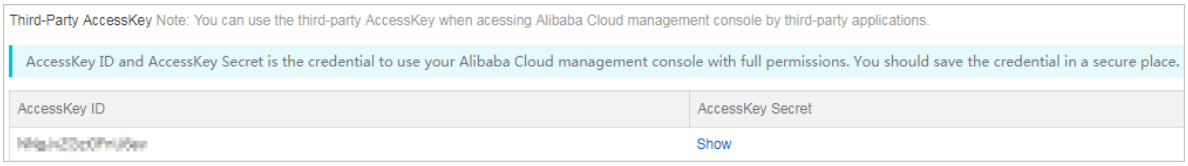
AccessKey ID and AccessKey Secret are keys for you to access the cloud resources with full permissions. Keep them properly.

10.3 View third-party AccessKey

If a third-party application calls the cloud control platform, you must obtain the AccessKey ID and AccessKey Secret used to authorize the logon of the third-party application.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the upper-right corner, click your avatar and select **Personal Information**.
3. In the **Third-Party AccessKey** section, view the third-party AccessKey information.



Note:
AccessKey ID and AccessKey Secret are keys for you to call the cloud control platform with full permissions. Keep them properly.

10.4 Change your avatar

You can change your avatar in the system by selecting a default avatar or uploading a custom avatar.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the upper-right corner, click your avatar and select **Personal Information**.
3. On the **Personal Information** page, click **Change Profile Picture** under the avatar.
4. In the **Change Profile Picture** dialog box, change your avatar.
   - Click the **Default Avatar** tab. Click the avatar and then click **OK**. The avatar is changed.
   - Click the **Custom Avatar** tab. Click **Upload Files**. Select the picture and then click **Open**. The picture appears in the preview section. Crop the picture as required. Then, click **OK**. The avatar is changed.

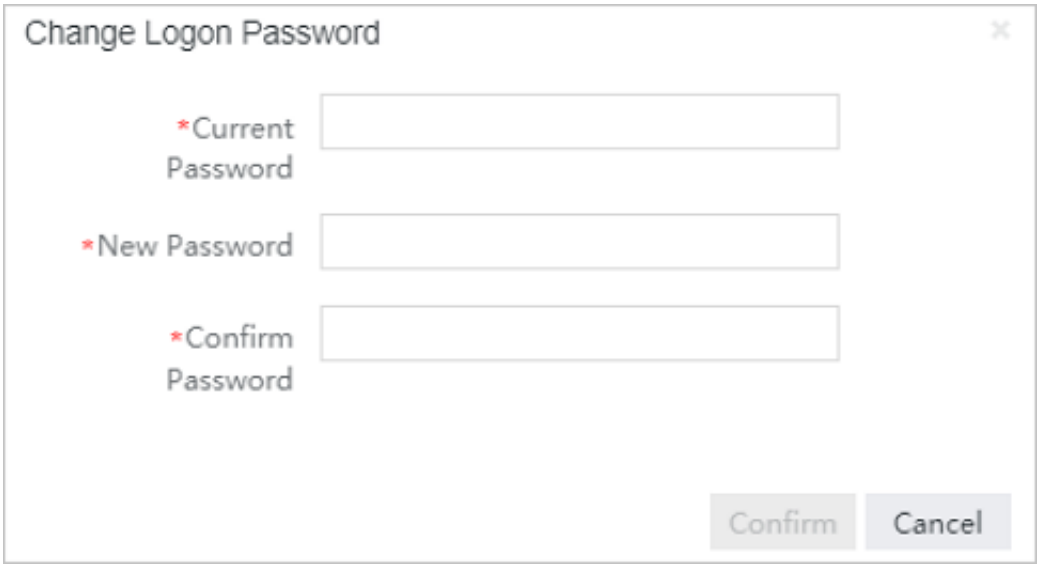
10.5 Change your logon password

To improve security, change your logon password in time.

Procedure

1. *Log on to the Apsara Stack console*.

2. In the upper-right corner, click your avatar and select **Personal Information**.
3. On the **Personal Information** page, click **Change Logon Password** under the avatar.
4. In the **Change Logon Password** dialog box, enter the current password, new password, and confirm password.



5. Then, click **OK**.

11 Elastic Compute Service (ECS)

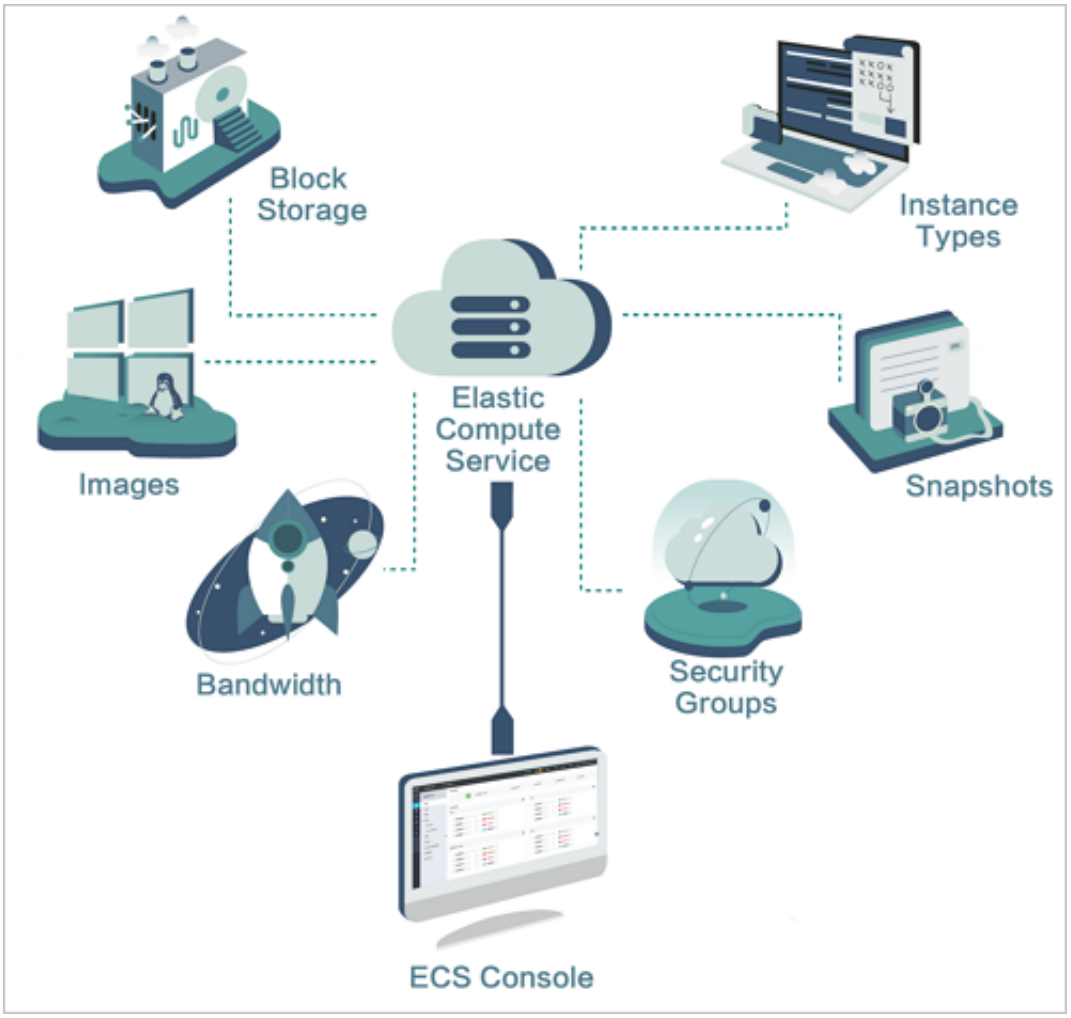
11.1 What is ECS

11.1.1 Overview

Elastic Compute Service (ECS) is a type of computing service that features elastic processing capabilities. As compared with the physical servers, ECS is more user-friendly and can be managed more efficiently. You can create instances, resize disks, and add or release any number of ECS instances any time according to your business demands.

As a virtual computing environment made up of the basic components such as CPU, memory, and storage, an ECS instance is provided by ECS for you to carry out relevant operations. It is the core concept of ECS and you can perform actions on ECS instances on the ECS console. As for other resources such as block storage, images, and snapshots, they cannot be used until being integraed with ECS instances. *Figure 11-1: Concept of an ECS instance* illustrates the services supported by an ECS instance.

Figure 11-1: Concept of an ECS instance



11.1.2 Instance types

Instance is the minimum unit for providing computing services, and its type reflects the computing capacity.

For an ECS instance, its type specifies two attributes, its CPU (including model and clock speed), and memory. To determine the scenario, however, you must select the image, disk, and network service at the same time.

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	ENI (including 1 primary elastic NIC)
n4	ecs.n4. small	1	2.0	N/A	0.5	50	1	1
	ecs.n4. large	2	4.0	N/A	0.5	100	1	1
	ecs.n4. xlarge	4	8.0	N/A	0.8	150	1	2
	ecs.n4. 2xlarge	8	16.0	N/A	1.2	300	1	2
	ecs.n4. 4xlarge	16	32.0	N/A	2.5	400	1	2
	ecs.n4. 8xlarge	32	64.0	N/A	5.0	500	1	2
mn4	ecs.mn4. small	1	4.0	N/A	0.5	50	1	1
	ecs.mn4. large	2	8.0	N/A	0.5	100	1	1
	ecs.mn4. xlarge	4	16.0	N/A	0.8	150	1	2
	ecs.mn4. 2xlarge	8	32.0	N/A	1.2	300	1	2
	ecs.mn4. 4xlarge	16	64.0	N/A	2.5	400	1	2
	ecs.mn4. 8xlarge	32	128.0	N/A	5.0	500	2	8
xn4	ecs.xn4. small	1	1.0	N/A	0.5	50	1	1
e4	ecs.e4. small	1	8.0	N/A	0.5	50	1	1
	ecs.e4. large	2	16.0	N/A	0.5	100	1	1

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	ENI (including 1 primary elastic NIC)
	ecs.e4. xlarge	4	32.0	N/A	0.8	150	1	2
	ecs.e4. 2xlarge	8	64.0	N/A	1.2	300	1	3
	ecs.e4. 4xlarge	16	128.0	N/A	2.5	400	1	8
sn1ne	ecs. sn1ne. large	2	4.0	N/A	1.0	300	2	2
	ecs. sn1ne. xlarge	4	8.0	N/A	1.5	500	2	3
	ecs. sn1ne. 2xlarge	8	16.0	N/A	2.0	1,000	4	4
	ecs. sn1ne. 3xlarge	12	24.0	N/A	2.5	1,300	4	6
	ecs. sn1ne. 4xlarge	16	32.0	N/A	3.0	1,600	4	8
	ecs. sn1ne. 6xlarge	24	48.0	N/A	4.5	2,000	6	8
	ecs. sn1ne. 8xlarge	32	64.0	N/A	6.0	2,500	8	8
sn2ne	ecs. sn2ne. large	2	8.0	N/A	1.0	300	2	2

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	ENI (including 1 primary elastic NIC)
	ecs. sn2ne. xlarge	4	16.0	N/A	1.5	500	2	3
	ecs. sn2ne. 2xlarge	8	32.0	N/A	2.0	1,000	4	4
	ecs. sn2ne. 3xlarge	12	48.0	N/A	2.5	1,300	4	6
	ecs. sn2ne. 4xlarge	16	64.0	N/A	3.0	1,600	4	8
	ecs. sn2ne. 6xlarge	24	96.0	N/A	4.5	2,000	6	8
	ecs. sn2ne. 8xlarge	32	128.0	N/A	6.0	2,500	8	8
	ecs. sn2ne. 14xlarge	56	224.0	N/A	10.0	4,500	14	8
se1ne	ecs. se1ne. large	2	16.0	N/A	1.0	300	2	2
	ecs. se1ne. xlarge	4	32.0	N/A	1.5	500	2	3
	ecs. se1ne. 2xlarge	8	64.0	N/A	2.0	1,000	4	4

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	ENI (including 1 primary elastic NIC)
	ecs.se1ne.3xlarge	12	96.0	N/A	2.5	1,300	4	6
	ecs.se1ne.4xlarge	16	128.0	N/A	3.0	1,600	4	8
	ecs.se1ne.6xlarge	24	192.0	N/A	4.5	2,000	6	8
	ecs.se1ne.8xlarge	32	256.0	N/A	6.0	2,500	8	8
	ecs.se1ne.14xlarge	56	480.0	N/A	10.0	4,500	14	8
se1	ecs.se1.large	2	16.0	N/A	0.5	100	1	2
	ecs.se1.xlarge	4	32.0	N/A	0.8	200	1	3
	ecs.se1.2xlarge	8	64.0	N/A	1.5	400	1	4
	ecs.se1.4xlarge	16	128.0	N/A	3.0	500	2	8
	ecs.se1.8xlarge	32	256.0	N/A	6.0	800	3	8
	ecs.se1.14xlarge	56	480.0	N/A	10.0	1,200	4	8
ebmg5	ecs.ebmg5.24xlarge	96	384.0	N/A	10.0	4,000	8	32

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	ENI (including 1 primary elastic NIC)
i2	ecs.i2.xlarge	4	32.0	1 * 894	1.0	500	2	3
	ecs.i2.2xlarge	8	64.0	1 * 1,788	2.0	1,000	2	4
	ecs.i2.4xlarge	16	128.0	2 * 1,788	3.0	1,500	4	8
	ecs.i2.8xlarge	32	256.0	4 * 1,788	6.0	2,000	8	8
	ecs.i2.16xlarge	64	512.0	8 * 1,788	10.0	4,000	16	8
d1	ecs.d1.2xlarge	8	32.0	4 * 5,500	3.0	300	1	4
	ecs.d1.3xlarge	12	48.0	6 * 5,500	4.0	400	1	6
	ecs.d1.4xlarge	16	64.0	8 * 5,500	6.0	600	2	8
	ecs.d1.6xlarge	24	96.0	12 * 5,500	8.0	800	2	8
	ecs.d1-c8d3.8xlarge	32	128.0	12 * 5,500	10.0	1,000	4	8
	ecs.d1.8xlarge	32	128.0	16 * 5,500	10.0	1,000	4	8
	ecs.d1-c14d3.14xlarge	56	160.0	12 * 5,500	17.0	1,800	6	8
	ecs.d1.14xlarge	56	224.0	28 * 5,500	17.0	1,800	6	8
scch5ib	ecs.scch5ib.16xlarge	64	192.0	N/A	10.0	4,500	8	32

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	ENI (including 1 primary elastic NIC)
sccg5ib	ecs. sccg5ib. 24xlarge	96	384.0	N/A	10.0	4,500	8	32
re5	ecs.re5. 15xlarge	60	990.0	N/A	10.0	1,000	16	8
	ecs.re5. 30xlarge	120	1,980.0	N/A	15.0	2,000	16	15
	ecs.re5. 45xlarge	180	2,970.0	N/A	30.0	4,500	16	15
sn1	ecs.sn1. medium	2	4.0	N/A	0.5	100	1	2
	ecs.sn1. large	4	8.0	N/A	0.8	200	1	3
	ecs.sn1. xlarge	8	16.0	N/A	1.5	400	1	4
	ecs.sn1. 3xlarge	16	32.0	N/A	3.0	500	2	8
	ecs.sn1. 7xlarge	32	64.0	N/A	6.0	800	3	8
sn2	ecs.sn2. medium	2	8.0	N/A	0.5	100	1	2
	ecs.sn2. large	4	16.0	N/A	0.8	200	1	3
	ecs.sn2. xlarge	8	32.0	N/A	1.5	400	1	4
	ecs.sn2. 3xlarge	16	64.0	N/A	3.0	500	2	8
	ecs.sn2. 7xlarge	32	128.0	N/A	6.0	800	3	8

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	ENI (including 1 primary elastic NIC)
	ecs.sn2.14xlarge	56	224.0	N/A	10.0	1,200	4	8

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	Instance type family	FPGA
f1	ecs.f1-c8f1.2xlarge	8	60.0	N/A	3.0	400	4	4	Intel ARRIA 10 GX 1150
	ecs.f1-c8f1.4xlarge	16	120.0	N/A	5.0	1,000	4	8	2 * Intel ARRIA 10 GX 1150
	ecs.f1-c28f1.7xlarge	28	112.0	N/A	5.0	2,000	8	8	Intel ARRIA 10 GX 1150
	ecs.f1-c28f1.14xlarge	56	224.0	N/A	10.0	2,000	14	8	2 * Intel ARRIA 10 GX 1150
f3	ecs.f3-c16f1.4xlarge	16	64.0	N/A	5.0	1,000	4	8	1 * Xilinx VU9P

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	Instance type family	FPGA
	ecs.f3-c16f1.8xlarge	32	128.0	N/A	10.0	2,000	8	8	2 * Xilinx VU9P
	ecs.f3-c16f1.16xlarge	64	256.0	N/A	20.0	2,500	16	8	4 * Xilinx VU9P

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	Instance type family	GPU
gn5	ecs.gn5-c4g1.xlarge	4	30.0	440	3.0	300	1	3	1 * NVIDIA P100
	ecs.gn5-c8g1.2xlarge	8	60.0	440	3.0	400	1	4	1 * NVIDIA P100
	ecs.gn5-c4g1.2xlarge	8	60.0	880	5.0	1,000	2	4	2 * NVIDIA P100
	ecs.gn5-c8g1.4xlarge	16	120.0	880	5.0	1,000	4	8	2 * NVIDIA P100
	ecs.gn5-c28g1.7xlarge	28	112.0	440	5.0	1,000	8	8	1 * NVIDIA P100
	ecs.gn5-c8g1.8xlarge	32	240.0	1,760	10.0	2,000	8	8	4 * NVIDIA P100

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	Instance type family	GPU
	ecs.gn5-c28g1.14xlarge	56	224.0	880	10.0	2,000	14	8	2 * NVIDIA P100
	ecs.gn5-c8g1.14xlarge	54	480.0	3,520	25.0	4,000	14	8	8 * NVIDIA P100
gn4	ecs.gn4-c4g1.xlarge	4	30.0	N/A	3.0	300	1	3	1 * NVIDIA M40
	ecs.gn4-c8g1.2xlarge	8	30.0	N/A	3.0	400	1	4	1 * NVIDIA M40
	ecs.gn4.8xlarge	32	48.0	N/A	6.0	800	3	8	1 * NVIDIA M40
	ecs.gn4-c4g1.2xlarge	8	60.0	N/A	5.0	500	1	4	2 * NVIDIA M40
	ecs.gn4-c8g1.4xlarge	16	60.0	N/A	5.0	500	1	8	2 * NVIDIA M40
	ecs.gn4.14xlarge	56	96.0	N/A	10.0	1,200	4	8	2 * NVIDIA M40
ga1	ecs.ga1.xlarge	4	10.0	1 * 87	1.0	200	1	3	0.25 * AMD S7150
	ecs.ga1.2xlarge	8	20.0	1 * 175	1.5	300	1	4	0.5 * AMD S7150

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	Instance type family	GPU
	ecs.ga1.4xlarge	16	40.0	1 * 350	3.0	500	2	8	1 * AMD S7150
	ecs.ga1.8xlarge	32	80.0	1 * 700	6.0	800	3	8	2 * AMD S7150
	ecs.ga1.14xlarge	56	160.0	1 * 1,400	10.0	1,200	4	8	4 * AMD S7150
gn5i	ecs.gn5i-c2g1.large	2	8.0	N/A	1.0	100	2	2	1 * NVIDIA P4
	ecs.gn5i-c4g1.xlarge	4	16.0	N/A	1.5	200	2	3	1 * NVIDIA P4
	ecs.gn5i-c8g1.2xlarge	8	32.0	N/A	2.0	400	4	4	1 * NVIDIA P4
	ecs.gn5i-c16g1.4xlarge	16	64.0	N/A	3.0	800	4	8	1 * NVIDIA P4
	ecs.gn5i-c16g1.8xlarge	32	128.0	N/A	6.0	1,200	8	8	2 * NVIDIA P4
	ecs.gn5i-c28g1.14xlarge	56	224.0	N/A	10.0	2,000	14	8	2 * NVIDIA P4

Instance type family	Instance type	vCPU (Core)	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)	NIC queues	Instance type family	GPU
gn5e	ecs. gn5e-c11g1. 3xlarge	10	58.0	N/A	2.0	150	1	6	1 * NVIDIA P4
	ecs. gn5e-c11g1. 5xlarge	22	116.0	N/A	4.0	300	1	8	2 * NVIDIA P4
	ecs. gn5e-c11g1. 11xlarge	44	232.0	N/A	6.0	600	2	8	4 * NVIDIA P4
	ecs. gn5e-c11g1. 22xlarge	88	464.0	N/A	10.0	1,200	4	15	8 * NVIDIA P4

The following instance types are only applicable for environments that upgrade from V2 to V3.

Instance type family	Instance type	vCPU (Core)	Memory (GiB)
n1	ecs.n1.tiny	1	1.0
	ecs.n1.small	1	2.0
	ecs.n1.medium	2	4.0
	ecs.n1.large	4	8.0
	ecs.n1.xlarge	8	16.0
	ecs.n1.3xlarge	16	32.0
	ecs.n1.7xlarge	32	64.0
n2	ecs.n2.small	1	4.0
	ecs.n2.medium	2	8.0
	ecs.n2.large	4	16.0
	ecs.n2.xlarge	8	32.0

Instance type family	Instance type	vCPU (Core)	Memory (GiB)
	ecs.n2.3xlarge	16	64.0
	ecs.n2.7xlarge	32	128.0
e3	ecs.e3.small	1	8.0
	ecs.e3.medium	2	16.0
	ecs.e3.large	4	32.0
	ecs.e3.xlarge	8	64.0
	ecs.e3.3xlarge	16	128.0
c1	ecs.c1.small	8	8.0
	ecs.c1.large	8	16.0
c2	ecs.c2.medium	16	16.0
	ecs.c2.large	16	32.0
	ecs.c2.xlarge	16	64.0
m1	ecs.m1.medium	4	16.0
	ecs.m1.xlarge	8	32.0
m2	ecs.m2.medium	4	32.0
s1	ecs.s1.small	1	2.0
	ecs.s1.medium	1	4.0
	ecs.s1.large	1	8.0
s2	ecs.s2.small	2	2.0
	ecs.s2.large	2	4.0
	ecs.s2.xlarge	2	8.0
	ecs.s2.2xlarge	2	16.0
s3	ecs.s3.medium	4	4.0
	ecs.s3.large	4	8.0
t1	ecs.t1.small	1	1.0

11.1.3 Instance lifecycle

The lifecycle of an instance begins with creation and ends with release. This section introduces such information of an instance as its status, status attributes, and corresponding API status.

Table 11-1: Lifecycle description shows the different states of an instance during its entire lifecycle.

Table 11-1: Lifecycle description

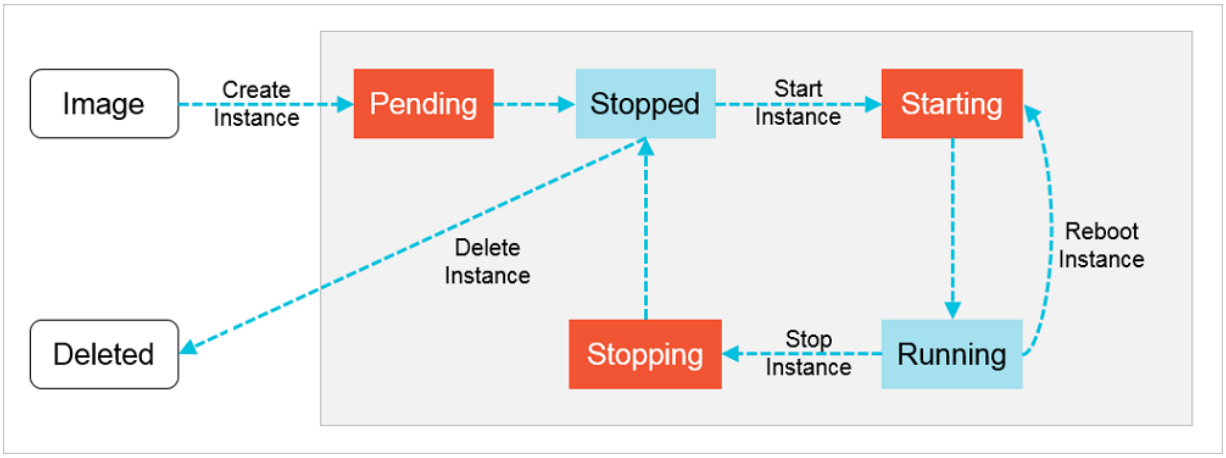
Status	Status attribute	Description	Corresponding API status
Creating task	Intermediate status	Instance creation in progress. Waiting for start. If an instance is in this status for a long time, an exception occurs.	**Pending**
Starting	Intermediate status	It is the state entered by an instance prior to the **Running** state before you perform a restart or start operation for that instance on the console or via an API. If an instance is in this status for a long time, an exception occurs.	**Starting**
Running	Stable status	Indicates that the instance is running smoothly. The instance in this state can accommodate your business needs.	**Running**
Stopping	Intermediate status	An instance is in this status after the Stop operation is performed on the console or using an API but before the instance enters the **Stop** state. If an instance is in this status for a long time, an exception occurs.	**Stopping**
Stop	Stable status	Indicates the instance has been stopped normally. In this status, the instance cannot accommodate external services.	**Stopped**
Re-initializing	Intermediate status	An instance is in this status after the system disk and/or data disk is re-initialized in the console or	**Stopped**

Status	Status attribute	Description	Corresponding API status
		using an API until it is **Running**. If an instance is in this status for a long time, an exception occurs.	
Replacing System Disk	Intermediate status	An instance is in this status after the operating system is replaced or another such operation is performed on the console using an API until, and before the instance enters the **Running** state. If an instance is in this status for a long time, an exception occurs.	**Stopped**

Table 11-1: Lifecycle description describes mappings between console statuses and API statuses.

The API status chart is shown in *Figure 11-2: API status chart*.

Figure 11-2: API status chart



11.2 Instructions before use

11.2.1 Overview

Learn about the precautions or restrictions before using ECS.

11.2.2 Prohibitions

Avoid the following don'ts when using an ECS instance.

- Do not upgrade the ECS kernel or operating system without prior authorization.
- Do not start SELinux for the other Linux systems except CentOS and RedHat.

- Do not detach PVDriver.
- Do not arbitrarily modify the MAC address of the network adapter.

11.2.3 Suggestions

To better use ECS, please consider the following suggestions.

- For ECS instances with memory above 4 GB, use a 64-bit operating system (a 32-bit operating system has a 4 GB limitation in memory addressing).
- A 32-bit Windows operating system supports CPUs with up to four cores.
- To guarantee service continuity and avoid service unavailability due to downtime migration, we recommend that you configure service applications to automatically start at system startup.

11.2.4 Restrictions

Learn about the restrictions on instance type families before using ECS.

General restrictions

- Windows does not support instance specifications higher than 64vCPU.
- Virtual application installation and subsequent virtualization (for example, using VMware) are not yet supported.
- Currently, ECS does not support sound card applications (only a GPU instance supports analog audio cards) and cannot connect to external hardware devices (such as hardware dongles, USB drives, external hard disks, and USB keys of banks).
- Currently, ECS does not support multicast protocols. If you want to use multicast services, we recommend that you use the unicast point-to-point method.

Instance type family ga1

To create instance type family ga1, you need to use the following images pre-installed with drivers :

- Ubuntu16.04 pre-installed with the AMD GPU driver
- Windows Server 2016 Chinese Edition pre-installed with the AMD GPU driver
- Windows 2008 R2 Chinese Edition pre-installed with the AMD GPU driver

Notes:

- The ga1 instance uses a driver optimized by Alibaba Cloud and AMD. The driver is included in the image provided by Alibaba Cloud. Driver download link is not provided and driver installation by the client is not supported.

- If the GPU driver fails to work properly because its related components are detached or removed, you need to restore GPU-related functions by *Change System Disks*.

  Note:
  Changing system disks may cause data loss.
- If the driver of a visual compute ga1 instance with GPUs fails to work properly because an incorrect image is selected, you need to reselect an image pre-installed with the AMD GPU driver through *Change System Disks*.
- If you use an image of Windows 2008 or earlier versions, the function of **Connect to Management Terminal** on the Alibaba Cloud console is unavailable after the installed GPU driver takes effect. The management terminal is unresponsive with a black screen or stuck at the startup interface. You can use other protocols to access the system, such as the remote desktop protocol (RDP) of Windows.
- The RDP of Windows does not support DirectX, OpenGL, and other related applications. You need to install VNC and a client or configure other supported protocols, such as PCOIP and XenDeskop HDX 3D.

Instance type family gn4

- Bandwidth: Select a bandwidth.

  Note:
  If you use an image of Windows 2008 R2, the function of **Connect to Management Terminal** on the Alibaba Cloud console cannot be used to connect to the gn4 instance after the installed GPU driver takes effect. Therefore, you need to set the bandwidth to a non-zero value or bind the created instance to an elastic public IP.
- Image: Select an image.

  If pre-installation of the NVIDIA GPU driver is unnecessary, you can select any image, and *Install the CUDA and GPU drivers for a Linux instance*.

Instance type families gn5i and gn5

- Bandwidth: Select a bandwidth.

  Note:

If you use an image of Windows 2008 R2, the function of **Connect to Management Terminal** on the Alibaba Cloud console cannot be used to connect to the gn5i or gn5 instance after the installed GPU driver takes effect.

- Image: Select an image.

  If pre-installation of the NVIDIA GPU driver is unnecessary, you can select any image, and *Install the CUDA and GPU drivers for a Linux instance*.

11.2.5 Precautions for using ECS instances in Windows

Consider the following precautions before using a Windows instance.

- If an instance uses a local disk as its data disk, there will be a risk of data loss. If you are unable to ensure the reliability of the data architecture, we strongly recommend you to use a cloud disk to establish your instance.
- Do not close the built-in shutdownmon.exe process. Otherwise, it may take longer to restart your Windows server.
- For normal use of the server, do not rename, delete, or disable administrator accounts.
- The use of virtual memory is not recommended.
- If you have changed your computer name, you must synchronize the related key values in the registry; otherwise, the computer name cannot be modified, causing a failure to install certain third-party programs. The following key values must be modified in the registry:

```
HKEY_LOCAL_MACHINE\SYSTEM\ControlSet001\Control\ComputerName\
ActiveComputerName
```

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\ComputerName\
ComputerName
```

11.2.6 Precautions for using ECS instances in Linux

Consider the following precautions before using a Linux instance.

- Do not modify the default /etc/issue file of a Linux instance. Otherwise, the system release version of the custom image created on the basis of the instance will be unidentifiable, and the instance created by using this image cannot be started.
- Do not change the permission of any directory in the partition in which the root directory resides , particularly, the permissions of /etc, /sbin, /bin, /boot, /dev, /usr, and /lib. Improper modification of permissions may cause exceptions.
- Do not rename, delete, or disable Linux root accounts.
- Do not compile or perform any other operations on the Linux kernel.

- The use of swap for partitioning is not recommended.
- Do not enable the NetWorkManager service. This service conflicts with the system's internal network service, causing network exceptions.

11.2.7 DDoS protection

To benefit from the DDoS protection capability, you need to purchase the Alibaba Cloud Security Advanced Edition. For details, see *Alibaba Cloud Security Overview*.

11.3 Quick start

11.3.1 Overview

This chapter introduces the preparations before using an ECS instance, such as logging on to the ECS console, creating a security group and creating instances.

Before using an ECS instance, do the following:

1. *Create a security group*

   A security group is a virtual firewall that controls the inbound and outbound traffic of an instance. An instance must belong to at least one security group. When creating instances, you need to select a security group for network access control. If no default security group exists, please create one in advance.

2. *Create an instance*

   An ECS instance is equivalent to a virtual machine, including the fundamental components like CPU, memory, operating system, network and disks. After creating a security group, you can select an *instance type* to finish instance creation.

3. *Connect to an instance*

   The way to connect to an instance is determined by the network settings and operating system of the ECS instance, as well as the operating system of your local device. After connecting to an instance, you can install your applications on it.

11.3.2 Log on to the ECS console

This section introduces how to log on to the ECS console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The

access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.

- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.
5. In the menu bar at the top of the page, select **Console** > **Compute, Storage&Networking** > **Elastic Compute Service**.

11.3.3 Create a security group

You must create a security group before you can create an ECS instance in a VPC. A security group is an important component of network security and isolation. It can be used to set network access control for one or more ECS instances.

Procedure

1. *Log on to the ECS console*.
2. Click the **Security Groups** tab. On the Security Groups tab page, click **Create Security Group**.
3. On the **Create Security Groups** page that appears, configure security group information.

   Table 11-2: Security group parameters describes the parameter configurations.

Table 11-2: Security group parameters

Area	Parameter	Description
Region	**Region**	Required. The region to which the security group belongs. It must be the same region as the VPC.
Basic Settings	**Department**	Required. The department to which the security group belongs. It must be the same department as the VPC.
	Project	Optional. The project for the security group to join.
	Network Type	Required. The network type of the security group. The default network type is **VPC**. It must be the VPC to which the security group belongs.
	Security Group Name	Required. The name of the security group.
	Description	Optional. The description of the security group.

4. Click **OK**.

11.3.4 Create an instance

Create an instance after setting up a security group.

Prerequisites

The following should be noted before creating an insance:

- Before creating an instance, complete the creation of a VPC and switch.
- Before you create an instance, check that a security group is available. If not, *Create a security group* first.
- Before creating a GPU instance, refer to *Restrictions*.

To create an instance, do the following:

Procedure

1. *Log on to the ECS console* and go to the **Instance** page.
2. Click **Create Instance**.
3. On the **Create Cloud Server (ECS)** page, complete the following configurations:

Table 11-3: Instance configuration

Item	Description
Region	• **Region**: select a region for the ECS instance to be created. • **Zone**: Zones are physical areas with independent power grids and networks within a region. Intranet communication and fault isolation are both enabled between different zones. If you want to improve application availability, we recommend that you create instances in different zones.
Configurations	• **Department**: select a department for the ECS instance. • **Project**: select a project for the ECS instance.
Network	• **Network Type**: It is a required parameter and set to VPC by default. Select a specific VPC and switch name. • **Security Groups**: It is a required parameter. **Note:** Before creating an ECS instance, you must create a security group. • **Configure Private IP**: It is an optional parameter. Determine an IP address segment based on the CIDR block where the switch is located. **Note:** — If it is null, the system will designate a private IP Address automatically. — If a private IP address is configured, instances cannot be created in batch.
Instances	• **Instance Series**: select **Series 3** or **Series 4** . • **I/O Optimized**: It is an optional parameter. By default, it is an I/O optimized instance. • **Instance Specifications**: It is an optional parameter. Select CPU and memory based on application requirements. Windows-based images are not applicable to some CPU and memory combinations. See *Suggestions*.
Images	**Image Type**: It is a required parameter. Select **Public Image** or **Custom Image** as the image type for your operating system.

Item	Description
Storage	• **`System Disk`**: It is a required parameter. Select an SSD cloud disk or ultra cloud disk as the system disk for installing the operating system. • **`Data Disk`**: It is an optional parameter. You can choose between SSD cloud disks and ultra cloud disks. Up to 16 data disks can be added. The maximum capacity of each data disk is 32 TB. You can select **Release with Instance** or **Encrypt**. **Note:** — If you check **`Release with Instance`**, this disk will be released together with the instance and the data is not recoverable; if Release with Instance is not checked, the system will detach this disk from the instance when this instance is released, and the disk will not be released. — If you check **`Encrypt`**, the disk created is an encrypted disk; if it is not checked, the disk created is a non-encrypted disk. — If you do not add data disks here, you can add them later by following the procedure described in *Create disks*.
Password	It is a required parameter. You can select **Configure Now** or **Configure After** to set your password on the console by using the password re-setting function. **Note:** **`Login Password`** can be 8 to 30 characters in length. It must contain at least three of the following character types: uppercase letters, lowercase letters, numbers, and special characters.
Deployment Set	Select the name of the deployment set to which the instance will be added.
Instance Name	It is an optional parameter. We recommend that you set a recognizable name for your instance. **Note:** The instance name can be 2 to 114 characters in length and can contain letters, Chinese characters, numbers, underscores (_), and hyphens (-). It must start with a letter or Chinese character.
Custom Data	You can enter the corresponding custom data encoding schemes. If the data to be entered adopt Base64 encoding schemes, check **Enter Base64 encoded finformation**.

Item	Description
	Note: Windows supports two formats: bat and powershell. Before Base64 codes, the first line is [bat] or [powershell]. Linux supports shell scripts.
Quantity	The default value is 1. If you apply for more than one instance, these instances are created in batch based on your current settings. **Note:** You can create up to 50 ECS instances in batch. If a private IP address is configured, batch creation is not supported.

4. Click **Create**.

Result

You can view the created instance in the instance list. Check whether the instance created is in the **`Running`** state. If so, the instance is successfully created.

11.3.5 Connect to an instance

11.3.5.1 Overview

After creating an ECS instance, you can connect to the instance and install application software in the instance.

Note:
The username is "root" in Linux and "administrator" in Windows.

You can connect to and manage your ECS instance in either of the following ways:

- Use a remote connection tool to connect to the ECS instance.

  Note:
  Check that your ECS instance has an elastic Internet IP address.

- Connect to the ECS instance by using the function of **Connect to Management Terminal** on the cloud console.

11.3.5.2 Connect to a Linux instance using the SSH command in Linux or Mac OS X

This section introduces how to connect to a Linux instance using the SSH command.

Prerequisites

Finish the creation of security groups and instances.

Procedure

1. Enter the following command: `ssh root@instance IP`.
2. Enter the password of the *root* user for this instance upon logon.

11.3.5.3 Connect to a Linux instance using a remote connection tool in Windows

This section introduces how to connect to a Linux instance using the PuTTY tool.

Prerequisites

The usage of different remote connection tools is similar. This article describes how to connect to a remote instance through PuTTY. Download PuTTY at *http://www.chiark.greenend.org.uk/~sgtatham/putty/* .

To connect to an instance, do the following:

Procedure

1. Download and install PuTTY for Windows.
2. Start the PuTTY client and complete the following settings:
   - **`Host Name (or IP address)`**: enter the public IP address for the instance.
   - **`Port`**: Set it to the default port number 22.
   - **`Connection Type`**: select **`SSH`**.
   - **`Saved Session`**: the name of the session. Click **Save**. After the settings are saved, PuTTY remembers the name and IP address of the instance. You do not have to enter the IP address every time you connect to the instance.
3. Click **Open** to connect to the instance.

   When you connect to the instance for the first time, a PuTTY Security Alert dialog box appears. Click **Yes**.
4. Enter the username **`root`** and press the **Enter** key.
5. Enter the password of your instance and press the **Enter** key.

If a message similar to the following appears, a connection is successfully established to the instance.

```
Welcome to aliyun Elastic Compute Server!
```

11.3.5.4 Connect to a Windows instance using the remote desktop connection function in Windows

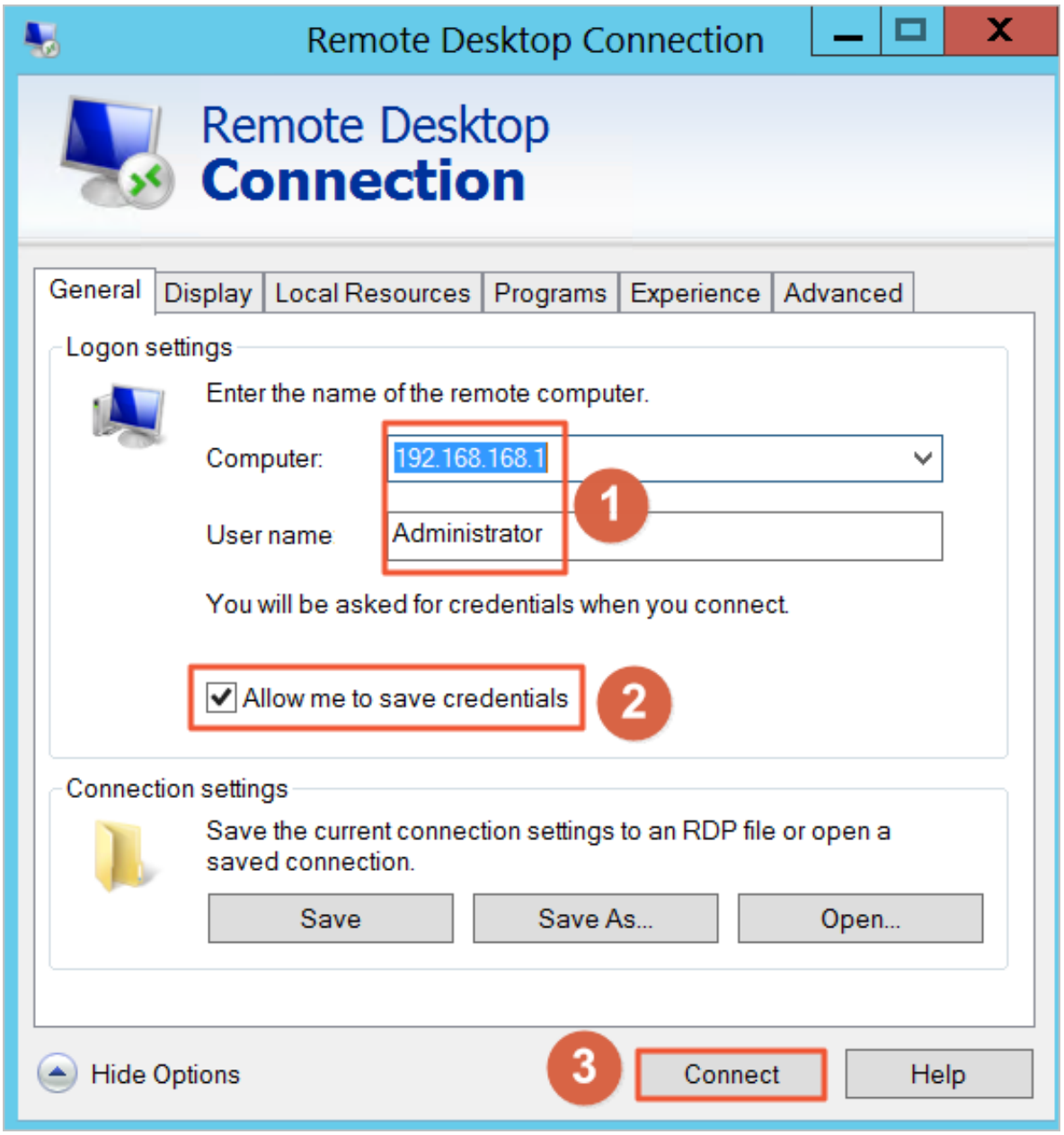
This article describes how to connect to a Windows instance through the remote desktop connection function of Windows.

Prerequisites

Finish the creation of security groups and instances.

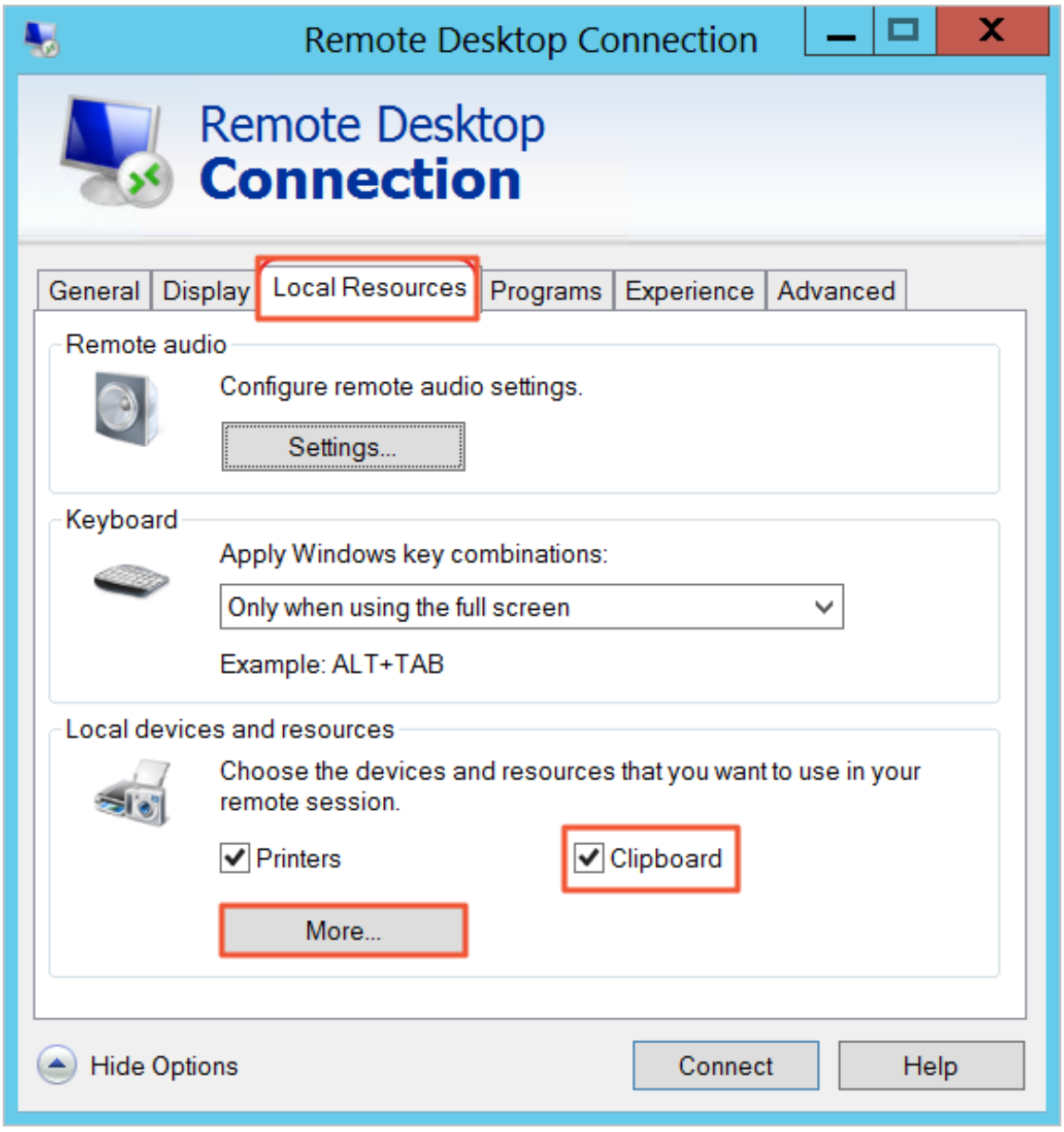
Procedure

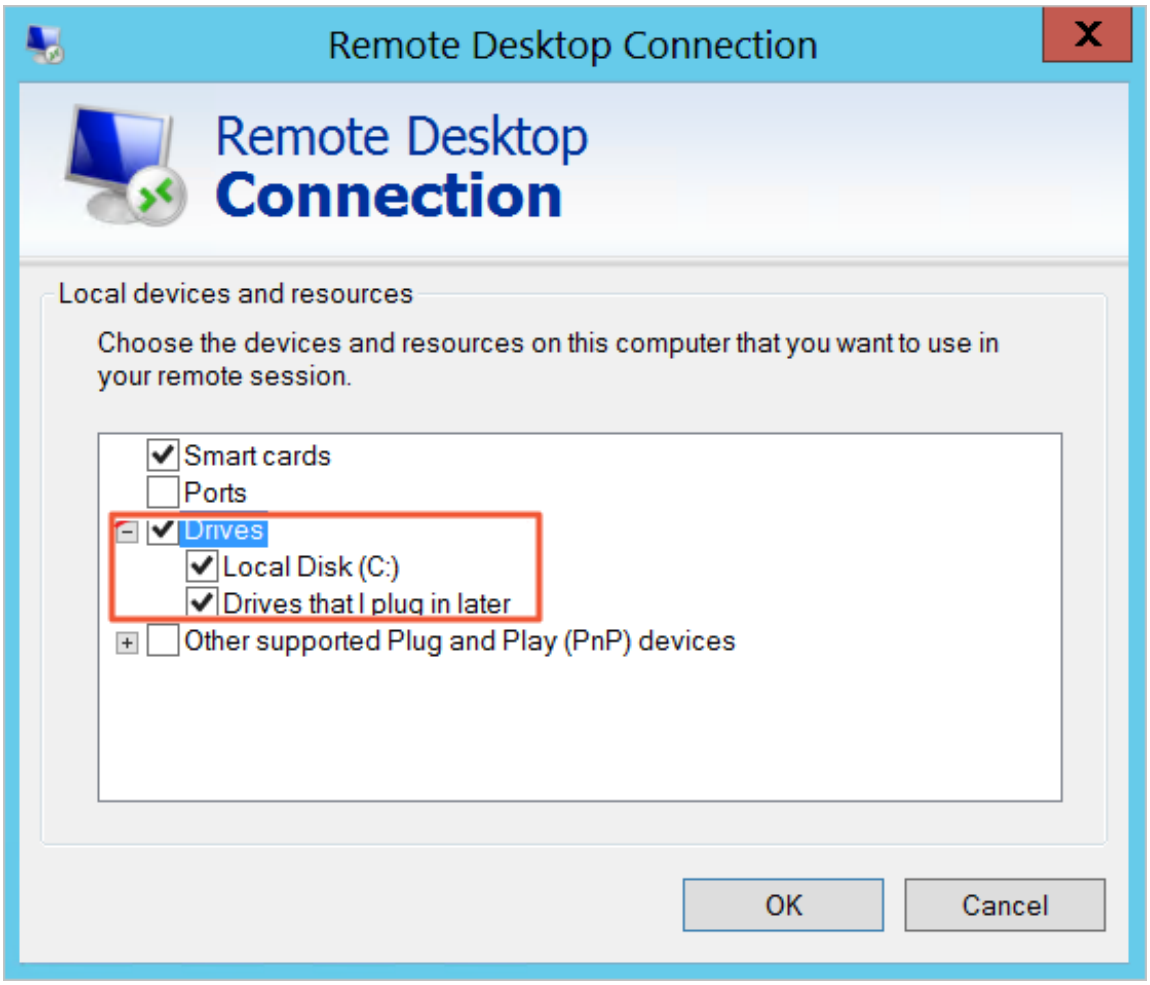
1. Start the remote desktop connection function in any of the following ways:
   - Click **Start**, enter **Remote Desktop Connection** in the search box, and click **Remote Desktop Connection** in the list that appears.
   - Enter **`mstsc`** in the search box and click **mstsc** in the list that appears.
   - Press the shortcut key **`Windows key+R`** to bring up the **Start** dialog box, enter **`mstsc`**, and press the Enter key to start the remote desktop connection function.
2. In the **Remote Desktop Connection** dialog box, enter the public IP address of the instance and click **Show Options (O)**.
3. Enter the username. The default value is **`Administrator`**. If you do not want to enter the password upon subsequent logon, select **Allow me to save credentials (R)**. After completing the settings, click **Connect** to connect to the instance.



You can also complete the following settings before connecting to the instance.

- If you want to copy local text to the instance, click the **Local Resource** tab and select **Clipboard**.
- If you want to copy a local file to the instance, click the **Local Resource** tab and then **Details**. Select **Drivers** and then the drive letter of the data disk where the file is stored. After completing the settings, click **OK**.


Remote Desktop Connection
Remote Desktop
Connection
General
Display
Local Resources
Programs
Experience
Advanced
Remote audio
Configure remote audio settings.
Settings...
Keyboard
Apply Windows key combinations:
Only when using the full screen
Example: ALT+TAB
Local devices and resources
Choose the devices and resources that you want to use in your remote session.
Printers
Clipboard
More...
Hide Options
Connect
Help



- You can click the **Display** tab to adjust the size of the remote desktop. Normally you can use full screen mode.

4. In the dialog box that appears, enter the password of the **`Administrator`** account of the Windows instance and click **OK** to connect to the instance.

Result

If the **Remote Desktop Connection** window displays a Windows desktop, a connection is successfully established to the instance.

11.3.5.5 Connect to an ECS instance by using the Management Terminal

If remote connection tools such as PuTTY, Xshell, and SecureCRT are unavailable, you can use the Management Terminal (also known as VNC) to connect to an ECS instance.

Prerequisites

- You have created a security group and an ECS instance.

- To use the Management Terminal, you must import the root certificate to the web browser. For more information, see *Install a certificate*.
- Before logging on to the Management Terminal, *Change the VNC password*.

> **Note:**
> The VNC password is used to connect to the Management Terminal from the ECS console. The instance password is used to log on to the instance.

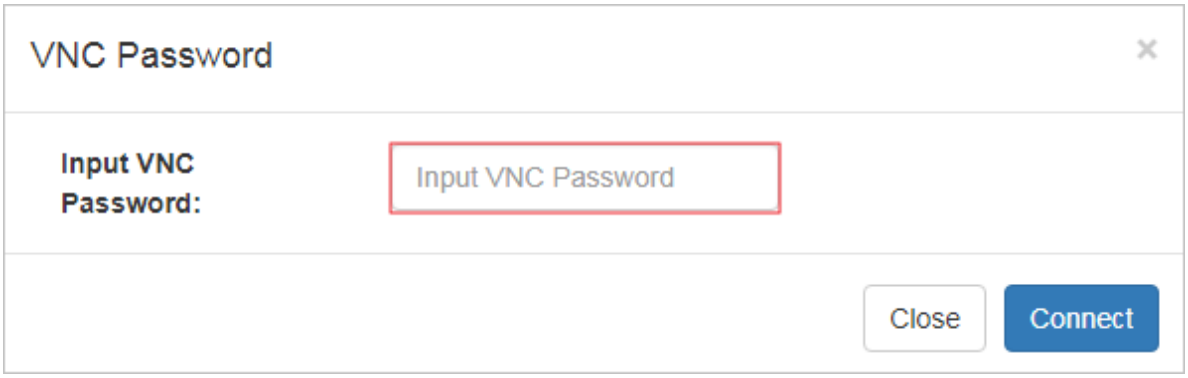
Context

Connect to Management Terminal applies to the following scenarios (including but not limited to):

- If it takes a long time to boot an instance (for example, self-test is initiated), you can view the progress by clicking **Connect to Management Terminal**.
- If a software-based remote connection fails due to incorrect instance configurations (for example, the firewall has been enabled by accident), you can connect to the instance by clicking **Connect to Management Terminal** and disable the firewall.
- If a remote connection fails due to high CPU or bandwidth usage (for example, botnet attacks occur), you can connect to the instance by clicking **Connect to Management Terminal** and terminate abnormal processes.

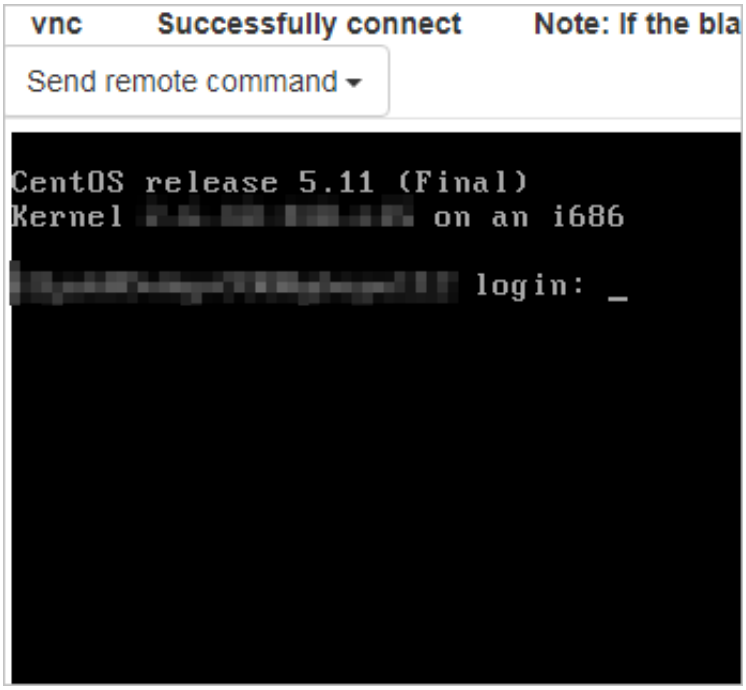
Procedure

1. *Log on to the ECS console*.
2. Click the **Instances** tab. On the Instances tab page, click the icon in the **Actions** column corresponding to an instance and choose **View Details** from the shortcut menu.
3. On the **Instance Details** tab page that appears, click **Connect to Management Terminal**.
4. In the dialog box that appears, enter the VNC password and click **Connect**.



5. The logon page appears after you connect to the Management Terminal.

The following figure shows the logon page in Linux.



6. Enter the username and password to log on to the instance.
   - For Linux instances, enter the *root* username and the logon password.
   - For Windows instances, enter the *administrator* username and the logon password.

> **Note:**
> There are no visual indicators when you enter the logon password in Linux. Press Enter after you enter the password.

11.4 Instances

11.4.1 Overview

An ECS instance is the minimal unit that can provide compute services for your business. It provides computing capabilities of specific specifications.

11.4.2 View an instance

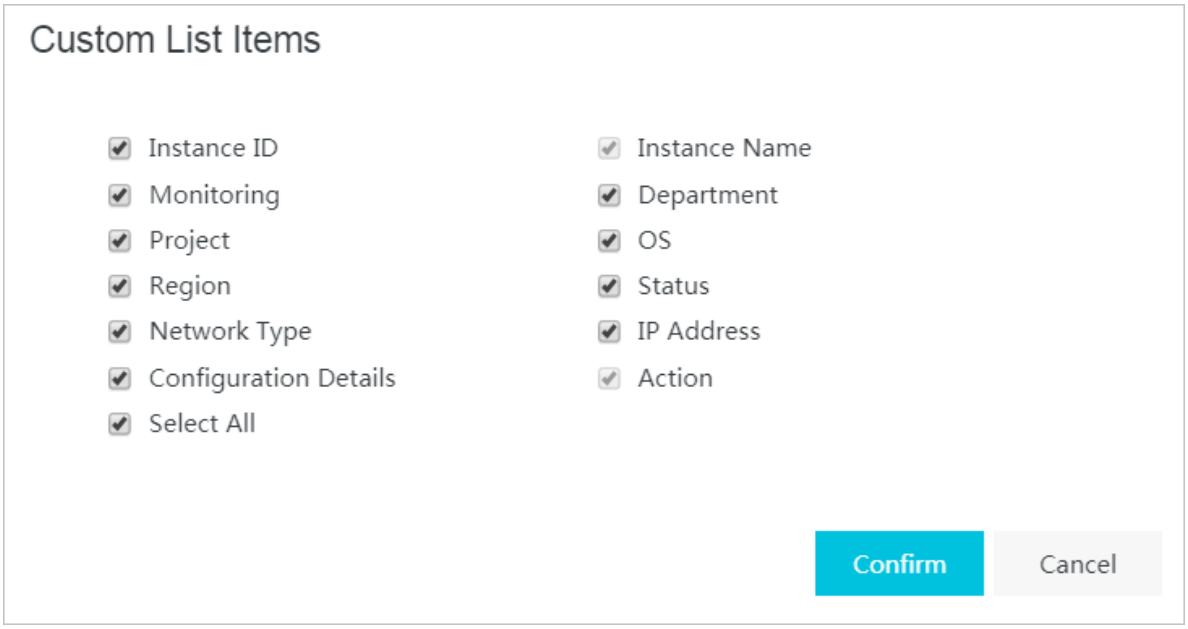
You can log on to the ECS console to view all your instances and their details.

Procedure

1. *Log on to the ECS console*.

   On the upper-right corner of the **Instances** page, click **Set**, select the items to be displayed in the **Custom List Items** box, and click **Confirm**. As shown in *Figure 11-3: Set the custom list*.

Figure 11-3: Set the custom list



2. On the **Instances** page, select a **`Department`** and a **`Region`** or enter an **`Instance Name`** and click **Search** to find the target instance.

> **Note:**
> Click **`Instance Name`** and you can choose other filtering conditions from the drop-down menu: **`Instance ID`**, **`Instance Status`**, **`VPC ID`**、**`IP Address`**, and **`Project Name`**.

3. In the **`Action`** column of the instance, click the icon and choose **View Details** to enter the **Instance Details** page and view the details of the instance.

> **Note:**
> On the **Instances** page, you can also click an instance ID to go to the **Instance Details** page.

11.4.3 Edit an instance

You can change the name and description of an instance on the ECS console.

Prerequisites

The instance has been created. For how to create an instance, see *Create an instance*.

Procedure

1. *Log on to the ECS console* and find the ECS instance you want to edit.

2. In the **Action** column of the target instance, click the icon  and select **Edit** from the drop-down list.
3. In the dialog box that appears, edit the **Name**, **Description**, and **Custom Data** of the instance and click **Confirm**.

> **Note:**
> For the custom data, Windows supports two formats: bat and powershell. Before a Base64 code, the first line is [bat] or [powershell]. Linux supports shell scripts.

11.4.4 Start, stop, or reboot an instance

On the ECS Console, you can start, stop, or reboot an instance just like on a real server.

Prerequisites

The instance has been created. For how to create an instance, see *Create an instance*.

Procedure

1. *Log on to the ECS console*.
2. On the **Instances** page, in the **Action** column of the corresponding instance, click  and select **Reboot**, **Stop**, or **Start** from the drop-down list.
   - You can stop or restart an instance only when the instance is in the **Running** state. You can start an instance only when the instance is in the **Stopped** state.

     > **Note:**
     > The Stop and Restart operations will stop your instance and interrupt your business. Therefore, exercise caution when performing these operations.

   - You can click the  icon behind a corresponding instance and choose **View Details** on the drop-down menu. On the **Instance Details** page, click **Reboot**, **Stop** or **Start** on the upper-right corner of the page to restart, stop, or start that instance.

11.4.5 Delete an instance

You can delete unneeded instances on the ECS console.

Prerequisites

The instances to be deleted are in the **Stopped** state.

Procedure

1. *Log on to the ECS console* and go to the **Instances** page.
2. Identify the ECS instance to be deleted. In the **Action** column of the instance, click the  icon and select **Delete** from the drop-down list and click **OK** on the dialog box popped up.

> **Note:**
> Alternatively, in the **Action** column of the instance, click the icon  and choose **View Details**. On the **Instance Details** page, click **Delete** and then click **OK** in the dialog box popped up to delete the instance.

11.4.6 Modify configurations

You can modify instance configurations on the ECS console.

Prerequisites

The instances to be modified are in the **Stopped** state.

Procedure

1. *Log on to the ECS console*, go to the **Instances** page, and find the instance whose configurations need to be changed.
2. In the **Action** column of the instance, click the  icon, choose **Change Configurations** from the drop-down menu, select new CPU and memory specifications, and then click **OK**.

> **Note:**
> After the configurations of the instance are modified, *Restart the instance* on the console for the new configurations to take effect.

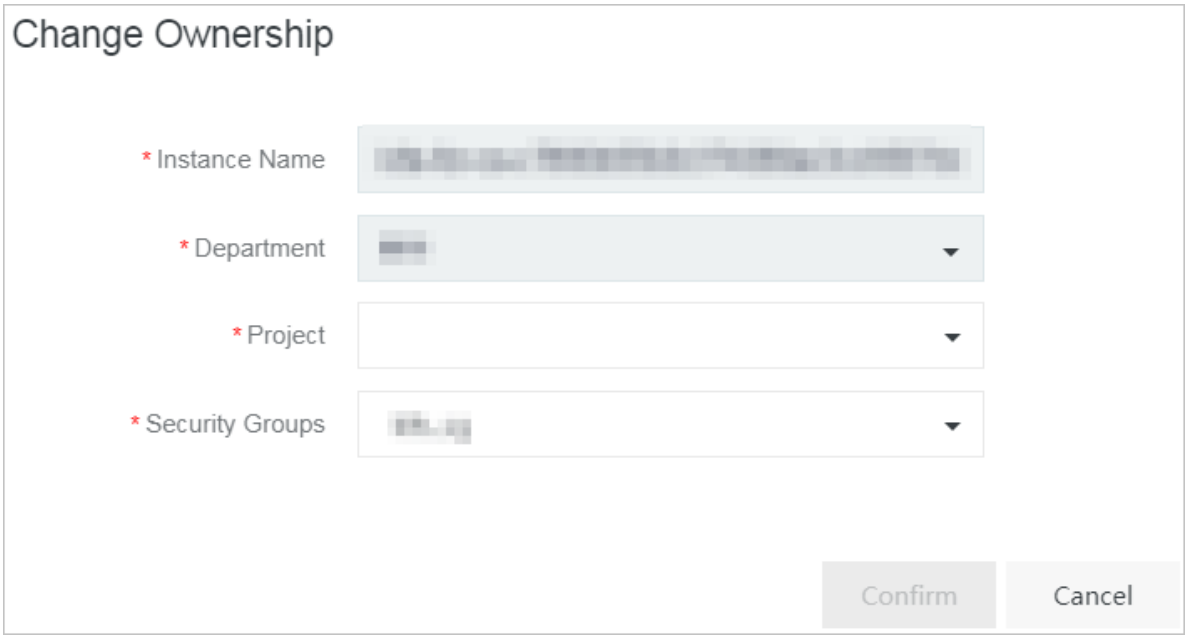
11.4.7 Change ownership

You can modify the department and project to which an instance belongs on the ECS console.

Procedure

1. *Log on to the ECS console*.
2. In the **Action** column of the instance, click the icon  and select **Change Ownership** from the drop-down list.
3. In the dialog box for **Change Ownership**, select new **Department** and **Project** and then click **Confirm**. See *Figure 11-4: Change ownership*.

Figure 11-4: Change ownership



11.4.8 Change the ECS instance logon password

You can modify the instance logon password on the ECS console.

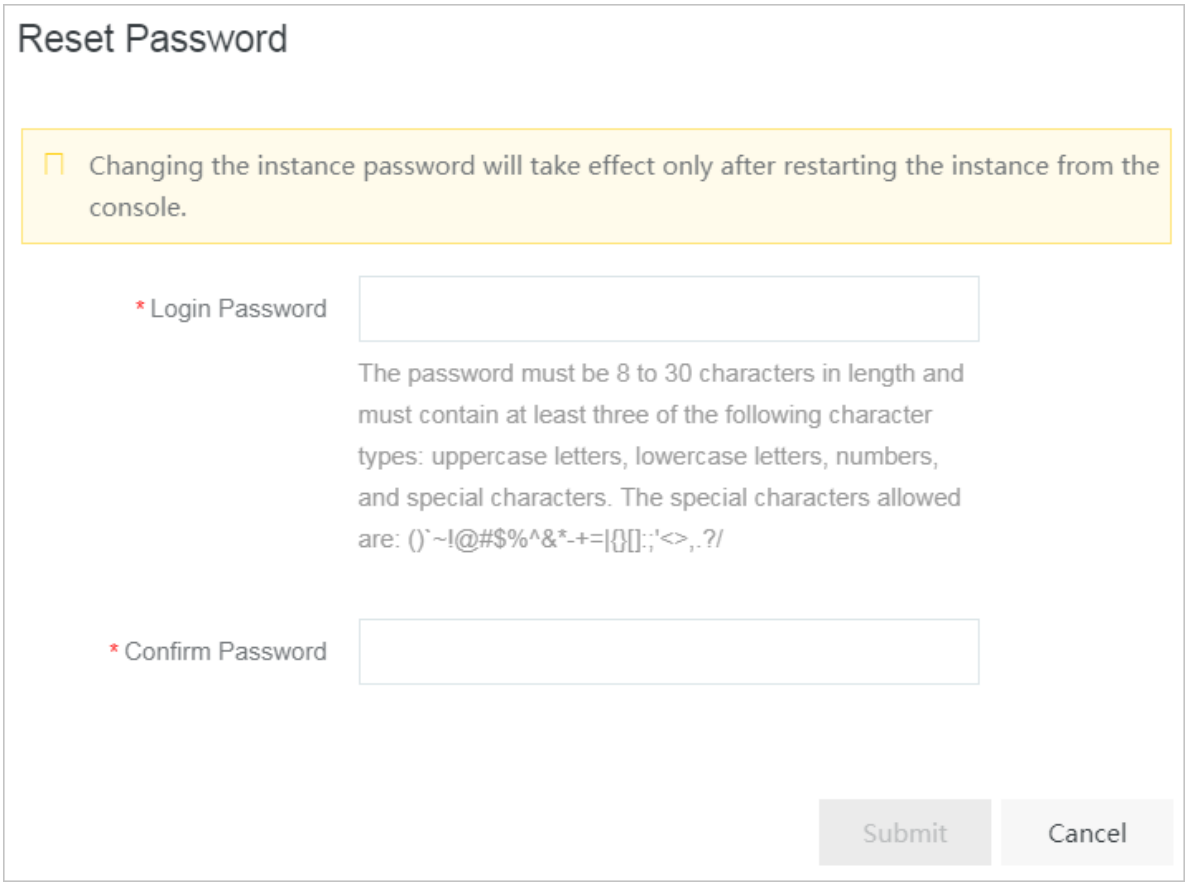
Context

If you do not set the logon password upon instance creation,you can do it via the **Reset Password** function on the ECS console.

Procedure

1. *Log on to the ECS console*.
2. In the **Action** column of the instance, click the icon and select **View Details** from the drop-down list.
3. On the **Instance Details** page, click **Change Password**.
4. In the dialog box that appears, enter a new **`Login Password`** and **`Confirm Password`**. Then click **Submit**. See *Figure 11-5: Reset your password*.

Figure 11-5: Reset your password



5. *Reboot the instance* on the console for the new instance password to take effect.

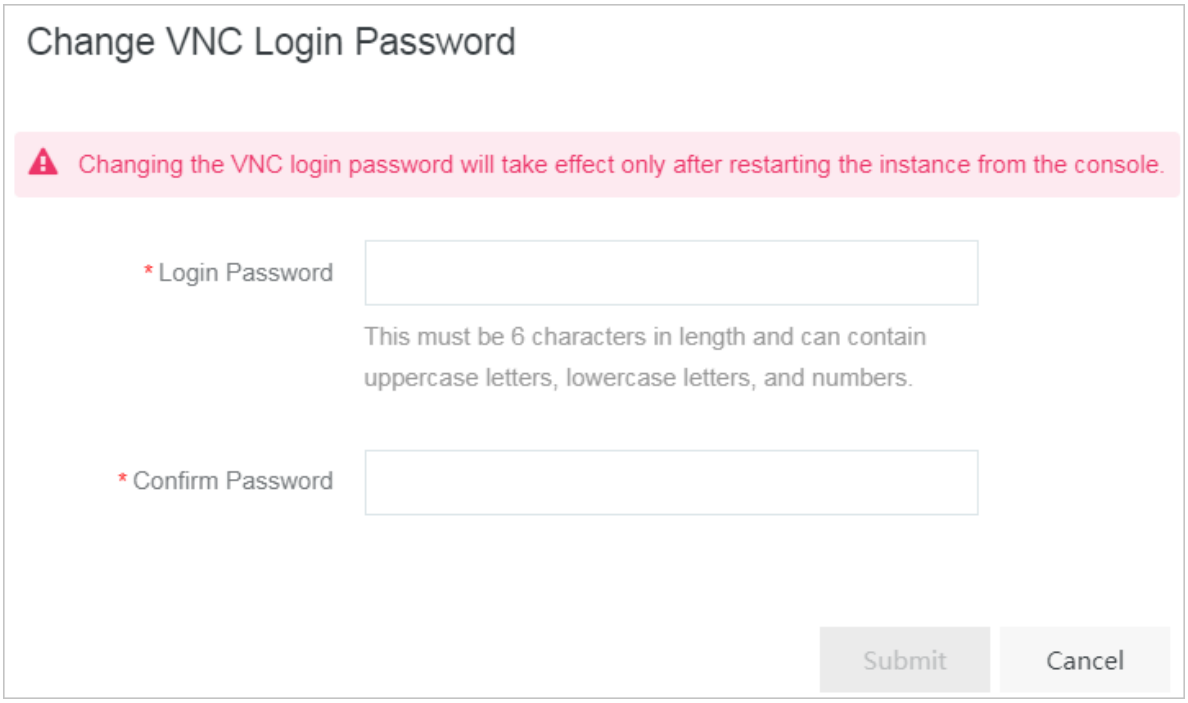
11.4.9 Change the VNC password

You can modify the VNC password on the ECS console.

Procedure

1. *Log on to the ECS console*.
2. In the **Action** column of the instance, click the icon and choose **View Details** from the drop-down menu.
3. On the **Instance Details** page, click **Change Management Terminal Password**.
4. In the dialog box that appears, enter a new **`Login Password`** and **`Confirm Password`**. Then click **Submit**. As shown in *Figure 11-6: Change the VNC password*.

Figure 11-6: Change the VNC password



Note:

Reboot the instance on the console for the new password of the VNC to take effect.

11.4.10 Join a security group

You can add an instance to a security group with either of the two methods on the ECS console.

Context

A security group is a virtual firewall that controls the inbound and outbound traffic of instances. You can add an instance to up to five security groups on the ECS console. Once the security group rules are changed, they apply to the instances automatically.

Method 1: Join a security group on the instance page

1. *Log on to the ECS console*.
2. Enter the **Instances** page. In the **Action** column of an instance, click the icon and select **View Details** from the drop-down list.
3. In the top navigation bar, click the **Instance Security Group** tab, and click **Join Security Group**.
4. In the **Move to Security Group** dialog box that appears, select the target security group and click **Confirm**.

Method 2: Add an instance to a security group on the security group page

You can also add an instance to a security group in the following procedure:

1. *Log on to the ECS console* and go to the **Security Groups** page.
2. Click the security group ID to go to the **ECS Instances** page.

> **Note:**
> Alternatively, in the **Action** column of a security group, click the icon and select **View Details** to enter the **ECS Instances** page.

3. In the upper-right corner of the **ECS Instances** page, click **Import ECS Instances**.
4. In the **Move to Security Group** dialog box, select an instance and click **Confirm**.

11.4.11 Customize instance data

ECS allows you to run the instance customization script upon startup and import data into instances.

Context

The instance data customization feature is applicable to both Windows and Linux instances. It allows you to:

- Run the instance customization script upon startup.
- Import data into instances.

Usage notes

- **Limits**

  The instance data customization feature can be used only when an instance meets all the following conditions:

  — Network type: VPC
  — Image: a system image or a custom image inheriting from the system image
  — Operating system: one type included in *Table 11-4: Supported operating systems*

  Table 11-4: Supported operating systems

Windows	Linux
■ Windows Server 2016 64-bit ■ Windows Server 2012 64-bit	■ CentOS ■ Ubuntu

Windows	Linux
■ Windows Server 2008 64-bit	■ SUSE Linux Enterprise ■ OpenSUSE ■ Debian ■ Aliyun Linux

— When you configure instance customization scripts, you must enter custom data based on the type of operating system and script.

> **Note:**
> Only English characters are allowed.

■ If your data is Base64 encoded, select **Enter Base64 Encoded Information**.

> **Note:**
> The size of the customization script cannot exceed 16 KB before the data is Base64 encoded.

■ For Linux instances, the script format must meet the requirements described in *Types of Linux instance customization scripts*.

■ For Windows instances, the script can only start with `[bat]` or `[powershell]`.

- After starting an instance, run a command to view the following information:

  — Execution result of the instance customization script

  — Data imported to instances

- **Console**: You can modify the custom instance data in the console. Whether the modified instance customization script needs to be re-executed depends on the script type. For example, if the `bootcmd`-type script in **`Cloud Config`** is modified for Linux instances, the script is automatically executed every time instances are restarted.
- **Open APIs**: You can also use open APIs to customize instance data. For more information about the operation method, see **CreateInstance** and **ModifyInstanceAttribute** in *ECS Developer Guide*.

Linux instance customization scripts

Linux instance customization scripts provided by Alibaba Cloud are designed based on the cloud-init architecture. They are used to automatically configure parameters of Linux instances. Customization script types are compatible with the cloud-init.

Description of Linux instance customization scripts

- Linux instance customization scripts are executed after instances are started and before `/etc/init` is executed.
- Linux instance customization scripts are executed with root permissions by default.

Types of Linux instance customization scripts

- **User-Data Script**
  - Description: A script, such as shell script, is used to customize data.
  - Format: The first line must include `#!`, such as `#! /bin/sh`.
  - Limit: The script size (including the first line) cannot exceed 16 KB before the data is Base64 encoded.
  - Frequency: The script is executed only when instances are started for the first time.
  - Example:

```
#! /bin/sh
```

```
echo "Hello World.  The time is now $(date -R)!" | tee /root/output10.txt
```

- **Cloud Config Data**
  - Description: Predefined data is used to configure services, such as specifying yum sources or importing SSH keys.
  - Format: The first line must be `#cloud-config`.
  - Limit: The script size (including the first line) cannot exceed 16 KB before the data is Base64 encoded.
  - Frequency: The script execution frequency varies with the specific service.
  - Example:

```
#cloud-config
```

```
apt:
```

```
primary:
```

```
- arches: [default]
```

```
uri: http://us.archive.ubuntu.com/ubuntu/
```

- **Include**

- Description: The configuration content can be saved in a text file and imported into cloud-init as a URL.
- Format: The first line must be `#include`.
- Limit: The script size (including the first line) cannot exceed 16 KB before the data is Base64 encoded.
- Frequency: The script execution frequency depends on the script type in the text file.
- Example:

```
#include
```

```
http://ecs-image-test.oss-cn-hangzhou.aliyuncs.com/userdata/
cloudconfig
```

- **GZIP format**
  - Description: Cloud-ini limits the size of customization scripts to 16 KB. You can compress and import the script file into the customization script if the file size exceeds 16 KB.
  - Format: The .gz file is imported into the customization script as a URL under `#include`.
  - Frequency: The script execution frequency depends on the script content contained in the GZIP file.
  - Example:

```
#include
```

```
http://ecs-image-test.oss-cn-hangzhou.aliyuncs.com/userdata/config
.gz
```

View the custom data of a Linux instance

Run the following command in the instance:

```
curl http://100.100.100.200/latest/user-data
```

Windows instance customization scripts

Windows instance customization scripts independently developed by Alibaba Cloud can be used to initialize Windows instances.

There are two types of Windows instance customization scripts:

- Batch processing program: starts with `[bat]` and serves as the first line. The script size must be smaller than 16 KB before the data is Base64 encoded.

- PowerShell script: starts with `[powershell]` and serves as the first line. The script size must be smaller than 16 KB before the data is Base64 encoded.

View the custom data of a Windows instance

Run the following PowerShell command in the instance:

```
Invoke-RestMethod http://100.100.100.200/latest/user-data/
```

11.4.12 Change private IP

You can change the private IP address on the ECS console.

Prerequisites

Before you change the private IP address of an instance, make sure that the instance is in the **Stopped** state. For how to stop an instance, see *Start, stop, or reboot an instance*.

Context

Each instance is assigned a private network interface that is bound with a private IP address. Private IP addresses are determined by the switch IP segment.

Procedure

1. *Log on to the ECS console*.
2. *Stop the instance*.
3. In the **`Action`** column of the instance, click the  icon and select **Change Private IP** from the drop-down list.
4. In the **Change Private IP** dialog box that appears, enter a new **`Private IP`** and click **Confirm**.

11.4.13 Install a certificate

Before you log on to the Management Terminal, you must export the certificate from the site and install it in your local web browser.

Context

The Management Terminal feature is provided by the backend VNC proxy service. The VNC proxy service uses a different certificate from that of Apsara Infrastructure Management Framework. Therefore, you need to import the certificate separately.

Procedure

1. Export the certificate from the site.

a) Log on to Apsara Stack Management Console. Press **F12** or **Fn + F12** to view and select the certificate, as shown in the following figure. For example, in the Chrome browser, press **F12** to open the developer tools.

Figure 11-7: View the certificate

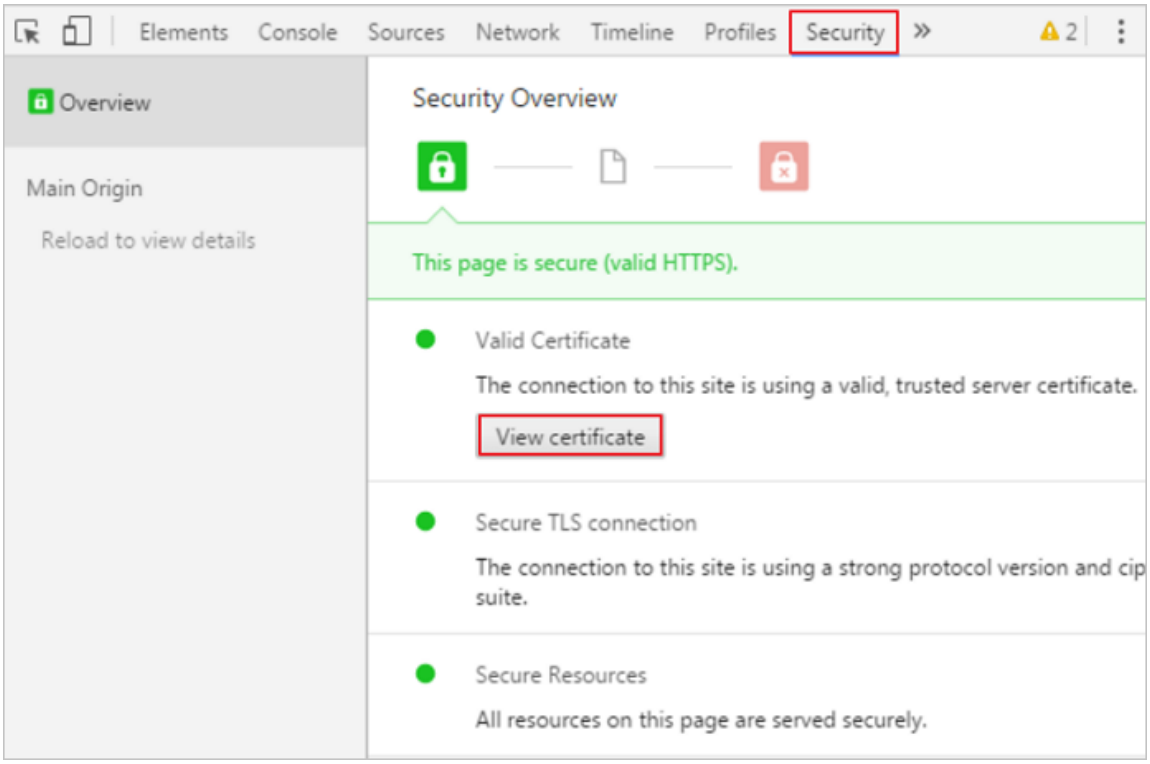
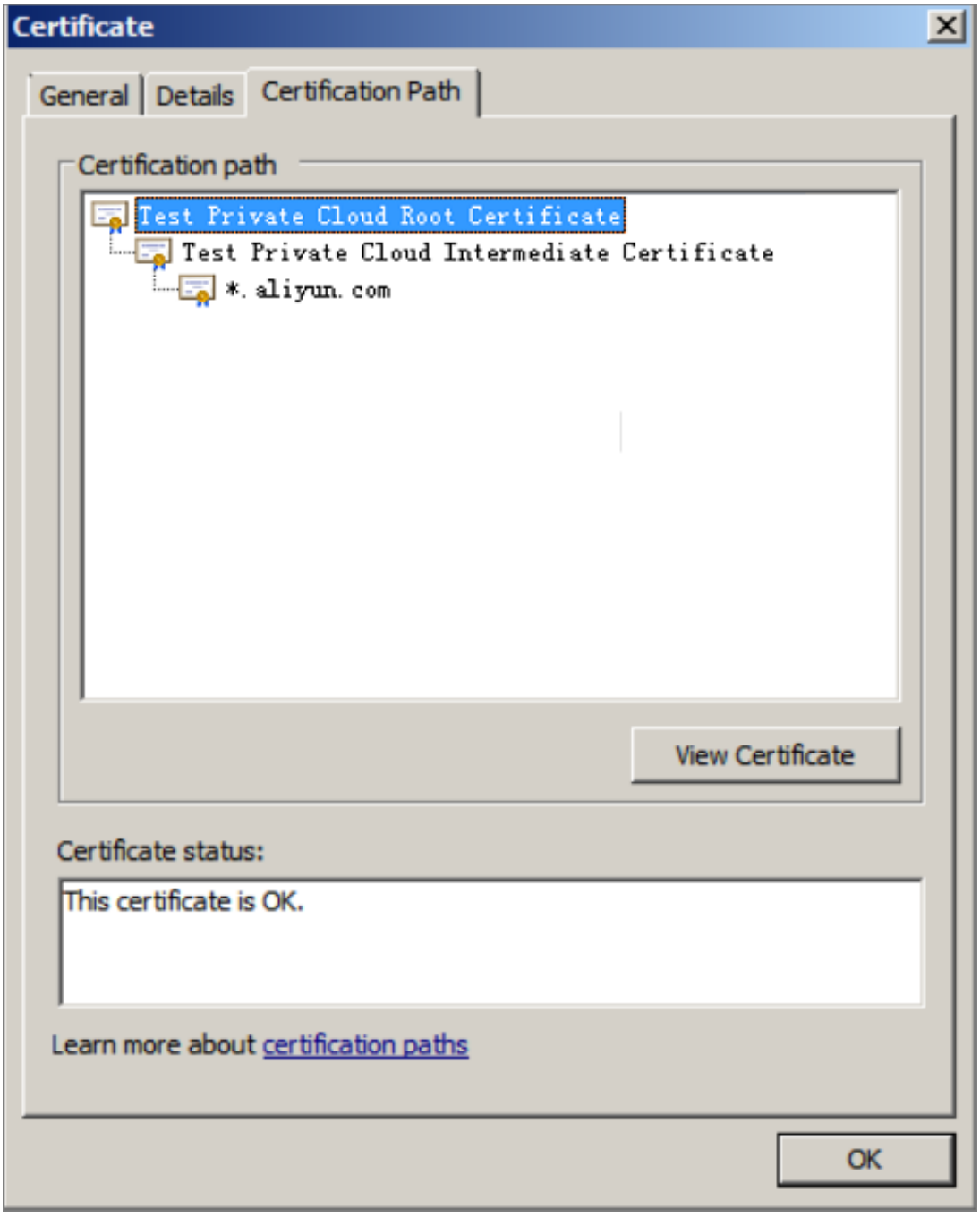


Figure 11-8: Select the certificate



b) In the Certificate dialog box, click **Copy to File**. Follow the instructions shown in the following figures. Enter a name and save the certificate to your local machine.

Figure 11-9: Copy the certificate to a file

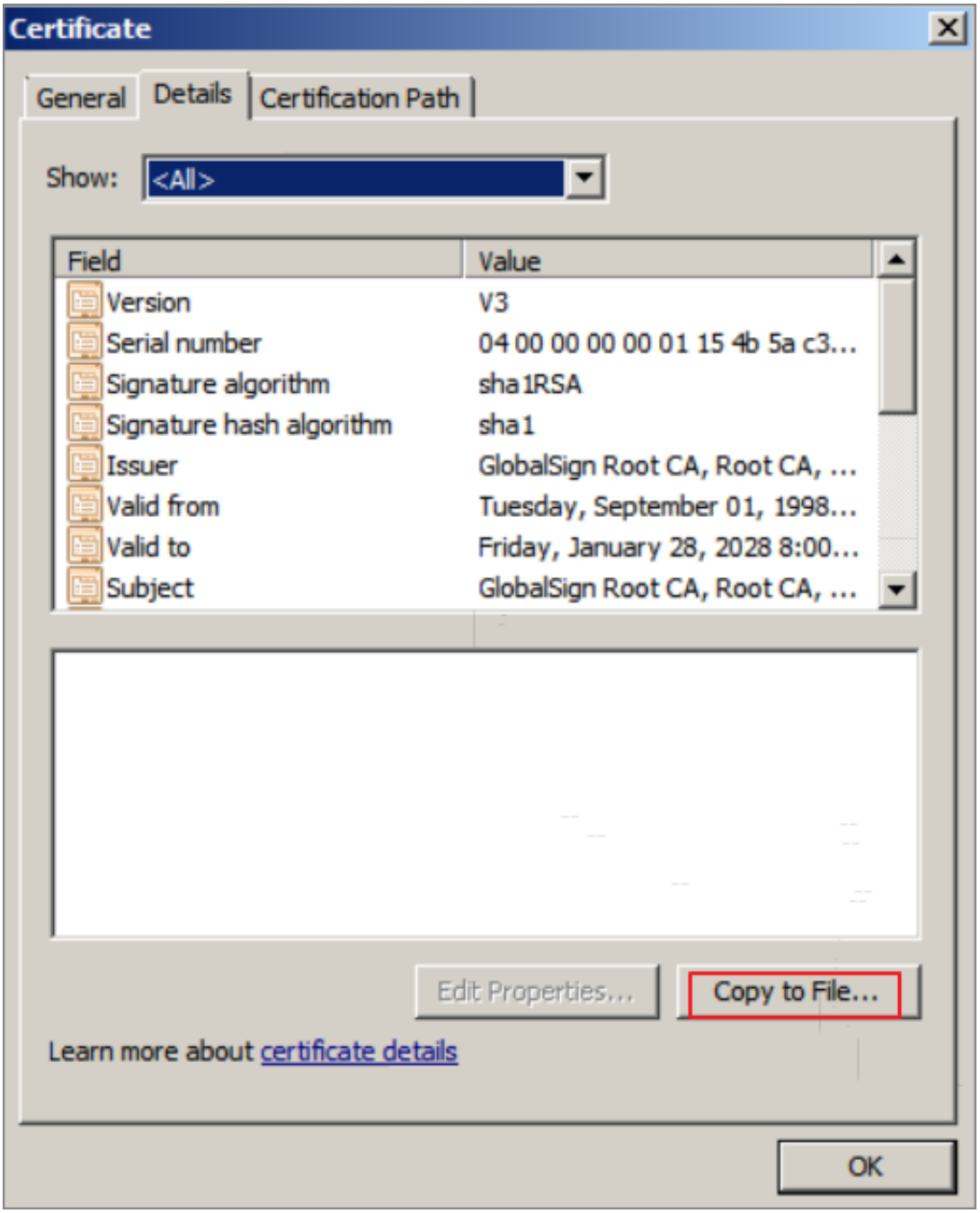


Figure 11-10: Certificate export wizard

Figure 11-11: Select the file export format

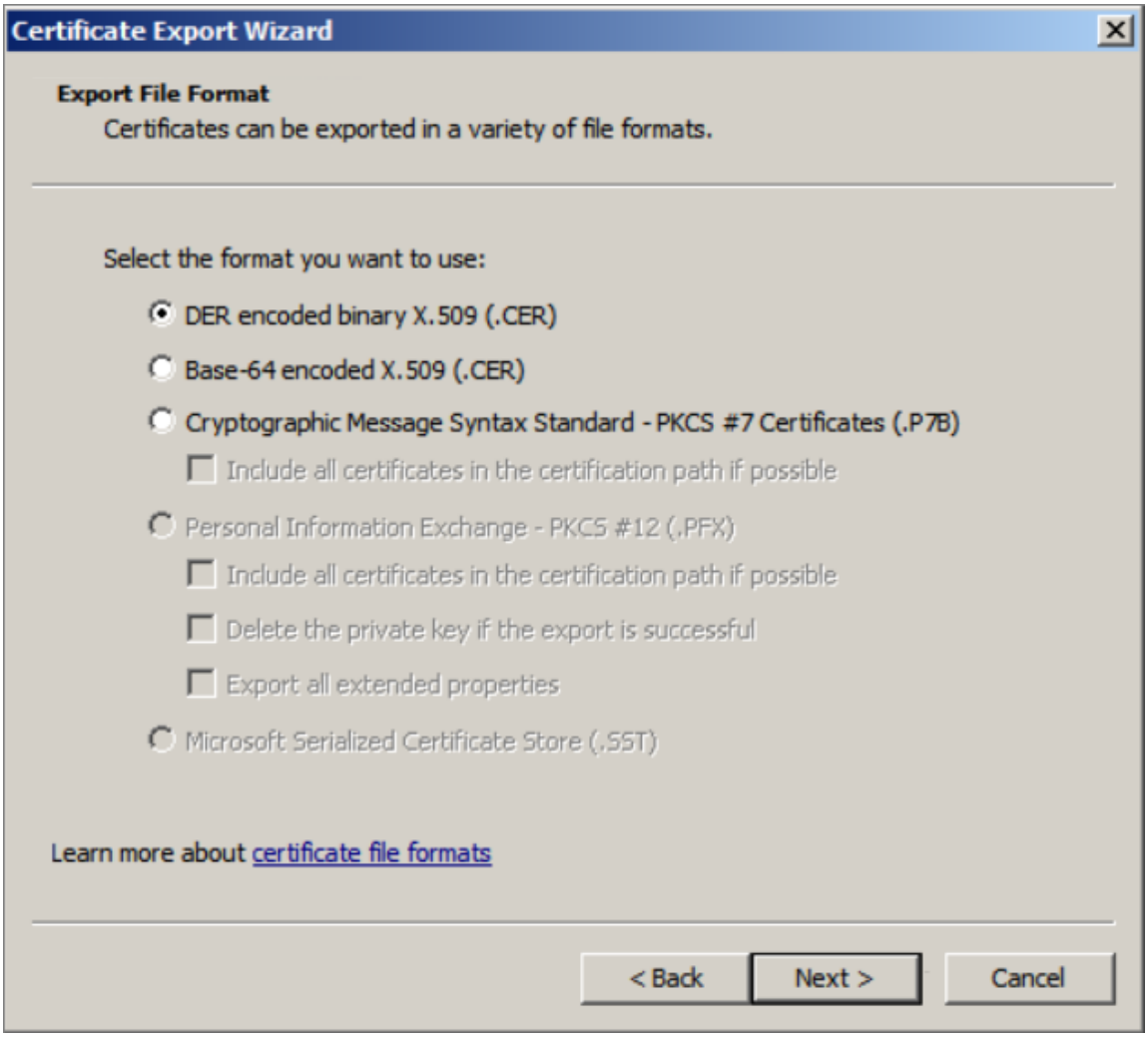


Figure 11-12: Select the file to be exported

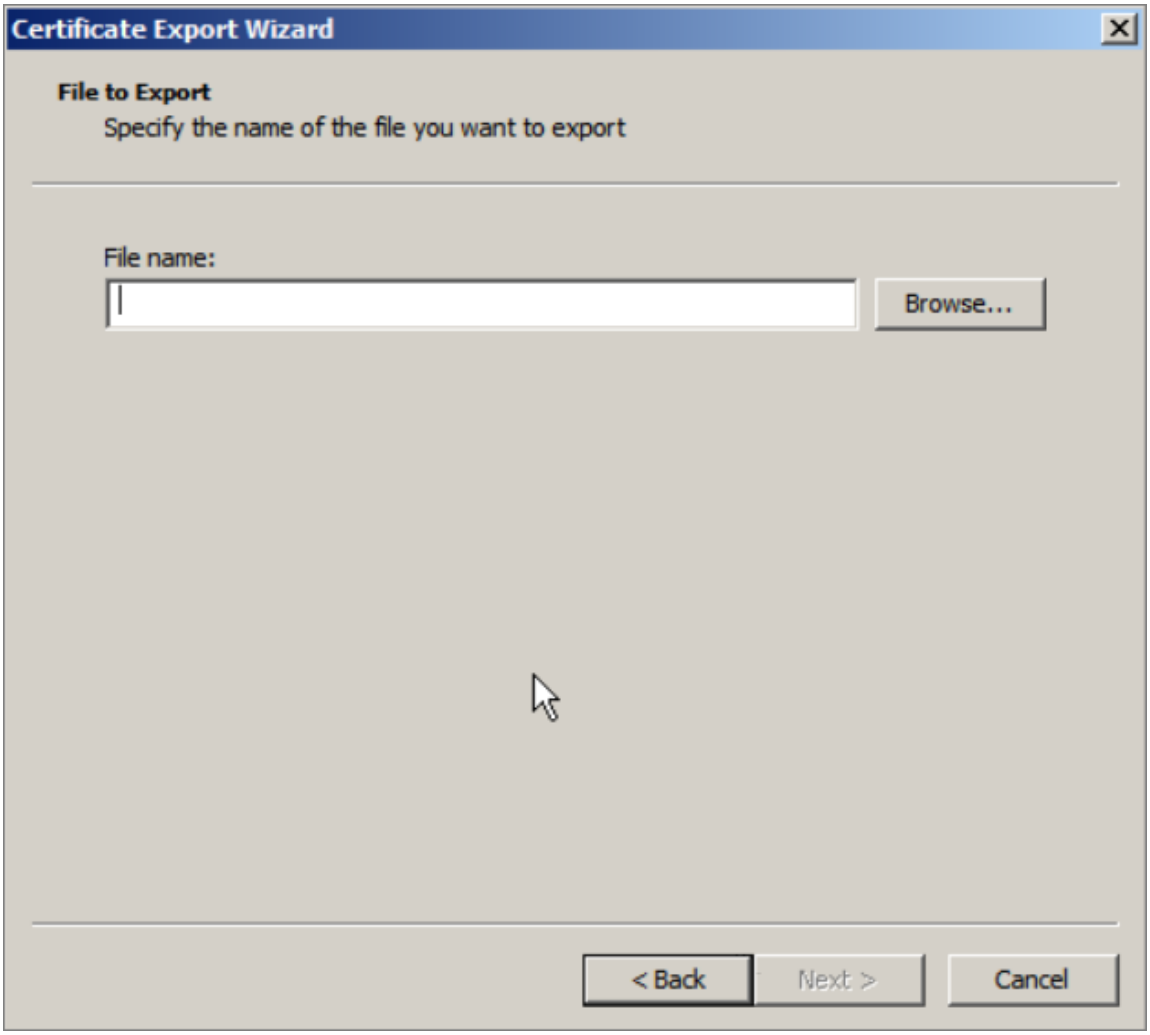
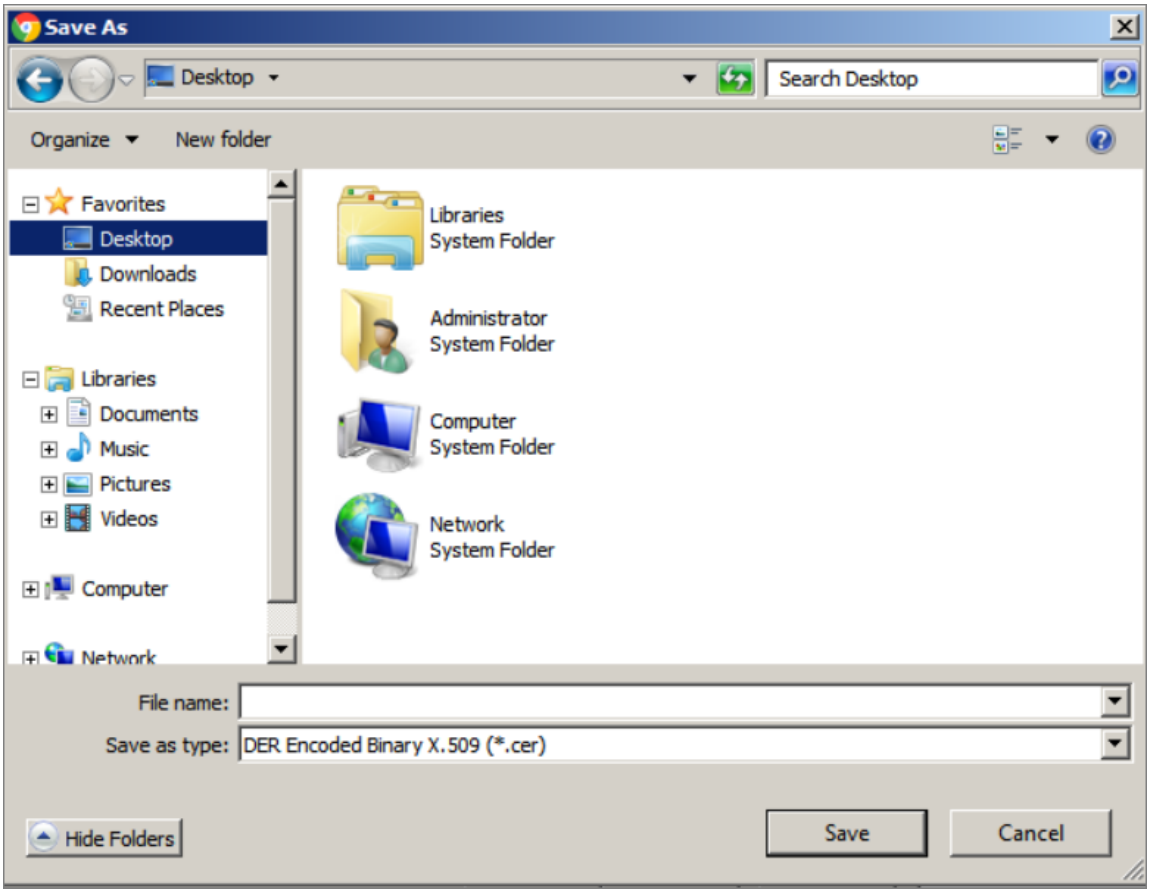


Figure 11-13: Customize the file name

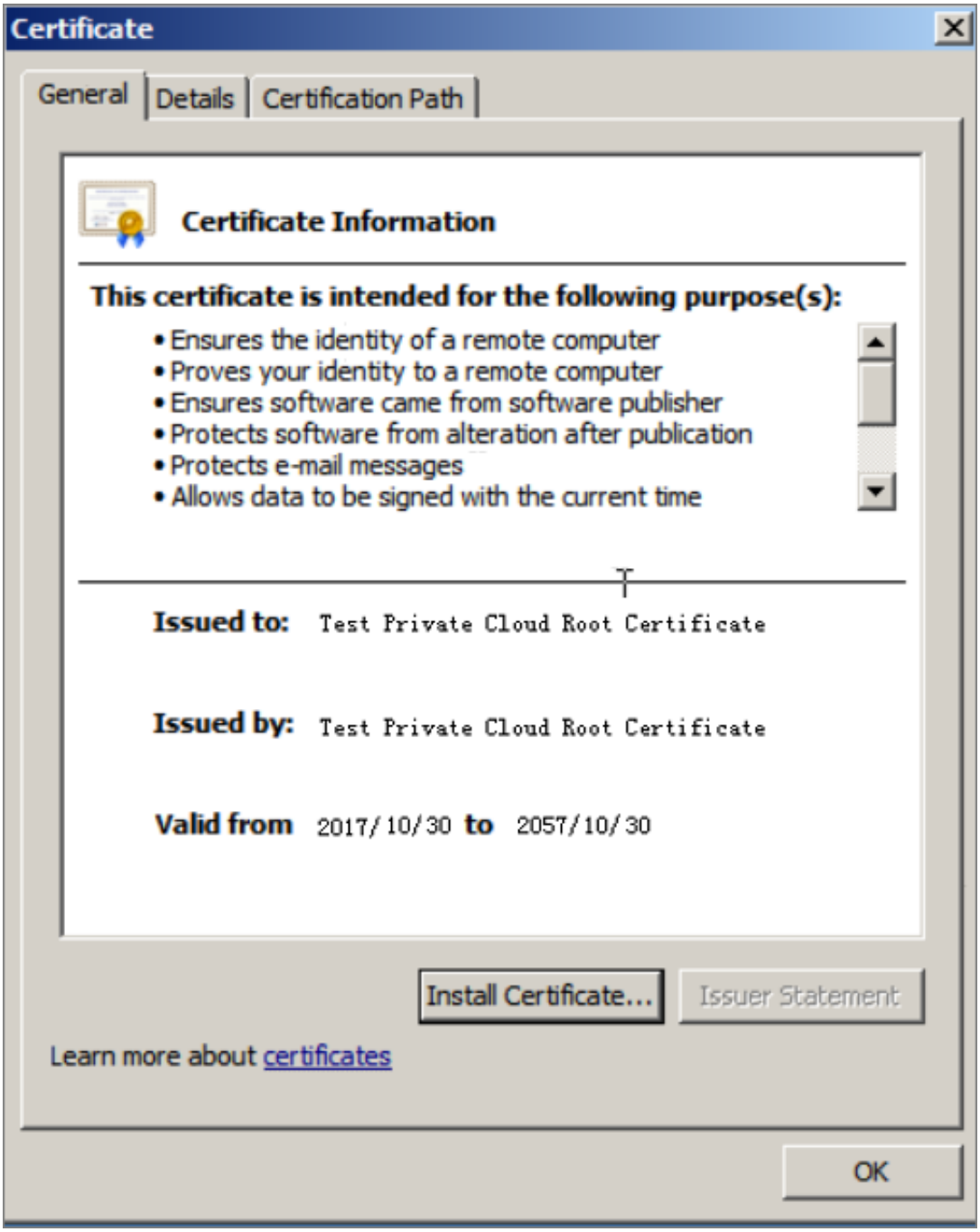


c) Click **OK**. A message appears, indicating that the certificate is exported.

2. Install the certificate in your local web browser.

   a) Double-click the certificate saved in your local machine. In the dialog box that appears, click **Install Certificate**.

Figure 11-14: Install a certificate



b) In the dialog box that appears, click **Put All Certificate to Following Storage** and **Browse**.

In the dialog box that appears, select **Trusted Root Certificate Authority** and click **OK**.

Figure 11-15: Store certificates

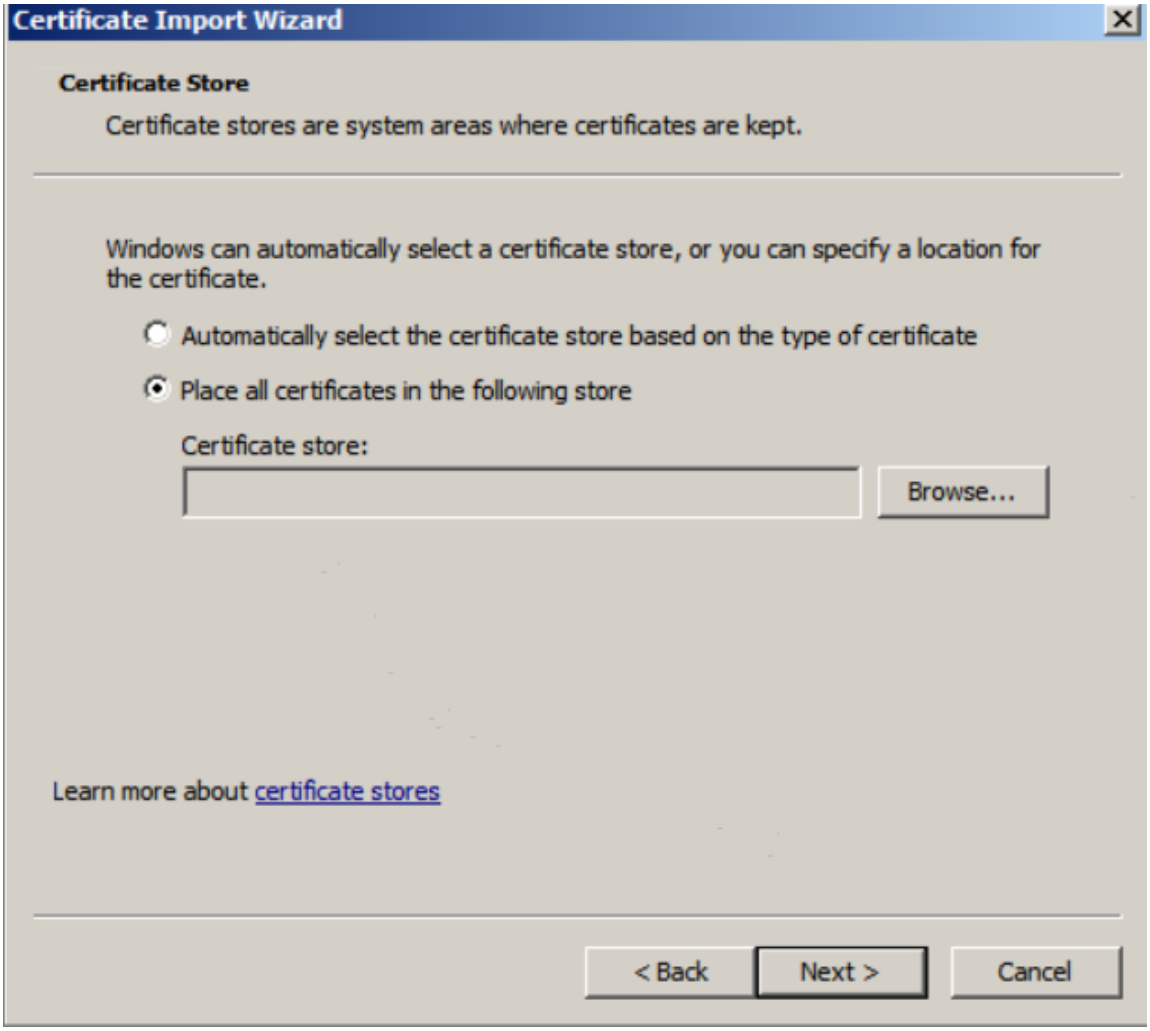
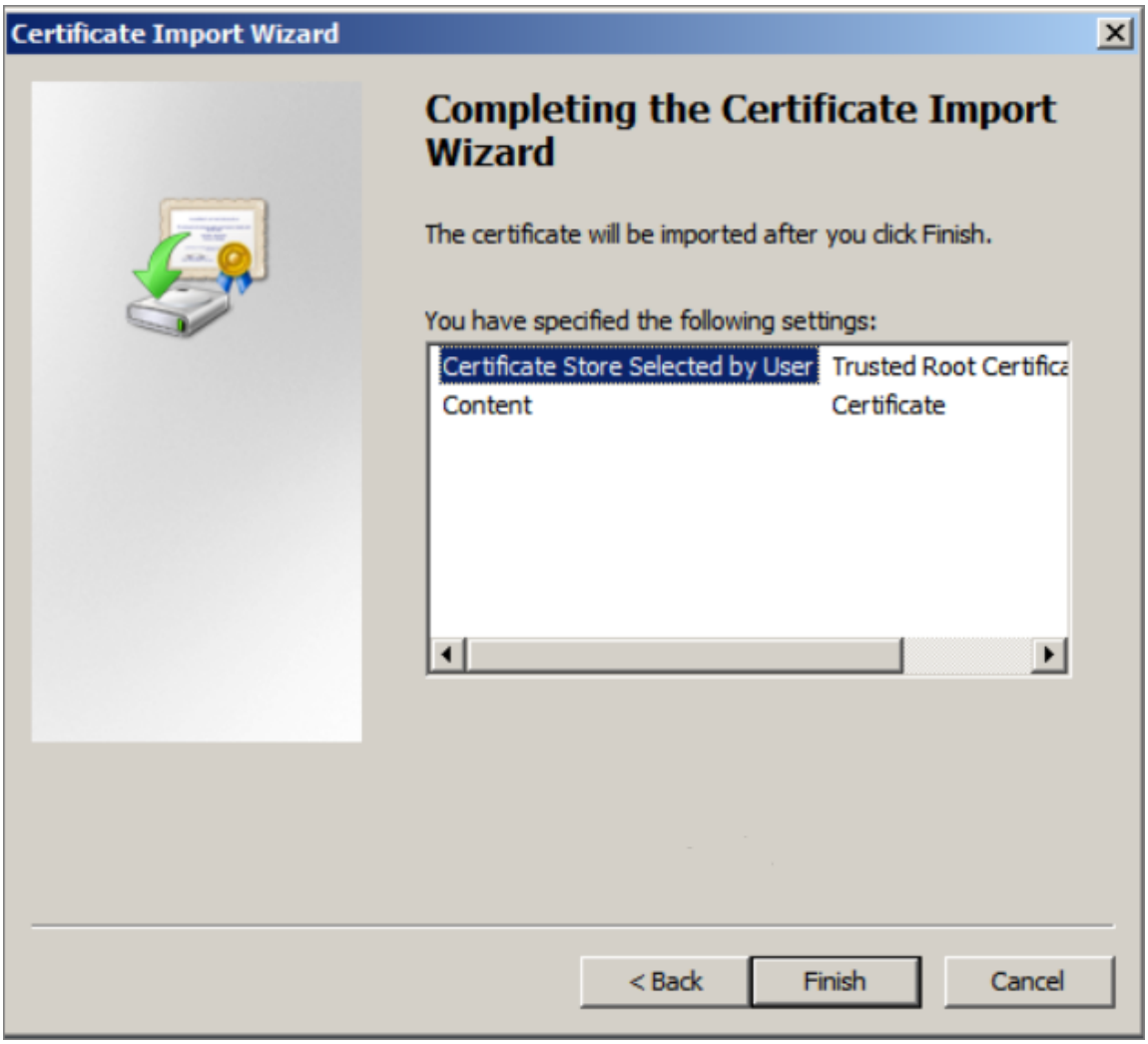


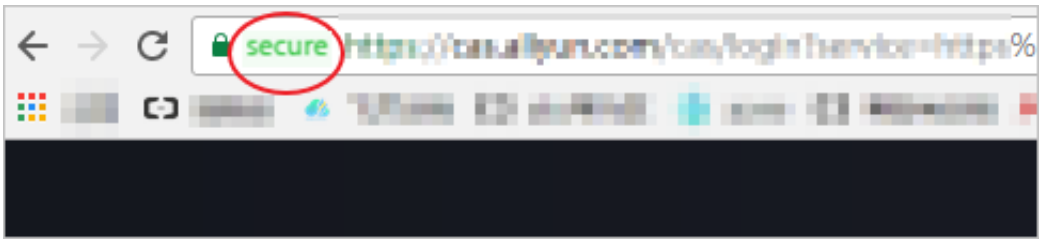
Figure 11-16: Certificate storage location

c) After importing the certificate, click **Finish**.



3. Restart your web browser and log on to Apsara Stack Management Console. A **Secure** indicator in green is displayed in the left part of the address bar, indicating that the certificate is installed, as shown in *Figure 11-17: Restart the web browser*.

Figure 11-17: Restart the web browser



11.4.14 Install the CUDA and GPU drivers for a Linux instance

The device where a GPU instance runs must be installed with the GPU driver. If the image you use does not contain a pre-installed GPU driver, you must install the CUDA and GPU drivers for the instance.

Context

When installing NVIDIA drivers, you must install the kernel package that contains the kernel header file before you install the CUDA and GPU drivers.

Procedure

1. Install the kernel package.
   a) Run the **`uname -r`** command to view the current kernel version.

      Output example:

      - CentOS: `3.10.0-862.14.4.el7.x86_64`
      - Ubuntu: `4.4.0-117-generic`

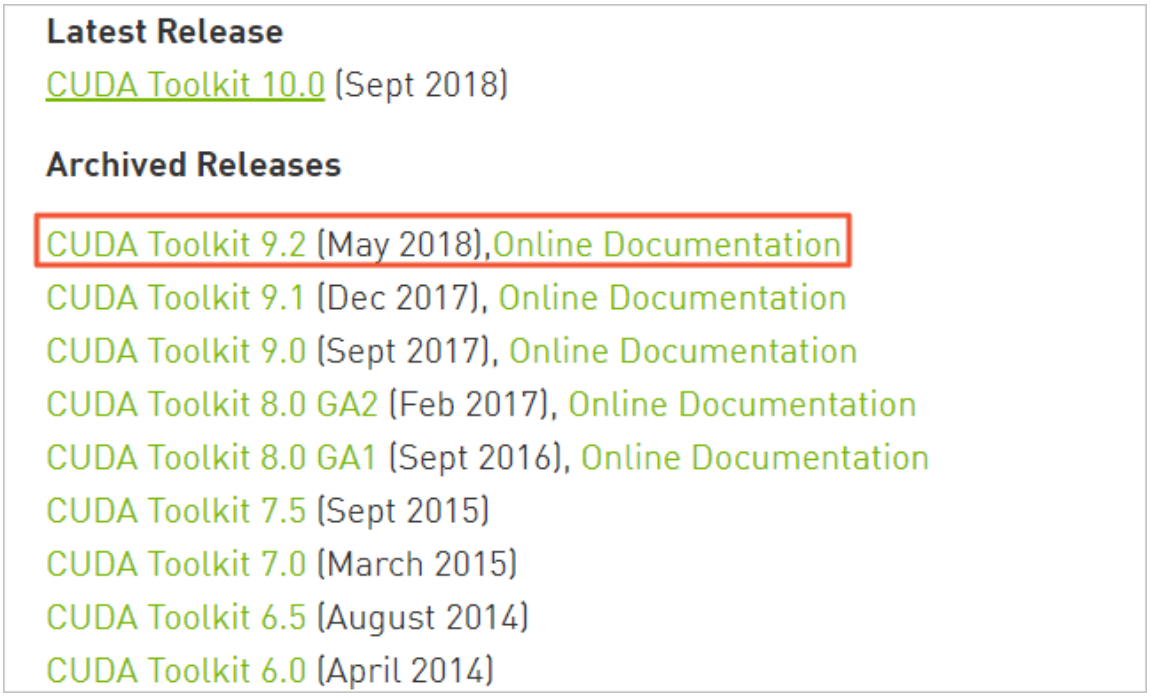
   b) Copy the kernel package of the corresponding version to the instance and install the package.

      - CentOS: Copy the RPM package of the `kernel-devel` component and the **`rpm -ivh 3.10.0-862.14.4.el7.x86_64.rpm`** command to install the package. In the preceding command, *`3.10.0-862.14.4.el7.x86_64.rpm`* is used as an example. Enter the actual name of the package you want to install.
      - Ubuntu: Copy the DEB package of the `linux-headers` component and run the **`dpkg -i 4.4.0-117-generic.deb`** command to install the package. In the preceding command, *`4.4.0-117-generic.deb`* is used as an example. Enter the actual name of the package you want to install.

2. Download the CUDA Toolkit.
   a) Access the *official download page*. Choose a suitable version based on the GPU application requirements for CUDA.

      You can choose *CUDA Toolkit 9.2*.

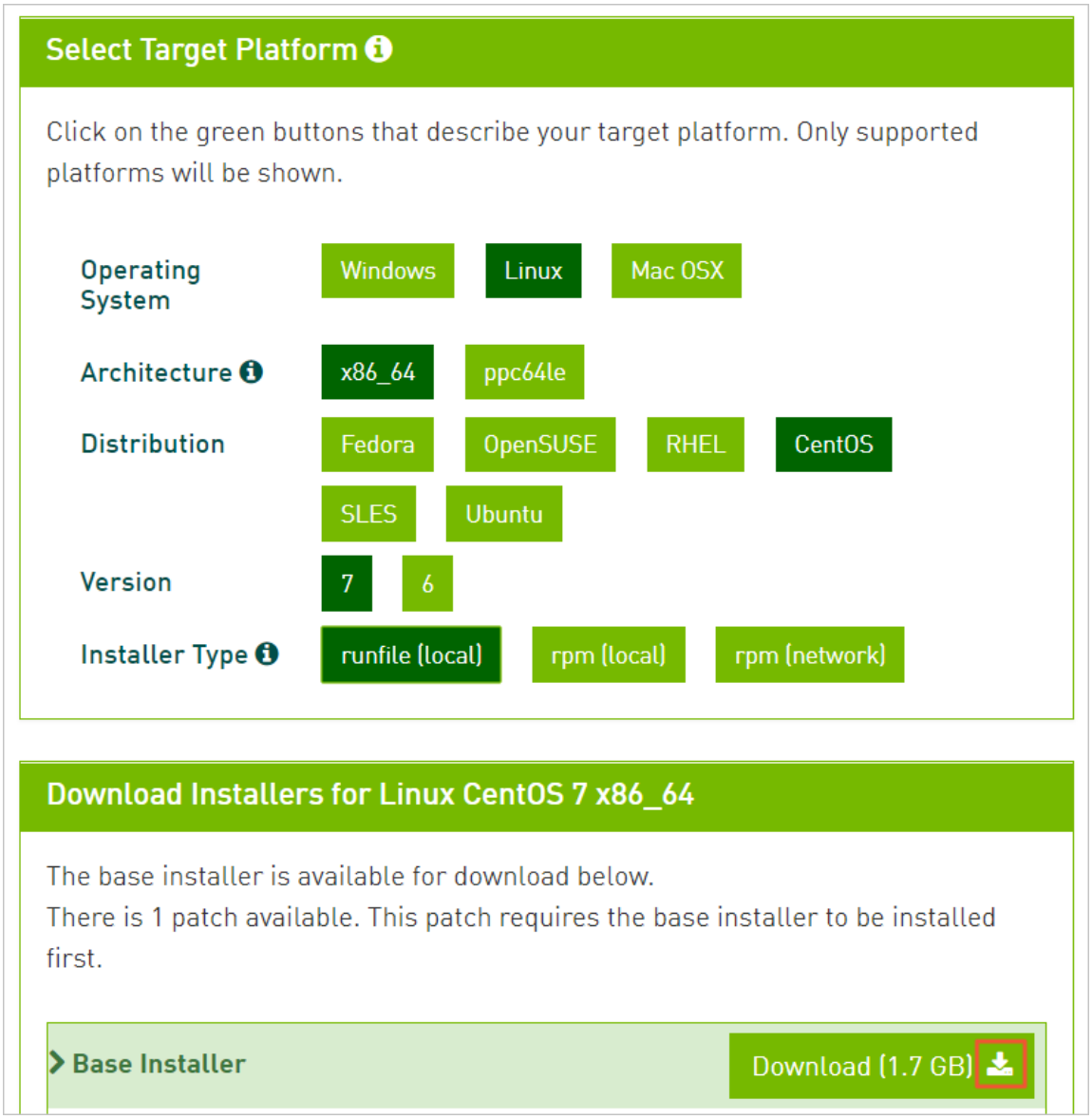
Figure 11-18: Download the CUDA Toolkit



b) Choose a platform based on your operating system. Set **Installer Type** to **runfile (local)** and click **Download**.

NVIDIA drivers are already contained in the CUDA Toolkit. You do not need to download them separately.

Figure 11-19: Download drivers



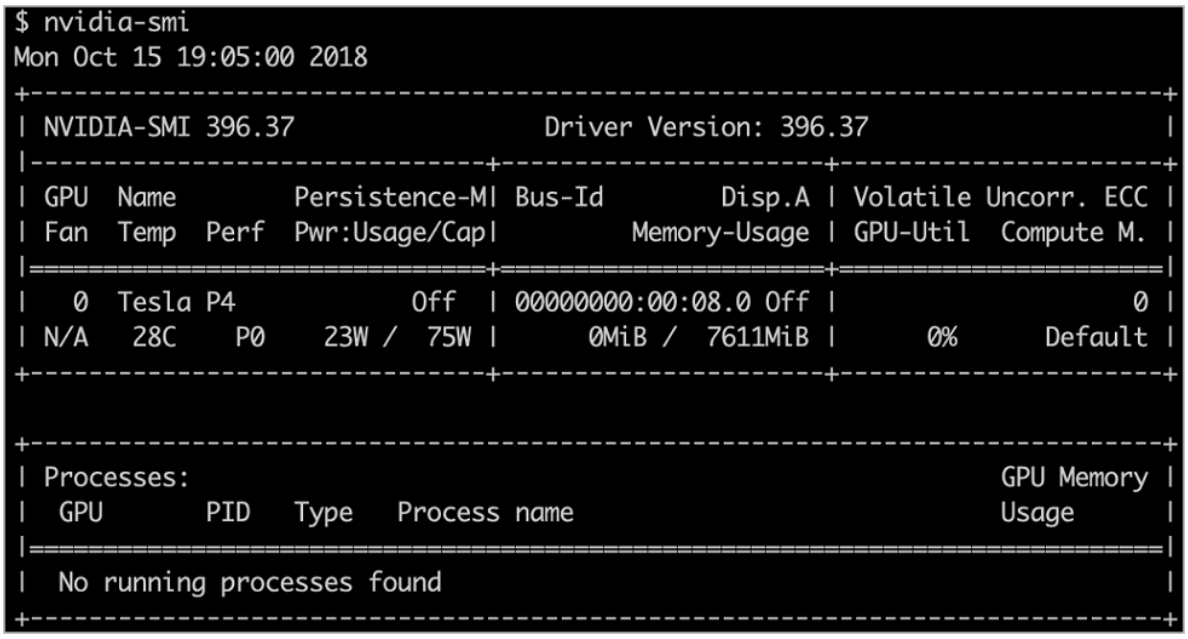
3. Copy the downloaded *cuda_9.2.148_396.37_linux.run* file to the instance. *cuda_9.2.148_396.37_linux.run* is used as an example. Use the actual name of the downloaded file.
4. Run the **sudo sh ./cuda_9.2.148_396.37_linux.run --silent --verbose --driver --toolkit --samples** command to install the CUDA. *cuda_9.2.148_396.37_linux.run* is used as an example. Use the actual name of the downloaded file.
   The installation takes 10 to 20 minutes. When `Driver: Installed` is displayed, the installation succeeds.

Figure 11-20: CUDA installation result

5. Run the **`nvidia-smi`** command to view the GPU driver status.

   If GPU driver details are displayed, the driver is in the normal state.

Figure 11-21: View the GPU driver status



What's next

If you want to run OpenGL programs, you must purchase a license and GRID drivers. For more information about the installation procedure, see official NVIDIA documentation.

11.4.15 Install the CUDA and GPU drivers for a Windows instance

The device where a GPU instance runs must be installed with the GPU driver. If the image you use does not contain a pre-installed GPU driver, you must install the CUDA and GPU drivers for the instance.

Context

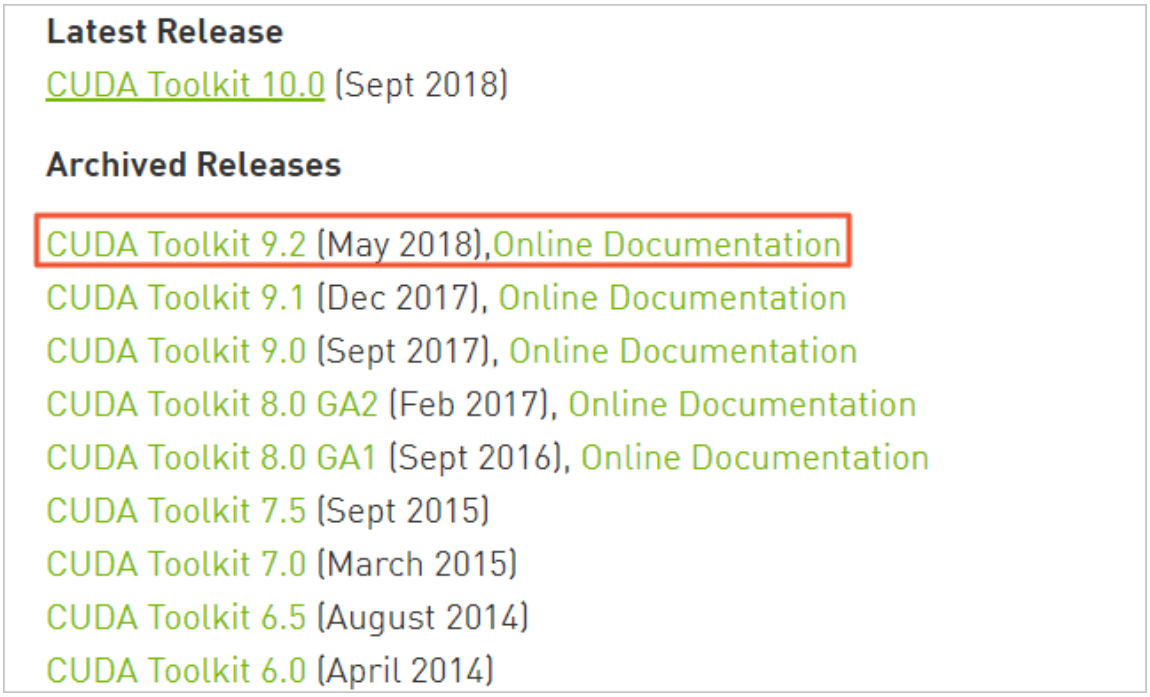
If you want to compile CUDA programs, first install a Windows compiling environment, such as Visual Studio 2015.

Procedure

1. Download the CUDA Toolkit.
   a) Access the *official download page*. Choose a suitable version based on the GPU application requirements for CUDA.

      You can choose *CUDA Toolkit 9.2*.

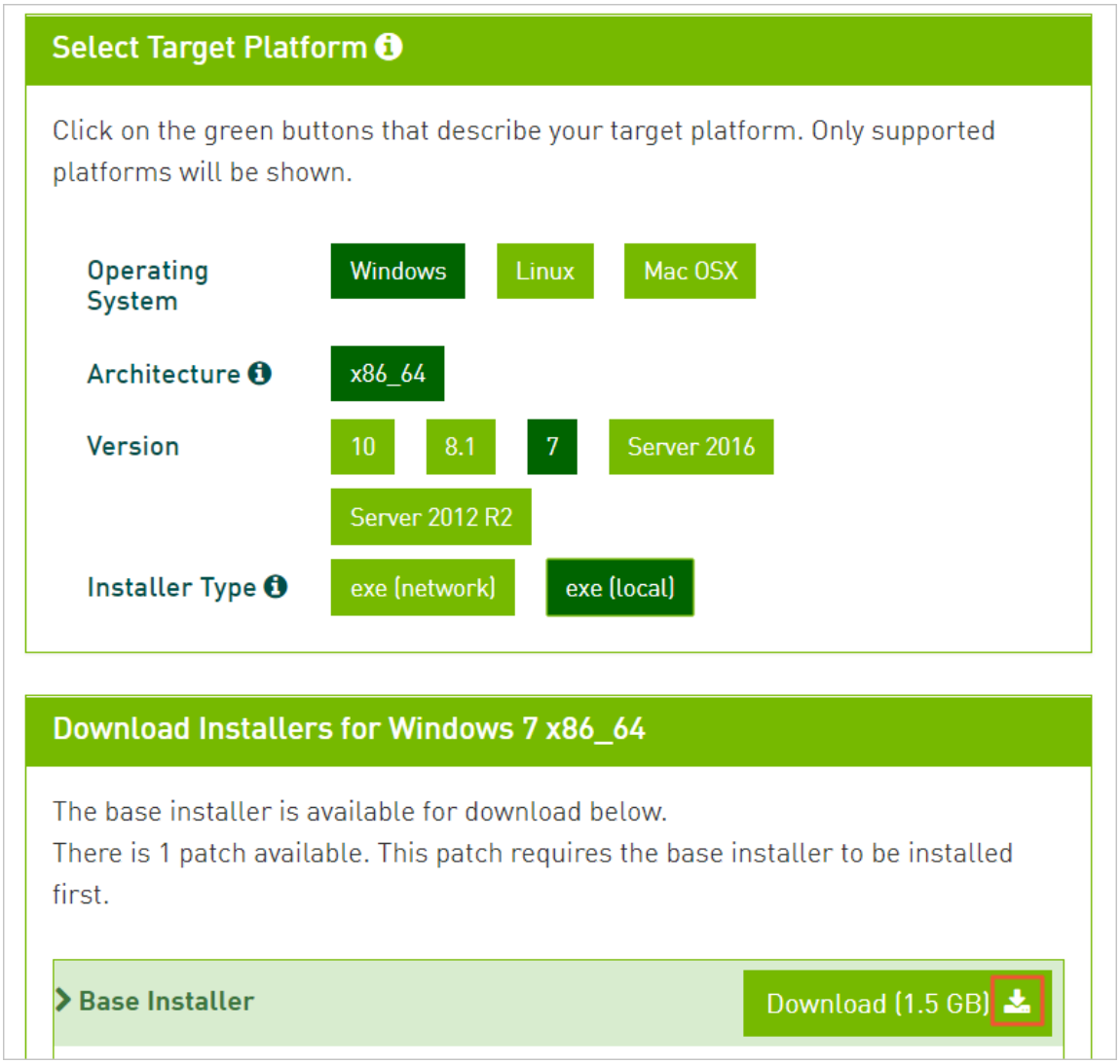
      Figure 11-22: Download the CUDA Toolkit

   b) Choose a platform based on your operating system. Set **Installer Type** to **exe (local)** and click **Download**.

      NVIDIA drivers are already contained in the CUDA Toolkit. You do not need to download them separately.

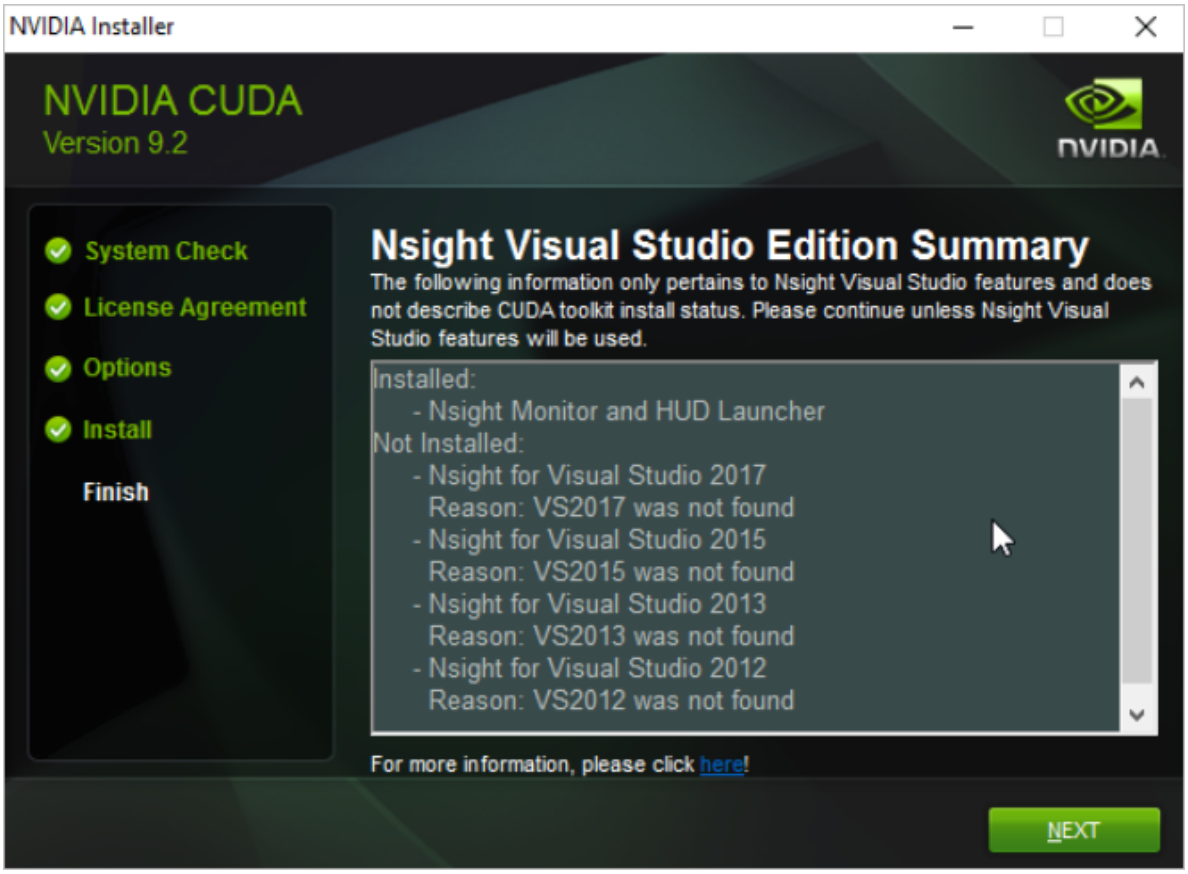
Figure 11-23: Download drivers



2. Copy the downloaded *cuda_9.2.148_windows.exe* file to the instance. *cuda_9.2.148_windows.exe* is used as an example. Use the actual name of the downloaded file.
3. Double-click *cuda_9.2.148_windows.exe* and follow the installation wizard to install the CUDA. *cuda_9.2.148_windows.exe* is used as an example. Use the actual name of the downloaded file.

   The installation takes 10 to 20 minutes. When `Installed: - Nsight Monitor and HUD Launcher` is displayed, the installation succeeds.

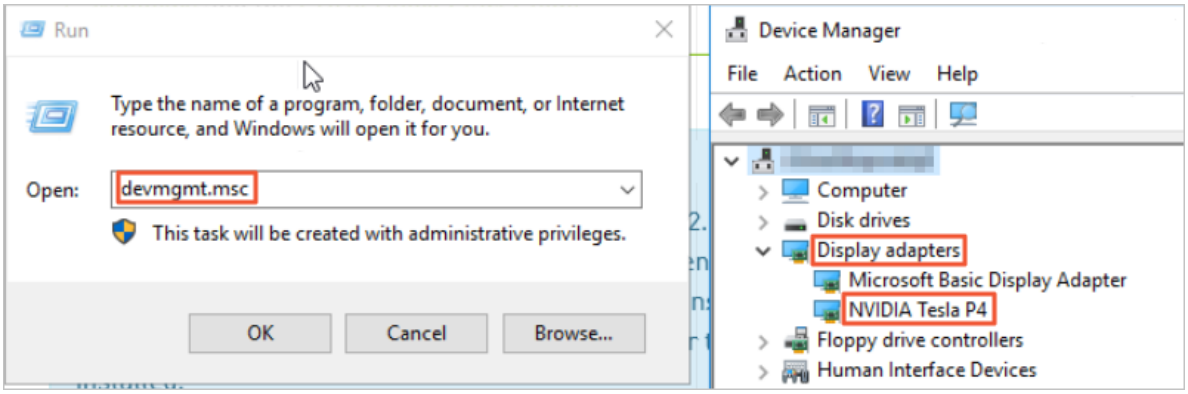
Figure 11-24: CUDA installation result



4. Press `Win+R` and enter **`devmgmt.msc`**.

   The NVIDIA device is displayed under **Display Adapter**.

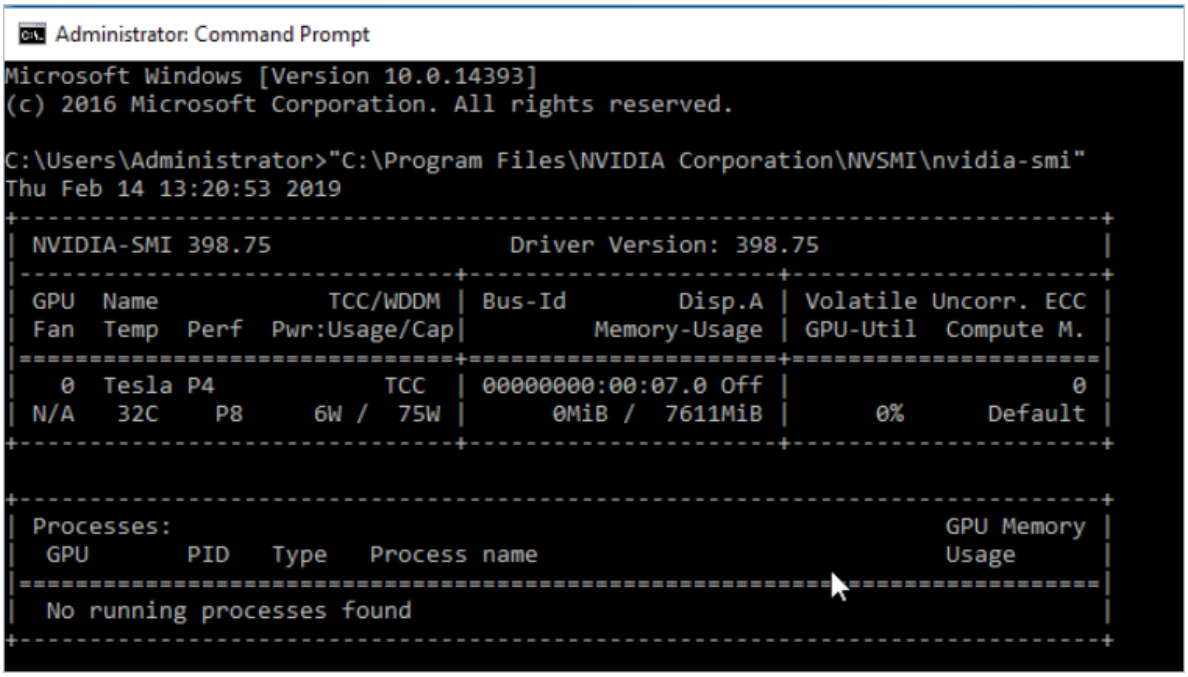
Figure 11-25: CUDA installation result



5. Press `Win+R`, enter **`cmd`**, and run **`"C:\Program Files\NVIDIA Corporation\NVSMI\nvidia-smi"`**.

   If GPU driver details are displayed, the driver is in the normal state.

Figure 11-26: View the GPU driver status



What's next

If you want to run OpenGL and DirectX programs, you must purchase the licenses and GRID drivers. For more information about the installation procedure, see official NVIDIA documentation.

11.5 Disks

11.5.1 Overview

For ECS instances, a cloud disk can be seen as a physical disk. You have to mount and format a cloud disk before using it.

Disk types

ECS disks can be classified into basic cloud disks, SSD cloud disks, and ultra cloud disks. A mount point is the position of an ECS disk on the disk controller bus. The selected mount point corresponds to the disk device number in Linux, and is consistent with the disk sequence in the disk manager in Windows.

Distributed storage

Snapshots and images are stored on the OSS in the same region. To guarantee the service flexibility and resource utilization, when you create a disk from a snapshot or image, the distribute d storage does not copy all the data to the new disk at one time. Rather, a data block is copied

when it is needed, thus optimizing the storage utilization. A data block is usually several MiBs in size and next time when the block is read or written, operations are performed on the disk directly.

Meanwhile, to guarantee the optimal I/O experience, the distributed storage copies the data of a snapshot or image to the disk step by step and block by block via backend duplication when there are less I/O operations.

Therefore, the first time a cloud disk is read or written, its I/O performance may decrease noticeably. However, once the disk has been accessed, its I/O performance will return to normal. Before high-load running, it is recommended to access the entire disk, such as reading the disk.

Disk features

A cloud disk has the following features:

- High data security.
- High IOPS for random and sequential reads/writes.
- Up to 17 disks can be added to an instance (incluiding the system disk and data disks).
- Upon downtime migration, data immediately before the downtime is saved.

11.5.2 Create disks

This section introduces how to separately create a new empty disk on the ECS console.

Context

You can create a block storage on the ECS console to resize the system storage space.

- One instance can have up to 16 data disks (this quota includes both cloud disks and shared block storages).
- A shared block storage can be attached to more than two ECS instances simultaneously. In the current version, one shared block storage can be attached to four ECS instances.
- Each ultra cloud disk (or ultra shared block storage) or SSD cloud disk (or SSD shared block storage) supports up to 32 TB capacity.

Note:

- Currently, ECS instances do not support combining multiple cloud disks. After creation, each cloud disk is an independent entity. The space of multiple cloud disks cannot be combined through formatting. We recommend that you plan the disk quantity and capacity in advance.

- A snapshot is intended for an independent disk, so data may be different after you perform snapshot rollback under the Logical Volume Management (LVM). Thus, if you have created multiple disks, we do not recommend that you configure LVM for them.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.
2. Click **Create Disk**.
3. You can view the following configuration on the **Create Disk** page:

Table 11-5: Disk configuration

Item	Description
Region	• **Region**: It is a required parameter. Select the region in which the disk resides. • **Zone**: It is a required parameter. Select the zone in which the disk resides.
Configurations	• **Name**: It is a required parameter. Enter the name of the disk. • **Department**: It is a required parameter. Select the department in which the disk resides. • **Project**: It is a required parameter. Select the project in which the disk resides. • **Storage**: It is a required parameter. Select the specific storage type of the disk, which can be **Disk** or **Shared block storage.** • **Type**: It is a required parameter. After selecting the storage type, you can specify the disk type as **Ultra cloud disk** (ultra block storage) or **SSD cloud disk** (SSD block storage) as needed. • **Encrypted**: It is an optional parameter. It specifies whether the created disk will be encrypted. • **Use Snapshots**: It is an optional parameter. After checking **Use Snapshots**, you still need to select the corresponding snapshot. **Note:** — If you have checked **Encryption** for the disk in the previous option, this option does not appear.

Item	Description
	— If the disk size specified by the user is smaller than the selected snapshot size, the disk size actually generated will be in line with the snapshot size; otherwise, the disk size will be the value specified by the user.

4. Click **Confirm**.

Result

In the disk list, check whether the disk is in the **`Available`** state. If so, the disk is successfully created.

What's next

The procedure varies depending on the operating system of the instance.

- If the Linux operating system is selected for the instance, you must *Attach a disk* and then *Partition, format, and attach data disks in Linux*.
- If a Windows operating system is selected for the instance, you must *Attach a disk* and then *Partition and format data disks in Windows*.

11.5.3 View disks

You can use the ECS console to view existing disks and related information.

Procedure

1. *Log on to the ECS console*.
2. Click the **Disks** tab to go to the **Disks** tab page.
3. Set **Department**, **Region**, and a specific filter criterion. Click **Search**.

   Note:
   Filter criteria include: **Disk Name**, **Disk ID**, **Disk Status**, **Disk Usage Type**, and **Project Name**.

4. Click the  icon in the **Actions** column corresponding to a disk and choose **View Details** from the shortcut menu. On the **Disk Details** tab page that appears, view disk details.

11.5.4 Roll back a disk

When you want to roll back the data on a disk to a previous time point, you can do so through disk rollback.

Prerequisites

Make sure that the instance of the target disk is in the **Stopped** state.

> **Note:**
> Snapshot rollback is irreversible. After rollback is finished, the original data cannot be restored. Exercise caution when performing this operation.

Procedure

1. *Log on to the ECS console* and go to the **Snapshots** page.
2. On the **Snapshots** page and in the **Action** column of the snapshot, click  and select **Rollback Disk** from the drop-down list.
3. In the confirmation box that appears, click **Confirm**.

> **Note:**
> If you select **Start instance right after rollback**, the instance will start automatically after the cloud disk is rolled back successfully.

11.5.5 Edit disk attributes

You can edit disk attributes on the ECS console.

Procedure

1. *Log on to the ECS console*.
2. On the **Disks** page, select the disk for which you want to modify the attributes. In the **Action** column of the disk, click the  icon and select **View Details** from the drop-down list to go to the **Disk Details** page.

> **Note:**
> You can also click the disk ID to go to the **Disk Details** page.

3. Click **Modify Properties**.

   You can set the following disk attributes:

- **Disk Name**: It is a string of 2 to 128 characters, including numbers, periods (.), underscores (_), and hyphens (-). It must begin with an uppercase or lowercase English letter or a Chinese character.
- **Disk Description**: It is a string of 2 to 256 characters. It cannot start with 'http://' or 'https://'.

4. Click **Confirm**.

11.5.6 Attach a disk

11.5.6.1 Overview

After creating a disk, you need to attach it to an instance. You can only attach independent cloud disks to ECS instances.

Note:

- When you attach a cloud disk to an ECS instance, you must check that the ECS instance is in the **Running** or **Stopped** state and the instance's security control indicator is not in the **Locked** state.
- When you attach an independent cloud disk, the cloud disk must be in the **Available** state.
- You can attach up to 16 data disks (including cloud disks and shared block storages) to one ECS instance.
- You must attach an independent cloud disk to an instance in the same zone.
- You can attach an independent cloud disk only as a data disk, but not a system disk.

You can attach a disk in either of the following ways:

- *Attach a disk on the Instance Details page*.
- *Attach a disk on the Disk List page*.

11.5.6.2 Attach a disk on the Instance Details page

To attach multiple disks to an instance on the ECS console, it is more convenient to do it on the **Instance Details** page.

Prerequisites

- Before you attach a disk, complete the following operation: *Create disks*.
- When you attach a data disk, ensure the cloud disk is in the **Available** state.

Context

- You can only attach data disks, not the system disk.
- You do not need to attach data disks that are created along with an instance.
- A disk can only be attached to an instance that is in the same zone under the same region as the disk is.
- An ECS instance can have up to 16 data disks attached. One disk can only be attached to one instance.
- An independently created disk can be attached to any instance in the same zone under the same region.

Procedure

1. *Log on to the ECS console* and go to the **Instances** page.
2. On the **Instances** list page, click the ID of the ECS instance to which you want to attach a disk.
3. Enter the **Instance Details** page and click the **Disks** tab.
4. Click **Attach**.
5. In the displayed dialog box, provide the following information:
   - **`Target Disk`**: It is a required parameter. Select an existing cloud disk that is in the **`Available`** state.
   - Select **`Deleted with Instance`**.

     Note:
     By default, the value of this option is **`No`**. If you select **`Yes`**, when the instance is deleted, the disk will be deleted together.
6. Click **Submit**.

11.5.6.3 Attach a disk on the Disk List page

To attach multiple disks to different instances on the ECS console, it is more convenient to do it on the **Disk List** page.

Prerequisites

- Before you attach a disk, complete the following operation: *Create disks*.
- When you attach a data disk, ensure the cloud disk is in the **`Available`** state.

Context

- You can only attach data disks, not the system disk.
- You do not need to attach data disks that are created along with an instance.

- A disk can only be attached to an instance that is in the same zone under the same region as the disk is.
- An ECS instance can have up to 16 data disks attached. One disk can only be attached to one instance.
- An independently created disk can be attached to any instance in the same zone under the same region.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.
2. In the **Action** column of the disk to be attached, click the icon  and choose **Attach**.
3. In the **Attach** dialog box, complete the following settings:
   - **`Destination Instance`**: Select the ECS instance to which you want to attach the selected cloud disk.
   - Select **`Are you sure you want to configure the disk to delete along with the instance?`**. The default value is **`No`**. Set it to **`Yes`** if you want to release the disk when the instance is deleted.
4. Click **Submit**.

11.5.7 Partition and format disks

11.5.7.1 Overview

You can Partition and format disks on the ECS console in the Windows or Linux environment.

ECS supports only secondary partitioning of **`Data Disks`**, but not **`System Disks`** (either in the Windows or Linux operating system). If you use a third-party tool to perform secondary partitioning of a system disk, you may encounter unknown risks, such as system crash and data loss.

Note:

Before you attach a data disk, *Create disks*.

11.5.7.2 Partition, format, and attach data disks in Linux

This section introduces how to partition, format, and attach data disks for Linux instances.

Prerequisites

- Complete *Connect to an instance*.
- Complete *Create disks* and *Attach a disk*.

Procedure

The data disks of Linux ECS instance are not partitioned or formatted. You can follow these steps to partition and format the data disks:

1. View the data disks. Before you partition and format the data disks, run the `fdisk -l` command (instead of `df –h`) to view the data disks.

   The output of the `fdisk -l` command shows information about the data disks, such as */dev/vdb* in the following figure. If */dev/vdb* is not displayed, the ECS instance has no data disks and you do not have to attach data disks.

```
[root@iZ********eZ ~]# fdisk -l

Disk /dev/vda: 42.9 GB, 42949672960 bytes
255 heads, 63 sectors/track, 5221 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x00078f9c

Device Boot      Start         End      Blocks   Id  System
/dev/vda1   *           1        5222    41940992   83  Linux

Disk /dev/vdb: 21.5 GB, 21474836480 bytes
16 heads, 63 sectors/track, 41610 cylinders
Units = cylinders of 1008 * 512 = 516096 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x00000000
```

2. Partition the data disks. Run the `fdisk /dev/vdb` command to partition the data disks, as shown in the preceding figure. Enter the following commands in sequence as prompted:

   a) `n` command: Creates a partition.

   b) `p`: Creates a primary partition.

   c) **`Partition number (1 to 4)`**: Number of the new partition, an integer in the range from 1 to 4. You can create up to four partitions. In this example, 1 is entered to indicate Partition 1.

   d) **`First cylinder`**: Start position of the partition. You can select the default value by pressing the **Enter** key. Also, you can enter a number in the range from 1 to 41610 and then press the **Enter** key. In this example, the default value 1 is used.

   e) **`Last cylinder`**: End position of the partition. You can select the default value by pressing **Enter**. Also, you can enter a number in the range from 1 to 11748 and then press the Enter key. In this example, the default value is used.

   f) Optional: If you want to create multiple partitions, you can repeat Steps a through e until all the four partitions are configured.

g) Run the `wq` command to start partitioning.

```
[root@iZ********eZ ~]# fdisk /dev/vdb
Device contains neither a valid DOS partition table, nor Sun, SGI or
 OSF disklabel
Building a new DOS disklabel with disk identifier 0x01ac58fe.
Changes will remain in memory only, until you decide to write them.
After that, of course, the previous content won't be recoverable.

Warning: invalid flag 0x0000 of partition table 4 will be corrected
 by w(rite)

WARNING: DOS-compatible mode is deprecated. It's strongly
 recommended to
        switch off the mode (command 'c') and change display units
 to
        sectors (command 'u').

Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-41610, default 1):
Using default value 1
Last cylinder, +cylinders or +size{K,M,G} (1-41610, default 41610):
Using default value 41610

Command (m for help): wq
The partition table has been altered!
```

3. View the new partition. Run the `fdisk -l` command to list all the partitions, as shown in *code*.

   If the command output shows */dev/vdb1*, the partition *vdb1* is successfully created.

```
[root@iZ********eZ ~]# fdisk -l

Disk /dev/vda: 42.9 GB, 42949672960 bytes
255 heads, 63 sectors/track, 5221 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x00078f9c

Device Boot      Start         End      Blocks   Id  System
/dev/vda1   *           1        5222    41940992   83  Linux

Disk /dev/vdb: 21.5 GB, 21474836480 bytes
16 heads, 63 sectors/track, 41610 cylinders
Units = cylinders of 1008 * 512 = 516096 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x01ac58fe
Device Boot      Start         End      Blocks   Id  System
/dev/vdb1               1       41610    20971408+  83  Linux
```

4. Format the new partition. For example, you can run the `mkfs.ext3 /dev/vdb1` command to format the new partition as ext3. The time required for formatting varies depending on the hard

disk size. You can also format the new partition as another file system type. For example, you can run the `mkfs.ext4 /dev/vdb1` command to format it as ext4.

Note:

Compared with ext2, ext3 only adds the log function. Compared with ext3, ext4 improves some important data structures. ext4 provides better performance and reliability, and more diverse functions.

```
[root@iZ********leZ ~]# mkfs.ext3 /dev/vdb1
mke2fs 1.41.12 (17-May-2010)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
Stride=0 blocks, Stripe width=0 blocks
1310720 inodes, 5242852 blocks
262142 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=4294967296
160 block groups
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
 32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632,
 2654208,
 4096000

Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 25 mounts or
180 days, whichever comes first.  Use tune2fs -c or -i to override.
```

5. Add partition information. Run the `echo '/dev/vdb1 /mnt ext3 defaults 0 0' >> /etc/fstab` command to add information about the new partition and then run the `cat /etc/fstab` command to view the partition information.

Note:

- This example adds partition information to the ext3 file system. You can add partition information to another file system type, such as ext4.
- Ubuntu 12.04 does not support barriers. Therefore, in Ubuntu 12.04, the `echo '/dev/vdb1 /mnt ext3 barrier=0 0 0' >> /etc/fstab` command is used to add partition information.

```
[root@iZ********eZ ~]# echo '/dev/vdb1 /mnt ext3 defaults 0 0' >> /etc
/fstab
[root@iZbp19cdhgdj0aw5r2izleZ ~]# cat /etc/fstab
```

```
#
# /etc/fstab
# Created by anaconda on Thu Aug 14 21:16:42 2014
#
# Accessible filesystems, by reference, are maintained under '/dev/
disk'
# See man pages fstab(5), findfs(8), mount(8) and/or blkid(8) for more
 info
#
UUID=94e4e384-0ace-437f-bc96-057dd64f42ee / ext4 defaults,barrier=0 1
 1
tmpfs                   /dev/shm                tmpfs   defaults
  0 0
devpts                  /dev/pts                devpts  gid=5,mode=620
  0 0
sysfs                   /sys                    sysfs   defaults
  0 0
proc                    /proc                   proc    defaults
  0 0
/dev/vdb1 /mnt ext3 defaults 0 0
```

To attach the data disk to a folder separately, for example, to store webpages separately, modify */mnt* of the preceding command.

6. Attach the new partition. Run the `Run mount -a` command to attach all the partitions listed in */etc/fstab* and then run the `df -h` command to check the attachment. If the following information is displayed, the partitions are successfully attached and the new partitions are available for use.

```
[root@iZ********eZ ~]# mount -a
[root@iZ********eZ ~]# df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/vda1        40G  5.6G   32G  15% /
tmpfs           499M     0  499M   0% /dev/shm
/dev/vdb1        20G  173M   19G   1% /mnt
```

11.5.7.3 Partition and format data disks in Windows

This section describes how to partition and format the data disks of a Windows instance.

Prerequisites

- Complete *Connect to an instance*.
- Complete *Create disks* and *Attach a disk*.

Context

The operations mentioned in this section only apply to Windows 2008.

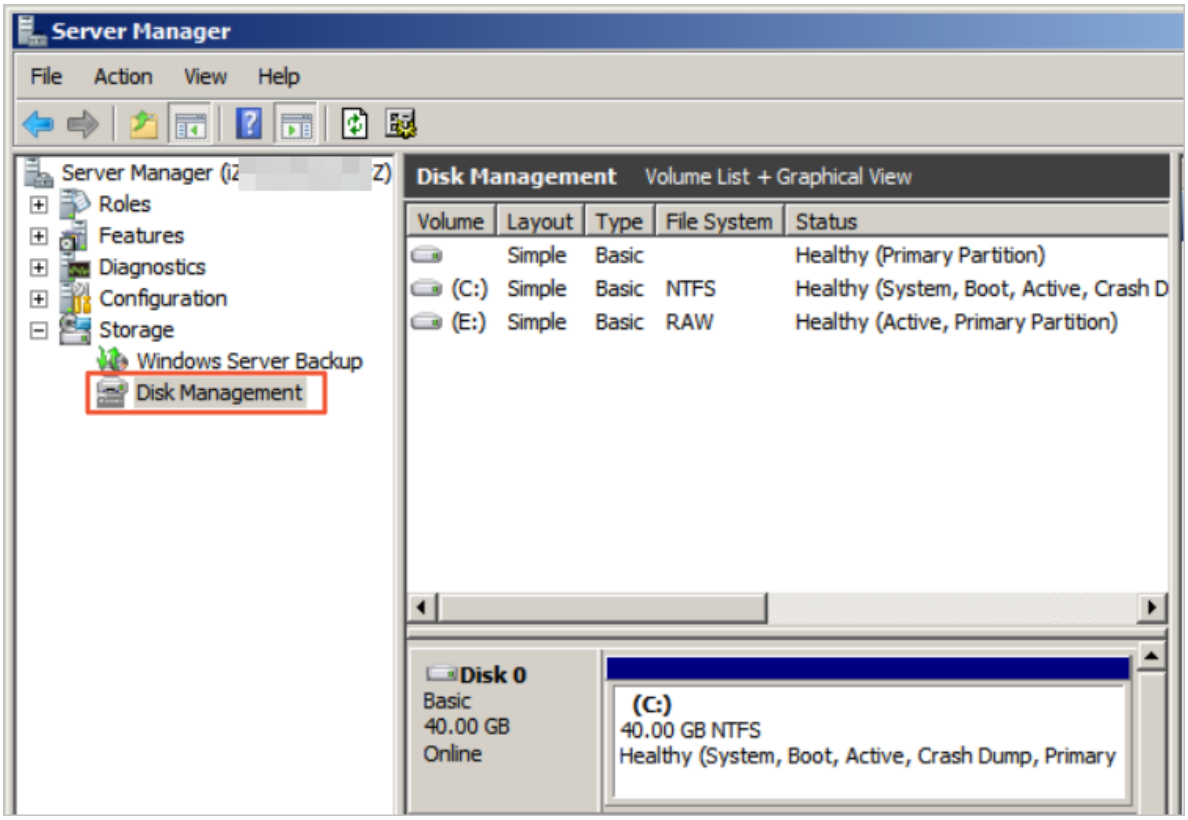
Procedure

Note:

If your data disks are in the **Offline** state, change them to the **Online** state before you allocate volume numbers and capacities to the data disks.

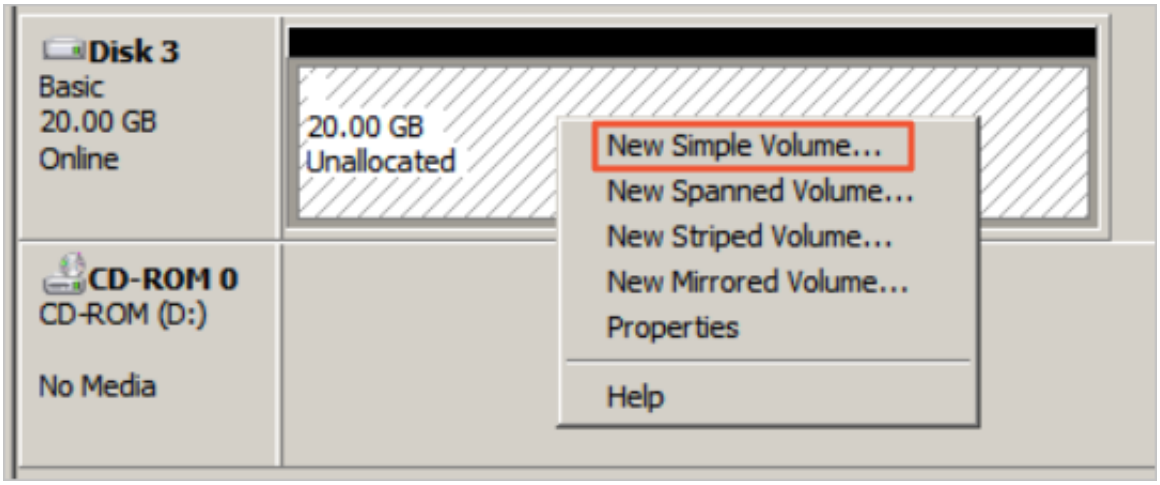
1. Click **Server Manager** on the toolbar in the lower-left corner to start the server manager.
2. On the left-side navigation bar of the **Server Manager** window, choose **Storage** > **Disk Management**.

Figure 11-27: Manage disks



3. Right-click an empty partition and choose **New Simple Volume** from the shortcut menu.

Figure 11-28: Choose the "New Simple Volume" option

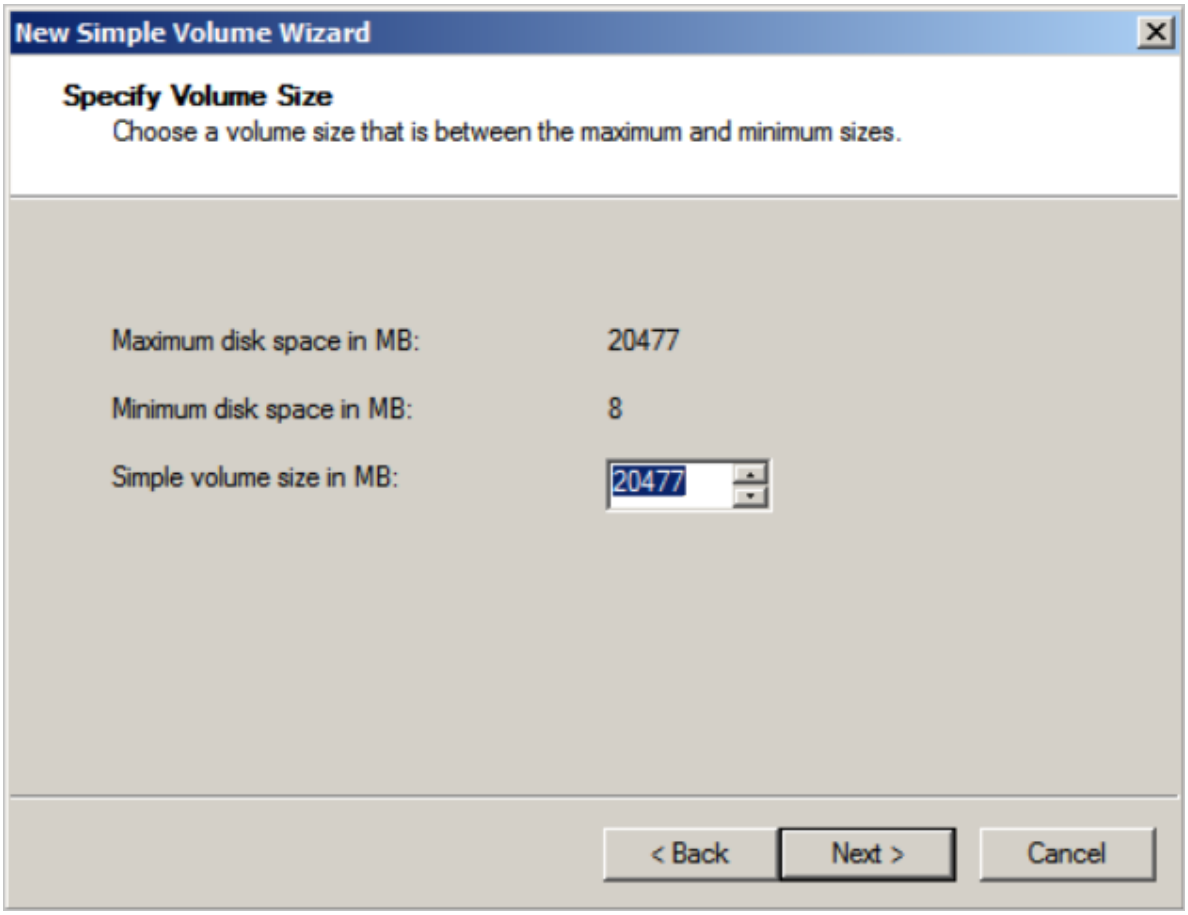


4. The "New Simple Volume" wizard appears. Click **Next**.

Figure 11-29: "New Simple Volume" wizard

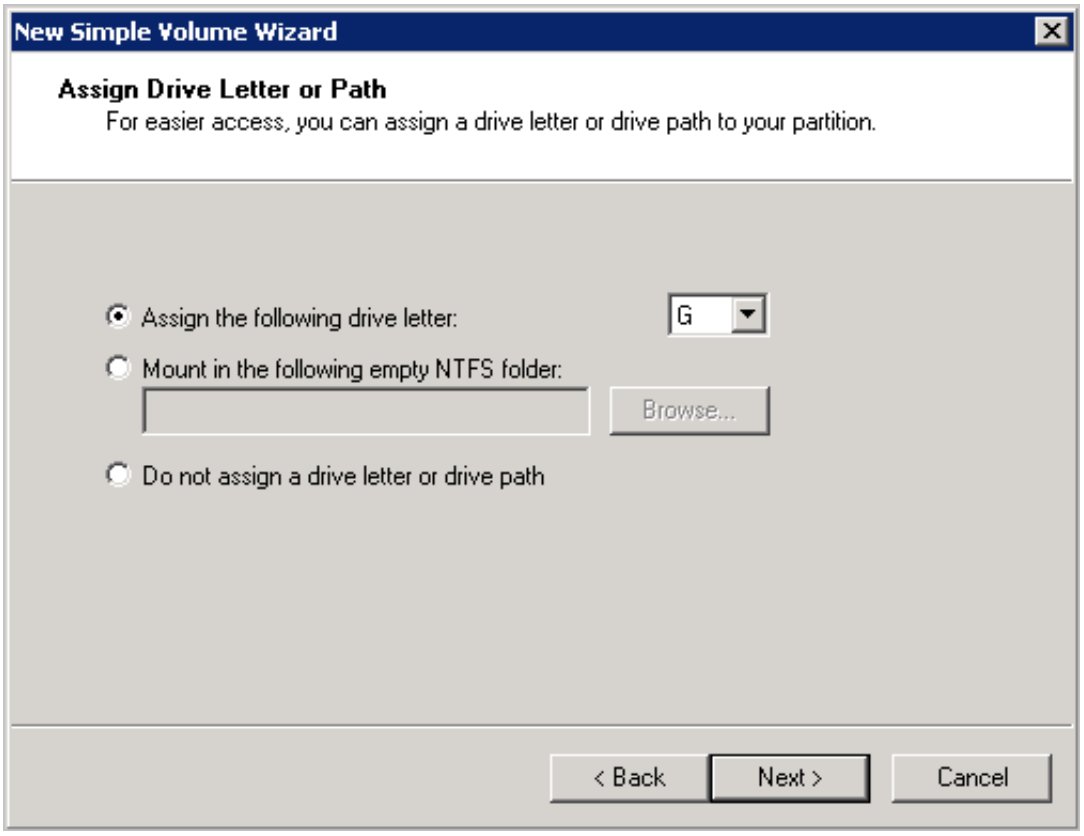
5. Set the size of the simple volume, that is, the partition size. The default value is **`Maximum Disk Space`**. You can specify the partition size as needed. After you complete the settings, click **Next**.

   Figure 11-30: Set the partition size



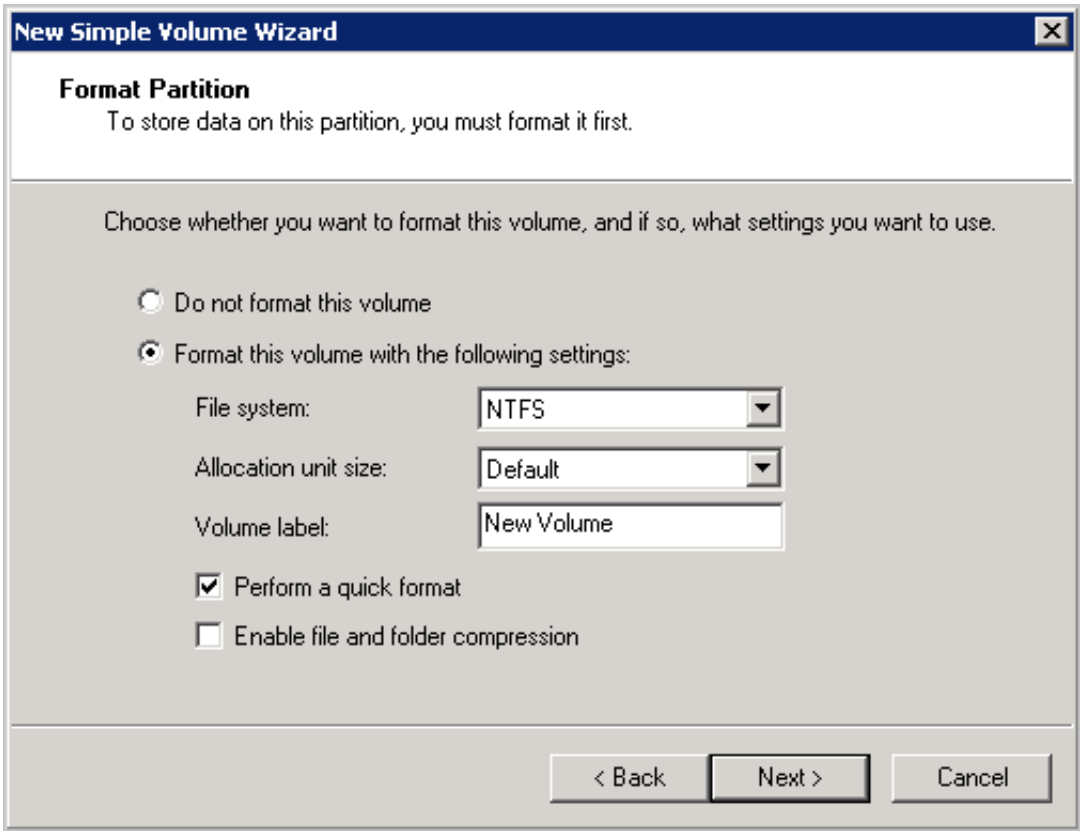
6. Specify the drive letter, which is listed in the alphabetic order by default. Click **Next**.

Figure 11-31: Allocate the drive letter



7. Format the partition. We recommend that you format the partition using the default settings of the wizard. After you complete the settings, click **Next** to start formatting.

Figure 11-32: Format the partition



8. When the wizard prompts that partitioning is complete, click **Finish** to close the wizard. The partition is successfully created.

11.5.8 Resize a system disk

11.5.8.1 Overview

As the business expands, you may need to increase the system disk size. In that case, you can do it by replacing the system disk.

Note:
Read the following Precautions before resizing the system disk.

Risks

- This operation requires you to stop your instance, which means interruption of your business.
- After replacement, you must redeploy the business runtime environment on the new system disk. There is a possibility of long interruption of your business. Extreme caution should be exercised when performing this operation.

- Your manually created snapshots are retained after the system disk is resized. However, because the disk ID is changed, you can no longer use the manually created snapshots on the original system disk to roll back the new system disk. The retained snapshots can still be used to create custom images.
- To retain enough snapshot quota for the automatic snapshot policy of the new disk, you can *delete unnecessary snapshots*.
- The original system disk is released after resizing.

Notes

- When resizing a system disk, you cannot reduce the disk capacity, but can only increase or keep the disk capacity.
- Resizing the system disk will not change the IP address and MAC address of your instance.
- The system disk type cannot be changed.
- Windows 2003 does not support system disk resizing.

Procedure of system disk resizing

If you are sure to resize a system disk, follow these steps:

1. *Create a snapshot for a system disk*.
2. *Create an image from a snapshot*.
3. *Change a system disk*.
4. *Set a snapshot policy for a system disk*.

11.5.8.2 Create a snapshot for a system disk

If it is necessary to save the system disk data before the expansion, please create a snapshot for it.

Prerequisites

Make sure that the instance of the target disk is in the **`Stopped`** state.

Context

If you do not want to retain the data on the system disk, skip this step and proceed to *Change a system disk*. To avoid impact on your business, do not create snapshots during traffic peak periods. It takes about 40 minutes to create a 40 GB snapshot for the first time. Reserve enough time. When you create a snapshot, check that the system disk has sufficient space. We recommend that you reserve **1 GB** space. Otherwise, the system may not start properly after the system disk is resized.

Procedure

1. *Log on to the ECS console* and go to the **Instances** page.
2. Select a **Department** and **Region** or enter an **Instance Name** and click "Search" to find the target instance.
3. Click the instance whose system disk you want to replace, or in the **Action** column of the instance, click  and choose **View Details** from the drop-down menu to go to the **Instance Details**page.
4. Click the **Disks** tab.
5. Find the system disk. In the **Action** column of the system disk, click the  icon and select **Create Snapshot** from the drop-down list.
6. In the **Create Snapshot** dialog box that appears, enter a name for the snapshot and click **Confirm** to create the snapshot.

   Note:
   The name of a snapshot cannot start with **auto** because **auto** has been reserved as the name prefix for the snapshot that the system automatically creates for you.
7. Click the **Instance Snapshot** tab, and you can view the snapshot creation progress and status. When the **Progress** is 100%, the snapshot is successfully created.

11.5.8.3 Create an image from a snapshot

To continue using the current operating system and keep its data, please create an image after the snapshot creation.

Context

- If you do not want to continue using the current operating system or retaining its data, skip this procedure and proceed to *Change a system disk*.
- If you want to continue using the current system disk, you need to make an image based on the current system disk. After the system disk is resized, you can completely copy all the data to a new environment.
- You can perform an alternative operation to create an image of the system disk. For details, see *Create custom images from snapshots*.

Note:

When you create an image, check that the system disk has sufficient space. We recommend that you reserve 1 GB space. Otherwise, the system may not start properly after the system disk is resized.

Procedure

1. *Log on to the ECS console*.
2. On the **Instances** page, select a **Department** and **Region** or enter an **Instance Name** and click **Search** to find the target instance.
3. In the **Action** column of the instance, click  and select **View Details** to go to the **Instance Details** page.

   Note:
   On the **Instances** page, you can also click an instance ID to go to the **Instance Details** page.
4. Click the **Instance Snapshot** tab. In the **Action** column of the target snapshot, click  and choose **Create Custom Image** from the drop-down menu.
5. In the **Create Custom Image** dialog box, enter a name and description for the custom image and click **Confirm**.

   Note:
   - Remember the image name. You must select the custom image when replacing the system disk.
   - Do not select "Add Data Disk Snapshot". Selection of data disks is not supported during the system disk replacement.

Result

After the image is successfully created, it is displayed on the **Image** page.

11.5.8.4 Change a system disk

Changing a system disk refers to allocating a new system disk. The system ID is updated and the original system disk is released.

Prerequisites

- To avoid data loss, back up all related data of the original system disk.
- Ensure that the instance where you want to change the system disk is in the **Stopped** state.

Context

Changing a system disk causes the following impacts:

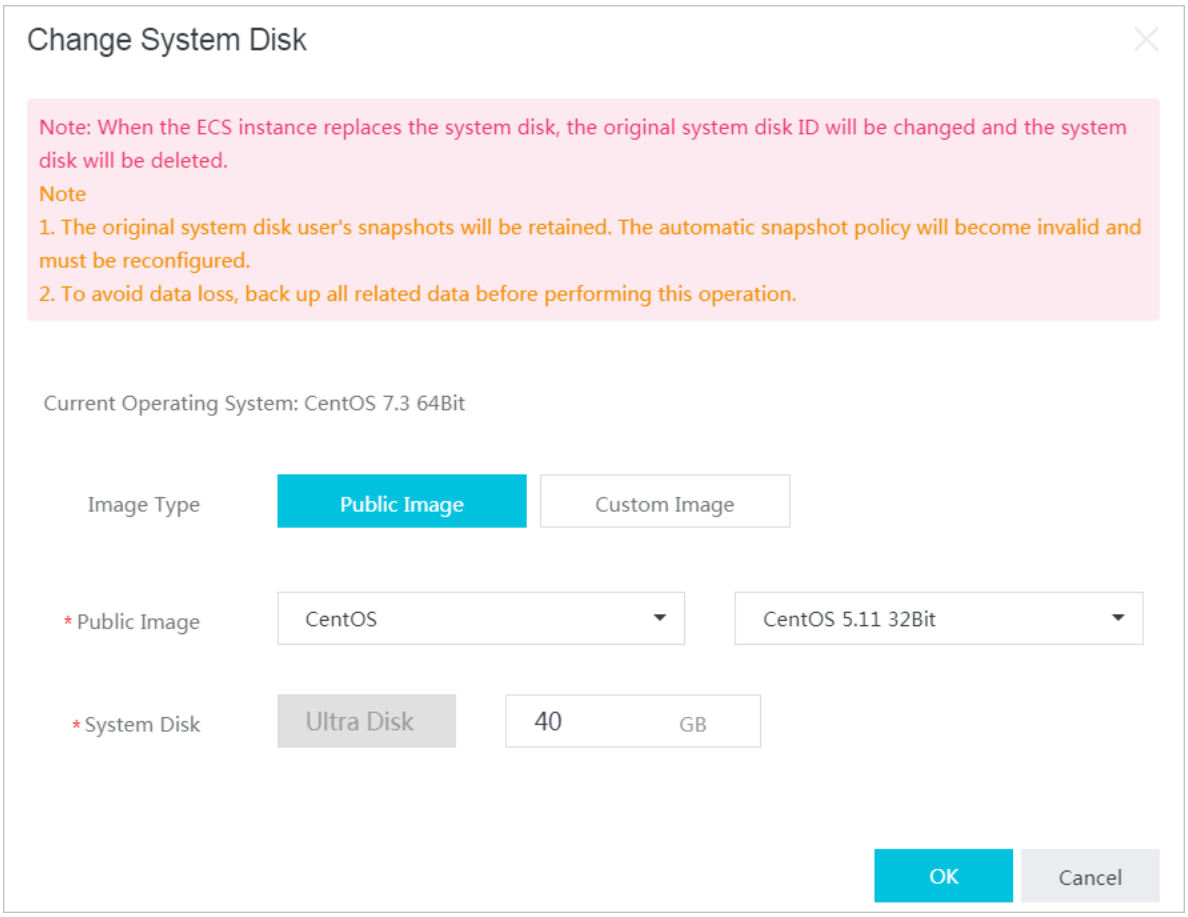
- After the system disk is changed, user snapshots of the original system disk are retained. The automatic snapshot policy becomes invalid and must be reconfigured.
- After the system disk is changed, the system disk ID changes and the original system disk is deleted.

Procedure

1. *Log on to the ECS console*.
2. Click the **Instances** tab to go to the **Instances** tab page.
3. Click the  icon in the **Actions** column corresponding to an instance and choose **View Details** from the shortcut menu.
4. On the **Instance Details** tab page that appears, click **Change System Disk**.
5. In the **Change System Disk** dialog box that appears, perform the following operations:

   Note:
   Before changing the system disk, read the prerequisites and background information carefully.

   - **Image Type**: If you want to reserve data in the original system disk, select the custom image created in *Create an image from a snapshot*. Otherwise, select a public image.
   - **System Disk**: You cannot change the disk type, but can specify a new disk size. The new disk size cannot be less than the original disk size. It can be set to a maximum of 500 GB.



6. Click **OK**. The system disk is scaled up.

> **Note:**
> Go back to the ECS console to view the task status. It may take 10 minutes to process the change. After the system disk is changed, the instance automatically starts.

11.5.8.5 Set a snapshot policy for a system disk

After you replace a system disk, you must set a snapshot policy for the new system disk if automatic snapshotting is needed.

For more information, see *Configure an automatic snapshot policy*.

11.5.9 Detaching a disk

On the ECS console, you can detach a data disk rather than a system disk. Local disks cannot be detached.

Prerequisites

Pay attention to the following before detaching a data disk:

- In Windows, you need to log on to the instance and perform **Offline** operation for the disk via disk management. After the command is executed successfully, you can enter the console to detach the disk.

  Note:
  To ensure data integrity, we recommend that you pause read/write operations for all the file systems in this disk. Otherwise, the data that is not read or written completely may be lost.

- In Linux, you need to log on to the instance and run the `unmount` command for the disk. After the command is executed successfully, you can enter the console to perform the detach operation for the disk.

  Note:
  If you have enabled automatically attaching data disk partitions during instance startup, before detaching the data disk, you must delete the attaching information of the data disk partitions from the /etc/fstab file first. Otherwise, you cannot connect the instance after the instance is restarted.

- The data disk to be detached must be in the **`Running`** state.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.
2. On the **Disks** page, select the disk you want to detach. In the **`Action`** column, click the management icon and select **Uninstall** from the drop-down list.
3. In the **Uninstall Disk** dialog box that appears, select the instance to which the disk is attached, confirm the information, and then click **Submit**.

Result

In the disk list, check whether the disk is in the **`Available`** state. If so, the disk is successfully detached from the instance.

11.6 Images

11.6.1 Overview

An ECS image is a template that contains the software configurations such as the operating system, application server, and application programs of the ECS instance. When creating an instance, you must specify an ECS image. The operating system and software provided by the

ECS image are installed in the created instance. You can create a custom image based on a created instance and then create more instances based on the custom image.

11.6.2 Select a suitable image

To create an instance, you must select a suitable image.

When selecting an image, consider the following factors:

- Region and zone.
- Select the Linux or Windows operating system.

  The 512 MB memory specifications do not support the Windows operating system, while the 4 GB memory specifications and above do not support the 32-bit operating system.
- Select the 32-bit or 64-bit operating system.

When creating an instance, you can select a custom image or public image.

11.6.3 Create a custom image

11.6.3.1 Overview

You can create a custom image, and then use it to create ECS instances or replace the system disk of an ECS instance.

11.6.3.2 Create custom images from snapshots

You can create custom images from snapshots on the system disk to fully load the operating system and data environment information in the snapshots to the images.

Prerequisites

- The disk attribute of the snapshot must be system disk, and data disks cannot be used to create a custom image.
- Make sure the system disk in your instance has available snapshots.

Procedure

1. *Log on to the ECS console* and go to the **Snapshots** page.
2. Select the system disk snapshot from which you want to create an image. In the **Action** column, click  and select **Create Custom Image**.
3. In the **Create Custom Image** dialog box, enter the following configuration information:

- **Custom Image Name**: It is a required parameter. It is a string of 2 to 128 characters, including special characters such as periods (.), hyphens (-), and underscores (_). It must start with an uppercase or lowercase English letter.
- **Custom Image Description**: It is a required parameter. It is a string of 2 to 256 characters. It cannot start with http:// or https://.

4. After completing the configuration information, click **Confirm**.

11.6.3.3 Create a custom image from an instance

By creating a custom image based on an instance, you can fully replicate all disks of the instance, including the data on the system disk and data disks, to the custom image (full image).

Prerequisites

To avoid data security risks, delete sensitive data before creating a custom image.

Context

When you create a full image from an instance, each disk of the instance creates a snapshot automatically, and all the snapshots constitute a complete custom image.

Procedure

1. *Log on to the ECS console*.
2. Find the instance from which you want to create a custom image. In the **Action** column of the instance, click the  icon and select **Create Custom Image** from the drop-down list.
3. In the **Create Custom Image** dialog box that pops up, provide the following configuration information:
   - **Custom Image Name**: It is a required parameter. It is a string of 2 to 128 characters, including special characters such as periods (.), hyphens (-), and underscores (_). It must start with an uppercase or lowercase English letter or a Chinese character.
   - **Custom Image Description**: It is an optional parameter. It is a string of 2 to 256 characters and can start with http:// or https://.
4. After you complete the settings, click **Confirm**.

11.6.4 View images

On the ECS console, you can view the created images and relevant information.

Procedure

1. *Log on to the ECS console*.

2. Select a **Department** and **Region** or enter an **Image Name** and click **Search** to find a particular image and view its details.

> **Note:**
> Click **Image Name** and you can select below filtering conditions for your query: **Image ID** and **Image Type**.

11.6.5 Copy images

To copy a custom image from one region to another, you can use the function of copying images, which is suitable for deploying an application across regions or running the same image environment on ECS instances in different regions.

Context

The time consumed for copying an image depends on the network status and concurrent tasks in the queue.

Procedure

1. *Log on to the ECS console*.
2. In the **Action** column of the image to be copied, click the icon  , then select **Copy Image** in the drop-down menu.
3. In the **Copy Image** pop-up, you can see the ID of the custom image to be copied. Now configure the following items:
   - **Name**: mandatory. Specify the name of the custom image shown in the target region.
   - **Description**: optional. Provide the description of the custom image shown in the target region..

     > **Note:**
     > The description is 2 to 256 characters long and cannot start with "http://" or "https://".
   - Click **Confirm**.
4. Switch to the target region and you can see that the custom image is in the status of **Creating**. When the status is **Available**, the image is copied successfully.

11.6.6 Share images

You can share your custom images with other departments.

Prerequisites

Only custom images can be shared.

Procedure

1. *Log on to the ECS console* and go to the **Images** page.
2. In the **Action** column of the image to be shared, click the icon  and select **Share Image** from the drop-down list.
3. In the dialog box that appears, select the target **Department** and click **Confirm**.

11.6.7 Import images

11.6.7.1 Overview

On the ECS console, you can import images as needed. You can create instances by using a custom image, which may be created from your on-premise server, virtual machines or a cloud server of other cloud platforms.

11.6.7.2 Notes for importing images

Pay attention to the following notes when importing images to ensure the availability of imported images and improve the efficiency of the import operation.

Notes for importing Linux and Windows images:

- *Import Linux images*
- *Import Windows images*

Import Linux images

Imported Linux images have the following limits:

- Multiple network interfaces are not supported.
- IPv6 addresses are not supported.
- The password must be 8 to 30 characters in length. It must contain uppercase or lowercase letters, numbers, and special symbols.
- The firewall is disabled, and port 22 is enabled by default.
- The Linux system disk size is between 40 GiB and 500 GiB.
- DHCP must be enabled in the image.
- SELinux cannot be enabled.
- The Kernel-based Virtual Machine (KVM) driver must be installed.
- We recommend that you install cloud-init to configure the hostname and NTP and yum sources .

- The imported Red Hat Enterprise Linux (RHEL) image must have a BYOL license.

Table 11-6: Notes

Item	Standard operating system image	Non-standard operating system image
Definition	Supported standard operating systems (including 32-bit and 64-bit) include: • CentOS • Ubuntu • SUSE • OpenSUSE • Red Hat • Debian • CoreOS • Aliyun Linux	**Non-standard operating systems** include: • Operating systems that are not supported by Alibaba Cloud • Standard operating systems that do not meet the requirements of **critical system configuration files**, **basic system environments**, and **applications**. If you want to use non-standard operating system images, you must select Others Linux when importing images. If you import non-standard operating system images, Alibaba Cloud does not pre-configure the created instances. After you create an instance, you must connect to the instance by clicking **Connect to Management Terminal** in the ECS console. You can then configure the IP address, route, and password.
Critical system configuration files	• Do not modify `/etc/issue*`. Otherwise, the system release cannot be identified, leading to system creation failure. • Do not modify `/boot/grub/menu.lst`. Otherwise, the system may fail to start up. • Do not modify `/etc/fstab`. Otherwise, partitions cannot	Requirements of standard operating system images are not supported.

Item	Standard operating system image	Non-standard operating system image
	be loaded, leading to system startup failure. • Do not change `/etc/shadow` to **read-only**. Otherwise, the password file cannot be modified, leading to system creation failure. • Do not enable SELinux by modifying `/etc/selinux/config`. Otherwise, the system may fail to start up.	
Basic system environment requirements	• Do not adjust the system disk partition. Only a single root partition is supported. • Make sure that the system disk has a sufficient free space. • Do not modify critical system files, such as `/sbin`, `/bin`, and `/lib*`. • Before importing an image, confirm the integrity of the file system. • Only ext3 and ext4 file systems are supported.	
Applications	Do not install `qemu-ga` in a custom image. Otherwise, some of the services that Alibaba Cloud needs may become unavailable.	
Image file formats	Only RAW and VHD images can be imported. If you want to import images in other formats, use a tool to convert the format before importing the image. We recommend that you import images in VHD format which has a smaller transmission capacity.	

Item	Standard operating system image	Non-standard operating system image
Image file size	We recommend that you configure the disk size for importing images based on the virtual disk size (not the image file size). The disk size for importing images must be greater than or equal to 40 GiB.	

Import Windows images

Imported Windows images have the following limits:

- The password must be 8 to 30 characters in length. It must contain uppercase or lowercase letters, numbers, and special symbols.
- Imported Windows images do not provide the Windows activation service.
- The firewall must be disabled. Otherwise, remote logon is not possible. Port 3389 must be enabled.
- The Windows system disk size is between 40 GiB and 500 GiB.

Table 11-7: Notes

Item	Description
Operating system versions	The following operating system images can be imported (including 32-bit and 64-bit): • Microsoft Windows Server 2016 • Microsoft Windows Server 2012 R2 (Standard Edition) • Microsoft Windows Server 2012 (Standard Edition and Datacenter Edition) • Microsoft Windows Server 2008 R2 (Standard Edition, Datacenter Edition, and Enterprise Edition) • Microsoft Windows Server 2008 (Standard Edition, Datacenter Edition, and Enterprise Edition) • Microsoft Windows Server 2003 R2 (Standard Edition, Datacenter Edition, and Enterprise Edition) • Microsoft Windows Server 2003 (Standard Edition, Datacenter Edition, and Enterprise Edition) or later versions, including Service Pack 1 (SP1)

Item	Description
	• Windows 7 (Professional Edition and Enterprise Edition)
Basic system environment requirements	• Multi-partition system disks are supported. • Make sure that the system disk has a sufficient free space. • Do not modify critical system files. • Before importing an image, confirm the integrity of the file system. • The NTFS file system with the MBR partition type is supported.
Applications	Do not install qemu-ga in an imported image. If it is installed, some of the services that Alibaba Cloud needs may become unavailable.
Image file formats	• RAW • VHD We recommend that you configure the system disk size for importing images based on the virtual disk size (not the image file size). The system disk size for importing images must be between 40 GiB and 500 GiB. **Note:** We recommend that you import images in VHD format which has a smaller transmission capacity.

11.6.7.3 Convert image file format

Only image files in RAW or VHD format can be imported. To import images in other formats, use a tool to convert the format before importing the images. You can use the qemu-img tool to convert image files into VHD or RAW from other formats, such as RAW, Qcow2, VMDK, VDI, VHD (vpc), VHDX, qcow1, or QED. You can also use qemu-img to convert image files between RAW and VHD formats.

Install qemu-img and convert image file format

You can use different methods to install qemu-img and convert the image file format based on operating system of your local computer:

- *Windows*
- *Linux*

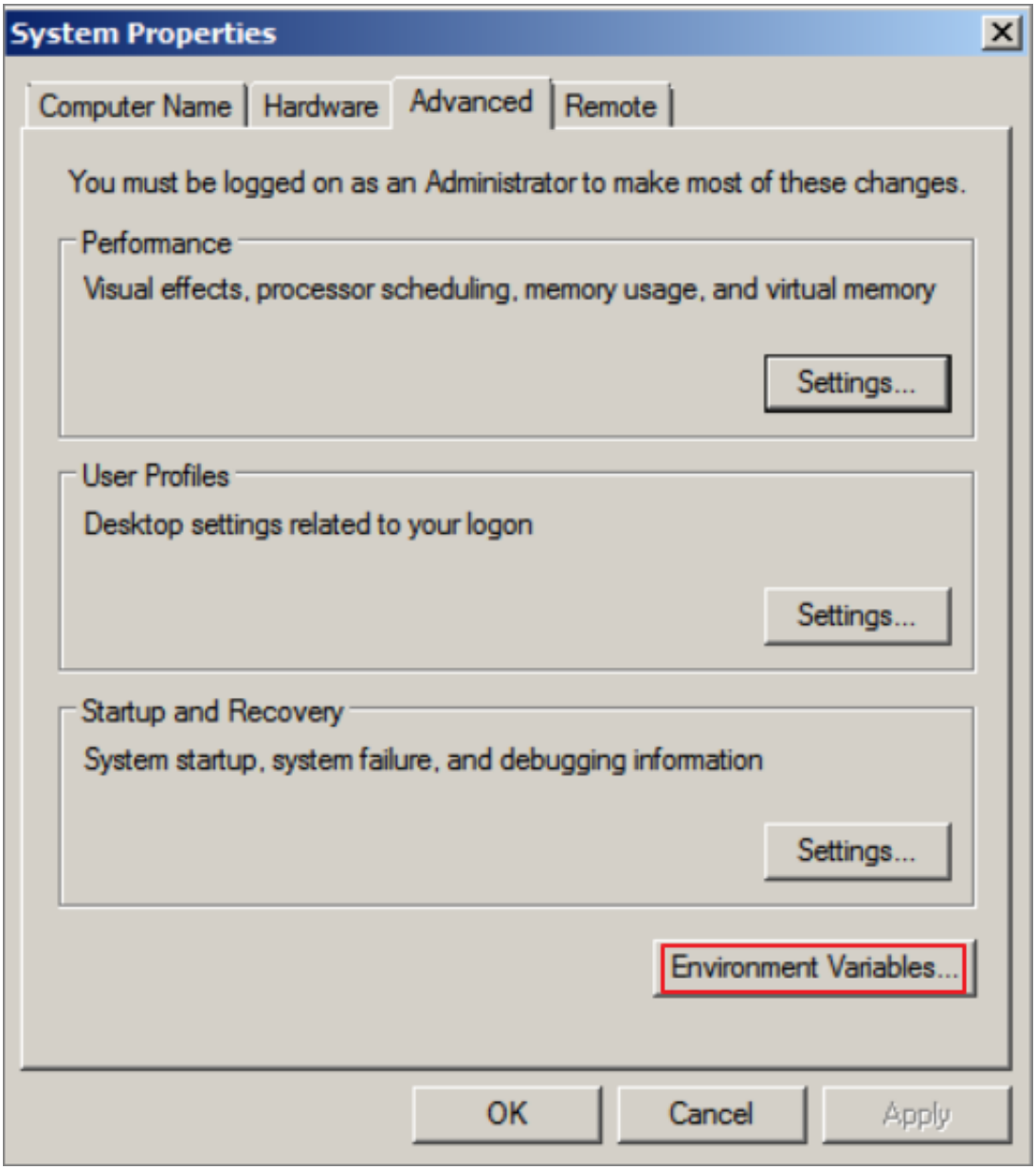
Windows

To install qemu-img and convert the image file format, follow these steps:

Procedure

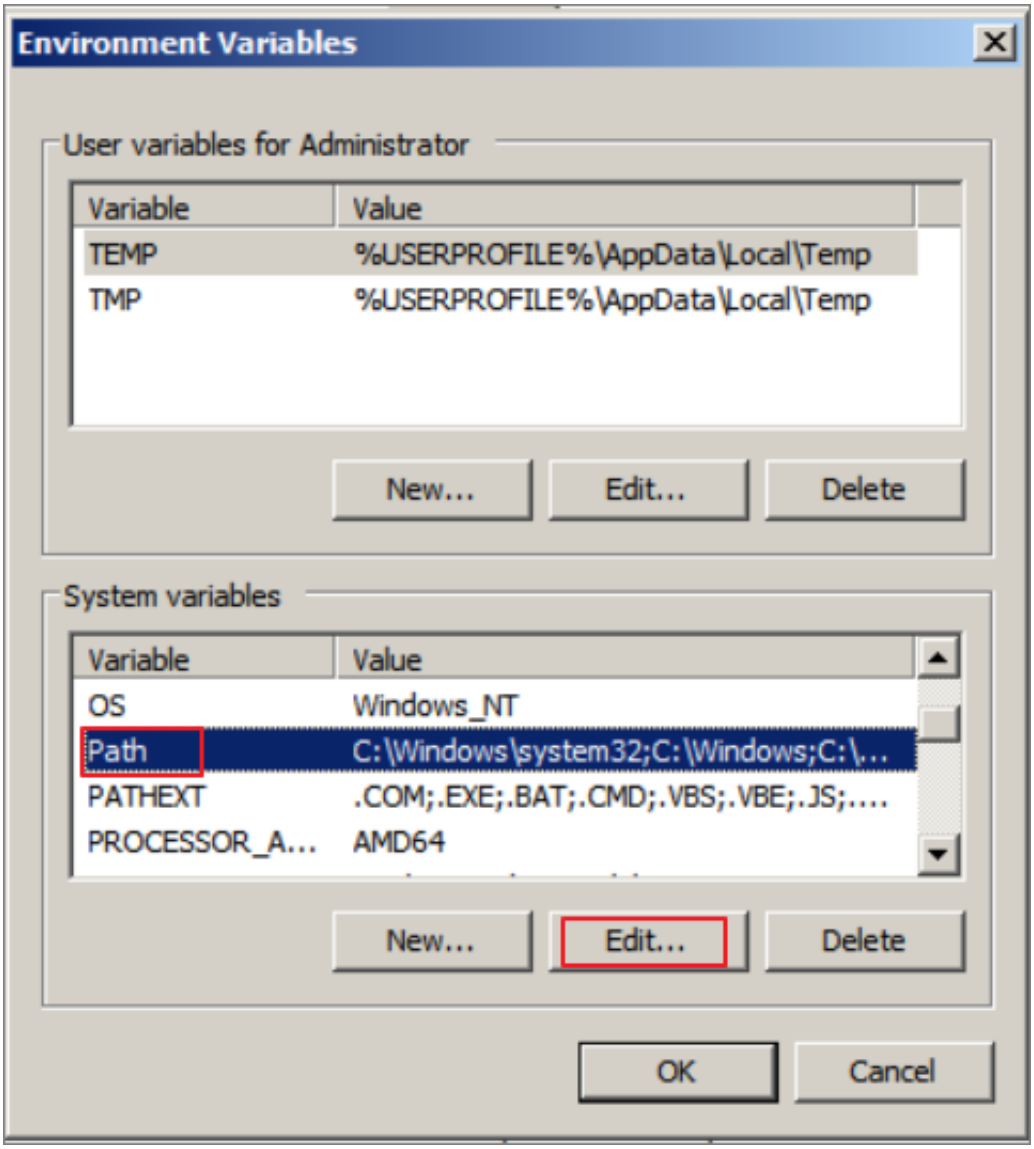
1. Download and *install qemu*. Installation path: `C:\Program Files\qemu`.
2. Do as follows to configure environment variables:
   a) Select **StartComputer**, and right-click **Properties**.
   b) In the left-side navigation pane, click **Advanced System Settings**.
   c) In the **System Properties** dialog box, click the **Advanced** tab and click **Environment Variables**.

Figure 11-33: System properties



d) In the **Environment Variables** dialog box, find the *Path* variable in **System Variables**, and click **Edit**. If the *Path* variable does not exist, click **New**.

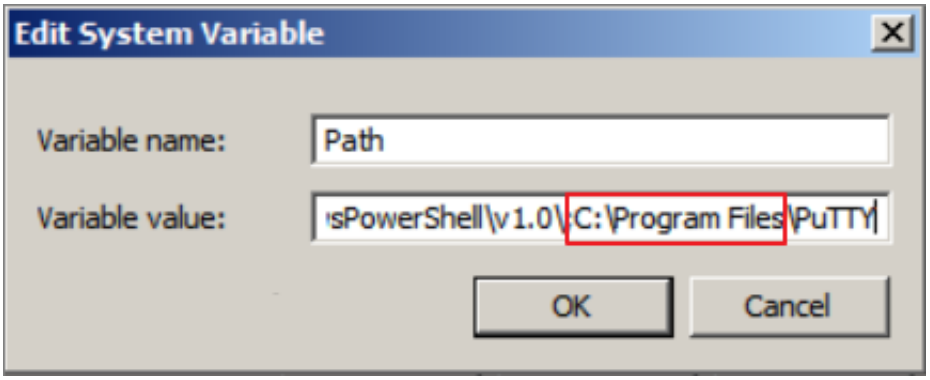
Figure 11-34: Edit variables



e) Add a variable value.

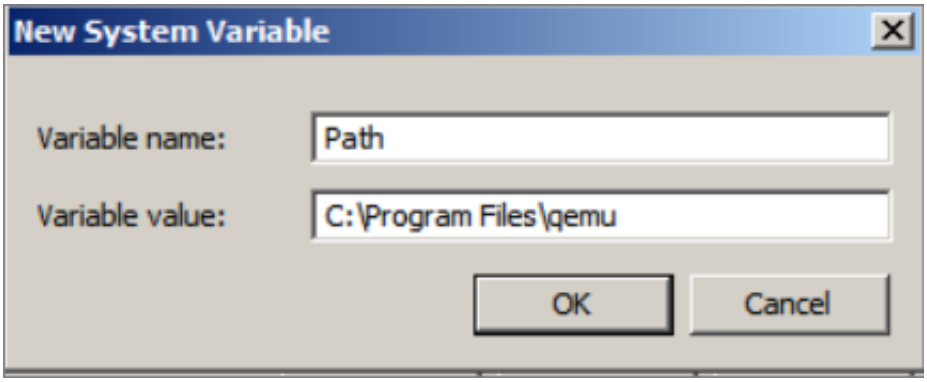
- In the **Edit System Variables** dialog box, add *C:\Program Files\qemu* to **Variable Value**. Different variable values are separated with semicolon (**;**).

Figure 11-35: Edit system variable



- In the **New System Variable** dialog box, enter *Path* in **Variable Name**, and *C:\Program Files\qemu* in **Variable Value**.

Figure 11-36: New system variable



3. Open the **command prompt** in Windows and run the `qemu-img --help` command. If it is displayed successfully, the installation succeeds.
4. In the **command prompt**, run the `cd [directory of the source image file]` command to change the file directory, for example, `cd D:\ConvertImage`.
5. Run the following command in the **command prompt** to convert the image file format: `qemu-img convert -f raw -O qcow2 centos.raw centos.qcow2`.

   The command parameters are described as follows:

   - `-f` is followed by the source image format.
   - `-O` (uppercase is required) is followed by the converted image format, the source file name, and the target file name.

   After the file format is converted, the target file appears in the directory of the source image file.

Linux

To install qemu-img and convert the image file format, follow these steps:

Procedure

1. Install qemu-img, for example:
   - For Ubuntu, run the `apt install qemu-img` command.
   - For CentOS, run the `yum install qemu-img` command.
2. Run the following command to convert the image file format.

   ```
   qemu-img convert -f raw -O qcow2 centos.raw centos.qcow2
   ```

   The command parameters are described as follows:
   - `-f` is followed by the source image format.
   - `-O` (uppercase is required) is followed by the converted image format, the source file name, and the target file name.

11.6.8 Export images

On the ECS console, you can export the custom images as needed.

Prerequisites

You have been authorized to export images. For the authorization procedure, see **RAM Management** in *Apsara Stack Console User Guide*.

Procedure

1. *Log on to the ECS console* and go to the **Images** page.
2. In the **Action** column of the image to be exported, click the icon  and select **Export Image** from the drop-down list.
3. In the dialog box that appears, select an **OSS Bucket**, set an **OSS Prefix**, and then click **Confirm**.

    **Note:**

   OSS Prefix: It is an optional parameter, the value of which ranges from 1 to 30 characters and is composed of letters and numbers.

11.6.9 Delete images

On the ECS console, you can delete images that are no longer needed.

Prerequisites

Currently, public images only support the batch delete function.

Procedure

1. *Log on to the ECS console* and go to the **Images** page.
2. Select a **Department** and **Region** or enter an **Image Name** and click **Search** to find the target image.
3. In the **Action** column of the image, click the icon and select **Delete** from the drop-down list.

   Note:
   When you select multiple images, you can click **Delete** in the upper-right corner of the page.
4. In the confirmation box that appears, click **Confirm**.

11.7 Snapshots

11.7.1 Overview

A snapshot saves the disk data of a certain point in time. It is used for data backup or making custom images.

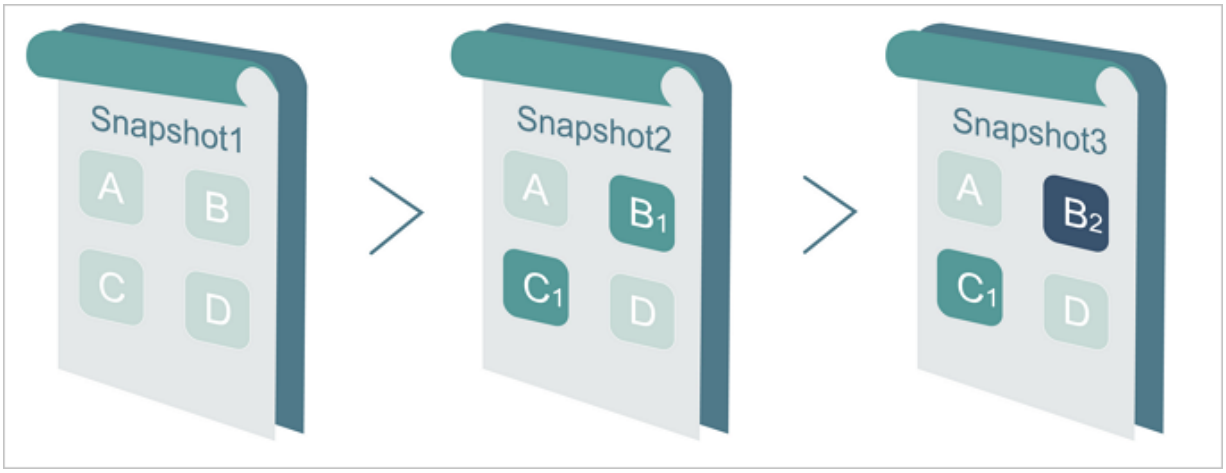
You may have the following needs when working with disks:

- Use the data on a disk as the basic data of another disk when writing or saving data to that disk .
- Restore your data to the expected status. Cloud disks provide a secure storage for your data. However, if the data stored on the disk is incorrect, for example, due to an application error or malicious tampering as a result of an application vulnerability, you may need to restore your data.

Alibaba Cloud allows you to create a snapshot to retain a copy of the data on a disk at a certain time point. You can create disk snapshots on a scheduled basis to guarantee your business continuity.

Snapshots use an incremental backup scheme. Two snapshots are compared so that only the changed data is copied, as shown in the following figure.

Figure 11-37: Operating principles of snapshots



Snapshot 1, Snapshot 2, and Snapshot 3 are the first, second, and third disk snapshots of the disk . The file system checks the disk data block by block. When a snapshot is created, only the blocks with changed data are copied to the snapshot.

Because Snapshot 1 is the first disk snapshot, it copies all of the data on the disk. Snapshot 2 only copies the changed data blocks B1 and C1 and references blocks A and D in Snapshot 1 as its data blocks A and D. Similarly, Snapshot 3 copies the changed data block B2, references blocks A and D in Snapshot 1 as its data blocks A and D, and references data block C1 in Snapshot 2 as its data block C1.

If you restore the disk to the status at the time of Snapshot 3, you can perform snapshot rollback to copy data blocks A, B2, C1, and D in Snapshot 3 to the disk.

If Snapshot 2 is deleted, data block B1 in the snapshot is deleted but data block C1 is not. In this way, when you restore the disk to the status at the time of Snapshot 3, data block C1 can also be restored.

Snapshots are stored in your Object Storage Service (OSS) instance, but the OSS console does not allow you to query, manage, or calculate the bucket space used by the snapshot files. You can operate snapshots only by using the ECS console or APIs.

11.7.2 Create a snapshot

On the ECS console, you can create a snapshot for your disk manually.

Context

Alibaba Cloud provides the snapshot function for each user and imposes a quota on the number of snapshots that can be created. Currently, you can create up to 64 snapshots for each disk.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.
2. Find the disk for which you want to create a snapshot. In the **Action** column of the disk, click  and select **Create Snapshot** from the drop-down list.
3. Enter a **Snapshot Name** and **Snapshot Description**, and click **Confirm** .
4. Click the **Snapshots** tab, and you can view the snapshot creation progress and status. When the **Progress** is 100%, the snapshot is successfully created.

> **Note:**
>
> - The first time you create a snapshot for a disk, it takes a relatively long time because this is a full snapshot.
> - When you create a snapshot for a disk with existing snapshots, it takes a relatively short time. The specific duration depends on the volume of data changed between the snapshot to be created and the previous snapshot. The greater the changed volume, the longer the duration.
> - Avoid creating snapshots during peak business hours.

11.7.3 View snapshots

On the ECS console, you can view the created snapshots and relevant information.

Procedure

1. *Log on to the ECS console* and go to the **Snapshots** page.
2. Select a **Department** and **Region** or enter a **Snapshot Name** and click **Search** to find the target snapshot.

> **Note:**
>
> Click **Snapshot Name** and you can select other filtering conditions: **Snapshot ID**, **Disk Name**, and **Project Name**.

11.7.4 Delete snapshots

You can use the ECS console to delete unnecessary snapshots.

Prerequisites

- Deleted snapshots cannot be restored. Exercise caution when deleting snapshots.

- If a system disk snapshot has been used to create a custom image, the snapshot cannot be deleted.

Procedure

1. *Log on to the ECS console*.
2. Click the **Snapshots** tab. Set **Department**, **Region**, and a specific filter criterion. Click **Search**.

   Note:
   Filter criteria include: **Snapshot Name**, **Snapshot ID**, **Disk Name**, **Project Name**, and **Created At**.

3. Click the icon in the **Actions** column corresponding to a snapshot and choose **Delete** from the shortcut menu.

   Note:
   To delete multiple snapshots at the same time, select the snapshots and click **Delete**.

4. In the message that appears, click **OK**.

11.7.5 Application scenarios

This section introduces how to use snapshots to create images and data disks.

In addition to rolling back the source disks, you can also use snapshots in the following situations:

- Create custom images.
- Create data disks when you create an instance.

Create custom images

If you want to use one of the instances as a template, you can quickly create a custom image. For the procedure, see *Create custom images from snapshots*.

Note:
Data disk snapshots cannot be used in the creation of custom images.

Use snapshots to create data disks for instances

You can use a snapshot to create a data disk for an instance so that the new data disk includes the data on the data disk of another instance.

Operation procedure

When you *Create disks*, select **Use Snapshots** to create a disk using the snapshot of another data disk in the same region. The capacity of the new data disk is determined by the snapshot capacity and cannot be changed.

Note:
If you reset a data disk that was created from a snapshot, the data disk restores the data in the snapshot.

11.8 Automatic snapshot policies

11.8.1 Overview

Automatic snapshot policies include such items as when to create a snapshot in a day, on which days in a week to repeat the creation, and how long to retain the snapshots. You can set them as needed.

11.8.2 Create an automatic snapshot policy

On the ECS console, you can create an automatic snapshot policy as needed.

Procedure

1. *Log on to the ECS console* and go to the **Snapshot Policies** page.
2. Click **Create**.
3. Define configuration information for the automatic snapshot policy:

Table 11-8: Automatic snapshot policy configuration

Item	Description
Name	Name of the automatic snapshot policy. It is a string of 2 to 128 characters, including numbers, underscores (_), and hyphens (-). It must start with an uppercase/ lowercase letter or a Chinese character.
Region	Sets the region to which the automatic snapshot policy is applied.
Department	Sets the department to which the automatic snapshot policy is applied.
Created At	Time of the day for starting automatic snapshot creation . The value must be on the hour and ranges from 00:00 to 23:00 (24 time points in total). You can select multiple time points.

Item	Description
Repeat Date	Days of the week for automatic snapshot creation, ranging from Monday to Sunday. You can select multiple days.
Snapshot Retention Period(Days)	By default, the snapshot will be retained permanently. You can enter: 1 to 65535.

4. After completing the settings, click **Confirm**.

What's next

After the automatic snapshot policy is successfully created, you must *Configure an automatic snapshot policy*.

11.8.3 View automatic snapshot policies

On the ECS console, you can view the created automatic snapshot policies and relevant information.

Procedure

1. *Log on to the ECS console* and go to the **Snapshot Policies** page.
2. Select a **Region** or enter a **Snapshot Policy ID** and click **Search** to view the automatic snapshot policy.

> **Note:**
> Click **Snapshot Policy ID** and you can select filtering conditions from the drop-down menu: **Automatic Snapshot Policy Name**.

11.8.4 Edit an automatic snapshot policy

On the ECS console, you can edit the information of an automatic snapshot policy.

Procedure

1. *Log on to the ECS console* and go to the **Snapshot Policies** page.
2. In the **Action** column for the target policy, click  and select **Edit**.
3. In the dialog box that appears, you can modify **Snapshot Policy Name**, **Created At**, **Repeat Date**, and **Snapshot Retention Period**. Then click **Confirm**.

11.8.5 Configure an automatic snapshot policy

On the ECS console, you can configure an automatic snapshot policy for a disk.

Prerequisites

- Make sure that the disk to which you want to apply an automatic snapshot policy is in the **Running** state.
- The automatic snapshot command takes the following format: auto_yyyyMMdd_1, for example , auto_20140418_1.

> **Note:**
>
> - We recommend that you select periods with a low service load for automatic snapshot policy execution.
> - The snapshots you have manually created do not conflict with automatic snapshots. However, if a disk is taking an automatic snapshot, you must wait for it to finish before you can manually create a snapshot.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.
2. Find the disk for which you want to set an automatic snapshot policy. In the **Action** column of the disk, click the  icon and select **Run Automatic Snapshot** from the drop-down list.
3. In the **Implement Automated Snapshot Policy** dialog box that appears, select **Snapshot Policies**, and click **Confirm**.

11.8.6 Configure an automatic snapshot policy for multiple disks

On the ECS console, you can configure an automatic snapshot policy for multiple disks.

Procedure

1. *Log on to the ECS console* and go to the **Snapshot Policies** page.
2. Click the  icon to the right of the target snapshot policies and select **Configure**.
3. Select one or more disks to which you want to configure automatic snapshot policies.
   - In **Optional Disk**, specify multiple disks and click → to choose the disks.
   - In **Selected Disk**, specify multiple disks and click ← to remove the disks.
   - On the top of **Optional Disk**, click **Select All** and click → to choose all disks.

- On the top of **Selected Disk**, click **Select All** and click ← to remove all disks.

4. Click **Confirm**.

11.8.7 Delete an automatic snapshot policy

On the ECS console, you can delete an automatic snapshot policy that is no longer needed.

Procedure

1. *Log on to the ECS console* and go to the **Snapshot Policies** page.
2. Select a **Region**, enter a **Snapshot Policy ID** or **Automatic Snapshot Policy Name**, and click **Search** to find the target snapshot policy.
3. In the **Action** column of the automatic snapshot policy to be deleted, click the  icon and select **Delete** from the drop-down list.
4. Click **Confirm**.

11.9 Security groups

11.9.1 Overview

A security group is a virtual firewall that controls the inbound and outbound traffic of an instance. An instance must belong to at least one security group. As an important means of network isolatin, a security group is used to divide the security domans on the cloud.

Security group restrictions

- A single security group cannot contain more than 1,000 instances. If you require Intranet mutual access among more than 1,000 instances, you can allocate them to different security groups and permit mutual access through mutual authorization.
- Each instance can join a maximum of five security groups.
- Each security group can have a maximum of 100 rules.
- Each user can have a maximum of 100 security groups.
- Adjusting security groups does not affect the continuity of your services.
- Security groups are stateful. If packets are permitted in the outbound direction, packets transmitted over this connection are also permitted in the inbound direction.

Security group rules

Security group rules can be set that permit or forbid ECS instances in a security group from accessing a public network or intranet in the inbound and outbound directions.

You can create or delete security group rules at any time. Once changes are made, the updated security group rules are automatically applied to ECS instances in the security group.

When setting security group rules, make sure they are concise. If you add an ECS instance to multiple security groups, hundreds of rules may apply to the instance, which may cause connection errors when you access the instance.

11.9.2 View security groups

You can use the ECS console to view existing security groups and related information.

Procedure

1. *Log on to the ECS console*.
2. Click the **Security Groups** tab. Set **Department**, **Region**, and a specific filter criterion. Click **Search**.

   Note:
   Filter criteria include: **Security Group Name**, **Security Group ID**, **VPC ID**, and **Project Name**.
3. Click the  icon in the **Actions** column corresponding to a security group and choose **View Details** from the shortcut menu. On the ECS Instances tab page that appears, view instances in the security group and security group rule details.

   Note:
   You can also click the ID of a security group to view security group details.

11.9.3 Add security group rules

You can add rules for a security group as required.

Procedure

1. *Log on to the ECS console*.
2. Click the **Security Groups** tab. Set **`Department`**, **`Region`**, and **`Security Group Name`**. Click **Search** to search for a security group.
3. Click the ID of the security group to go to the **ECS Instances** tab page.
4. Click the **Rules** tab. On the Rules tab page, click **Create Security Group Rule**.
5. In the dialog box that appears, configure security group parameters and click **OK**.

Parameter description:

- **Authorization Policy**: allows you to select **Allow** or **Block**.

  The Block policy discards data packets without any response. If two security groups have the same rules but different authorization policies, the **Block** policy takes effect while the **Allow** policy does not.

- **Rule Direction**:
  - **Outbound**: Your ECS instances access other ECS instances in the internal network or resources in the public network.
  - **Inbound**: Other ECS instances in the internal network or resources in the public network access your ECS instances.
- **Protocol Type** and **Port Range**: The port range setting is affected by the selected protocol type. *Table 11-9: Parameter description* describes the relationship between protocol types and port ranges.

  Table 11-9: Parameter description

Protocol type	Port range	Scenario
All	-1/-1 is displayed, indicating all ports. You cannot set a port range for this protocol type.	It is used in all trusted scenarios.
TCP	The port range can be customized. Valid values: 1 to 65535, in the format of **Start Port**/**End Port**. Even for a single port, you need to set the port range in the standard format. For example, use 80/80 to indicate port 80.	It can be used to allow or block one or several successive ports.
UDP		
ICMP	-1/-1 is displayed, indicating all ports. You cannot set a port range for this protocol type.	You can run the `ping` command to check network connection status between instances.
GRE	-1/-1 is displayed, indicating all ports. You cannot set a port range for this protocol type.	It is used for VPN service.

- **Priority**: allows you to set the priority. It is 1 by default, indicting the highest priority.

- **`Authorization Type`** and **`Authorized IP Addresses`**: The authorized IP address setting is affected by the authorization type. *Table 11-10: Authorization description* describes the relationship between authorization types and authorized IP addresses.

  Table 11-10: Authorization description

Authorization type	Authorized IP address
IP Address Segment Access	Use the IP or CIDR block formats such as 12.1.1.1 or 13.1.1.1 /25. Only IPv4 addresses are supported. 0.0.0.0/0 indicates to allow or block all IP addresses. Exercise caution when setting 0 .0.0.0/0.
Security Group Access	Security Group Access applies to internal IP addresses only. For public network access, you must set the authorization type to **IP Address Segment Access**.

- **`Description`**: allows you to add more information.

11.9.4 Remove an instance from a security group

On the ECS console, you can remove an instance from a security group that is no longer needed.

Procedure

1. *Log on to the ECS console* and go to the **Security Groups** page.
2. Select a **`Department`** and **`Region`**, or enter a **`Security Group Name`**, and click **Search** to find the target security group.
3. Click the security group ID to go to the **ECS Instances** page.
4. Click the  icon next to the target instance, select **Remove from Security Group**, and click **Confirm**.

11.9.5 Delete a security group

On the ECS console, you can remove a security group that is no longer needed.

Prerequisites

Ensure all the instances are removed from the security group; otherwise, this security group cannot be deleted.

Procedure

1. *Log on to the ECS console* and go to the **Security Groups** page.

2. Select a **Department** and **Region** or enter a **Security Group Name** and click **Search** to find the target security group.
3. In the **Action** column of the security group, click the icon and select **Delete** from the drop-down list.

> **Note:**
> On the **Security Groups** page, you can select multiple security groups and click **Delete**. In the confirmation box that appears, select **Yes** and click **Confirm**.

4. In the confirmation box that appears, click **Confirm**.

11.10 Elastic network interfaces

11.10.1 Overview

Elastic network interfaces (ENIs) are divided into primary ENIs and secondary ENIs. The primary ENI is created by default upon the creation of an instance in a VPC. The lifecycle of the primary ENI is the same as that of the instance and you are not allowed to detach the primary ENI from the instance. The secondary ENI is created independently. You can attach it to an instance or detach it from an instance. This topic describes secondary ENIs.

11.10.2 Create an ENI

On the ECS console, you can create an ENI as needed.

Procedure

1. *Log on to the ECS console* and go to the **Elastic NIC** page.
2. Click **Create NIC** in the upper-right corner of the page.
3. On the **Create NIC** page, configure the following ENI information and click **Confirm**.

Table 11-11: Configure an NIC

Item	Description
Region	• **Region**: Required. Select a region where the target ENI resides. • **Zone**: Required. Select a physical zone with independent power grids and networks within a region. A zone can communicate with other zones using the Intranet, and is not affected by faults in other zones. If you want to improve

Item	Description
	application availability, we recommend that you create instances in different zones.
Configurations	• **Department**: Required. Select a department to which the ENI belongs. • **Project**: Required. Select a project to which the ENI belongs. • **VPC**: Required. Select a VPC where your instance resides. The ENI can only be attached to an instance on the same VPC. **Note:** You cannot change the VPC where the created ENI resides. • **Security Groups**: Required. Select a security group of the current VPC. • **ENI Name**: Optional. Set the name of the ENI. • **IP Address**: Optional. Enter the primary Intranet IPv4 address of the ENI. The IPv4 address must be an idle address in the CIDR network segment of VSwitches. If you do not specify an IPv4 address when creating an ENI, the system automatically allocates an idle private IPv4 address for you. • **Description**: Optional. Enter a description for the ENI for future management.

11.10.3 View ENIs

In the ECS Console, you can view the created ENIs and relevant information.

Procedure

1. *Log on to the ECS console*.
2. Enter the **Network Interfaces** page, select **Department** and **Zone**, or type in the**NIC ID**.

   Click **Search** to find the specified ENI. You can view the details in the ENI list.

   Note:

   By clicking **ID**, you can select other filters in the drop-down list, such as **NIC Name**, **Project Name**, **VPC ID**, **VSwitch ID** and **Instance ID**.

11.10.4 Edit an ENI

On the ECS console, you can modify the information of an ENI.

Procedure

1. *Log on to the ECS console* and go to the **Elastic NIC** page.
2. Find the target secondary ENI, click the  icon in the **Action** column, and select **Edit** from the drop-down list.
3. In the dialog box that appears, modify the **NIC Name**, **Security Groups**, and **Description** of the ENI, and click **Confirm**.

11.10.5 Attach an ENI to an instance

After creating an ENI, you can attach it to an instance.

Prerequisites

Pay attention to the following limits when you attach an ENI to an instance:

- You can only attach a secondary ENI to an instance.
- You have created an ENI as described in *Create an ENI*. The NEI must be in the **Available** state.
- The instance must be in the **Running** or **Stopped** state. For more information about how to start or stop an instance, see *Start, stop, or reboot an instance*.
- The ENI must be in the same VPC as the instance.
- The VSwitches of the ENI and the instance can be different, but they must be in the same zone .
- An ENI can only be attached to one ECS instance at a time. However, an ECS instance can be associated with multiple ENIs. For more information about the maximum number of ENIs that can be attached to each instance type, see **Instance types** in *ECS Product Introduction*.

Procedure

1. *Log on to the ECS console*.
2. Click the **Elastic Network Interfaces (ENIs)** tab. Click the  icon in the **Actions** column corresponding to a secondary ENI and choose **Attach to ECS Instance** from the shortcut menu.
3. In the Attach to ECS Instance dialog box that appears, select a **Destination Instance** and click **OK**.

11.10.6 Detach an ENI from an instance

You can use the ECS console to detach a secondary ENI (but not primary ENI) from an instance.

Prerequisites

- The secondary ENI must be in the **InUse** state.
- The instance must be in the **Running** or **Stopped** state. For more information about how to start or stop an instance, see *Start, stop, or reboot an instance*.

Procedure

1. *Log on to the ECS console*.
2. Click the **Elastic Network Interfaces (ENIs)** tab. Click the  icon in the **Actions** column corresponding to a secondary ENI and choose **Detach from ECS Instance** from the shortcut menu.
3. In the message that appears, click **OK**.

11.10.7 Delete an ENI

On the ECS console, you can delete an ENI that is no longer needed.

Context

You can only delete ENIs one by one, but not multiple ENIs at a time.

Prerequisites

An ENI has been detached from an instance and must be in the **Available** status.

Procedure

1. *Log on to the ECS console* and go to the **Elastic NIC** page.
2. Select a **Department** or **Region**, or enter an **NIC ID**, and click **Search** to find the target ENI.
3. In the **Action** column of the secondary ENI, click the  icon and select **Delete** from the drop-down list.
4. In the prompt box that appears, click **Confirm**.

11.11 Deployment sets

11.11.1 Overview

Provided by ECS, a deployment set allows you to understand the topology of host machines, racks and switches. Also, it allows you to select a deployment strategy based on your business needs, thus improving the reliability and performance of your services.

11.11.2 Create a deployment set

On the ECS console, you can create a deployment set as needed.

Procedure

1. *Log on to the ECS console* and go to the **Deployment Set** page.
2. Click **Create Deployment Set**.
3. On the **Create Deployment Set** page, make configuration for a deployment set.

Table 11-12: Configure a deployment set

Item	Description
Region	• **`Region`**: Required. Select a region where the target deployment set resides. • **`Zone`**: Required. Select a physical zone with independent power grids and networks within a region. A zone can communicate with other zones using the Intranet, and is not affected by faults in other zones. If you want to improve application availability, we recommend that you create instances in different zones.
Configurations	• **`Department`**: Required. Select a department to which the target deployment set belongs. • **`Project`**: Required. Select a project to which the target deployment set belongs. • **`Deployment Domain`**: Required. Set the deployment domain to **Default** or **Switch**. • **`Deployment Granularity`**: Required. Set the deployment granularity to **Host Machine**, **Rack**, or **Switch**. • **`Deployment Strategy`**: Required. Set the deployment policy to **Loose Dispersion** or **Strict Dispersion**. • **`Deployment Set Name`**: Optional. Enter the name of the deployment set. • **`Description`**: Optional. Enter a description for the deployment set.

4. After completing the settings, click **Confirm**.

Result

You can view the created deployment set with the preceding attributes in the list of deployment sets.

11.11.3 View a deployment set

On the ECS console, you can view the created deployment sets and the relevant information.

Procedure

1. *Log on to the ECS console* and go to the **Deployment Set** page.
2. Select a **Department**, **Region**, or enter a **Deployment Set Name**, and click **Search** to find the target deployment set. View detailed information about the deployment set in the list of deployment sets.

> **Note:**
> Click **Deployment Set Name**, and you can select **Deployment Set ID** or **Project Name** from the drop-down list.

11.11.4 Edit a deployment set

On the ECS console, you can modify the information of a deployment set.

Procedure

1. *Log on to the ECS console* and find the deployment set you want to edit.
2. In the **Action** column of the deployment set, click the icon and select **Edit** from the drop-down list.
3. In the dialog box that appears, edit the **Name** or **Description** of the deployment set and click **Confirm**.

11.11.5 Delete a deployment set

On the ECS console, you can delete a deployment set that is no longer needed.

Prerequisites

ECS instances have been completely removed from the deployment set.

Procedure

1. *Log on to the ECS console* and find the deployment set you want to delete.

2. In the **Action** column of the deployment set, click the  icon and select **Delete** from the drop-down list.
3. In the prompt box that appears, click **Confirm**.

11.12 Install FTP software

11.12.1 Overview

The File Transfer Protocol (FTP) is used to transfer files between two computers by establishing two TCP connections. One of them is the command connection for transferring commands between an FTP client and an FTP server. The other is the data connection for uploading or downloading data. Before uploading files to an instance, you need to build an FTP site for instances.

11.12.2 Install VSFTP in CentOS

This section introduces how to install VSFTP on CentOS instances.

Procedure

1. Install VSFTP. Run the `yum install vsftpd -y` command to install VSFTP.
2. Add an FTP account and directory.
   a. Check the location of *nologin*. It is usually in */usr/sbin/nologin* or */sbin/nologin*.
   b. Create an account. Run the following command to create an account with */alidata/www/wwwroot* as your PWFTP home directory. To customize your account name and directory, run `useradd -d /alidata/www/wwwroot -s /sbin/nologin pwftp`.
   c. Run the following command to change the account password:

      ```
      passwd pwftp
      ```

   d. Run the following command to modify the permissions on the specified directory:

      ```
      chown -R pwftp.pwftp /alidata/www/wwwroot
      ```

3. Configure VSFTP.
   a. Run the `vi /etc/vsftpd/vsftpd.conf` command to edit the VSFTP configuration file.
   b. Change **anonymous_enable=YES** to **anonymous_enable=NO** in the configuration file.
   c. Remove the comment tag # before the following configuration:

      ```
      local_enable=YES
             write_enable=YES
             chroot_local_user=YES
      ```

d. To save the changes, press the **ESC** key and enter the following command: `wq`.

4. Modify the shell configuration by editing */etc/shells* in the `vi` editor. If the file does not contain */usr/sbin/nologin* or */sbin/nologin* (depending on the current system configuration), add either one to the file.
5. Start VSFTP and test logon.
   a. Run the `service vsftpd start` command to start VSFTP.
   b. Use the account pwftp to test FTP logon. In this example, the directory is */alidata/www/wwwroot*.

11.12.3 Install VSFTP in Ubuntu and Debian

This section introduces how to install VSFTP on Ubuntu or Debian instances.

Procedure

1. Run the `apt-get install vsftpd -y` command to install VSFTP.
2. Add an FTP account and directory.
   a. Check the location of nologin. It is usually in */usr/sbin/nologin* or */sbin/nologin*.
   b. Create an account. Run the following command to create an account with */alidata/www/wwwroot* as your PWFTP directory. To customize your account name and directory, run `useradd -d /alidata/www/wwwroot -s /sbin/nologin pwftp`.
   c. Run the `passwd pwftp` command to change the account password.
   d. Run the `chown -R pwftp.pwftp /alidata/www/wwwroot` command to modify the permissions on the specified directory.
3. Configure VSFTP.
   a. Run the `vi /etc/vsftpd.conf` command to edit the VSFTP configuration file.
   b. Change **anonymous_enable=YES** to **anonymous_enable=NO** in the configuration file.
   c. Remove the comment tag # before the following configuration:

```
local_enable=YES
        write_enable=YES
        chroot_local_user=YES
        chroot_list_enable=YES
        chroot_list_file=/etc/vsftpd.chroot_list
```

   d. Save and exit.
   e. Edit the */etc/vsftpd.chroot_list* file to add the FTP account to this file. Save and exit.

4. Modify the shell configuration by editing */etc/shells* in the `vi` editor. If the file does not contain */usr/sbin/nologin* or */sbin/nologin* (depending on the current system configuration), add either one to the file.
5. Restart VSFTP and test logon.
   a. Run the `service vsftpd restart` command to restart VSFTP.
   b. Use the account pwftp to test FTP logon. The directory is */alidata/www/wwwroot*.

11.12.4 Configure FTP through IIS in Windows 2003

This section introduces how to configure FTP via IIS in Windows 2003.

Procedure

1. After you connect to your ECS instance remotely, right-click **My Computer** and choose **Manage** from the shortcut menu.
2. In the navigation pane, double-click **Users** to open the user list. Right-click the blank area of the user list and click **New User** from the drop-down list.
3. Enter your FTP username and password, and click **Create**.
4. In the navigation pane, expand **Internet Information Services (IIS) Manager** and delete the default FTP site. Right-click **FTP Site**, click **New** > **FTP Site (F)...** on the pop-up menu. Then, you can create a new FTP site according to the instructions in **FTP Site Creation Wizard**.
5. Enter the path to the main directory and set access permission. In this example, the path is *D:\websoft\www*.
6. Configure the permissions for the www folder of the new FTP site.
   a. Right-click the www folder and choose **Attributes** from the shortcut menu.
   b. Click the **Security** tab. In **Group or User Name**, select **Users** and set permissions.
   c. Click **Advanced**, complete advanced security settings, and click **Apply**.
   d. In the dialog box that appears, click **Yes**.
   e. Wait until the operation is completed.
   f. On the **Security** tab, add the permissions of the pwftp account, and add a user or group.
   g. Click **OK**. The permission setting page appears. Click **Advanced** and complete advanced security settings for the pwftp account. The setting method is the same as that for Users.
   h. After completing the settings, click **Apply** and then **OK**.

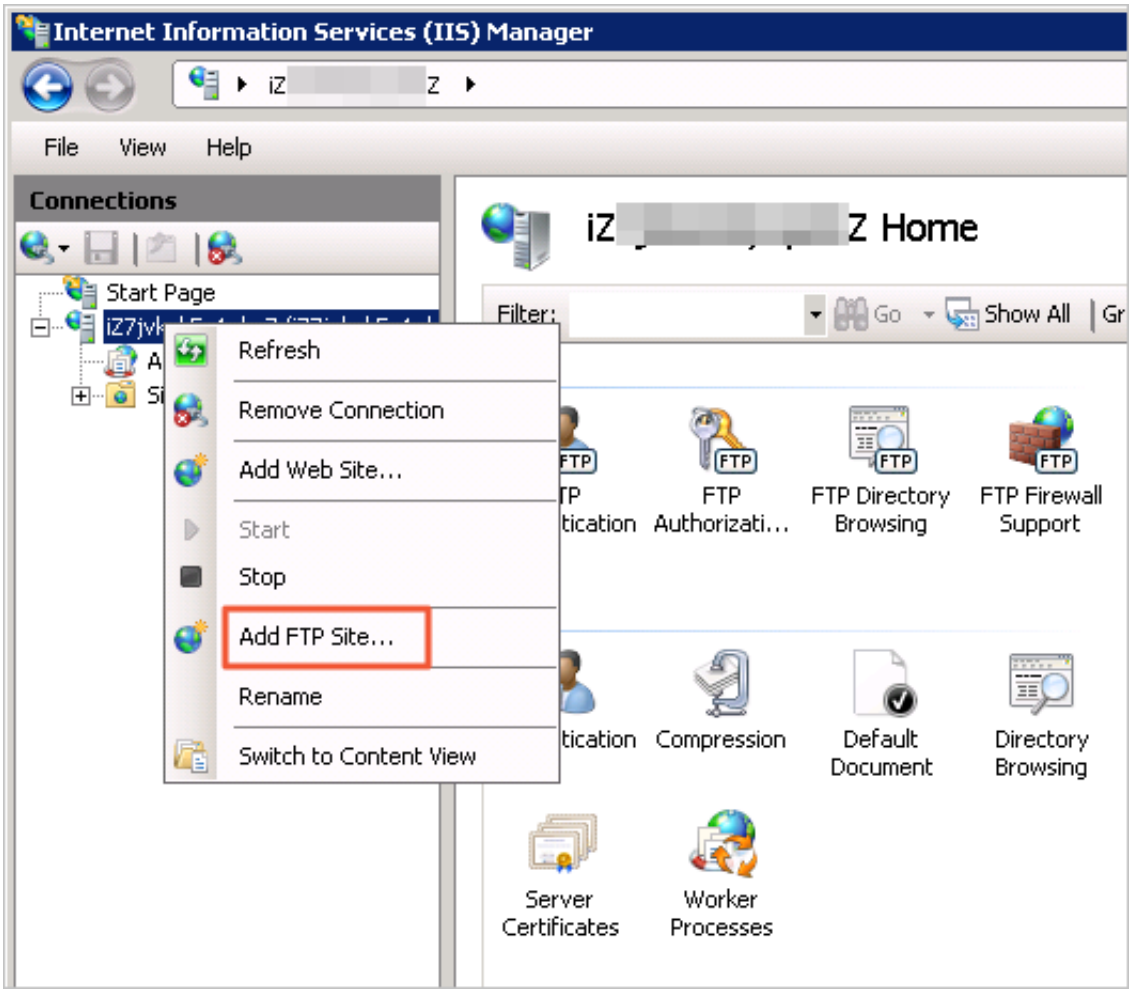
11.12.5 Install and configure FTP in Windows 2008

This section introduces how to build FTP sites on instances in Windows Server 2008 or higher.

Procedure

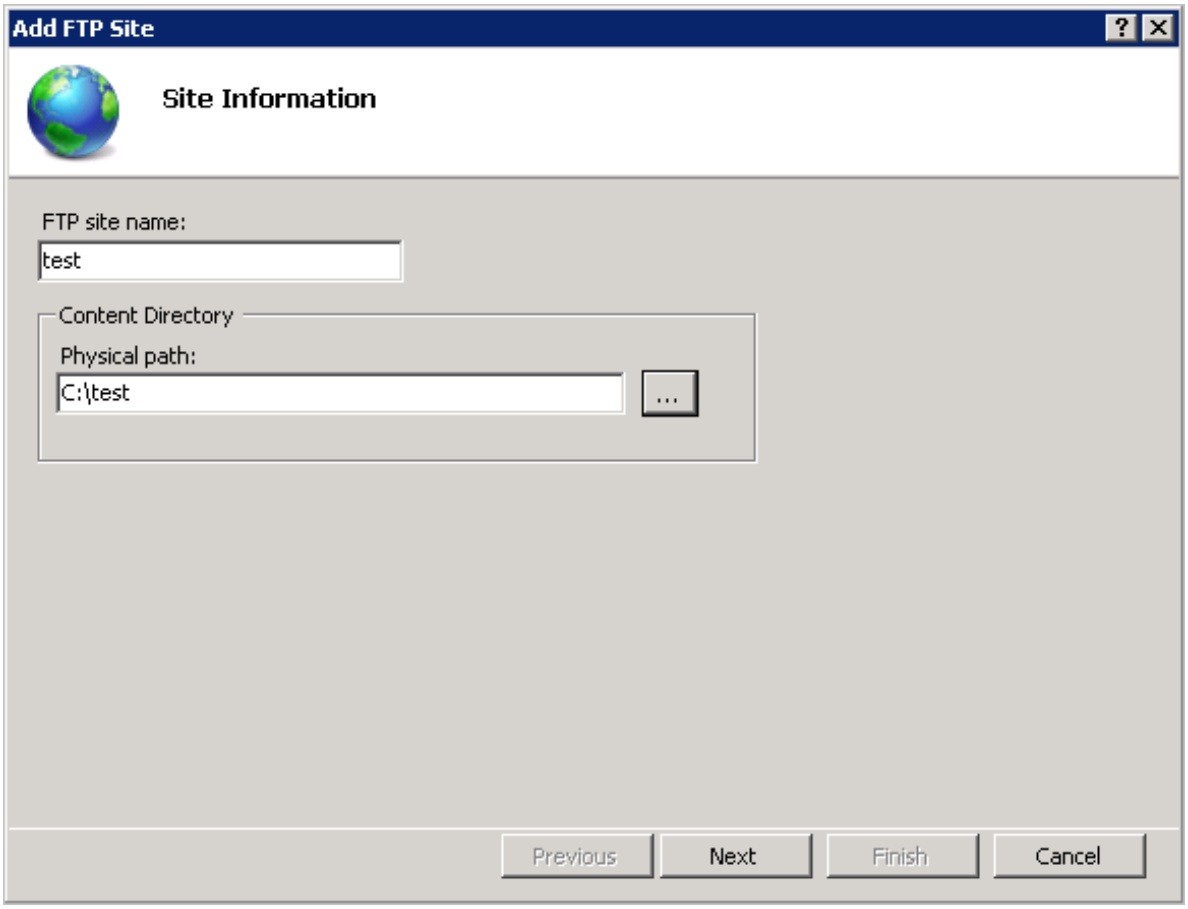
1. After you connect to your ECS instance remotely, choose **Start** > **Management Tools** > **Internet Information Services (IIS) Manager**, right-click the server name, and choose **Add FTP Site** from the shortcut menu.

   Figure 11-38: Add an FTP site



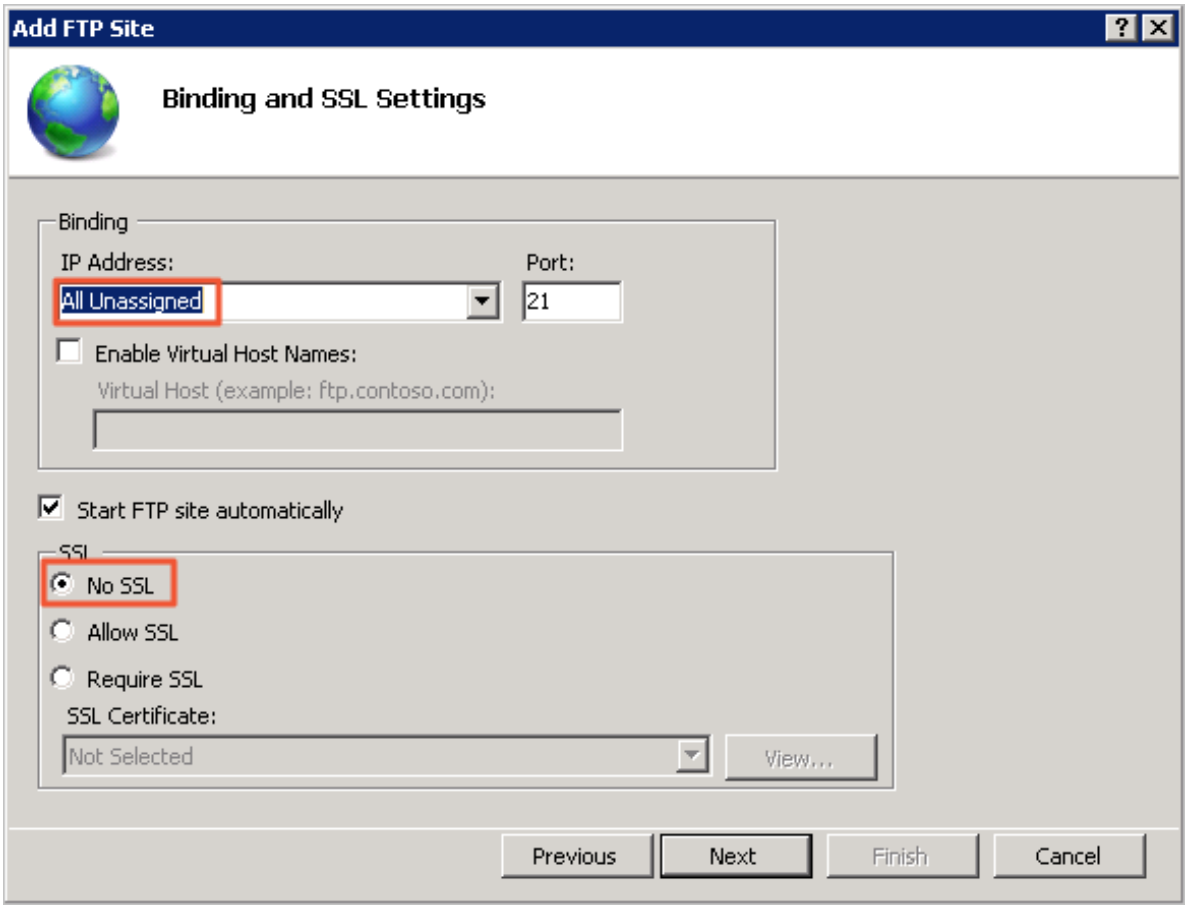
2. Enter the FTP site name and specified path, and click **Next**.

Figure 11-39: Add an FTP site



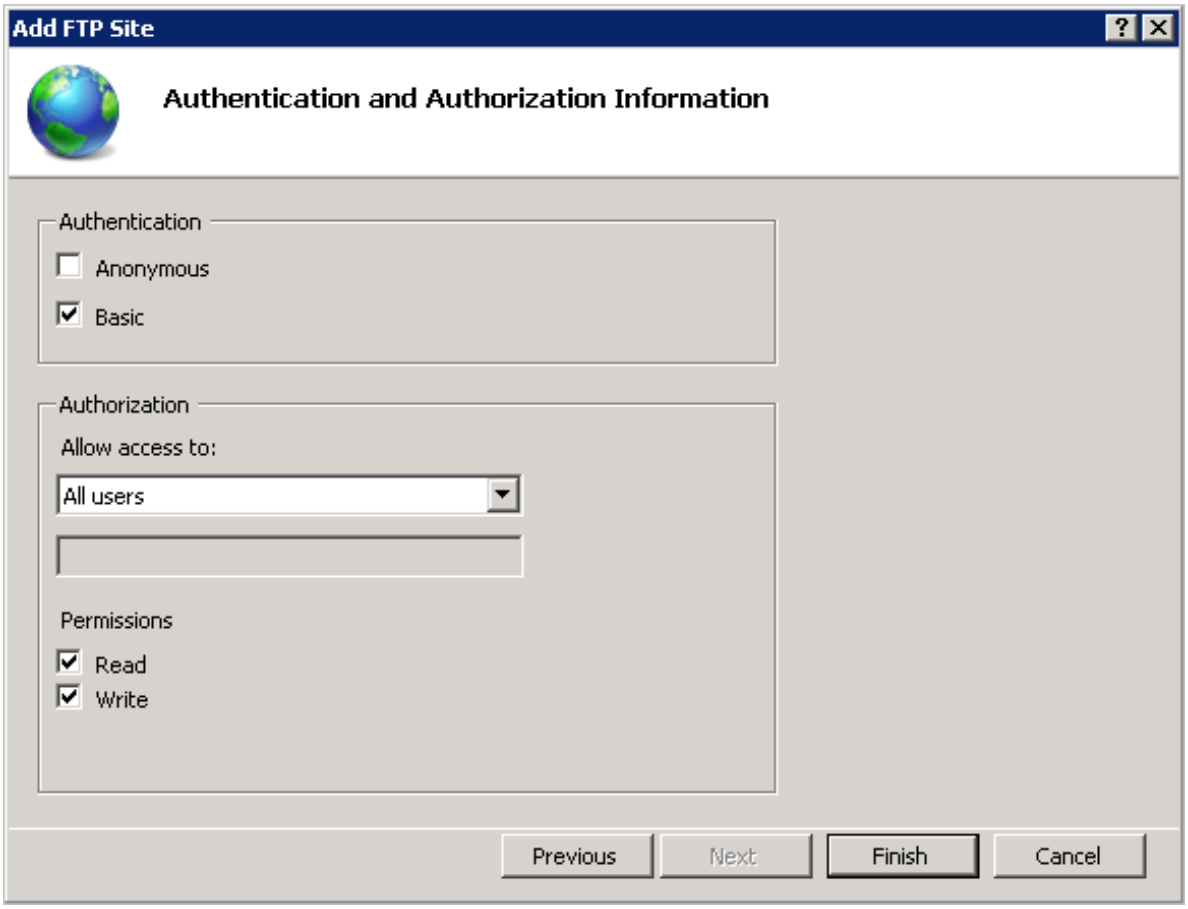
3. As shown in Figure *Figure 11-40: Bind an IP address*, set **IP Address** to **All Unassigned** and **SSL** to **None**.

Figure 11-40: Bind an IP address



4. Set **Authentication** to **Basic**, **Authorization** to **All Users**, and **Permissions** to **Read** and **Write**.

Figure 11-41: Set authentication and authorization



5. Click **Finish** after completing FTP settings. You can use the administrator account and password to upload and download files through FTP.

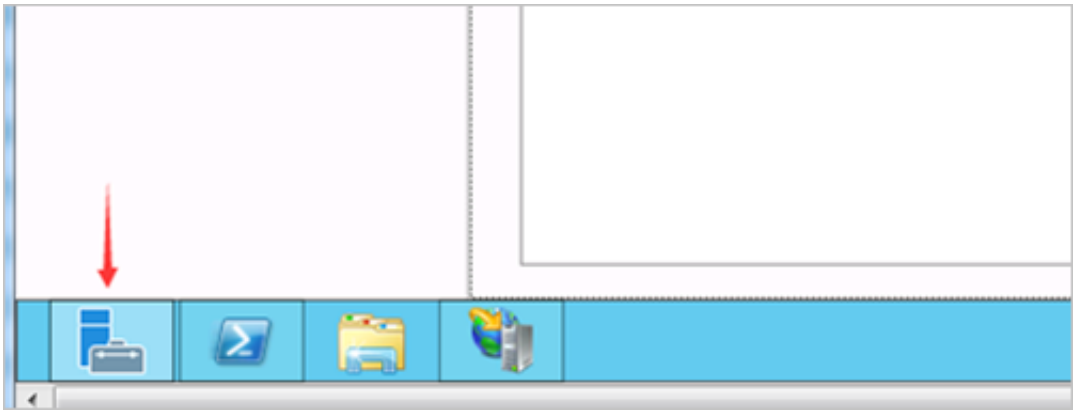
11.12.6 Install and configure IIS and FTP in Windows 2012

This section introduces how to install and configure IIS and FTP in Windows 2012.

Procedure

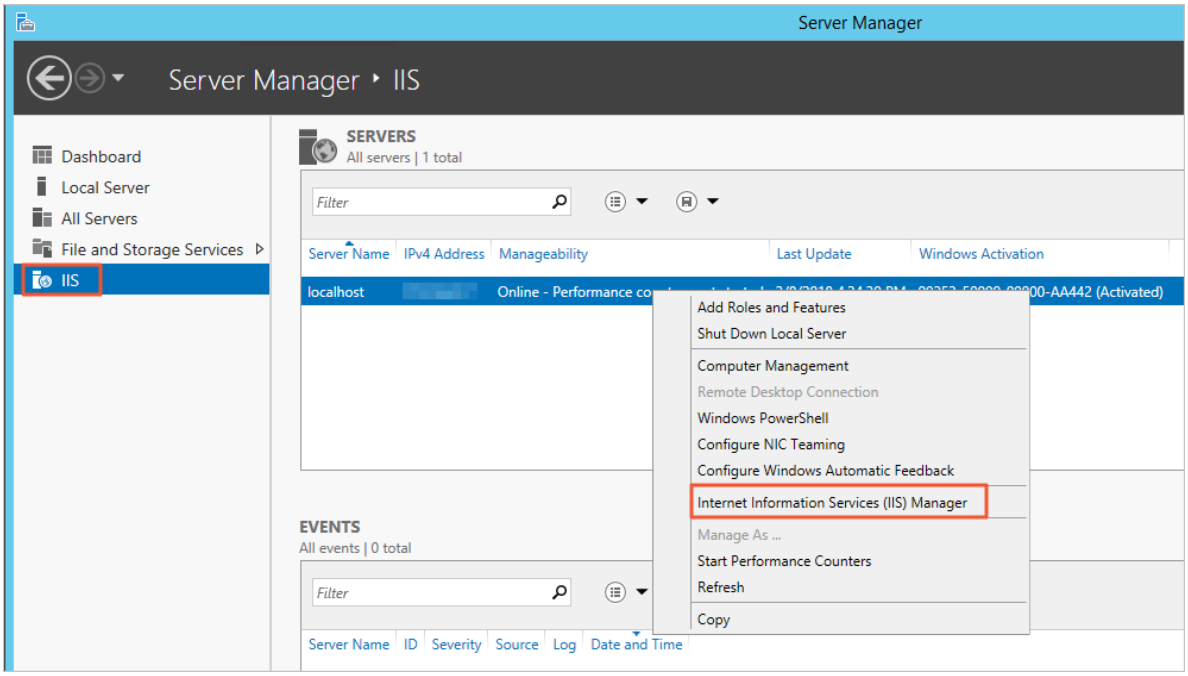
1. In the lower-left corner of the server interface, click **Server Manager** to start the **Server Manager**.

Figure 11-42: Start the server manager



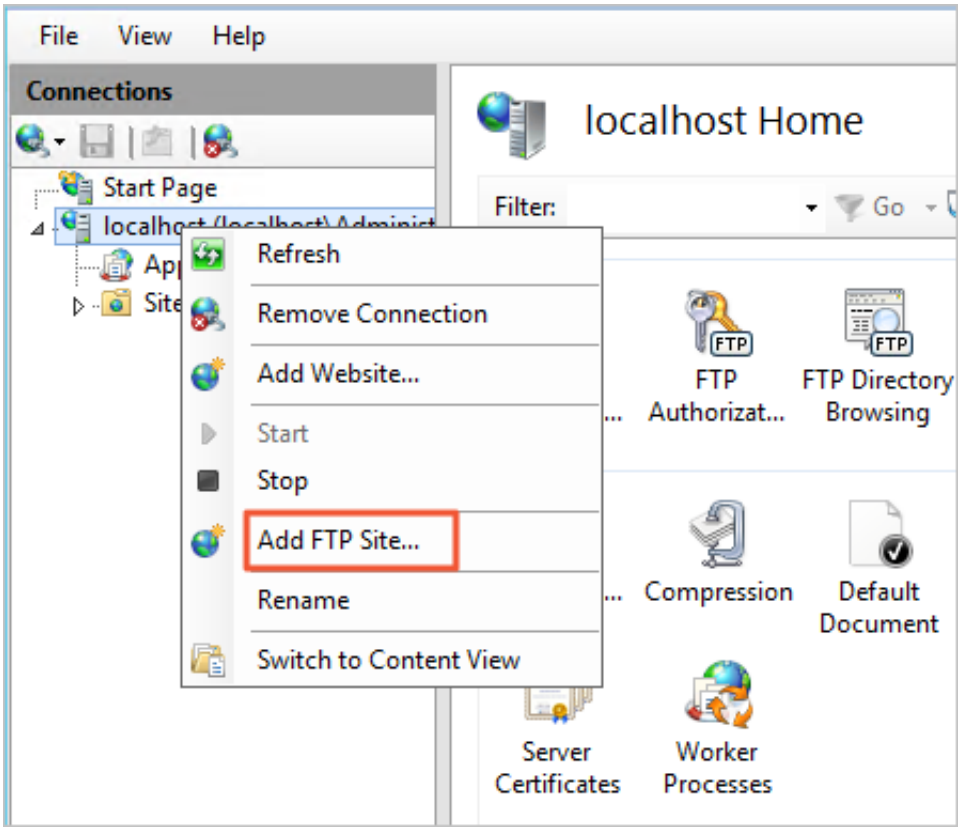
2. Start the IIS manager, as shown in the following figure.

Figure 11-43: Start the IIS manager



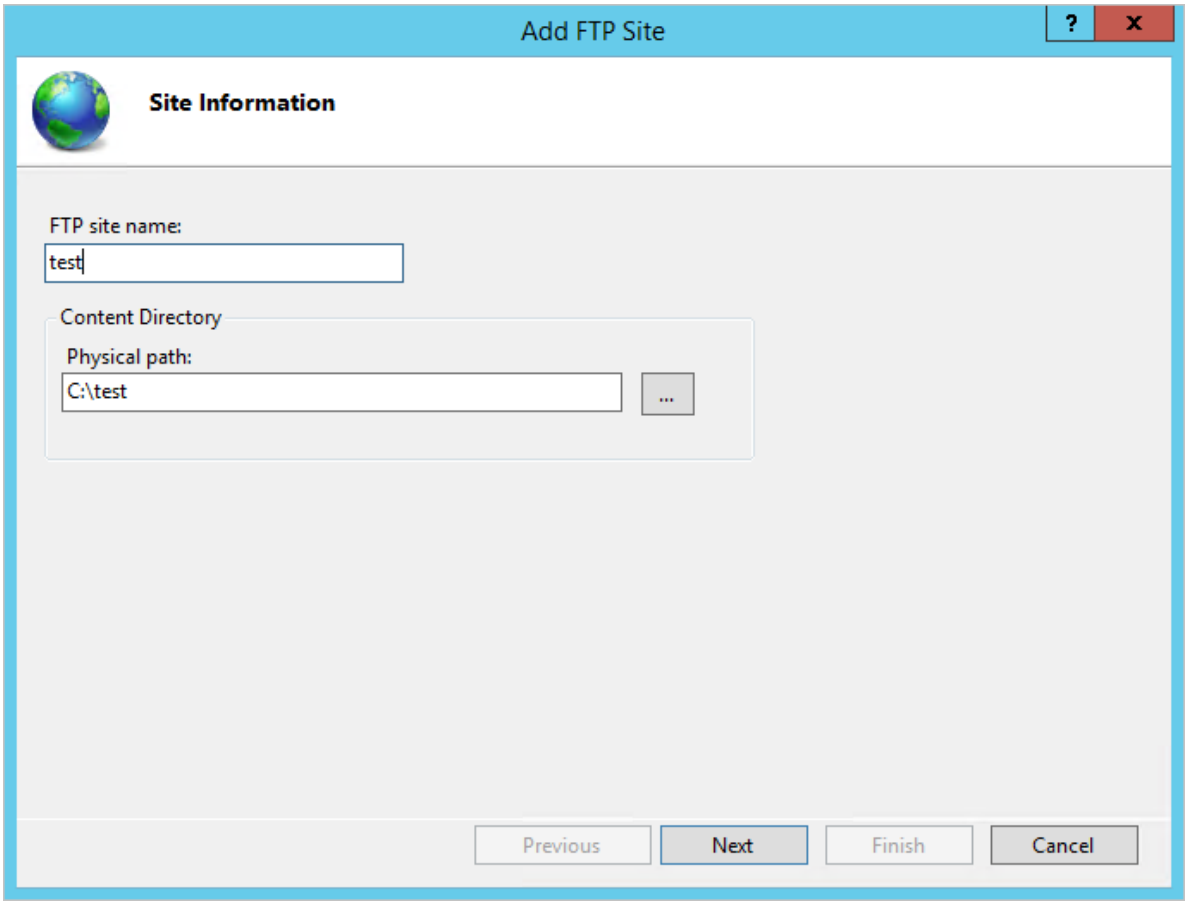
3. Add an FTP site to the IIS manager, as shown in the following figure.

Figure 11-44: Add an FTP site



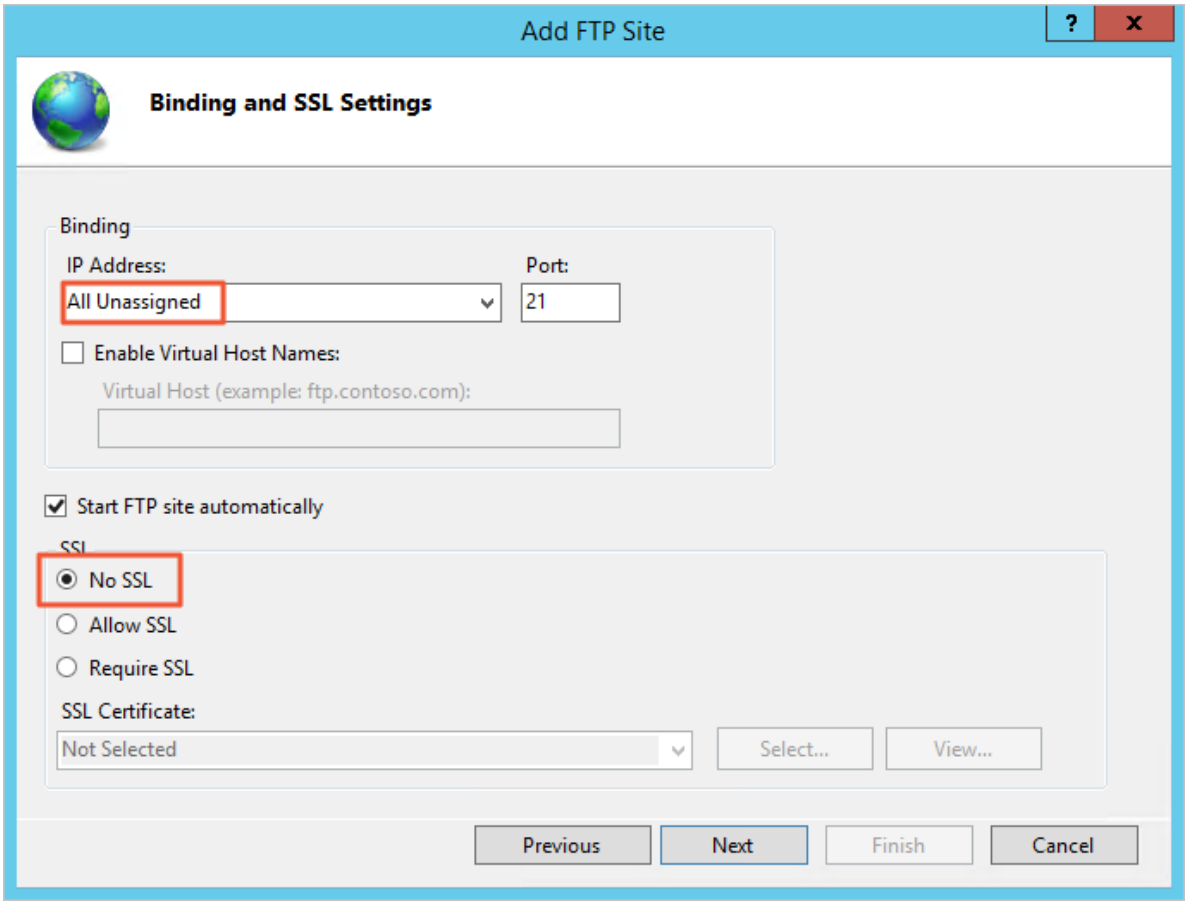
4. Enter the FTP site name and specify the FTP path.

Figure 11-45: Enter site information



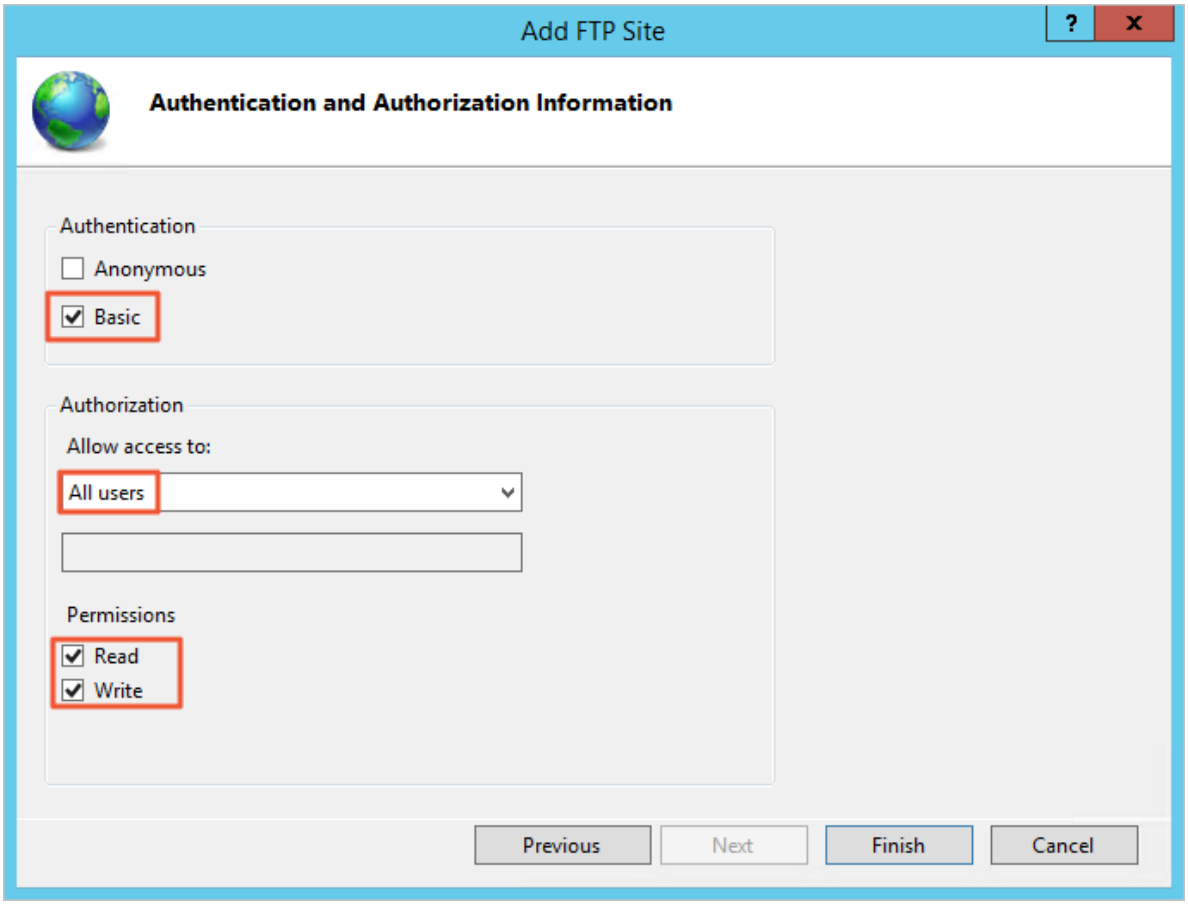
5. As shown in the following figure, set **IP Address** to **All Unassigned** and **SSL** to **None**.

Figure 11-46: Bind an IP address and set SSL



6. Set **Authentication** to **Basic**, **Authorization** to **All Users**, and **Permissions** to **Read** and **Write**.

Figure 11-47: Set authentication and authorization



7. After completing FTP settings, use the default administrator account and password to test logon. Then you can upload and download files.

12 Auto Scaling (ESS)

12.1 What is ESS

Auto Scaling (ESS) is a management service that automatically adjusts the number of elastic computing resources based on your business demands and strategies.

Based on user-defined scaling rules, ESS automatically adds ECS instances as business loads increase to ensure sufficient computing capabilities. When your business loads decrease, ESS automatically removes ECS instances to reduce running costs.

ESS provides the following functions:

- **Elastic scale-out**

  When business loads surge, ESS automatically increases underlying resources. This helps maintain access speed and ensure that resources are not overloaded. For example, if the CPU utilization of ECS instances exceeds 80%, ESS scales out ECS resources based on the rules you defined. During the scale-out process, ESS automatically creates and adds ECS instances to a scaling group, and adds the new instances to the SLB instance and RDS whitelist.

- **Elastic scale-in**

  When business loads decrease, ESS automatically releases underlying resources. This prevents resource wastage and helps to reduce cost. For example, if the CPU utilization of ECS instances in a scaling group falls below 30%, ESS scales in ECS resources based on the rules you defined. During the scale-in process, ESS removes the ECS instances from the scaling group, the SLB instance, and RDS whitelist.

- **Elastic recovery**

  The health status of ECS instances in a scaling group is determined based on the life cycle of the instances. If an ECS instance is in an unhealthy state, ESS automatically releases the instance and creates a new one. ESS adds the new instance to the SLB instance and RDS whitelist. This process is called elastic recovery. It ensures that the number of healthy ECS instances in a scaling group will not fall below the threshold that you defined.

12.2 Usage

12.2.1 Overview

Before you use ESS, you must understand its usage limitations and take necessary precautions.

12.2.2 Precautions

This topic describes the precautions when you use ESS.

Scaling rules

During calculation and execution, a scaling rule can automatically adjust the number of ECS instances that need to be increased or decreased based on the MinSize and MaxSize values of the scaling group. For example, if the number of ECS instances to be increased that is specified by a scaling rule is 50 but MaxSize of the scaling group is 45, the scaling rule will be adjusted to increase the number of instances to a maximum of 45 instances.

Scaling activities

- Only one scaling activity can be executed at a time in a scaling group.
- A scaling activity cannot be interrupted. For example, if a scaling activity to create 20 ECS instances is being executed but only five have been created, the scaling activity cannot be forcibly terminated.
- When a scaling activity fails to complete, the system prioritizes the integrity of the ECS instances over the scaling activity. The system will roll back the ECS instances that fail to be added or removed, but not the scaling activity. For example, if a scaling group has 20 ECS instances, out of which 19 instances are added to SLB, only the one ECS instance that failed to be added is automatically released.

Cooldown period

- During the cooldown period, if you manually execute a trigger task such as scaling rule or scheduled task, the task is executed immediately without being affected by the cooldown period.
- The cooldown period starts after the last ECS instance is added to or removed from the scaling group by a scaling activity.

12.2.3 Manual intervention

If you manually intervene with ESS operations, ESS will process the intervention accordingly.

ESS does not prevent you from performing manual intervention, such as deleting automatically created ECS instances through the ECS console. The following table describes how ESS processes manual intervention.

Resource	Manual intervention	Processing method
ECS	A user deletes an ECS instance from a scaling group through the ECS console or open API.	ESS determines whether the ECS instance is in an unhealthy state through health check. If it is, ESS removes the instance from the scaling group. The intranet IP address of the ECS instance is not automatically deleted from the RDS access whitelist. When the number of ECS instances (Total Capacity) in the scaling group is smaller than MinSize, ESS automatically creates and adds ECS instances to the group until the number of instances is equal to MinSize.
ECS	A user revokes the ECS open API permissions granted to ESS.	ESS rejects all scaling activity requests.
SLB	A user manually removes an ECS instance from an SLB instance through the SLB console or open API.	ESS does not automatically detect this action or handle such exceptions. The ECS instance remains in the scaling group. If this instance is selected based on the removal policy during a scaling activity, it is released.
SLB	A user manually deletes an SLB instance or disables its health check function through the SLB console or open API.	ESS does not add ECS instances to scaling groups that are associated with this SLB instance. Scaling tasks can trigger scaling rules to remove ECS instances from the scaling group. ECS instances determined to be unhealthy by the health check function are also removed.
SLB	An SLB instance is unavailable because of system-related reasons.	All scaling activities fail except for instance removal tasks that are manually executed.

Resource	Manual intervention	Processing method
SLB	A user revokes the SLB open API permissions granted to ESS.	ESS rejects all scaling activity requests for scaling groups associated with SLB instances.
RDS	A user manually removes the IP address of an ECS instance from an RDS whitelist through the RDS console or open API.	ESS does not automatically detect this action or handle such exceptions. The ECS instance remains in the scaling group. If this instance is selected based on the removal policy during a scaling activity, it is released.
RDS	A user manually deletes an RDS instance through the RDS console or open API.	ESS does not add ECS instances that are associated with this RDS instance to scaling groups. Scaling tasks can trigger scaling rules to remove ECS instances from the scaling group. ECS instances determined to be unhealthy by the health check function are also removed.
RDS	An RDS instance is unavailable because of system-related reasons.	All scaling activities fail except for instance removal tasks that are manually executed.
RDS	A user revokes the RDS open API permissions granted to ESS.	ESS rejects all scaling activity requests for the scaling groups associated with RDS instances.

12.2.4 Quantity limits

Before you use ESS, you need to understand the quantity limits of ESS.

You can only create a limited number of scaling groups, scaling configurations, scaling rules, scaling ECS instances, and scheduled tasks.

Table 12-1: Quantity limits

Item	Description
Scaling group	You can create a maximum of 20 scaling groups.
Scaling configuration	You can create a maximum of 10 scaling configurations in a scaling group.
Scaling rule	You can create a maximum of 10 scaling rules in a scaling group.
ECS instances for scaling	You can configure a maximum of 100 ECS instances for automatic scaling.

Item	Description
	Note: This limit applies to the ECS instances that are automatically created, but does not apply to manually added ones.
Scheduled task	You can create a maximum of 20 scheduled tasks.

12.2.5 Scaling group statuses

Before you manage a scaling group, you need to understand the scaling group statuses.

A scaling group can be in Active, Inactive, or Deleting state. *Table 12-2: Scaling group statuses* describes the details.

Table 12-2: Scaling group statuses

Status	Status in open API
Creating	Inactive
Created	Inactive
Enabling	Inactive
Running	Active
Disabling	Inactive
Stopped	Inactive
Deleting	Deleting

12.2.6 Scaling activity process

Before you use ESS, you need to understand the processes related to the scaling activity.

Automatic scaling of a scaling group

Automatic scale-out

1. Check the health status and other prerequisites for scaling.
2. Allocate the activity ID and execute the scaling activity.
3. Create ECS instances.
4. Modify Total Capacity.
5. Allocate IP addresses to the created ECS instances.
6. Add ECS instances to the RDS whitelist.

7. Start ECS instances.
8. Associate the ECS instances to an SLB instance and set the weight to the **SLB weight** value when the scaling configuration is created.
9. The scaling activity completes, and the cooldown period starts.

Automatic scale-in

1. Check the health status and other prerequisites for scaling.
2. Allocate the activity ID and execute the scaling activity.
3. Remove ECS instances from the SLB instance.
4. Stop ECS instances.
5. Remove ECS instances from the RDS whitelist.
6. Release ECS instances.
7. Modify Total Capacity.
8. The scaling activity completes, and the cooldown period starts.

Manually adding or removing existing ECS instances

Manually adding

1. Check the health status and other prerequisites for scaling, and check the status and type of ECS instances.
2. Allocate the activity ID and execute the scaling activity.
3. Add ECS instances.
4. Modify Total Capacity.
5. Add ECS instances to the RDS whitelist.
6. Associate ECS instances to an SLB instance and set the weights to the **SLB weight** value of the active scaling configuration.

> **Note:**
> When you need to manually add an instance, the instance type must be the same as that specified in the active scaling configuration of the scaling group. Therefore, you must set the weight to the **SLB weight** value specified in the scaling configuration.

7. The scaling activity completes, and the cooldown period starts.

Manual removal

1. Check the health status and boundary conditions of a scaling group.

2. Allocate the activity ID and execute the scaling activity.
3. SLB stops forwarding traffic to ECS instances.
4. Remove ECS instance from SLB after 60 seconds.
5. Remove ECS instances from the RDS whitelist.
6. Modify Total Capacity.
7. Remove ECS instances from the scaling group.
8. The scaling activity completes, and the cooldown period starts.

> **Note:**
> The life cycle of a scaling activity starts at **checking the health status and other prerequisites for scaling.** and ends at **starting the cooldown time**.

12.2.7 Removal of unhealthy ECS instances

Before you use ESS, you need to read information about the removal of unhealthy ECS instances.

After an ECS instance has been successfully added to a scaling group, ESS periodically scans its status. If the ECS instance is not in **Running** state, ESS removes the ECS instance from the scaling group.

- If an ECS instance is created automatically, ESS immediately removes and releases it.
- If the ECS instance is added manually by a user, ESS immediately removes it, but does not stop or release it.

The MinSize attribute of a scaling group does not limit the removal of unhealthy ECS instances. That is, the total number of ECS instances can fall below MinSize after the removal. ESS automatically creates ECS instances based on the difference between the actual instance number and MinSize to ensure the total number is equal to MinSize.

12.2.8 Instance rollback after a scaling activity failure

Before you use ESS, you need to understand the mechanism of instance rollback after a failed scaling activity.

When a scaling activity fails to complete, the system prioritizes the integrity of the ECS instances over the scaling activity. The system will roll back the ECS instances that fail to be added or removed, but not the scaling activity. That is, the system rolls back ECS instances, not the scaling activity.

Example

If a scaling group has created 20 ECS instances, out of which 19 instances are added to SLB, only the one ECS instance that failed to be added is automatically released.

12.2.9 Instance life cycle management

Before you use ESS, you need to understand the concepts related to instance life cycles.

ECS instances in a scaling group can be created automatically or added manually.

Automatically created ECS instances

ECS instances are automatically created by ESS based on user-defined scaling configurations and rules.

ESS manages the entire life cycle of this type of ECS instances. ESS creates this type of ECS instances during scale-out, and stops and release them during scale-in.

Manually added ECS instances

ECS instances are manually added to a scaling group.

ESS does not manage the entire life cycle of this type of ECS instances. Such instances are not created by ESS, but are manually added by a user to a scaling group. When the ECS instances are removed from a scaling group manually or as the result of an automatic scale-in, ESS removes the instances but does not stop or release them.

Instance status

An ECS instance in a scaling group undergoes the following status during its life cycle:

- Pending: The ECS instance is being added to a scaling group. For example, ESS is creating the instance or adding it to an SLB instance or RDS whitelist.
- In Service: The ECS instance has been successfully added to a scaling group and is providing services normally.
- Removing: The ECS instance is being removed from a scaling group.

Instance health statuses

An ECS instance in a scaling group has the following health statuses:

- Healthy
- Unhealthy

If an ECS instance is not in Running state, it is considered as an unhealthy instance. ESS automatically removes unhealthy ECS instances from a scaling group.

- ECS instances that are **automatically created** are stopped and released by ESS.
- ECS instances that are **manually added** are not stopped and released by ESS.

12.3 Quick start

12.3.1 Overview

This topic describes how to quickly create scaling groups, configurations, and rules. It is designed to guide you through the process of automatic scaling creation.

Follow these steps:

1. *Create a scaling group*

   Configure information such as MinSize and MaxSize attributes of scaling resources, as well as SLB and RDS instances to be associated with a scaling group.
2. *Create a scaling configuration*

   Configure ECS instance configurations for automatic scaling, such as Image ID and Instance Type.
3. *Enable a scaling group*

   Enable the scaling group created in step 2.
4. *Create a scaling rule*

   Create a scaling rule based on actual conditions. ESS executes scaling based on the specified rule, such as **adding N ECS instances**.
5. *Create a scheduled task*

   Create a scheduled task based on actual conditions. ESS executes scaling rules at a specified point in time. For example, you can create a scheduled task to execute the scaling rule created in step 4 at 12:00.

12.3.2 Log on to the ESS console

This section describes how to log on to the ESS console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.

- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.
5. From the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Auto Scaling**.

12.3.3 Create a scaling group

You must create scaling groups before you can use the services provided by ESS.

Prerequisites

- The scaling groups must be in the same region as the SLB and RDS instances that they will be associated with.
- A maximum of 20 scaling groups can be created by a user.

Procedure

1. *Log on to the ESS console*.
2. Navigate to the **Scaling Groups** page, and click **Create Scaling Group**. On the page that appears, configure relevant parameters.

   Table 12-3: Parameters for creating a scaling group describes the configurations of each parameter.

Table 12-3: Parameters for creating a scaling group

Type	Parameter	Description
Region	**Region**	Required. The region to which the scaling group belongs.
Basic Settings	**Department**	Required. The department to which an instance belongs.
	Project	Required. The project to which an instance belongs.
	Scaling Group Name	Required. A scaling group name must be 2 to 40 characters in length. It can contain numbers, uppercase letters, lowercase letters, underscores (_), hyphens (-), and periods (.). It must start with a number or letter.
	Minimum Instances	Required. The minimum number of instances that a scaling group must contain to ensure availability. After you have completed this scaling group configuration, the system creates a group containing the number of instances as specified here. Value range: 0–100.
	Maximum Instances	Required. The maximum number of instances that a scaling group can contain, to control costs. Value range: 0–100.
	Default Cooldown Time	Required. The cooldown period for a scaling group. After a scaling activity has been successfully executed and the last ECS instance is added to or removed from the group, the cooldown period starts immediately. During the cooldown period, this scaling group cannot execute any new scaling activities. The value must be an integer with a minimum value of 0. Value range: 0–86400.
	Removal Policy	Optional. This policy is used to filter and remove ECS instances from a scaling group using multiple filtering conditions.
Network type	**VPC**	Virtual Private Cloud (VPC): VPC helps you build an isolated network environment in Apsara Stack. You can customize routing tables, IP address segments, and gateways in

Type	Parameter	Description
		a VPC. We recommend that you set Network Type to VPC to improve security. Before you set Network Type to VPC, ensure that you have created a VPC and VSwitch. For more information, see **Create a VPC** and **Create a VSwitch** in *VPC User Guide*.
	Classic Network	Cloud services in a classic network are not isolated. Unauthorized access to cloud services is blocked only by the security group or whitelist policy.
Whitelist Configuration	**SLB Instances**	Optional. The SLB instance to be associated with the scaling group. If an SLB instance is specified for a scaling group, the scaling group automatically adds its ECS instances to the specified SLB instance.  **Note:** • The specified SLB instance must be active . • An ECS instance that is added for load balancing has a default weight of 50.
	Databases	• Optional. If an RDS instance is specified for a scaling group, the scaling group automatically adds the intranet IP addresses of its ECS instances to the whitelist of the specified RDS instance.  **Note:** — The specified RDS instance must be in the **`Running`** state. — The number of IP addresses in the whitelist of the specified RDS instance cannot exceed the whitelist upper limit. — A scaling group does not take effect immediately after creation. It must be enabled before scaling rules can be triggered and scaling activities can be executed.

3. After you complete the parameter configurations, click **Create**.

12.3.4 Create a scaling configuration

You can create a scaling configuration to customize the specifications of the ECS instances that are to be automatically added to a scaling group.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, click the ID of the scaling group for which you want to create a scaling configuration.
3. Click the **Scaling Configurations** tab, and then click **Create Scaling Configuration**. On the page that appears, configure relevant parameters.

   Table 12-4: Parameters for creating a scaling configuration describes the configurations of each parameter.

   Table 12-4: Parameters for creating a scaling configuration

Type	Parameter	Description
Basic Settings	**Configuration Name**	The name of the scaling configurat ion. The name must be 2 to 40 characters in length. It can contain numbers, uppercase letters, lowercase letters, underscores (_), hyphens (-), and periods (.). It must start with a number or letter.
	Security Groups	The security group to which an instance belongs.
Instances	**Instance Series**	The default value is Series 3.
	I/O Optimized	The default value is I/O-Optimized Instances.
	Instance Specification	The instance specification that you need.
Network	**Public Bandwidth**	The method used for calculate billing for public bandwidth. The values include **Pay By Traffic** and **By Fixed Bandwidth**.

Type	Parameter	Description
	Bandwidth	The bandwidth that you need. You can adjust the slider to set the bandwidth.
Images	**Image Type**	• Click **Public Image** to select an operating system and version as required. • If you need to configure automatic Web server startup, code download, and script download, select **Custom Image**.
Storage	**System Disk**	The type and size of the cloud disk that is used to install the system image.
	Data Disk	The data disks to be added. You can select the type, size, and mount point of the data disk. In the current ESS version, you can add up to four data disks to each ECS instance.

4. After you complete the parameter configurations, click **Create**.

12.3.5 Enable a scaling group

Before you use a scaling group, you must manually enable the group.

Prerequisites

- The scaling group is in the **Inactive** state.
- The scaling group has an active scaling configuration.
- A single scaling group can only have one **Active** scaling configuration at a time.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group that you want to enable, click the  icon in the **Actions** column, and choose **Enable** from the shortcut menu.
3. In the message that appears, click **OK**.

Result

The scaling group status will change from **Inactive** to **Active**.

12.3.6 Create a scaling rule

After you create a scaling group, you must create scaling rules to be applied within the group.

Context

- Up to 10 scaling rules can be created within a scaling group.
- If the number of ECS instances in a scaling group is smaller than **Minimum Instances** or greater than **Maximum Instances** after the execution of a rule, ESS automatically adds or removes a number of ECS instances. This ensures that the number of instances is always within the preset value range.
- After a scaling rule is created, a unique identifier is generated. The unique identifier can be used by the following open APIs:
  - ScalingRuleAri: You can specify the identifier when calling this API to manually execute a scaling rule.
  - ScheduledAction: You can specify the identifier when calling this API to execute the scaling rule at scheduled intervals.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group for which you want to create a scaling rule, and click the group ID. The **Basic Information** page appears.
3. Click the **Scaling Rules** tab, and then click **Create Scaling Rule**.
4. In the **Create Scaling Group Rules** dialog box that appears, configure relevant parameters.

   Table 12-5: Parameters for creating a scaling rule describes the configurations of each parameter.

   Table 12-5: Parameters for creating a scaling rule

Parameter	Description
Rule Name	The name of a scaling rule. The name must be 2 to 40 characters in length. It can contain numbers, uppercase letters, lowercase letters, underscores (_), hyphens (-), and periods (.). It must start with a number or letter.
Rule	Select **Change To**, **Add**, or **Remove** from the drop-down list, and enter a number in the text box to specify the number of ECS instances.
Cooldown Time	The cooldown period.

Parameter	Description
	Note: If this parameter is left empty, the cooldown period of the scaling group will be used by default.

5. After you have completed the parameter configuration, click **OK**.

12.3.7 Create a scheduled task

In the ESS console, you can directly create a scheduled task.

Prerequisites

- A scheduled task is created based on input parameters. You can create up to 20 scheduled tasks.
- If multiple tasks are scheduled for the same point in time, the most recent scheduled task is executed.

Procedure

1. *Log on to the ESS console*.
2. Click the **Scheduled Task** tab. Then, click **Create Scheduled Task** and configure parameters.

   Table 12-6: Parameters for scheduled task configuration describes the configurations of each parameter.

   Table 12-6: Parameters for scheduled task configuration

Type	Parameter	Description
Region	**Region**	Required. It is automatically set by the system and cannot be changed.
Basic Configuration	**Department**	Required. It specifies the department to which an instance belongs.
	Project	Required. It specifies the project to which an instance belongs.
	Task Name	Required. The name must be 2 to 40 characters in length. It can contain numbers, uppercase letters, lowercase letters, underscores (_), hyphens (-), and periods (.). It must start with a number or letter.

Type	Parameter	Description
	Description	Optional. It is a description of the instance. It must be at least two characters in length.
	Execution Time	Required. It specifies the execution time of a task.
	Scaling rule	Required. It specifies the scaling group and rule of the task to be executed.
	Retry Expiry Time (Seconds)	Optional. It specifies a period of time during which the system retries to execute a task.
Repeat Cycle	**Settings**	It specifies whether to set a repeat cycle. The default value is **`No Watermark`**. If you select **`Settings`**, you must set **`Repeat Cycle`** and **`Duplicate End Time`** simultaneously.

3. After you complete the parameter configurations, click **Create**.

12.4 Scaling group

12.4.1 Overview

A scaling group is a group of ECS instances that is dynamically scaled based on the configured scenario. You can specify the maximum and minimum number of ECS instances in a scaling group, as well as SLB and RDS instances associated with the group.

12.4.2 Query a scaling group

You can query created scaling groups and their related information in the ESS console.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, set filtering conditions such as **`Department`**, **`Region`**, or **`Instance Name`**, and click **Search**.

> **Note:**
>
> - Click **Instance Name**, you can select other filtering conditions from the drop-down list, including **`Instance ID`**, **`Status`**, and **`Project Name`**.
> - On the **Scaling Groups** tab page, you can click an instance ID to go to the **Basic Information** page.

3. Click the  icon in the **Actions** column corresponding to an instance, and choose **View Details** from the shortcut menu. The **Basic Information** page appears.

12.4.3 Modify scaling group information

You can modify scaling group information in the ESS console.

Context

When the number of ECS instances in a scaling group is not within the range specified by MinSize and MaxSize, ESS automatically adds or removes instances from the group. This ensures that the number of instances in the group is within the values specified for MaxSize and MinSize.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group to be modified, click the  icon in the **Actions** column, and choose **Change** from the shortcut menu.
3. In the **Edit Scaling Group Information** dialog box that appears, modify the parameters of the scaling group.
4. After you have completed the parameter configurations, click **Change**.

12.4.4 Disable a scaling group

You can disable scaling groups in the ESS console.

Prerequisites

- You can disable a scaling group only when it is not executing any scaling activity.
- You can successfully disable a scaling group only when it is in the **Active** state.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group to be disabled, click the  icon in the **Actions** column, and choose **Disable** from the shortcut menu.
3. In the message that appears, click **OK**.

Result

The scaling group status changes from **Active** to **Inactive**.

12.4.5 Delete a scaling group

You can delete scaling groups that you no longer use in the ESS console.

Context

- If you delete a scaling group through the ESS console, the group is forcibly deleted.
- Deleting a scaling group also deletes all scaling configurations, rules, activities, and requests related to the group.
- Deleting a scaling group does not delete scheduled tasks, SLB instances, or RDS instances related to the group.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group to be deleted, click the  icon in the **Actions** column, and choose **Delete** from the shortcut menu.
3. In the message that appears, click **OK**.

12.4.6 Query ECS instances

You can query ECS instances in the ESS console.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group whose ECS instance you want to query, click the  icon in the **Actions** column, and choose **View Details** from the shortcut menu.
3. Click the **ECS Instances** tab to view instance details.
   - **Query ECS instances in a scaling group**

     There are two types of ECS instances in a scaling group:

     — **Automatically created**: ECS instances that are created automatically based on scaling configurations and rules.

     — **Manually added**: ECS instances that are added manually to a scaling group by a user.
   - **Life cycle of ECS instances in a scaling group:**

     — **Adding**: An ECS instance is being added to a scaling group. For example, the instance is being created or added to an SLB or RDS whitelist.

- **`In Service`**: An ECS instance has been successfully added to a scaling group and is providing services normally.
- **`Removing`**: An ECS instance is being removed from a scaling group.

- **ECS health status**

  Health status of ECS instances:

  - Healthy
  - Unhealthy

  ESS automatically removes unhealthy ECS instances from a scaling group. ECS instances that are **automatically created** are stopped and released by ESS. ECS instances that are **manually added** are not stopped or released by ESS.

12.5 Scaling configuration

12.5.1 Overview

Scaling configurations specify the specifications of ECS instances used for automatic scaling. When automatically adding ECS instances to a scaling group, ESS will create ECS instances based on the scaling configurations.

12.5.2 Query a scaling configuration

You can query created scaling configurations and their related information in the ESS console.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group of which you want to query the scaling configuration, and click a configuration in the **`Scaling Configuration`** column corresponding to the group. The **Scaling Configurations** tab page appears.
3. On the **Scaling Configurations** tab page, click the  icon in the **`Actions`** column and choose **View Details** from the shortcut menu.

> **Note:**
> You can also click a scaling configuration name on the Scaling Configurations tab page to view configuration details.

12.6 Scaling rule

12.6.1 Overview

Scaling rules define specific scaling actions executed by ESS, such as scaling in and out ECS instances.

12.6.2 Query a scaling rule

You can query created scaling rules and their related information in the ESS console.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group for which you want to query a scaling rule, and click the group ID. The **Basic Information** page appears.
3. Click the **Scaling Rules** tab, locate the scaling rule you are searching for, and click the  icon in the **Actions** column.
4. Choose **View Details** from the shortcut menu. The scaling rule details are displayed.

12.6.3 Edit a scaling rule

You can edit scaling rule information in the ESS console.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group for which you want to edit scaling rules, and click the group ID. The **Basic Information** page appears.
3. Click the **Scaling Rules** tab, locate the scaling rule to be modified, click the  icon in the **Actions** column, and choose **Change** from the shortcut menu.
4. In the **Edit Scaling Group Rules** dialog box that appears, edit **Rule Name**, **Rule**, or **Cooldown Time**, and click **OK**.

 **Note:**

- **Rule Name**: The name must be 2 to 40 characters in length. It can contain numbers, uppercase letters, lowercase letters, underscores (_), hyphens (-), and periods (.). It must start with a number or letter.

- **`Cooldown Time`**: If this parameter is left empty, the cooldown period of the scaling group will be used by default.

12.6.4 Delete a scaling rule

You can delete scaling rules that you no longer use in the ESS console.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group that contains the scaling rule to be deleted, and click the group ID. The **Basic Information** page appears.
3. Click the **Scaling Rules** tab, locate the scaling rule to be deleted, click the  icon in the **`Actions`** column, and choose **Delete** from the shortcut menu.
4. In the **Delete Scaling Rule** message that appears, click **OK**.

12.7 Trigger tasks

12.7.1 Overview

In the ESS console, you can perform automatic scaling by manually executing scaling rules or adding ECS instances.

12.7.2 Manually execute a scaling rule

This topic describes how to manually execute a scaling rule.

Prerequisites

If you need to execute a scaling rule, note the following limits:

- The status of the scaling group including the scaling rule must be **`Active`**.
- The scaling group including the scaling rule is not executing any scaling activity.
- An Apsara Stack tenant account can automatically scale up to a maximum of 1,000 ECS instances across all scaling groups in all regions. This limit applies to the ECS instances that are automatically created, but does not apply to manually added ECS instances.
- ESS automatically scales ECS instances to ensure that the actual number of instance does not exceed the limits.

Procedure

1. *Log on to the ESS console*.

2. On the **Scaling Groups** tab page, locate the scaling group for which you want to execute a scaling rule, and click the group ID. The **Basic Information** page appears.
3. Click the **Scaling Rules** tab, locate the scaling rule to be executed, click the  icon in the **Actions** column, and choose **Execute** from the shortcut menu.

12.7.3 Add an ECS instance

You can add ECS instances to a specific scaling group in the ESS console.

Prerequisites

Only ECS instances that meet the following conditions can be added to the scaling group:

- It must be in the same region as the scaling group.
- The instance type must be the same as that specified in the active scaling configuration.
- It must be in the **Running** state.
- It cannot belong to any other scaling group at the same time.
- If the network type of the scaling group is VPC, only instances belonging to the same VPC as the scaling group can be added.

Before adding an ECS instance, you must ensure that:

- The scaling group is in the **Active** state.
- No scaling activities are being executed in the scaling group.

Note:

- If no scaling activities are being executed in the scaling group, you can immediately remove ECS instances and do not need to wait after the cooldown period.
- If a manual addition operation would cause the number of instances to exceed MaxSize, the add operation fails.
- Manually added ECS instances are not associated with the active scaling configuration in the scaling group.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group to which you want to add ECS instances, and click the Group ID. The **Basic Information** page appears.
3. Click the **ECS Instances** tab, and click **Add an Existing Instance**.

4. In the dialog box that appears, select the target instances in the right-side **`Available Instances`** list, click the  icon to add instances to the left-side **`Selected Instances`** list, and click **OK**.

 **Note:**

- You can click **All** to select all instances in a list.
- Click the  icon to remove selected instances.

12.7.4 Remove an ECS instance

You can remove ECS instances from a scaling group in the ESS console.

Prerequisites

- When an automatically created ECS instance is removed from a scaling group, the instance is stopped and released.
- When a manually added ECS instance is removed from a scaling group, the instance is not stopped or released.

Prerequisites to remove an ECS instance:

- The scaling group must be **`Active`**.
- No scaling activities are being executed in the scaling group.

 **Note:**

- If no scaling activities are being executed in the scaling group, you can immediately remove ECS instances and do not need to wait after the cooldown period.
- If a manual remove operation would cause the number of instances be less than MinSize, the remove operation fails.

Procedure

1. *Log on to the ESS console*.
2. On the **Scaling Groups** tab page, locate the scaling group from which you want to remove ECS instances, and click the group ID. The **Basic Information** page appears.
3. Click the **ECS Instances** tab. Click **Auto Created** or **Manually Added** on the tab page.

4. Locate the instance to be removed, click the  icon in the **Actions** column, and choose **Remove and Release** from the shortcut menu.
5. In the message that appears, click **OK**.

12.8 Scheduled tasks

12.8.1 Overview

If a scaling group is disabled or executing a scaling activity, a scheduled task fails to execute a scaling rule. The scheduled task is automatically retried within LaunchExpirationTime. After LaunchExpirationTime expires, the task is abandoned. If multiple tasks in the same group are scheduled at similar points in time, the earliest task executes its scaling activity first. A scaling group can execute only one scaling activity at a time. Other tasks attempt to execute the rule within LaunchExpirationTime. If a scaling activity is completed within LaunchExpirationTime, the completed activity will trigger the next scheduled scaling rule and execute the scaling activity.

12.8.2 Query a scheduled task

In the ESS console, you can query created scaling rules and related information.

Procedure

1. *Log on to the ESS console*.
2. Click the **Scheduled Task** tab, locate a task to be queried, and click the  icon in the **Actions** column.
3. Choose **View Details** from the shortcut menu. The task details are displayed.

12.8.3 Edit a scheduled task

In the ESS console, you can modify scheduled task information.

Procedure

1. *Log on to the ESS console*.
2. Click the **Scheduled Task** tab, locate a scheduled task to be edited, click the  icon in the **Actions** column, and choose **Change** from the shortcut menu.
3. In the displayed **Change Scheduled Task** dialog box, modify the following parameters:

Table 12-7: Parameters for scheduled task modification

Parameter	Description
Task Name	It specifies the name of a task. The name must be 2 to 40 characters in length. It can contain numbers, uppercase letters, lowercase letters, underscores (_), hyphens (-), and periods (.). It must start with a number or letter.
Description	It specifies supplementary information about a task. It must be at least two characters in length.
Execution Time	It specifies the execution time of a task.
Scaling Rule	It specifies the scaling rule of a task.
Retry Expiry Time	It specifies a period of time during which the system retries to execute a task.
Duplicate Setting	It defaults to Setting.
Repeat Cycle	It specifies the repetition cycle of a scheduled task.
Duplicate End Time	It specifies the end time of a repeating scheduled task.

4. After you complete the parameter configurations, click **OK**.

12.8.4 Disable a scheduled task

In the ESS console, you can disable scheduled tasks.

Prerequisites

The scheduled task must be in **Running** state.

Procedure

1. *Log on to the ESS console*.
2. Click the **Scheduled Task** tab, locate the scheduled task to be disabled, click the  icon in the **Actions** column, and choose **Disable** from the shortcut menu.
3. In the displayed message, click **OK**.

12.8.5 Enable a scheduled task

In the ESS console, you can enable scheduled tasks in **Stopped** state as needed.

Prerequisites

The scheduled task is in **Stopped** state.

Procedure

1. *Log on to the ESS console*.
2. Click the **Scheduled Task** tab, locate the target scheduled task, click the  icon in the **Actions** column, and choose **Enable** from the shortcut menu.
3. In the displayed message, click **OK**.

12.8.6 Delete a scheduled task

You can delete scheduled tasks that you no longer use in the ESS console.

Procedure

1. *Log on to the ESS console*.
2. Click the **Scheduled Task** tab, locate a scheduled task to be deleted, click the  icon in the **Actions** column, and choose **Delete** from the shortcut menu.
3. In the message that appears, click **OK**.

13 Object Storage Service (OSS)

13.1 What is OSS

Alibaba Cloud Object Storage Service (OSS) is a massive, secure, low-cost, and highly reliable cloud storage service provided by Alibaba Cloud.

It can be considered as an out-of-the-box storage solution with unlimited storage capacity. Compared with the user-created server storage, OSS has many outstanding advantages in reliability, security, cost, and data processing capabilities. Using OSS, you can store and retrieve a variety of unstructured data files, such as text files, images, audios, and videos, over the network at any time.

OSS uploads data files as objects to buckets. OSS is an object storage service that uses a key-value pair format. You can retrieve object content based on unique object names (keys).

On OSS, you can:

- Create a bucket and upload objects to the bucket.
- Obtain an object URL from OSS to share or download an object.
- Complete the ACL settings of a bucket or object by modifying its properties or metadata.
- Perform basic and advanced OSS tasks through the OSS console.
- Perform basic and advanced OSS tasks using the Alibaba Cloud SDKs or directly calling the RESTful APIs in your application.

13.2 Instructions

Before you use OSS, you need to understand the following content.

- To allow other users to use all or part of OSS functions, you need to create RAM users and grant permissions to the users by assigning RAM policies to them. For more information, see contents related to RAM users and RAM policies in RAM User Guide.
- Before you use OSS, you also need to understand the following service limits:

Item	Description
Bucket	— You can create a maximum of 10 buckets. — You cannot change the name or region of a bucket after it is created.

Item	Description
Object upload	— Objects larger than 5 GB cannot be uploaded by using the following modes: console upload, simple upload, form upload, or append upload. To upload an object that is larger than 5 GB, you must use the multipart upload mode. The size of an object that you want to upload in multipart upload mode cannot exceed 48.8 TB. — You can upload an object with the same name as an existing object. This object will overwrite the existing one.
Object deletion	— Deleted objects cannot be restored. — You can delete up to 50 objects at one time in the OSS console. To delete more than 50 objects at one time, you must use an API or SDK.
Lifecycle	You can configure up to 1,000 lifecycle rules for each bucket.
Image processing	— For source images: ■ Only JPG, PNG, BMP, GIF, WEBP, and TIFF formats are supported. ■ The object size cannot exceed 20 MB. ■ To use the image rotation or cropping function, the width or height of an image cannot exceed 4,096 pixels. — For thumbnails: ■ The object dimensions cannot exceed 4,096 x 4,096 pixels. ■ The length of each side cannot exceed 4,096 pixels.

13.3 Quick start

13.3.1 Log on to the OSS console

This topic describes how to log on to the OSS console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The

access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.

- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.
5. Log on to Apsara Stack Management Console. Choose **Console** > **Compute, Storage, and Networking** > **Object Storage Service** from the top navigation bar.

13.3.2 Create a bucket

Objects uploaded to OSS are stored in buckets. Before you upload an object to OSS, you need to create a bucket.

Context

Properties of a bucket include geographic region, ACL settings, and other metadata.

Procedure

1. *Log on to the OSS console*.
2. Click **Create Bucket**. On the **Add Bucket** page that appears, set required parameters.

   Table 13-1: Parameters for creating a bucket lists the parameters for creating a bucket.

Table 13-1: Parameters for creating a bucket

Parameter	Description
Department	Select a department from the drop-down list.
Project	Select a project from the drop-down list.
Region	Select the data center where the bucket is deployed from the drop-down list. **Note:** • After a bucket is created, the region cannot be changed. • If you want to access OSS through the ECS intranet, select the same region where your ECS instance is deployed.
Permissions	Select permissions of the bucket. The following options are available: • **Private**: Only the owner of the bucket and the authorized users can perform read, write, and delete operations on objects in the bucket. Other users cannot access objects in the bucket. • **Public (Read-Only)**: Only the owner of the bucket and the authorized users can perform write operations on objects in the bucket. All users (including anonymous users) can perform read operations on objects in the bucket. • **Public**: All users (including anonymous users) can perform read and write operations on objects in the bucket. Fees incurred by such operations are paid by the owner of the bucket. Configure this permission only when necessary. **Note:** After a bucket is created, you can modify its permissions. For more information, see *Change ACL settings*.
Bucket Name	Enter the name of the bucket.

Parameter	Description
	Note: • The bucket name must comply with the naming conventions. • The bucket name must be globally unique in Alibaba Cloud OSS. • The bucket name cannot be changed after the bucket is created.
Bucket Capacity	Configure the capacity of the bucket.
Quantity	Enter the number of buckets that you apply for. You can create a maximum of 10 buckets at a time.

3. Click **Create**. The bucket is created successfully.

13.3.3 Upload objects

After you create a bucket, you can upload objects to it.

Context

You can upload an object of any format to a bucket. You can use the OSS console to upload an object no larger than 500 MB to a bucket. To upload an object larger than 500 MB, use an SDK or API.

Procedure

1. *Log on to the OSS console*.
2. On the OSS homepage, click the name of the target bucket. The bucket details page is displayed.
3. Click the **Object Management** tab. The Object Management tab page is displayed.
4. Click **Upload Files**.
5. In the dialog box that appears, select the object that you want to upload and click **Open**.
6. After the object is successfully uploaded, refresh the Object Management tab page to view the uploaded object.

   You can view the upload progress and result on the **Task Management** tab page.

13.3.4 Obtain an object URL

You can obtain the URL of an object uploaded to a bucket. This URL can be used to share or download the object.

Prerequisites

Before you obtain an object URL, you need to create a bucket and upload an object to it.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket. The bucket details page is displayed.
3. Click the **Object Management** tab. The Object Management tab page is displayed.
4. Click the  icon in the Actions column corresponding to the object and click **Get URL**. The **Get Object URL** dialog box is displayed.

    **Note:**
   To obtain the URL of a private bucket, you need to configure a validity period for the URL. Click **Get URL** to get the object URL. The validity period of a signed URL is calculated based on NTP. You can share the object URL to other users so that they can use the URL to access the object within the validity period. If a bucket is private, the obtained URL is a signed URL.
5. Copy the object URL and send it to other users so that they can view or download the object.

13.4 Bucket

13.4.1 View a bucket

You can view the details of created buckets in the OSS console.

Prerequisites

Before you view a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket or the  icon in the Actions column and then click **Details**.

   On the bucket details page that appears, click the **Bucket Information** tab. On the Bucket Information tab page, view bucket details such as Service IP Address and Creation Time.

13.4.2 Delete a bucket

You can delete a bucket in the OSS console.

Prerequisites

Before you delete an object, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

> **Note:**
> To delete a bucket, make sure that all objects in it are deleted, including parts generated by incomplete multipart upload. Otherwise, the bucket cannot be deleted.

Procedure

1. *Log on to the OSS console*.
2. Click the icon in the Actions column corresponding to the target bucket and click **Delete**.
3. In the **Delete Bucket** dialog box, click **OK**.

> **Note:**
> To delete multiple buckets at a time, select these buckets and click **Delete**.

13.4.3 Change the capacity

During actual usage, you may need to scale up or down the capacity of a bucket. You can change the capacity of a bucket in the OSS console.

Prerequisites

Before you change the capacity of a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Procedure

1. *Log on to the OSS console*.
2. Click the icon in the Actions column of the target bucket and click **Change Capacity**.
3. In the **Change Capacity** dialog box that appears, change the capacity of the bucket. Click **OK**.

13.4.4 Change the ownership

You can change the department or project to which a bucket belongs in the OSS console.

Prerequisites

Before you change the department or project to which a bucket belongs, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Procedure

1. *Log on to the OSS console*.
2. Click the  icon in the Actions column of the target bucket and click **Change Ownership**.
3. In the **Change Ownership** dialog box that appears, change the department or project to which the bucket belongs. Click **OK**.

13.4.5 Change ACL settings

You can change Access Control List (ACL) settings of a bucket in the OSS console to control access to the bucket.

Prerequisites

Before you change the ACL settings of a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

OSS provides the ACL function to control bucket access permissions. After a bucket is created, the ACL of the bucket is set to private by default. You can change ACL settings after creating a bucket.

The OSS ACL function provides bucket-level access control. Three ACL settings are available for a bucket:

- Private: Only the bucket owner and authorized users can perform read and write operations on objects in the bucket. Other users cannot access objects in the bucket without authorization.
- Public (Read-Only): Only the bucket owner and authorized users can perform write operations on objects in the bucket. Other users (including anonymous users) can perform read operations on objects in the bucket.
- Public: All users (including anonymous users) can perform read and write operations on objects in the bucket. Fees incurred by these operations are paid by the bucket owner. Configure this permission type only when necessary.

Procedure

1. *Log on to the OSS console*.

2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Bucket Properties** tab. On the Bucket Properties tab page, click **Read/Write Permissions**.
4. Select an option for Read/Write Permissions.
5. Click **Set** to save your modifications.

13.4.6 Configure static website hosting

You can configure static website hosting in the OSS console so that users can access the static website from the bucket endpoint.

Prerequisites

Before you configure static website hosting for a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

If the default page is blank, static website hosting is disabled.

The default homepage is displayed if you directly access the root domain name of the static website or any URL ending with a forward slash (/) under this domain name.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Bucket Properties** tab. On the Bucket Properties tab page, click **Website Settings**.

   Configure the following parameters:

   - **Default Homepage** indicates the index page (equivalent to index.html of a website). You must enter the name of an HTML object that is stored in the bucket.
   - **Default 404 Page** indicates the default 404 page that is displayed when you access an incorrect path. You must enter the name of an HTML object that is stored in the bucket. If this field is left empty, the default 404 page is disabled.
4. Click **Set** to save your settings.

13.4.7 Enable the logging function

You can enable or disable the logging function for a bucket in the OSS console.

Prerequisites

Before you enable or disable the logging function for a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

You can store logs of a bucket in the current bucket or a new bucket.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Bucket Properties** tab. On the Bucket Properties tab page, click **Logging Settings**.

   Configure the following parameters:

   - Select the name of the bucket where logs are to be stored from the **Log Storage Location** drop-down list. The selected bucket must be owned by you and in the same region as the bucket in which logging is enabled. Select **Do Not Save** to directly disable logging.
   - Enter a prefix in **Log Prefix**. This parameter corresponds to <*TargetPrefix*> in the following naming conventions. Logs are stored in the root directory. You can also add a folder path in front of <*TargetPrefix*>, such as *log/*<*TargetPrefix*>. Logs are stored in the *log/* directory.

   The naming conventions for objects that store access logs are as follows:

   <TargetPrefix><SourceBucket>YYYY-MM-DD-HH-MM-SS-<UniqueString>

   - <*TargetPrefix*>: indicates the specified log prefix.
   - <*SourceBucket*>: indicates the name of the source bucket.
   - *YYYY-MM-DD-HH-MM-SS*: indicates the time in CST (UTC+8) when the log is created. YYYY indicates a 4-digit year, MM indicates a 2-digit month, DD indicates a 2-digit day, HH indicates a 2-digit hour, MM indicates a 2-digit minute, and SS indicates a 2-digit second.
   - <*UniqueString*>: indicates a string generated by OSS.

For example, the name of an object that stores OSS access logs is MyLog-OSS-example2015-09-10-04-00-00-0000.

MyLog- is the specified log prefix. oss-example is the name of the source bucket. 2015-09-10-04-00-00 is the time in CST (UTC+8) when the log is created. 0000 is a string generated by OSS.

4. Click **Set** to save your settings.

13.4.8 Configure hotlinking protection

You can configure hotlinking protection for a bucket in the OSS console to prevent other domains from directly linking the data in your bucket.

Prerequisites

Before you configure hotlinking protection for a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

OSS provides the hotlinking protection function to prevent other domains from linking to your data stored in OSS. You can configure the referer field in the HTTP header to realize the protection. In the OSS console, you can configure a whitelist for the referer field and configure whether to allow requests with an empty referer field. For example, for a bucket named oss-example, configure its referer whitelist to http://www.aliyun.com. Then only requests with the referer http://www.alicloud.com can access objects in the oss-example bucket.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Bucket Properties** tab. On the Bucket Properties tab page, click **Hotlink Protection Settings**.
4. Enter whitelist URLs in **Referer**.
5. Configure whether to allow requests with an empty referer field.

   Select Allow Empty referer Field if you do not need to restrict access requests.
6. Click **Submit** to save your settings.

13.4.9 Configure CORS

You can configure cross-origin resource sharing (CORS) in the OSS console to enable cross-region access.

Prerequisites

Before you configure CORS for a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

OSS provides HTML5 CORS settings to enable cross-region access. When OSS receives a CORS request (or OPTIONS request), it reads the CORS rules of the target bucket and checks its ACL settings. OSS matches the rules one by one. When OSS finds the first match, it returns a corresponding header. If no match is found, OSS does not attach any CORS-related headers.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Bucket Properties** tab. On the Bucket Properties tab page, click **CORS Settings**.
4. Click **Add Rules**
5. In the **Add CORS Settings** dialog box that appears, set required parameters.

   Table 13-2: Parameters for CORS settings lists parameters for CORS settings.

   Table 13-2: Parameters for CORS settings

Parameter	Description
Source	Specifies the sources of allowed CORS requests. You can configure multiple matching rules separated by a carriage return . Each matching rule can contain up to one asterisk (*) wildcard.
Method	Specifies the allowed CORS request methods .
Allowed Header	Specifies the allowed CORS request headers . You can configure multiple matching rules separated by a carriage return. Each

Parameter	Description
	matching rule can contain up to one asterisk (*) wildcard.
Expose Header	Specifies the response headers that allow access from applications.
Cache Time	Specifies the duration when the browser catches the response to a prefetch (OPTIONS) request to a specific resource.

Note:
You can configure up to 10 rules for each bucket.

6. Click **Confirm** to save your settings. You can also modify or delete existing rules.

13.4.10 Manage lifecycle rules

You can define and manage lifecycle rules for a bucket in the OSS console.

Prerequisites

Before you manage lifecycle rules of a bucket, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

You can define a rule for a full set or a subset (by specifying the prefix keyword) of objects in a bucket. A rule is automatically applied to all objects that match the rule. You can use lifecycle management to perform operations, such as batch object management and automatic part deletion.

- If an object matches a rule, data of the object is cleared within two days from the effective date.
- Data that is deleted in batches based on a lifecycle rule cannot be restored. Configure a rule only when necessary.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Bucket Properties** tab. On the Bucket Properties tab page, click **LifeCycle Settings**.

4. Click **Add Rules**.
5. In the **Add LifeCycle Rule** dialog box, configure required parameters.

   Table 13-3: Parameters for lifecycle settings lists the parameters for lifecycle settings.

   Table 13-3: Parameters for lifecycle settings

Parameter	Description
Status	Select the status of this rule. You can select Enable or Disable.
Policy	Select an object matching policy, including Configure for Entire Bucket and Configure by Prefix.
Prefix	If image objects with the prefix img/ are stored in the bucket, you can enter `img/` in this field to manage the lifecycle of these objects.
Expired	Configure an expiration date or days to expiration for objects. • **Set by Date**: indicates the **expiration date** of objects. All objects that are created before this date are deleted. Perform this operation only when necessary. • **Set by Number of Days**: indicates **days to expiration**, that is, a lifecycle of objects in days. When the number of days from the last modification of an object exceeds the specified number of days, the object is deleted based on the rule. If this parameter is set to 30 days, objects last modified on January 1, 2016 are deleted on January 31, 2016.

6. Click **Confirm** to save your settings. After the rule is saved, you can view it on the LifeCycle Settings tab page. You can click **Edit** or **Delete** in the Actions column to edit or delete the rule.

13.4.11 Configure cross-cloud data synchronization

You can synchronize data in a bucket to another cloud in the OSS console.

Prerequisites

Before you configure cross-cloud data synchronization, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region and there is data uploaded to the bucket.

Context

You can specify a prefix to allow only objects with this prefix to be replicated to another cloud. You can configure a data synchronization policy by specifying a synchronization type and historical data synchronization.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Bucket Properties** tab. On the Bucket Properties tab page, click **Copy Cross-Cloud Server Settings**.
4. On the Copy Cross-Cloud Server Settings tab page, click **Enable Data Synchronization**. The **Enable Data Synchronization** dialog box is displayed.
5. Configure parameters for cross-cloud data synchronization.

   Table 13-4: Parameters for data synchronization lists the parameters for data synchronization.

   Table 13-4: Parameters for data synchronization

Parameter	Description
Synchronization Target Cloud	Select the target cloud for data synchroniz ation.
Synchronization Target Cloud Address	Enter the address of the target cloud.
Synchronization Target Bucket	Enter the bucket name in the target cloud. **Note:** Cross-cloud data synchronization replicates data in the source bucket to the target bucket in the target cloud. Therefore, you need to specify the target bucket in the target cloud. The target bucket must have the same name as the source bucket.

Parameter	Description
Data Synchronization Object	Select the object of data synchronization. You can select **Synchronize All Files**. You can also choose **Synchronize Files with** > **Add** and enter a prefix to synchronize objects with this prefix.
Data Synchronization Policy	Select a data synchronization policy. You can select **Write Synchronization** or **Add/Delete/Modify**.
Synchronize Historical Data	Select whether to synchronize historical data.

6. Click **Confirm** to save your settings. After your settings are saved, you can view this rule on the Copy Cross-Cloud Server Settings tab page. You can click **Edit** or **Delete** in the **Actions** column to edit or delete the rule.

13.5 Object

13.5.1 Search for objects

You can search buckets or folders for objects with a specific name prefix in the OSS console.

Prerequisites

Before you search for an object, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region and at least one object in the bucket.

Context

When you search for an object based on a prefix, the search string is case-sensitive and cannot contain a forward slash (/). The search range is limited to the root directory of the current bucket or the objects in the current folder (excluding subfolders and the objects in them).

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Object Management** tab.
4. On the Object Management tab page, enter a prefix in the search box, and press Enter or click **Search**.

To search a folder, open the folder and enter a prefix in the search box. The system lists the names of objects and folders matching the prefix in the root directory of the folder.

13.5.2 Delete objects

You can delete uploaded objects in the OSS console.

Prerequisites

Before you delete objects, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region and at least one object in the bucket.

Context

You can delete one or more objects at a time. Up to 50 objects can be deleted at a time. You can use an SDK or API to delete a specific object or more than 50 objects at a time.

Note:
Deleted objects cannot be restored. Perform this operation only when necessary.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Object Management** tab.
4. On the Object Management tab page, click the  icon in the Actions column corresponding to the target object and click **Delete**.

   Note:
   A folder may fail to be deleted if it contains an excessive number of objects.
5. In the **Delete Object** dialog box, click **Confirm**.

13.5.3 Configure ACL of an object

You can set the ACL for an object in the OSS console to control access to the object.

Prerequisites

Before you configure ACL of an object, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region and at least one object in the bucket.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Object Management** tab. The Object Management tab page is displayed.
4. Click the  icon in the Actions column corresponding to the target object and click **Set File ACL**.
5. In the **Set File ACL** dialog box that appears, select an option from the **Read/Write Permissions** drop-down list.
6. Click **OK**.

13.5.4 Create a folder

You can create a folder in a bucket in the OSS console.

Prerequisites

Before you create a folder, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

OSS does not use folders. All elements are stored as objects. In the OSS console, a folder is an object with a size of **0**, whose name ends with a forward slash (/). A folder is used to sort objects of the same type and process them in batches. The OSS console displays objects ending with a slash as folders by default. This object can be uploaded and downloaded normally. In the OSS console, you can use OSS folders similar to Windows folders.

Note:
The OSS console displays any object ending with a forward slash as a folder, no matter whether it contains data. You can download such objects only by using an application programming interface (API) or software development kit (SDK).

Procedure

1. *Log on to the OSS console*.

2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Object Management** tab.
4. On the Object Management tab page, click **Create Folder**.
5. In the **Create Folder** dialog box, enter a folder name in **Folder Name**.
6. Click **OK**.

13.6 Image service

13.6.1 Create a style

You can create an image style in the OSS console to define a processing rule for uploaded image objects.

Prerequisites

Before you create an image style, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Image Processing** tab.
4. On the Image Processing tab page, click **Create Style**.

   Table 13-5: Parameters for creating a style lists the parameters for creating a style.

   Table 13-5: Parameters for creating a style

Parameter	Description
Rule Name	Enter a style name, which must comply with the naming conventions.
Edit Type	You can select Basic to edit the image style based on the graphical user interface (GUI). You can also select Advanced to edit the image style using an SDK or parameters.
Preview	Select an image preview method.
Thumbnail Style	Select a thumbnail method.

Parameter	Description
	Note: Longer edge refers to the side with a bigger source size to target size ratio. Shorter edge refers to the side with a smaller source size to target size ratio. For example, for a source image that is scaled from 400 x 200 pixels to 800 x 100 pixels, the original-to-target ratios are 0.5 (400/800) and 2 (200/100). Because 0.5 is smaller than 2, the side that is 200 pixels is the longer side, and the side that is 400 pixels is the shorter side.
Thumbnail Width	Configure the thumbnail dimension in px.
Thumbnail Limit	Configure whether to disable enlarging.
Fit Direction	Configure the image fit direction, including **Original Image Default** and **Shrink after Rotating**.
Image Processing	Configure whether to perform special processing. You can select **Sharpening**.
Picture Quality	Configure image quality, including **Relative**, **Absolute**, and **No Compression**.
Save Format	Configure the format in which an image is saved, including **Original Format**, **jpg**, **png**, **webp**, and **bmp**.
Add Watermark	Configure the watermark mode of an image, including **No Watermark**, **Text Watermark**, and **Image Watermark**.

5. After editing the image style, click **Submit** to save your settings.
6. After the style is submitted, you can click **Export Style** to download the style to your local device.

13.6.2 Enable source image protection

You can enable source image protection in the OSS console to prevent unauthorized use of images.

Prerequisites

Before you enable source image protection, make sure that you have completed the procedure described in *Quick start*, or there is at least one bucket in the current region.

Context

To prevent unauthorized use of images in business systems, you need to prevent exposure of image URLs. Thus, unauthorized users can only obtain thumbnails or watermarked images. For this purpose, you can enable source image protection. After source image protection is enabled, source images are accessible only through URLs carrying stylenames or signature-based accesses. You are not allowed to access source images in OSS or specify image parameters to modify image styles.

Procedure

1. *Log on to the OSS console*.
2. Click the name of the target bucket, or click the  icon in the Actions column and then click **Details**.
3. On the bucket details page that appears, click the **Image Processing** tab. On the Image Processing tab page, click **Service Management**.
4. Click **Edit**. Set **Source Image Protection**.

   If you select Enable, you must also set **File Extensions for Source Image Protection** to restrict access to source images with one or more extensions.

   Source image protection is designed to protect image objects. You must specify the extensions of image objects to be protected. For example, if you enable source image protection for `.jpg` objects, you can still directly access the source images of `.png` objects.
5. In **Style Access Method**, set **Delimiter (Default: @!)**.
6. Click **Save** to save your settings.

13.7 Create single tunnels

You can create single tunnels between OSS and a VPC so that you can access OSS resources from the VPC.

Prerequisites

Before you can create single tunnels, you need to create a VPC and a VSwitch.

Procedure

1. *Log on to the OSS console*.
2. Click the **OSS Access Control for VPC** tab.
3. On the OSS Access Control for VPC tab page, click **Create Single Tunnel**.
4. In the **Create Single Tunnel** dialog box, set required parameters.

Table 13-6: Parameter for creating single tunnels lists the parameter configurations for creating single tunnels.

Table 13-6: Parameter for creating single tunnels

Parameter	Description
Region	Select a region.
Department	Select a department or all departments.
Description	Enter a description for the single tunnel.
VPC	Select a VPC. For more information about how to create a VPC, see "**Create a VPC and a VSwitch**" in *VPC User Guide*.
VSwitch	Select a VSwitch. For more information about how to create a VSwitch, see "**Create a VSwitch**" in *VPC User Guide*.

5. Click **Confirm**.

14 Table Store

14.1 What is Table Store

Table Store is a NoSQL database service independently developed by Alibaba Cloud. Table Store is a copyrighted software program that is certified by the relevant authority in China. Table Store is built on Alibaba Cloud's Apsara system, and can store and access large volumes of structured data in real time.

Table Store provides the following features:

- Supports a minimum of 10 PB of data in each cluster, and a minimum of 1 PB of data or 1 trillion records in each table. Table Store offers schema-free data structure storage. You do not need to define attribute columns before you use them. You do not require table-level changes to add or reduce attribute columns. You can enable Time To Live (TTL) on a table to delete expired data from the table.
- Adopts the triplicate technology to keep three copies of data on three servers placed on three different racks. Each cluster supports either pure SSD instances or mixed storage instances to meet different budget and performance requirements.
- Adopts a fully redundant architecture that prevents single point of failure (SPOF). With support for hot cluster upgrades and automatic data migration, you can dynamically add or delete nodes without service interruptions. The concurrent read and write throughput and storage capacity can be linearly scaled. Each cluster can have no less than 500 nodes.
- Supports highly concurrent read/write operations. Concurrent read/write operations scale with the number of hosts. Read/write performance is not directly related to the amount of data in a single table.
- Supports identity authentication and multi-tenancy; provides comprehensive permission authentication and isolation mechanisms to safeguard your data; supports VPC networks and access through HTTPS; supports RAM and account authorization; provides multiple authentica tion and authorization mechanisms so that you can define access permissions for individual tables and APIs.

14.2 Instructions

Before using Table Store, you need to understand the precautions or restrictions.

The following table describes the restrictions for Table Store. Some of the limit ranges indicate the maximum available values instead of the suggested values. For better performance, set the

table structure and data size in a single row properly based on actual conditions, and adjust the following configurations.

Item	Limit range	Description
Number of instances under an Alibaba Cloud account	1,024	To raise the upper limit, contact the technical support personnel.
Number of tables in an instance	1,024	To raise the upper limit, contact the technical support personnel.
Instance Name Length	3–16 bytes	The character set includes [a-z, A-Z, 0-9] and hyphens (-). It must start with a letter and cannot finish with a hyphen (-).
Table name length	1–255 bytes	The character set includes [a-z, A-Z, 0-9] and underlines (_). It must start with a letter or underline (_).
Column name length	1–255 bytes	The character set includes [a-z, A-Z, 0-9] and underline (_). The name must start with a letter or underline (_).
Number of primary key columns	1–4	A primary key can contain one to four columns.
Size of string type primary key column values	1 KB	The values of the String type columns in a single primary key column cannot exceed 1 KB.
Size of string type attribute column values	2 MB	The values of the string type columns in a single attribute column cannot exceed 2 MB.
Size of binary type primary key column values	1 KB	The values of the binary columns in a single primary key column cannot exceed 1 KB.
Size of binary type attribute column values	2 MB	The values of the binary columns in a single attribute column cannot exceed 2 MB.
Number of attribute columns in a row	Unlimited	The number of attribute columns in a single row is unlimited.
Number of attribute columns written in a single request	1,024	During the PutRow, UpdateRow, or BatchWriteRow operation, the number of attribute columns written in a single row cannot exceed 1,024.
Data size of a single row	Unlimited	The total size of all column names and column values for a single row are unlimited.

14.3 Quick start

14.3.1 Log on to the Table Store console

This topic describes how to log on to the Table Store console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.
5. In the top navigation bar, choose **Console** > **Compute, Storage&Networking** > **Table Store**.

14.3.2 Create an instance

An instance is a logical entity in Table Store used to manage tables. An instance is the basic unit in the resource management system of Table Store. Table Store provides application access control and resource measurement at the instance level.

Procedure

1. *Log on to the Table Store console*.
2. On the Table Store page, click **Create Instance**.

   Note:
   You can create different instances for different services to manage related tables, or create different instances for development, test, and production environments of the same service. Table Store allows you to create up to 1,024 instances under a cloud account and up to 1,024 tables in each instance by default.
3. Set **`Instance Name`**, **`Department`**, **`Project`**, **`Region`**, and **`Instance Specification`**.

   Note:
   - Table Store supports high-performance instance and capacity instance. The instance specifications depend on the cluster you deploy in.
   - Naming rules for an instance name: An instance name must be 3 to 16 characters in length and can contain letters, numbers, and hyphens (-). It must start with letters. It cannot finish with hyphens (-).
4. Click **OK**.

   The new instance is displayed in the list.

14.3.3 Create a table

After creating an instance, you can create, update, and delete tables in the instance.

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance you want to manage and click the instance name to go to the **Instance Details** page.
3. Click the **Data Tables** tab.
4. On the **Data Tables** tab page, click Create Data Table.

   Note:
   You can create a maximum of 64 data tables in an instance.
5. Enter the data table information.

   Configuration descriptions are provided, as shown in *Table 14-1: Data table parameters*

Table 14-1: Data table parameters

Parameter	Description
Data Table Name	A data table name can contain uppercase/ lowercase letters, numbers, and underlines (_). It must start with a letter or underline (_). The data table name must be unique at the instance level.
Reserved Read Throughput	The reserved read/write throughput can be set to 0. When the reserved read/ write throughput is larger than 0, Table Store allocates and reserves correspond ing resources for the table based on the configuration. The value ranges from 0 to 5000 and must be an integer. Capacity-type instances do not support this parameter.
Reserved Write Throughput	
Data Life Cycle	The minimum data life cycle is 86,400s (one day) or –1. (Data never expires.)
Maximum Data Version	A non-zero value. Maximum Data Version indicates the maximum number of data versions that can be stored in each attribute column of a data table. When the number of versions in an attribute column exceeds the parameter value, the earliest version will be deleted asynchronously.
Valid Data Version Margin	The offset of the version of all written data columns from the data write time must be within the range of the valid data version offset. Otherwise, data write may fail. The valid version range of an attribute column is calculated based on the formula: Valid data version margin range = `[Data write time – Valid data version margin, Data write time + Valid data version margin)`.

Parameter	Description
Table Primary Key	A maximum of four primary keys can be set. The first primary key is the partition key by default. Click **Add Primary Key** to add a new primary key. The primary key type can be **`Integer`** or **`String`**. Once set, the primary key configuration and the key order cannot be modified. The primary key name can contain uppercase /lowercase letters, numbers, and underlines (_). It must start with a letter or underline (_).

6. Click **OK**.

   After the table is created, it is displayed in the table list.

14.4 Manage instances

14.4.1 View an instance

On the Table Store console, you can view the region, creation time, and internal and external access URLs of an instance you have created.

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance you want to view, click the  icon in the Actions column, and choose

   View Details from the shortcut menu.

   The following information is displayed: the status, region, creation time, and internal and external access URLs of the instance, as well as whether the instance is bound to VPC.

14.4.2 Release an instance

You can release a Table Store instance you have created.

Prerequisites

Before releasing an instance, delete all tables from the instance. Otherwise, the instance cannot be released.

Procedure

1. *Log on to the Table Store console*.

2. Locate the instance you want to release, click the icon in the Actions column, and choose **Release** from the shortcut menu.
3. In the **Delete** page, click **OK**.

14.5 Manage tables

14.5.1 View table details

You can view the basic information and actual usage of a table on the table management page.

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance you want to view and click the instance name to go to the **Instance Details** page.
3. Click the **Data Tables** tab, locate the instance you want to view, click the icon in the Actions column, and choose **View Details** from the shortcut menu.

   You can view the data table name, reserved read/write throughput, last modification time, primary keys (sorted in the sequence specified during table creation) and stream information.

14.5.2 Update a table

You can change the parameter values of a Table Store table, such as reserved read/write throughput and data life cycle.

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance you want to manage and click the instance name to go to the **Instance Details** page.
3. Click the **Data Tables** tab, locate the table you want to update, click the icon in the Actions column, and choose **Adjust Data Table Parameters** from the shortcut menu.
4. On the **Adjust Data Table Parameters** page, enter the parameters you want to update, such as the reserved read/write throughput and data life cycle.
5. Click **Confirm**. The parameter values take effect immediately.

14.5.3 Delete a table

You can delete a table you have created in the Table Store console.

Context

> **Note:**
> After a data table is deleted, the data in the table cannot be restored.

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance you want to manage and click the instance name to go to the **Instance Details** page.
3. Click the **Data Tables** tab, locate the table you want to delete, click the icon in the Actions column, and choose **Release** from the shortcut menu.
4. In the **Confirm Deletion** dialog box, click **Confirm**.

   After the deletion is confirmed, the table and the data in the table are deleted permanently.

14.6 Bind VPC instances

Virtual Private Cloud (VPC) is an isolated network environment built on Apsara Stack.

Prerequisites

- You must create a VPC instance first. Select an appropriate node when creating a VPC instance and ensure that the VPC and Table Store instances are in the same node. For more information about how to create a VPC instance, see **Create a VPC Instance** in *VPC User Guide*.
- After the VPC instance is created, create an ECS instance in the VPC instance.

Context

You can take full control of your virtual network if a VPC is bound to Table Store. For example, you can select a private IP address range, allocate network segments, or configure a routing table and gateway. You can also connect a VPC instance to a traditional data center through a leased line or VPN to build an on-demand network environment, achieving smooth cloud migration.

Procedure

1. *Log on to the Table Store console*.

2. Locate the instance you want to bind and click the instance name to access the **Instance Details** page.
3. Click **Bind VPC** to access the instance and VPC instance binding page.
4. Enter the information and click **OK**.

   Before you use a sub-account to log on to the Table Store console and manage VPCs, ensure that the sub-account has relevant VPC permissions (AliyunVPCReadOnlyAccess) obtained from the RAM console. Otherwise, you cannot obtain relevant VPC information due to lack of permissions.
5. After the instance is bound to the VPC instance, the system automatically returns to the instance details page. Information about the bound VPC instance is displayed in the VPC instance list. Click the link in the VPC ID column. The Table Store instances bound to the VPC instance and the VPC information list are displayed.

   You can use the access URL of the VPC instance as the endpoint to access Table Store from the ECS instance in the VPC instance.

What's next

If the VPC instance is not needed, click the  icon in the Actions column, and choose **Unbind** from the shortcut menu to unbind the ECS instance from the VPC instance.

After the instance is unbound from the VPC instance, the ECS instance in the VPC instance cannot access Table Store through the preceding URL. To access Table Store, you need to bind the instance to the VPC instance again.

15 Network Attached Storage (NAS)

15.1 What is NAS?

Alibaba Cloud Network Attached Storage (NAS) is a file storage service that can be mounted to compute nodes such as ECS, E-HPC, and Container Service instances.

NAS allows you to use standard file access protocols to access distributed file systems without making any changes to your existing applications. NAS features unlimited capacity and performance expansion, single namespace, data sharing, high reliability, and high availability. Compared with traditional user-created storage, NAS greatly reduces maintenance costs and data security risks. In addition, a NAS instance can be mounted to multiple compute nodes at the same time, greatly reducing data replication and synchronization costs.

You can perform the following operations:

- Create NAS instances and mount points.
- Create permission groups and add rules to the permission groups in NAS instances. These rules allow access and grant different levels of access permissions to IP addresses or IP address segments.
- Use the standard NFS or SMB protocol to mount NAS instances to compute nodes such as ECS, E-HPC, and Container Service instances, and use the standard POSIX interface to access the NAS instances.
- Perform basic and advanced operations on NAS instances, mount points, and permission groups in the NAS console.
- Call NAS APIs to perform basic and advanced operations on NAS instances.

15.2 Instructions

Before you can use NAS, you need to understand the following content.

- NAS supports the NFSv3 and NFSv4 protocols.
- NFSv4.0 does not support the following attributes: FATTR4_MIMETYPE, FATTR4_QUOTA_AVAIL_HARD, FATTR4_QUOTA_AVAIL_SOFT, FATTR4_QUOTA_USED, FATTR4_TIME_BACKUP, and FATTR4_TIME_CREATE. The client displays an NFS4ERR_ATTRNOTSUPP error.
- NFSv4.1 does not support the following attributes: FATTR4_DIR_NOTIF_DELAY, FATTR4_DIRENT_NOTIF_DELAY, FATTR4_DACL, FATTR4_SACL, FATTR4_CHA

NGE_POLICY, FATTR4_FS_STATUS, FATTR4_LAYOUT_HINT, FATTR4_LAYOUT_TYPES , FATTR4_LAYOUT_ALIGNMENT, FATTR4_FS_LOCATIONS_INFO, FATTR4_MDS THRESHOLD, FATTR4_RETENTION_GET, FATTR4_RETENTION_SET, FATTR4_RET ENTEVT_GET, FATTR4_RETENTEVT_SET, FATTR4_RETENTION_HOLD, FATTR4_MOD E_SET_MASKED, and FATTR4_FS_CHARSET_CAP. The client displays an NFS4ERR_AT TRNOTSUPP error.

- NFSv4 does not support the following OPs: OP_DELEGPURGE, OP_DELEGRETURN, and NFS4_OP_OPENATTR. The client displays an NFS4ERR_NOTSUPP error.
- NFSv4 does not support Delegation.
- About UID and GID:
  - For NFSv3, if the file UID or GID exists in a Linux local account, the corresponding username and group name are displayed based on the mapping between the local UID and GID. If the file UID or GID does not exist in the local account, the UID and GID are displayed .
  - For NFSv4, if the version of the local Linux kernel is earlier than 3.0, "nobody" is displayed as the UID and GID of all files. If the kernel version is later than 3.0, the display rule is the same as that of NFSv3.

  Note:
  If you use NFSv4 to mount a NAS instance and the Linux kernel version is earlier than 3.0, we recommend that you do not change the owner or group of local files or directories. Such changes can cause the UIDs and GIDs of the files or directories to become "nobody."
- A NAS instance can be mounted to up to 10,000 compute nodes for parallel access.

15.3 Quick start

15.3.1 Log on to the NAS console

This topic describes how to log on to the NAS console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.
5. Log on to Apsara Stack Management Console. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Network Attached Storage**.

15.3.2 Create NAS instances

NAS instances are the running entities of NAS. Before you can use NAS, you must create a NAS instance.

Context

Before you create a NAS instance, note that:

- You can create up to 1,000 NAS instances.
- The upper limit of the NAS instance capacity is 10 PB.
- If you need to raise the limit, contact the administrator.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click **Create File System**.
3. On the **Create NAS File System** page that appears, set the parameters.

Table 15-1: Parameters for creating a NAS instance lists the parameters for creating a NAS instance.

Table 15-1: Parameters for creating a NAS instance

Category	Parameter	Description
Region	**Region**	Select a region from the drop-down list.
Basic configuration	**Department**	Select a department from the drop-down list.
	Project	Select a project from the drop -down list.
	File System Name	Enter the NAS instance name .
Storage configuration	**Storage Type**	Select **Capacity Type**.
	Protocol Type	Select **NFS** or **SMB**.

4. Click **OK** .

15.3.3 Create permission groups

NAS uses permission groups and permission group rules to manage NAS instance permissions. Before you can use a NAS instance, you must create permission groups and configure the required parameters.

Context

Each permission group in a NAS instance has an IP address whitelist. You can add rules to a permission group of a NAS instance to allow users from specified IP addresses or address segments to access the NAS instance with different permissions.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the **Permission Group** tab.
3. On the Permission Group tab that appears, click **Create Access Group**.

> **Note:**
> You can create up to 100 permission groups. If you need to raise the limit, contact the administrator.

4. In the **Create Permission Group** dialog box, set the parameters.

   Table 15-2: Parameters for creating a permission group lists the parameters for creating a permission group.

   Table 15-2: Parameters for creating a permission group

Parameter	Description
Region	Select a region from the drop-down list.
Department	Select a department from the drop-down list.
Project	Select a project from the drop-down list.
Permission Group Name	Enter the name of the permission group.
Network Type	Select **VPC** or **Classic Network**.

5. Click **OK**.

15.3.4 Create permission group rules

NAS uses permission groups and permission group rules to manage NAS instance permissions. Before you can use a NAS instance, you must create rules in its permission groups and configure the required parameters.

Context

Each permission group in a NAS instance has an IP address whitelist. You can add rules to a permission group of a NAS instance to allow users from specified IP addresses or address segments to access the NAS instance with different permissions.

Warning:
To ensure data security, we strongly recommend that you use caution when adding permission group rules and granting permissions to IP addresses.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the **Permission Group** tab. On the Permission Group tab that appears, click the name of a permission group to go to the **Rules List** page.
3. Click **Create Rule**.

Note:

You can create up to 1,000 permission group rules. If you want to raise the limit, contact the administrator.

4. In the **Add Rule** dialog box, set the parameters.

   Table 15-3: Parameters for creating a permission group rule lists the parameters for creating a permission group rule.

   Table 15-3: Parameters for creating a permission group rule

Parameter	Description
Authorized IP Address	The IP address or IP address segment of the object authorized by the rule. For a classic network, you can specify only one IP address.
Read/Write Permissions	Select **Read-Only** or **Read and Write** to allow the authorized object to perform read-only or read/write operations on the NAS instance.
User Permissions	Select **Do Not Limit root User (no_squash)**, **Limit root User (root_squash)**, or **Limit All Users (all_squash)** to specify whether to limit the access from the Linux system users of the authorized object to the NAS instance. Description: • **Do Not Limit root User (no_squash)** allows the root user to access the NAS instance. • **Limit root User (root_squash)** considers the root user as nobody. • **Limit All Users (all_squash)** considers all users including root as nobody.
Priority	The priority value ranges from 1 to 100. The value 1 indicates the highest priority. When an authorized object matches multiple rules, the rule with the highest priority takes effect.

5. Click **OK**.

15.3.5 Add mount points

After you create a NAS instance and its permission groups, you must add mount points to the NAS instance so that you can mount the NAS instance to compute nodes, such as ECS, E-HPC, or Container Service instances.

Context

A mount point is an access address of a NAS instance in a VPC or classic network. Each mount point corresponds to a domain name. NAS supports two types of mount points: VPC and classic network.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click a NAS instance ID to go to the **instance details** page.
3. Click the **Mount Point** tab.
4. On the Mount Point tab that appears, click **Add Mount Point**.

   Note:
   You can create up to 100 mount points. If you need to raise the limit, contact the administrator.
5. In the **Add Mount Point** dialog box, configure the parameters.
   - If you set **Mount Point Type** to **Classic Network**, select a permission group to be bound to the mount point from the **Permission Group** drop-down list.
   - If you set **Mount Point Type** to **VPC**, set **VPC** and **VSwitch**. Then, select a permission group to be bound to the mount point from the **Permission Group** drop-down list.

   Note:
   - If you set Mount Point Type to VPC, ensure that the corresponding VPC and VSwitch have been created.
   - If you set Mount Point Type to Classic Network, the NAS instance can be accessed only by ECS instances under the same account as the mount point.
   - You can use a single mount point to mount a NAS instance to multiple compute nodes such as ECS, E-HPC, or Container Service instances for parallel access.
6. Click **OK**.

15.3.6 Mount NAS instances

After you create a NAS instance and add a mount point to it, you can mount the NAS instance to a compute node such as an ECS node.

Prerequisites

The following conditions determine whether an ECS instance can access a NAS instance through a mount point:

- If the network type of the mount point is VPC, you can mount the NAS instance only to the ECS instances that are in the same VPC as the mount point. In addition, the VPC IP address of each of the ECS instances must match the authorized IP address of a rule in the permission group bound to the mount point.
- If the network type of the mount point is classic network, you can mount the NAS instance only to ECS instances under the same account as the mount point. In addition, ensure that the authorized IP address of a rule in the permission group bound to the mount point matches the intranet IP address of the ECS instance.

Before you can use NFS to mount a NAS instance, ensure that `nfs-utils` or `nfs-common` has been installed. If not, run the following command to install the software package:

- CentOS: `sudo yum install nfs-utils`
- Ubuntu or Debian: `sudo apt-get install nfs-common`

Context

NAS supports the NFSv3 and NFSv4 protocols. You can choose a protocol version for mounting a NAS instance based on your scenario.

Use NFSv4.0 to mount a NAS instance

Format

```
sudo mount -t nfs -o vers=4.0 <domain name of the mount point>:<NAS
 instance directory> <target local mounting directory>
```

Parameter description

- Domain name of the mount point: It is automatically generated when you create a NAS instance and a mount point.
- NAS instance directory: It is a directory of the NAS instance, which may be the root directory "/" or any subdirectory.
- Target local mounting directory: It is a directory on the local server, to which the NAS instance is to be mounted.

Examples

- Run the following command to mount the root directory of the NAS instance:

```
mount -t nfs -o vers=4.0 file-system-id-xxxx.regionid.nas.example.
com:/ /local/mntdir
```

- Run the following command to mount the subdirectory named sub1 of the NAS instance:

```
mount -t nfs -o vers=4.0 file-system-id-xxxx.regionid.nas.example.
com:/sub1 /local/mntdir
```

Use NFSv3 to mount a NAS instance

Format

```
sudo mount -t nfs -o vers=3,nolock,proto=tcp <domain name of the mount
 point>:<NAS instance directory> <target local mounting directory>
```

Examples

- Run the following command to mount the root directory of the NAS instance:

```
mount -t nfs -o vers=3,nolock,proto=tcp file-system-id-xxxx.regionid
.nas.example.com:/ /local/mntdir
```

- Run the following command to mount the subdirectory named sub1 of the NAS instance:

```
mount -t nfs -o vers=3,nolock,proto=tcp file-system-id-xxxx.regionid
.nas.example.com:/sub1 /local/mntdir
```

View the mount point information

After the directories are mounted, run the following command to check the mounted NAS instance:

```
mount -l
```

Run the following command to check the current capacity of the mounted NAS instance:

```
df -h
```

15.4 NAS instance

15.4.1 View the NAS instance details

You can view details of an existing NAS instance in the NAS console, including system details and mount points.

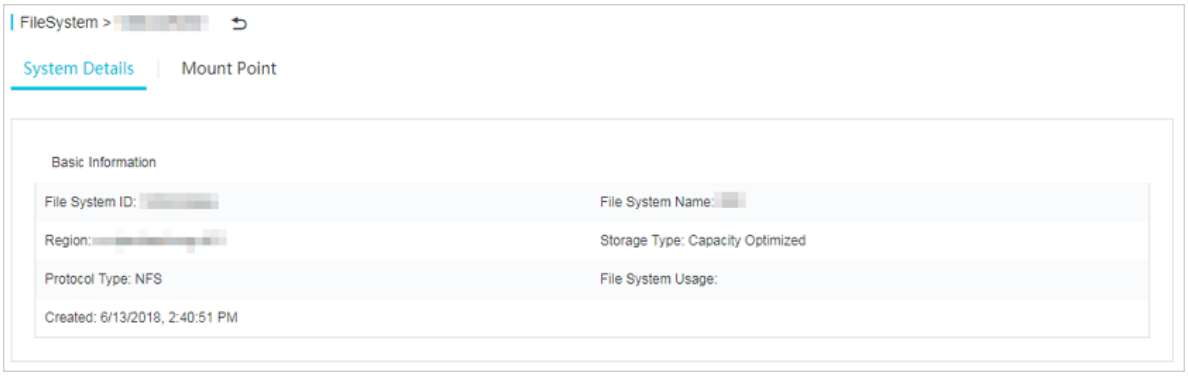
Prerequisites

Before you can view NAS instance details, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance has been created.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the ID of the NAS instance that you want to view or click > **Details** in the Actions column corresponding to the NAS instance ID. The NAS instance details page is displayed, as shown in *Figure 15-1: NAS instance details*.

   Figure 15-1: NAS instance details



   The NAS instance details page has two tabs:

   - The **System Details** tab displays the basic information about the NAS instance, including the NAS instance ID, region, and storage capacity.
   - The **Mount Point** tab lists the mount points of the NAS instance. You can manage the mount points on this tab.

15.4.2 Delete NAS instances

You can delete a NAS instance in the NAS console.

Prerequisites

Before you can delete a NAS instance, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance has been created and the NAS instance does not contain mount points.

Procedure

1. *Log on to the NAS console*
2. On the **File Storage NAS** page, click > **Delete** in the Actions column corresponding to the NAS instance that you want to delete.
3. In the **Delete File System Instance** message that appears, click **OK**.

15.5 Mount point

15.5.1 View the mount point list

You can view a list of existing mount points in the NAS console.

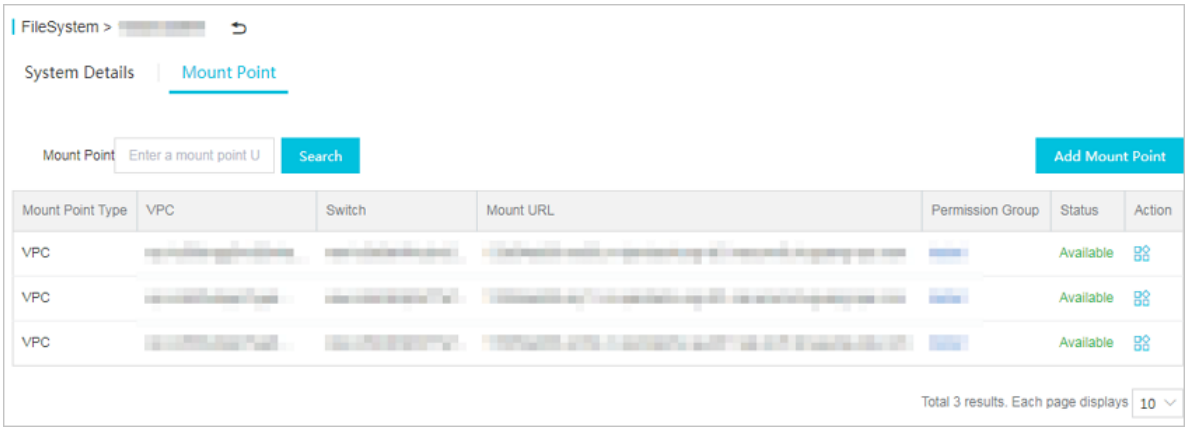
Prerequisites

Before you can view the mount point list, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one mount point have been created.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the ID of the NAS instance where the mount point that you want to view is located. On the instance details page that appears,
3. click the **Mount Point** tab. On the Mount Point tab that appears, you can view a list of all mount points in the NAS instance, as shown in *Figure 15-2: Mount point list*.

Figure 15-2: Mount point list



15.5.2 Enable or disable mount points

You can enable or disable a mount point in the NAS console.

Prerequisites

Before you can enable or disable a mount point, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one mount point have been created.

Procedure

1. *Log on to the NAS console*.

2. On the **File Storage NAS** page, click the ID of the NAS instance where the mount point that you want to view is located. On the instance details page that appears, click the **Mount Point** tab.
3. In the mount point list, you can perform the following operations:
   - Click  > **Disable** in the Actions column corresponding to the mount point that you want to disable. In the message that appears, click **OK** to disable access to the mount point from clients.
   - Click  > **Enable** in the Actions column corresponding to the mount point that you want to enable. In the message that appears, click **OK** to enable access to the mount point from clients.

15.5.3 Delete mount points

You can delete a mount point in the NAS console.

Prerequisites

Before you can delete a mount point, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one mount point have been created.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the ID of the NAS instance where the mount point that you want to view is located. On the instance details page that appears, click the **Mount Point** tab.
3. In the mount point list, click  > **Delete** in the Actions column corresponding to the mount point that you want to delete.
4. In the **Delete Mount Point** message that appears, click **OK**.

> **Note:**
> After a mount point is deleted, it cannot be restored. Use caution when you delete a mount point.

15.5.4 Modify the permission group of a mount point

You must bind a permission group to each mount point. You can change the permission group that is bound to a mount point in the NAS console.

Prerequisites

Before you can change the permission group that is bound to a mount point, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one mount point have been created in the region, and the mount point has been bound with a permission group.

Context

You must bind a permission group to each mount point. You can configure a source IP address list for the permission group to restrict access from ECS instances to the mount point. You can change the permission group that is bound to a mount point as required.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the ID of the NAS instance where the mount point that you want to change is located. On the instance details page that appears, click the **Mount Point** tab.
3. In the mount point list, click  > **Modify Permission Group** in the Actions column of the mount point of which the permission group is to be modified.
4. In the **Modify Mount Point Permission Group** dialog box that appears, set **Change to** and click **OK**.

> **Note:**
> The modification may require up to 1 minute to take effect.

15.6 Permission group

15.6.1 View the permission group list

You can view a list of existing permission groups in the NAS console.

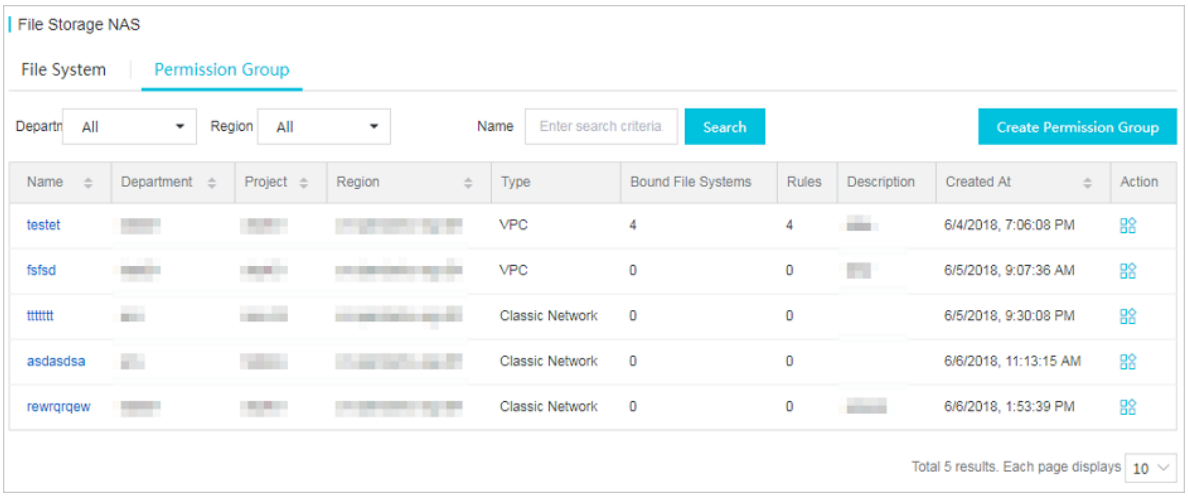
Prerequisites

Before you can view the permission group list, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one permission group have been created.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the **Permission Group** tab. On the Permission Group tab that appears, you can view a list of permission groups in the current region, as shown in *Figure 15-3: Permission group list*.

Figure 15-3: Permission group list



15.6.2 Delete permission groups

You can delete a permission group in the NAS console.

Prerequisites

Before you can delete a permission group, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one permission group have been created.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the **Permission Group** tab.
3. In the permission group list, click > **Delete** in the Actions column corresponding to the permission group that you want to delete.
4. In the **Delete Permission Group** message that appears, click **OK**.

Note:

Permission groups that are in use cannot be deleted. To delete a permission group in use, you must first disable it.

15.6.3 Manage permission group rules

You can manage the rules of a permission group in the NAS console, including modifying and deleting rules.

Prerequisites

Before you can manage the rules of a permission group, you must complete the procedure in *Quick start*, or ensure that at least one NAS instance and one permission group have been created, and the permission group has at least one rule.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the **Permission Group** tab.
3. On the Permission Group tab that appears, click the name of a permission group to go to the **Rules List** page.
4. On this page, you can modify or delete the rules of the permission group.
   - To modify a rule, click  > **Change** in the Actions column corresponding to the rule. In the **Modify Rule** dialog box that appears, modify **Authorization Address**, **Read/Write Permissions**, **User Permission**, or **Priority**. Click **OK**.
   - To delete a rule, click  > **Delete** in the Actions column corresponding to the rule. In the **Delete Rule** message that appears, click **OK**.

15.7 Migrate data

15.7.1 Data migration tool for Windows

The NAS data migration tool for Windows is available for use after you download and decompress it. This tool synchronizes files from object storage services (such as OSS) or local disks to NAS instances.

Context

Features of nasimport:

- Synchronizes local files, files stored in OSS, files stored in third-party cloud storage, and HTTP -linked files to NAS instances.
- Synchronizes stored data (files modified after a specified time point).
- Synchronizes incremental data automatically.

- Supports resumable data transfer.
- Uploads and downloads data in parallel.

Operation requirements

Run the nasimport tool on an ECS virtual machine (VM) where you can mount the desired NAS instance. To determine whether the NAS instance can be mounted to an ECS VM and how to mount the NAS instance, see *Mount a NAS instance*.

Supported operating systems

- Windows Server 2008 Standard edition SP2 32-bit
- Windows Server 2008 R2 Datacenter edition 64-bit
- Windows Server 2012 R2 Datacenter edition 64-bit
- Windows Server 2016 Datacenter edition 64-bit

Deployment and configuration

1. Download the *nasimport toolkit*.
2. Create a synchronization working directory (such as C:\NasImport) on your local server and download the nasimport toolkit to this directory.
3. Edit the configuration file named `config/sys.properties` in the working directory.

   We recommend that you use the default configurations. You can edit the configurations fields based on your requirements. For more information, see *Field description*.

   Table 15-4: Field description

Field	Description
workingDir	The working directory to which the nasimport toolkit is decompressed.
slaveTaskThreadNum	The number of working threads that run synchroniz ation simultaneously.
slaveMaxThroughput(KB/s)	The upper limit of migration traffic.
slaveAbortWhenUncatchedException	Whether to skip an unknown error or abort. The process skips an unknown error by default.
dispatcherThreadNum	The number of parallel threads in a dispatching task . Keep the default value.

Running

Nasimport commands

- Submit a job: `nasimport -c config/sys.properties submit <your-job-configuration>`
- Cancel a job: `nasimport -c config/sys.properties clean <job-name>`
- View a job: `nasimport -c config/sys.properties stat detail`
- Retry a job: `nasimport -c config/sys.properties retry <job-name>`
- Start nasimport: `nasimport -c config/sys.properties start`

1. Start nasimport.

   Enter the working directory and open the CLI. Run the following command in the CLI:

   ```
   nasimport –c config/sys.properties start
   ```

   Figure 15-4: Start nasimport

   Note:

   - Keep nasimport running. You can also configure nasimport as a background service in Windows.
   - When you start nasimport, you can redirect the log to a file for easy viewing in the future.

     ```
     nasimport -c config\sys.properties start > nasimport.log
         2>&1
     ```

2. Define a job.

Use the config\local_job.cfg template to define a job.

Table 15-5: Field description

Field	Description
jobName	The name that uniquely identifies the job. You can submit multiple jobs with different names.
jobType	The job type. Values: import and audit. Import synchronizes data while audit checks the source and target data for consistency.
isIncremental=false	Whether to enable the automatic incremental mode. If this field is set to true, incremental data is scanned at the interval specified by incrementalModeInterval (in seconds) and synchronized to the NAS instance.
incrementalModeInterval=86400	The synchronization interval in the incremental mode. Unit: second.
importSince	The start time. Incremental data that is generated on and after this time point is synchronized to the NAS instance. This parameter is in the UNIX timestamp format. Unit: second. Default value: 0.
srcType	The synchronization source type. You can synchronize local files, files stored in OSS, or files stored in third-party cloud storage.
srcAccessKey	The AccessKey ID of the data source. Specify this field if you have set srcType to OSS or to a third-party cloud storage.
srcSecretKey	The AccessKey Secret of the data source. Specify this field if you have set srcType to OSS or to a third-party cloud storage.
srcDomain	The endpoint of the data source.  **Note:** If the data source of a migration job is OSS, set srcDomain to the intranet domain name with "internal." With this setting, you can save the cost on downloading data from OSS and enjoy a faster migration service. You

Field	Description
	only pay for accessing OSS. You can obtain the intranet domain name of OSS in the OSS console. If your NAS instance is in a VPC and the data source is OSS, set srcDomain to the VPC domain name provided by OSS.
srcBucket	The source bucket name.
srcPrefix	The source prefix. Default value: null. If you have set srcType to local, enter the local directory to be synchronized. Note that the directory must be a full path ended with a forward slash (/). If you have set srcType to OSS or to a third-party cloud storage, enter the prefix of the object to be synchronized. To synchronize all files, set the prefix to null.
destType	The synchronization target type. Default value: NAS.
destMountDir	The local directory to which the NAS instance is mounted.
destMountTarget	The domain name of the NAS instance mount point.
destNeedMount=true	Whether nasimport performs automatic mounting. Default value: true. You can set this field to false and manually change the NAS instance mount point to the destMountDir directory.
destPrefix	The prefix of the synchronization target file. Default value: null.
taskObjectCountLimit	The maximum number of files that are processed by each task. This field affects the maximum number of parallel threads. It is usually set to the total number of files divided by the number of download threads that you have set. If you do not know the total number of files, you can keep the default value.
taskObjectSizeLimit	The maximum volume of the data downloaded by each task. Unit: byte.
scanThreadCount	The number of threads that scan files in parallel. This field affects file scan efficiency.
maxMultiThreadScanDepth	The maximum allowable depth of the directory in parallel scan. You can keep the default value.

Note:

- If you have configured the automatic incremental mode, the job runs periodically and permanently to scan the latest data.
- If you have set srcType to a third-party cloud storage, the List operation on files cannot implement checkpoints due to the API limits of third-party cloud storage. Killing the process before the List operation is complete may cause all the files to be relisted.

3. Submit the job.

   The following example shows how to copy the local *C:\Program Files\Internet Explorer* directory to a NAS instance.

   a. Edit the job: Copy *config\local_job.cfg* to the working directory and edit the items listed in the following table.

srcType	local
srcPrefix	C:\\Program Files\\Internet Explorer
destMountDir	h:
destNeedMount	true
destMountTarget	xxxx-yyy.cn-beijing.nas.aliyuncs.com

   Note:
   You must set destMountDir to a drive letter that does not exist. Otherwise, destMountDir may conflict with an existing drive. destMountTarget is the mount point of the NAS instance.

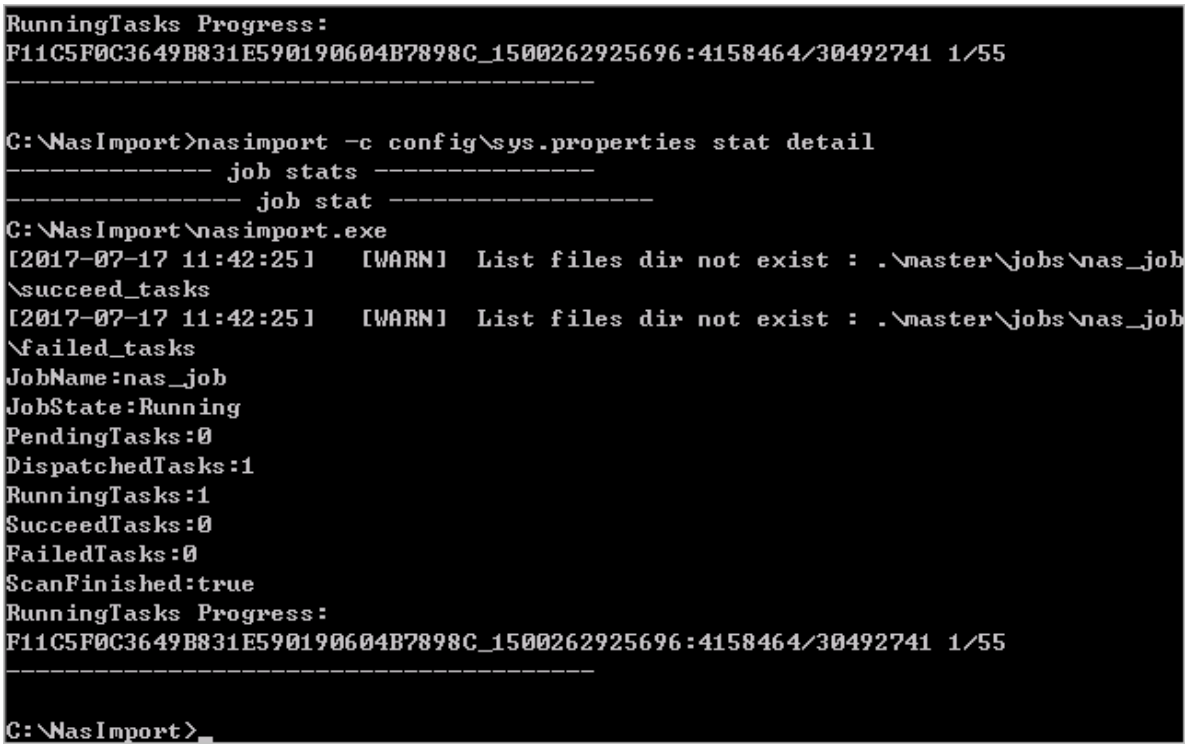
   b. Submit the job: Restart the CLI in the working directory and run the `nasimport -c config\sys.properties submit local_job.cfg command.`

   Note:

   - If the job that you want to submit has the same name as a job in progress, you cannot submit the job.
   - To pause a synchronization job, stop the nasimport process. You can restart the nasimport process to resume synchronization from where it was paused.

4. Check the job status. Run the following command on the CLI:

```
nasimport -c config\sys.properties stat detail
```



The preceding command output displays the overall progress of the current job and the progress of the current task. In this example, `4158464/30492741` indicates that the volume of data already uploaded is 4,158,464 bytes and the total volume of data to be uploaded is 30,492,741 bytes. `1/1` indicates that the total number of files to be uploaded is 1 and the number of files already uploaded is 1.

The migration tool splits each job that you submit into multiple tasks for parallel execution. After all the tasks are complete, the job is considered complete. After the job is complete, JobState displays `Succeed` or `Failed`, to indicate whether the job is successful or not. If the job fails, you can view the cause of failure for each task in the following file:

```
master/jobs/$jobName/failed_tasks/*/audit.log
```

We have already retried failed jobs in nasimport. If a failure is caused by the temporary unavailability of the source or target data, run the following command to retry the job:

```
nasimport  -c config/sys.properties retry <job-name>
```

Run the stat detail command again after a while.

```
PendingTasks:0
DispatchedTasks:1
RunningTasks:1
SucceedTasks:0
FailedTasks:0
ScanFinished:true
RunningTasks Progress:
F11C5F0C3649B831E590190604B7898C_1500262925696:4158464/30492741 1/55
----------------------------------------

C:\NasImport>nasimport -c config\sys.properties stat detail
-------------- job stats ---------------
---------------- job stat ------------------
JobName:nas_job
JobState:Succeed
PendingTasks:0
DispatchedTasks:0
RunningTasks:0
SucceedTasks:1
FailedTasks:0
ScanFinished:true
RunningTasks Progress:
----------------------------------------

C:\NasImport>_
```

`SucceededTasks` is **1**, indicating that the task is complete. Open the file explorer and you can see that the *H:* drive contains the file.

Common causes of failures

- The job configurations are incorrect, for example, the AccessKey ID is incorrect or permissions are insufficient. In this case, all tasks fail. To identify this cause, check the nasimport.log file in the working directory (ensure that you have redirected the log to nasimport.log when starting nasimport). You can also run the nasimport start command to check for the cause.

```
C:\NasImport\nasimport.exe
[2017-07-17 12:22:40]   [INFO]  JobDispatcher:Init
[2017-07-17 12:22:40]   [INFO]  job controller daemon start, working dir:.\
[2017-07-17 12:22:40]   [INFO]  watching job queue:.\master\jobqueue\
[2017-07-17 12:22:40]   [INFO]  JobDispatcher:Run
[2017-07-17 12:22:40]   [INFO]  try lock .\master\jobs\nas_job\.lock succeed
[2017-07-17 12:22:40]   [INFO]  start job:nas_job
[2017-07-17 12:22:40]   [INFO]  list checkpoint: .\master\jobs\nas_job\checkpoints\0, cpt
[2017-07-17 12:22:40]   [INFO]  scan task load checkpoint: [totalSize=0, totalCount=0, pre
[2017-07-17 12:22:40]   [INFO]  single thread scan start: nas_job
com.aliyun.oss.OSSException: The OSS Access Key Id you provided does not exist in our reco
        at com.aliyun.oss.common.utils.ExceptionFactory.createOSSException(ExceptionFactor
        at com.aliyun.oss.internal.OSSErrorResponseHandler.handle(OSSErrorResponseHandler.
        at com.aliyun.oss.common.comm.ServiceClient.handleResponse(ServiceClient.java:248)
        at com.aliyun.oss.common.comm.ServiceClient.sendRequestImpl(ServiceClient.java:130
        at com.aliyun.oss.common.comm.ServiceClient.sendRequest(ServiceClient.java:68)
        at com.aliyun.oss.internal.OSSOperation.send(OSSOperation.java:94)
        at com.aliyun.oss.internal.OSSOperation.doOperation(OSSOperation.java:149)
        at com.aliyun.oss.internal.OSSOperation.doOperation(OSSOperation.java:113)
        at com.aliyun.oss.internal.OSSBucketOperation.listObjects(OSSBucketOperation.java:
        at com.aliyun.oss.OSSClient.listObjects(OSSClient.java:526)
        at com.aliyun.ossimport2.master.scanner.OssLister.list(OssScanner.java:65)
        at com.aliyun.ossimport2.master.scanner.SingleThreadTask.run(SingleThreadTask.java
```

- The encoding method of source file names is inconsistent from the default file name encoding method of the system (GBK for Windows and UTF-8 for Linux). This is the typical cause of failure for NFS data sources.
- A file in the source directory is modified during the upload process. This cause is indicated by a `SIZE_NOT_MATCH` error in audit.log. In this case, the old file is uploaded, but the changes are not synchronized to the NAS instance.
- The source file is deleted during the upload process, causing file download to fail.
- An error occurs in the data source, causing source data download to fail.
- The Clean operation is performed before the nasimport process is killed, which may cause a program execution error.
- The nasimport tool aborts and the job status is Abort. If this failure occurs, contact Alibaba Cloud technical support engineers.

15.7.2 Migrate local files or files stored in OSS to NAS instances

The nasimport tool helps you synchronize files and data from your local data center, OSS, or third-party cloud storage to NAS instances.

Context

Functions of nasimport:

- Synchronizes local files, files stored in OSS, files in third-party cloud storage, or HTTP-linked files to NAS instances.
- Mounts NAS instances automatically.
- Synchronizes stored data (files modified after a specified time point).
- Synchronizes incremental data automatically.
- Supports resumable data transfer.
- Lists, uploads, and downloads data in parallel.

To migrate a large volume of data (exceeding 2 TB) to a NAS instance in a short period of time , you can contact Alibaba Cloud technical support for a parallel synchronization on multiple machines in addition to using nasimport.

Runtime environment

Run the nasimport tool on an ECS virtual machine (VM) where you can mount the desired NAS instance. To determine whether the NAS instance can be mounted to an ECS VM and how to mount the NAS instance, see *Mount a NAS instance*.

You must run nasimport in Java JDK 1.7 or later. We recommend the *Oracle version JDK*.

Note:
Before you start nasimport, run the **`ulimit -n`** command to view the number of files that the process allows you to open. If the number is smaller than 10,240, you must modify the configuration first.

Deployment and configuration

1. Create a working directory for synchronization on your local server, and download the nasimport toolkit to this directory.

   Example: Run the following commands to create */root/ms* as the working directory and download the toolkit to this directory:

   ```
   export work_dir=/root/ms
   wget http://docs-aliyun.cn-hangzhou.oss.aliyun-inc.com/assets/attach
   /45306/cn_zh/1479113980204/nasimport_linux.tgz
   tar zxvf ./nasimport_linux.tgz -C "$work_dir"
   ```

2. Run the following commands to edit the configuration file named config/sys.properties in the working directory named $work_dir:

   ```
   vim $work_dir/config/sys.properties
   workingDir=/root/ms
   slaveUserName=
   slavePassword=
   privateKeyFile=
   slaveTaskThreadNum=60
   slaveMaxThroughput(KB/s)=100000000
   slaveAbortWhenUncatchedException=false
   dispatcherThreadNum=5
   ```

   We recommend that you use the default configurations. When necessary, you can edit the configuration fields. For more information, see *Table 15-6: Field description*.

Table 15-6: Field description

Field	Description
workingDir	The working directory to which the nasimport toolkit is decompressed.
slaveTaskThreadNum	The number of working threads that run synchroniz ation simultaneously.
slaveMaxThroughput(KB/s)	The upper limit of migration traffic.
slaveAbortWhenUncatchedException	Whether to skip an unknown error or abort. The process skips an unknown error by default.
dispatcherThreadNum	The number of parallel threads in a dispatching task . Keep the default value.

Running

The nasimport tool supports the following commands:

- Submit a job:

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 submit $jobConfigPath
```

- Cancel a job:

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 clean $jobName
```

- View job status:

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 stat detail
```

- Retry the job:

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 retry $jobName
```

Perform the following procedure to run a migration job:

1. Run the following commands to start nasimport:

```
cd $work_dir
```

```
nohup java -Dskip_exist_file=false -jar $work_dir/nasimport.jar -c $
work_dir/config/sys.properties start > $work_dir/nasimport.log 2>&1
 &
```

Note:

The related log file is automatically generated in the directory where nasimport is started. We recommend that you start nasimport in the working directory named $work_dir. If the value of **`skip_exist_file`** is true when you start nasimport, nasimport skips the files that already exist in the NAS instance with the length the same as the source.

2. Edit the sample job description file named *`nas_job.cfg`*.

Table 15-7: Field description

Field	Description
jobName	The name that uniquely identifies the job. You can submit multiple jobs with different names.
jobType	The job type. Values: import and audit. Import synchroniz es data while audit checks the source and target data for consistency.
isIncremental=false	Whether to enable the automatic incremental mode. If this field is set to true, incremental data is scanned at the interval specified by incrementalModeInterval (in seconds) and synchronized to the NAS instance.
incrementalModeInterval=86400	The synchronization interval in the incremental mode. Unit: second.
importSince	The start time. Incremental data that is generated on and after this time point is synchronized to the NAS instance . This parameter is in the UNIX timestamp format. Unit: second. Default value: 0.
srcType	The synchronization source type. You can synchronize local files, files stored in OSS, or files stored in third-party cloud storage.
srcAccessKey	The AccessKey ID of the data source. Specify this field if you have set srcType to OSS or to a third-party cloud storage.
srcSecretKey	The AccessKey Secret of the data source. Specify this field if you have set srcType to OSS or to a third-party cloud storage.

Field	Description
srcDomain	The endpoint of the data source.  **Note:** If the data source of a migration job is OSS, set srcDomain to the intranet domain name with "internal." With this setting, you can save the cost on downloading data from OSS and enjoy a faster migration service. You only pay for accessing OSS. You can obtain the intranet domain name of OSS in the OSS console. If your NAS instance is in a VPC and the data source is OSS, set srcDomain to the VPC domain name provided by OSS.
srcBucket	The source bucket name.
srcPrefix	The source prefix. Default value: null. If you have set srcType to local, enter the local directory to be synchronized. Note that the directory must be a full path ended with a forward slash (/). If you have set srcType to OSS or to a third-party cloud storage, enter the prefix of the object to be synchronized. To synchronize all files, set the prefix to null.
destType	The synchronization target type. Default value: NAS.
destMountDir	The local directory to which the NAS instance is mounted.
destMountTarget	The domain name of the NAS instance mount point.
destNeedMount=true	Whether nasimport performs automatic mounting. Default value: true. You can set this field to false and manually change the NAS instance mount point to the destMountDir directory.
destPrefix	The prefix of the synchronization target file. Default value: null.
taskObjectCountLimit	The maximum number of files that are processed by each task. This field affects the maximum number of parallel threads. It is usually set to the total number of files divided by the number of download threads that you have set. If you do not know the total number of files, you can keep the default value.

Field	Description
taskObjectSizeLimit	The maximum volume of the data downloaded by each task. Unit: byte.
scanThreadCount	The number of threads that scan files in parallel. This field affects file scan efficiency.
maxMultiThreadScanDepth	The maximum allowable depth of the directory in parallel scan. You can keep the default value.

Note:

- If you have configured the automatic incremental mode, the job runs periodically and permanently to scan the latest data.
- If you have set srcType to a third-party cloud storage, the List operation on files cannot implement checkpoints due to the API limits of third-party cloud storage. Killing the process before the List operation is complete may cause all the files to be relisted.

3. Submit the job.

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
  submit $work_dir/nas_job.cfg
```

Note:

- If the job that you want to submit has the same name as a job in progress, you cannot submit the job.
- To pause a synchronization job, stop the nasimport process. You can restart the nasimport process to resume synchronization from where it was paused.
- To resynchronize all files, stop the nasimport process and run the following command to clear the current job. For example, the job name is nas_job (you can set the job name in the nas_job.cfg file).

```
ps axu | grep "nasimport.jar.* start" | grep -v grep | awk '{
print "kill -9 "$2}' | bash
java -jar $work_dir/nasimport.jar  -c $work_dir/conf/sys.
properties clean nas_job
```

4. Check the job status.

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
  stat detail
--------------job stats begin---------------
----------------job stat begin------------------
JobName:nas_job
```

```
JobState:Running
PendingTasks:0
RunningTasks:1
SucceedTasks:0
FailedTasks:0
ScanFinished:true
RunningTasks Progress:
FD813E8B93F55E67A843DBCFA3FAF5B6_1449307162636:26378979/26378979 1/1
----------------job stat end-----------------
--------------job stats end--------------
```

The preceding command output displays the overall progress of the current job and the progress of the current task. For example, `26378979/26378979` indicates that the total volume of data to be uploaded is 26,378,979 bytes and the volume of data already uploaded is 26,378,979 bytes. `1/1` indicates that the total number of files to be uploaded is 1 and the number of files already uploaded is 1.

The migration tool splits each job that you submit into multiple tasks for parallel execution. After all the tasks are complete, the job is considered complete. After a job is complete, JobState displays `Succeed` or `Failed`, to indicate whether the job is successful or not. If a job fails, run the following command to check the failure cause of each task.

In the following command, replace $jobName with the actual job name (you can set jobName in the local_job.cfg file).

```
cat $work_dir/master/jobs/$jobName/failed_tasks/*/audit.log
```

We have already retried failed jobs in nasimport. If a failure is caused by the temporary unavailability of the source or target data, run the following command to retry the job:

```
java -jar $work_dir/nasimport.jar  -c $work_dir/config/sys.
properties retry $jobNam
```

Common causes of job failures

- The job configurations are incorrect, for example, the AccessKey ID is incorrect or permissions are insufficient. In this case, all tasks fail. To identify this cause, check the *`$work_dir/nasimport.log`* file.
- The encoding method of source file names is inconsistent from the default file name encoding method of the system (GBK for Windows and UTF-8 for Linux). This is the typical cause of failure for NFS data sources.
- A file in the source directory is modified during the upload process. This cause is indicated by a `SIZE_NOT_MATCH` error in audit.log. In this case, the old file is uploaded, but the changes are not synchronized to the NAS instance.

- The source file is deleted during the upload process, causing file download to fail.
- An error occurs in the data source, causing source data download to fail.
- The Clean operation is performed before the nasimport process is killed, which may cause a program execution error.
- The nasimport tool aborts and the job status is Abort. If this failure occurs, contact Alibaba Cloud technical support.

15.8 Directory-level ACL

NAS allows you to configure an Access Control List (ACL) for a directory to control access to the directory and files in it.

Prerequisites

- You must use the NFSv4 protocol to mount a NAS instance on a client.
- You must use the alinas-acl tool to configure an ACL. To ensure correct permission settings, do not change the mode or run the chmod command to modify the file permissions.

Procedure

1. Run the **`sudo mount -t nfs -o vers=4.0 <mount point domain name>:<NAS directory> <target directory on the current server>`** command, such as **`mount -t nfs -o vers=4.0 014544bbf6-wdt41.cn-hangzhou.nas.aliyuncs.com:/ /mnt`**, to ensure that a NAS instance has been mounted by using the NFSv4 protocol.

   Note:

   - The value of the vers parameter varies with the client version. If an error occurs when you set vers to 4.0, set vers to 4 instead.
   - If a NAS instance with the ACL feature disabled has been mounted before, we recommend that you mount the instance again to ensure that the ACL feature takes effect.

2. Run the following command to install `nfs4-acl-tools` in CentOS:

   ```
   sudo yum install nfs4-acl-tools -y
   ```

3. Run the following command to ensure that Python 2.7.5 has been installed:

```
python --version Python 2.7.5
```

4. Use alinas-acl to make ACL settings.

```
./alinas_acl set ./foo --add --user Alice --rule r #Grant the read
permission on the foo file to user Alice.
./alinas_acl set ./foo -a -u Alice -r r #Abbreviated format of the
previous command
./alinas_acl set ./dir --add --group Staff --rule rwx #Grant the
read, write, and execute permissions on the dir directory to group
Staff.
./alinas_acl set ./foo --add --user EVERYONE@ --rule none #Grant no
permissions to user EVERYONE@.
./alinas_acl set ./foo --add --user 1001 --rule none #Grant no
permissions to the user whose uid is 1001.
./alinas_acl set ./dir -d -u Bob #Revoke the permissions of user Bob
 on the dir directory.
```

Note:

- To avoid performance deterioration, we recommend that you configure an ACL for a directory, instead of for files in the directory.
- The number of Access Control Entries (ACEs) for a single file must not exceed 10.

5. View the ACL.

```
./alinas_acl get ./foo #View the permissions on the foo file. # file
: foo/ # owner:
                root # group: root OWNER@::rw- GROUP@::r-- EVERYONE
@::--- Alice::r-- Staff:g:rwx
                1001::---
```

Note:

OWNER@, GROUP@, and EVERYONE@ are three special usernames that are automatically generated when you configure the ACL. They correspond to the user, group, and others classes in the mode operand. If there are conflicts between the permissions specified in the ACL and the mode operand, the actual permissions may vary based on the client version.

16 Distributed File System (DFS)

16.1 What is DFS

Alibaba Cloud Distributed File System (DFS) is a file storage service designed for computing resources, such as Elastic Compute Service (ECS) instances and Container Service.

DFS provides standard Hadoop Distributed File System (HDFS), enabling you to use it without making any changes to existing big data analysis applications. It offers features such as unlimited capacity and performance expansion, single namespace, multi-sharing, high reliability, and high availability. Compared with user-created HDFS storage, DFS greatly reduces maintenance costs and data security risks.

You can perform the following operations:

- Create DFS file system instances and mount points.
- Create permission groups for DFS file system instances and add rules to permission groups to allow access or grant different levels of access permissions to IP addresses or CIDR blocks.
- Access file system instances through standard HDFS protocol interfaces within computing resources such as ECS and Container Service.
- Perform basic and advanced operations to file systems, mount points, and permission groups in the DFS console.
- Perform basic and advanced operations on DFS by using Software Development Kits (SDKs) or Application Program Interfaces (APIs).

16.2 Limits

The restrictions on DFS are as follows:

Hadoop Distributed File System and Abstract File System

- Does not support the settings of directory modification time (mtime) and access time (atime), or settings of file mtime and atime attributes through setTimes.
- Does not support symbolic links.
- Does not support file truncation (truncate).
- Does not support file concatenation (concat).
- Does not support extended attributes (XAttrs) operations.
- Does not support snapshot operations.
- Does not support delegation token operations.

- Does not support checksum operations (setWriteChecksum and setVerifyChecksum).
- Does not support ACL operations.
- Does not support file block locations.

Hadoop fs command line tool

- Does not support snapshot commands (createSnapshot, deleteSnapshot, and renameSnap shot).
- Does not support ACL commands (setfacl and getfacl).
- Does not support XAttr commands (setfattr and getfattr).
- Does not support file truncation commands (truncate).

16.3 Quick start

16.3.1 Log on to the DFS console

This topic describes how to log on to the DFS console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/ manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at

least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

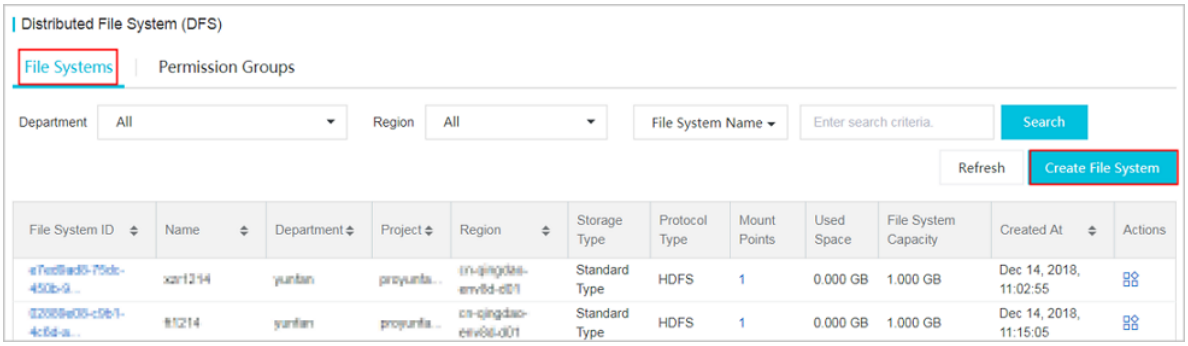
4. Click **LOGIN** to go to the **Dashboard** page.
5. In the top navigation bar of the page, choose **Console** > **Compute, Storage & Networking** > **Distributed File System**.

16.3.2 Create file systems

File systems are running entities of DFS. You need to create file systems before you can use DFS.

Procedure

1. *Log on to the DFS console*.
2. On the **Distributed File System (DFS)** page, click the **File Systems** tab, and click **Create File System**, as shown in the following figure.



Note:

- You can create up to 1,000 file systems.
- The upper limit of the file system capacity is 10 PB.

To raise the limit, contact the administrator.

3. On the **Create DFS File System** page, configure parameters.

Table 16-1: Parameters to create file systems describes the parameter configurations.

Table 16-1: Parameters to create file systems

Category	Parameter	Description
Region	**Region**	Select a region.

Category	Parameter	Description
	Zone	Select a zone from the drop-down list.
Basic Settings	**Department**	Select a department from the drop-down list.
	Project	Select a project from the drop-down list.
	File System Name	Enter a name for the file system. You must enter a file system name. The value must not exceed 100 bytes, and must be globally unique.
	Description	Enter the description of the file system.
Storage Configuration	**Protocol Type**	Select **HDFS**.
	Storage Type	Select **STANDARD**.
	File System Capacity (GB)	Enter the capacity of the file system.

4. Click **OK** to create the file system.

   If the newly created file system does not appear on the File Systems page, refresh the page.

16.3.3 Create permission groups

DFS enables you to manage permissions on file systems through permission groups. You need to create permission groups and configure parameters before you can use DFS.

Context

In DFS, the permission group acts as a whitelist. You can add rules to the permission group to allow access or grant different levels of access permissions to IP addresses or CIDR blocks.

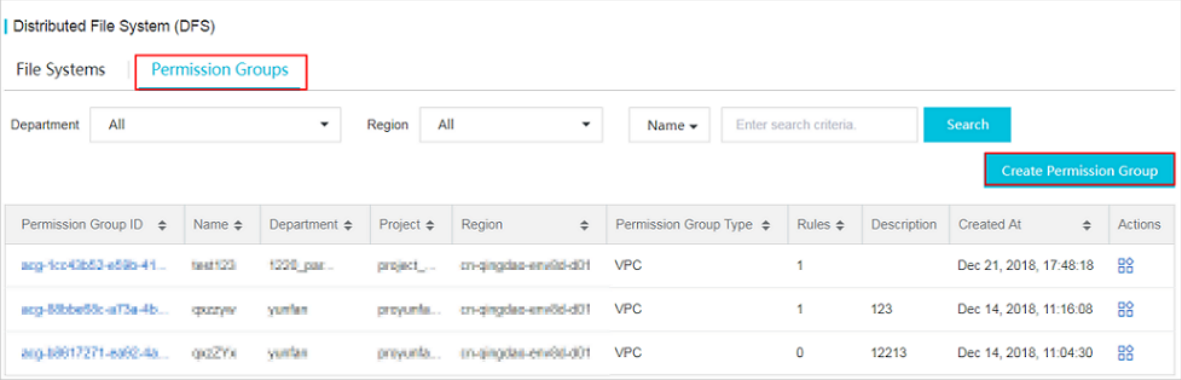
Warning:

To secure your data, We strongly recommend that you exercise caution when adding permission group rules and granting permissions to IP addresses.

Procedure

1. *Log on to the DFS console*.

2. On the **Distributed File System (DFS)** page, click the **Permission Groups** tab, and click **Create Permission Group**, as shown in the following figure.



Note:

You can create up to 100 permission groups. To raise the limit, contact the administrator.

3. In the **Create Permission Group** dialog box, configure parameters.

Table 16-2: Parameters to create permission groups describes the parameter configurations.

Table 16-2: Parameters to create permission groups

Parameter	Description
Region	Select a region.
Department	Select a department from the drop-down list.
Project	Select a project from the drop-down list.
Name	Enter a name for the permission group. You must enter a permission group name. The value must not exceed 100 bytes, and must be globally unique.
Network Type	Select **VPC**.
Description	Enter the description of the permission group.

4. Click **OK** to create the permission group.

16.3.4 Create permission group rules

DFS permission groups have various rules, which enable you to manage permissions of file system instances. Before using DFS, you need to create permission group rules and configure parameters for the created permission groups.

Context

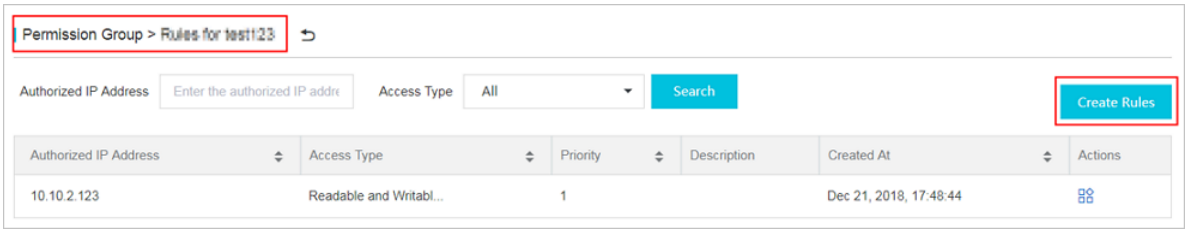
In DFS, the permission group acts as a whitelist. You can add rules to the permission group to allow access or grant different levels of access permissions to IP addresses or CIDR blocks.

Warning:

To secure your data, We strongly recommend that you exercise caution when adding permission group rules and granting permissions to IP addresses.

Procedure

1. *Log on to the DFS console*.
2. On the **Distributed File System (DFS)** page, select the permission group created in *Create permission groups*, click the permission group name to go to the **Rules** page, and click **Create Rules**, as shown in the following figure.



Note:

You can create up to 1,000 permission group rules. To raise the limit, contact the administrator.

3. In the **Create Rule** dialog box, configure parameters.

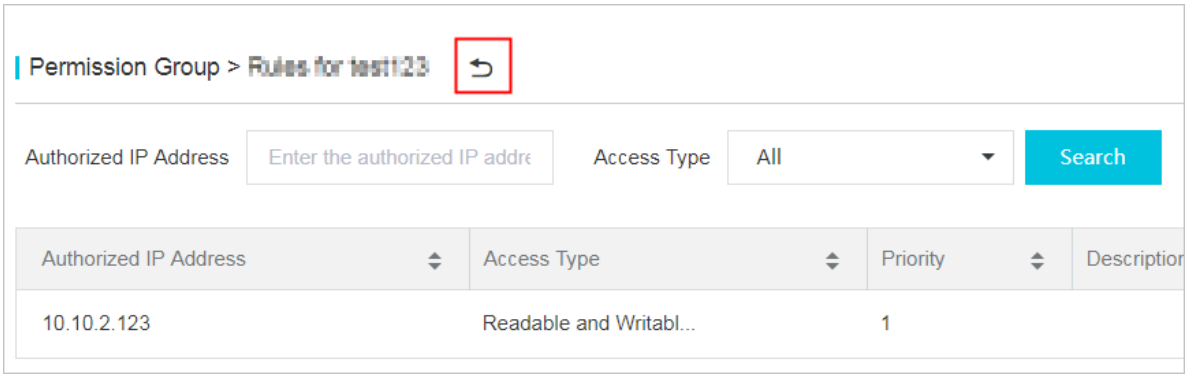
   Table 16-3: Parameters to create permission group rules describes the parameter configurations.

Table 16-3: Parameters to create permission group rules

Parameter	Description
Access Type	The optional value is **Readable and Writable**, allowing authorized object to read from and write to the file system.
Authorized IP Address	The authorized IP address is an IP address or CIDR block, such as 192.168.1.2 or 192.168.1.0/24. It is the authorized object of this rule.
Priority	The priority value ranges from 1 to 100. The value 1 indicates the highest priority. When an authorized

Parameter	Description
	object matches with multiple rules, the rule with the highest priority takes effect.
Description	Enter the description of the permission group rule.

4. Click **OK** to create the permission group rule.
5. Click the return button to return to the Distributed File System (DFS) page, as shown in the following figure.



16.3.5 Add mount points

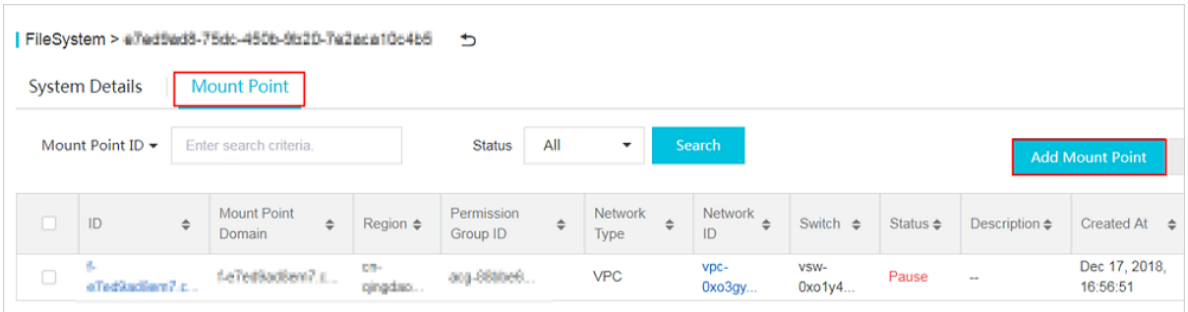
After a file system and its permission groups are created, you need to add mount points to the file system to mount computing nodes (ECS or Container Service instances) to the file system.

Context

A mount point is the access address of a file system instance in a VPC or classic network. Each mount point corresponds to a domain name. DFS supports only VPC mount points.

Procedure

1. *Log on to the DFS console*.
2. On the **Distributed File System (DFS)** page, click the **File Systems** tab.
3. Locate the file system created in *Create file systems* and click the file system ID to go to the **System Details** page.
4. Click the **Mount Point** tab and click **Add Mount Point**, as shown in the following figure.



Note:

You can create up to 100 mount points. To raise the limit, contact the administrator.

5. In the **Add Mount Point** dialog box, configure parameters.

   - Select **Network Type** to **VPC**.
   - Select corresponding **VPC Network** and **VPC VSwitch**.
   - Select the permission group that the mount point is bound to from the **Permission Group** drop-down list.
   - To create multiple mount points in one file system, ensure that the configuration of each mount point is unique. At least one of the configurations (permission groups, VPC networks , or VSwitches) of a mount point is different from the other mount points.

   Table 16-4: Parameters to add mount points describes the parameter configurations.

Table 16-4: Parameters to add mount points

Parameter	Description
Region	The region of a mount point.
File Systems	The file system to which the mount point is added.
Permission Groups	The permission group to which the mount point is bound.
Network Type	The network type, which must be set to **VPC**.
VPC Network	The VPC that corresponds to the mount point.
VSwitch	The VSwitch that corresponds to the mount point.
Description	The description of the mount point.

Note:

- Ensure that the VPC and VSwitch have been created.

- You can mount a mount point on multiple computing nodes (ECS or Container Service instances) for shared access.

6. Click **OK** to add the mount point.

16.3.6 Mount file systems

After creating a file system and adding a mount point for the file system, you can mount DFS instances through the mount point.

DFS currently supports Hadoop 2.7.x.

Prerequisites

The following conditions determine whether an ECS instance can access a file system through a mount point:

- If the mount point network type is VPC, you can mount the file system only on the ECS instance in the same VPC. In addition, ensure that the authorized IP address of a rule in the permission group bound to the mount point matches the VPC IP address of the ECS instance.
- DFS provides a UserGroupService interface based on Linux `/etc/passwd`. For more information, see *UserGroupService*.

Note:
If the UserGroupService interface is set to the default configuration, ensure that the content of `/etc/passwd` files is consistent across all computing nodes to guarantee permission control over files and directories.

  - You can define user and group information as needed through the UserGroupService interface and configure `core-site.xml` in **`alidfs.usergroupservice.impl`** to integrate the DFS SDK.
  - Before mounting a file system through HDFS, ensure that Java 1.8 is installed on the ECS instance.

UserGroupService

In Hadoop, user and group information associated with files and directories exist as strings. In DFS, user and group information associated with files and directories exist as 32-bit integers. When you create a file or directory in the DFS SDK, the user information obtained through UserGroupInformation is converted into a UID, and the group information obtained through

UserGroupService is converted into a GID. When you obtain file or directory information, the UID is converted into user name, and the GID is converted into group name.

DFS provides the UserGroupService interface, which enables you to:

- Maintain mappings between users and groups.
- Maintain mappings between user names and UIDs.
- Maintain mappings between group names and GIDs.

DFS URI format

URI format of DFS paths: dfs://DfsInstanceID.RegionID.alidfs.aliyun.com:10290

Example: dfs://f-63a47d43wh98.cn-neimeng-env10-d01.alidfs.aliyun.com:10290

Procedure

1. Configure core-site.xml: Add the following content to the *core-site.xml* file on a node and synchronize the file content to all nodes dependent on hadoop-common:

```
<property>
     <name>fs.oss.impl</name>
     <value>dfs://DfsInstanceID.RegionID.alidfs.aliyun.com:10290</
value>
</property>
<property>
     <name>fs.oss.impl</name>
     <value>com.alibaba.dfs.DistributedFileSystem</value>
</property>
<property>
     <name>fs.AbstractFileSystem.dfs.impl</name>
     <value>com.alibaba.dfs.DFS</value>
</property>
<property>
<name>alidfs.usergroupservice.impl</name>
<value>com.alibaba.dfs.security.LinuxUserGroupService.class</value>
</property>
```

Note:

- Replace RegionID and DfsInstanceID with the actual region ID and DFS instance ID.
- You must synchronize the content in *core-site.xml* to all nodes dependent on hadoop-common.

2. Deploy the DFS SDK: Download *alicloud.dfs-1.0.0.jar* and deploy it in the CLASSPATH of the Hadoop ecosystem component. We recommend that you deploy the package in the directory where dhadoop-common-X.YZ.jar is located.

For example, the following figure shows the directory structure of Spark 2.3.0 after decompression:

```
[root@Hadoop3 spark-2.3.0-bin-hadoop2.7]# ls
bin   data       examples  kubernetes  licenses  metastore_db  python  README.md  sbin  yarn
conf  derby.log  jars      LICENSE     logs      NOTICE        R       RELEASE    work
```

You need to copy *alicloud.dfs-1.0.0.jar* to the *jars* directory.

3. Configure optimization settings: The DFS SDK provides some configuration items, such as **io.file.buffer.size** and **dfs.connection.count**, to optimize application performance. After you configure the configuration items in *core-site.xml*, you must synchronize the file content to all nodes dependent on hadoop-common.

```
<property>
     <name>io.file.buffer.size</name>
     <value>4194304</value>
     <description>To achieve high throughput, no less than 1 MB, no
 more than 8 MB</description>
</property>
<property>
     <name>dfs.connection.count</name>
     <value>1</value>
     <description>If multi threads in the same process will read/
write to DFS, set to count of threads</description>
</property>
```

4. Verify installation: After deployment and configuration, use the hadoop fs command line tool to verify the installation:

```
hadoop fs -ls /
```

```
[hadoop@iZ5wf05xt7fvxpnkx15oy2Z ~/hadoop-2.7.2]$ bin/hadoop fs -ls /
Found 12 items
drw-r----T   - hadoop       hadoop         75498848 1970-01-01 08:00 /MR
drwxr----T   - alicloud-dfs alicloud-dfs   75498848 1970-01-01 08:00 /benchmarks
drw-r----T   - hadoop       hadoop         75498848 1970-01-01 08:00 /hadoop
drwxr----T   - hadoop       hadoop         75498848 1970-01-01 08:00 /tcpds
drw-r----T   - alicloud-dfs alicloud-dfs   75498848 1970-01-01 08:00 /tmp
```

If no error is reported, the deployment is successful.

16.4 File systems

16.4.1 View file system details

You can view the details of a file system in the DFS console.

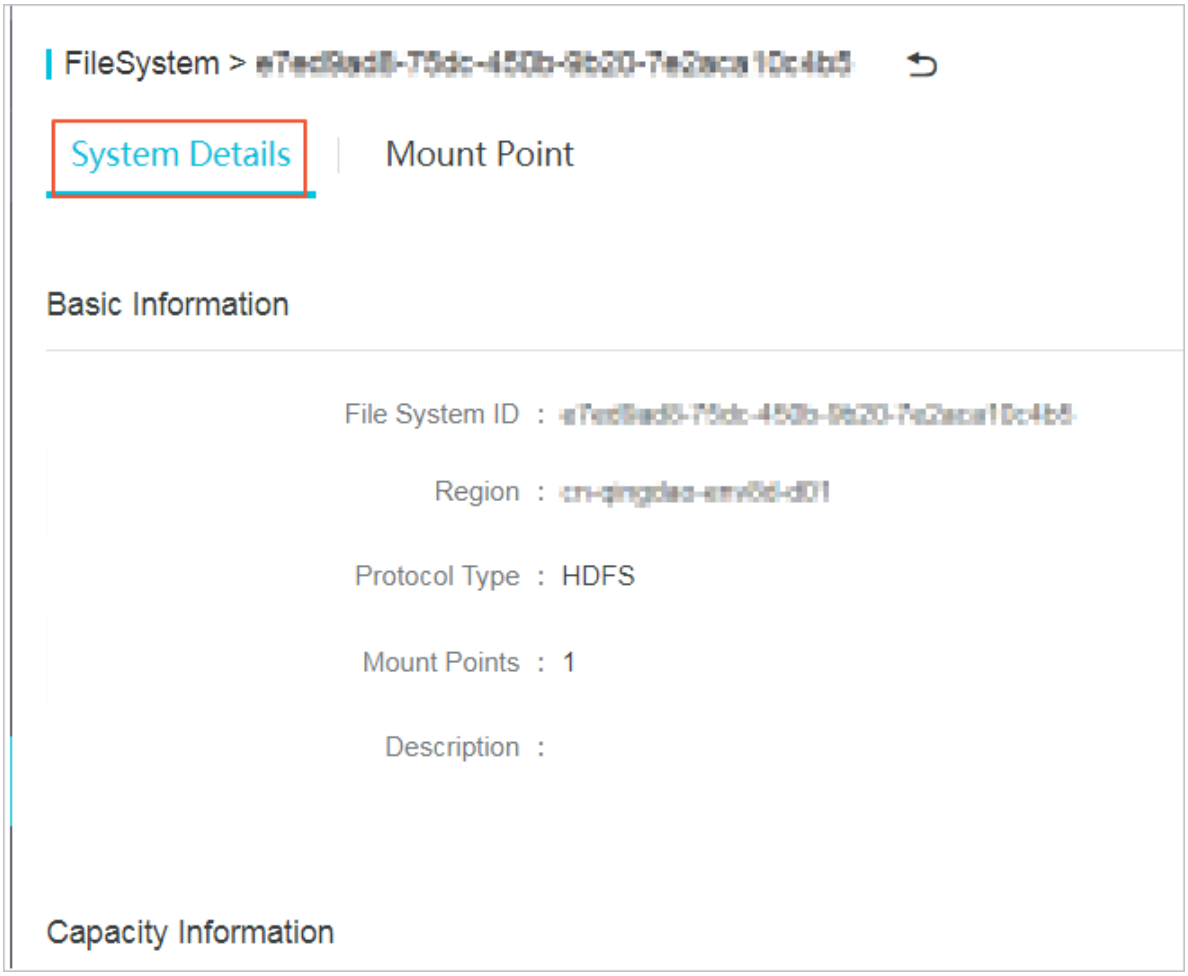
Prerequisites

Before viewing file system details, you need to complete the steps in *Quick start*, or ensure that at least one file system has been created.

Procedure

1. *Log on to the DFS console*.
2. On the **Distributed File System (DFS)** page, click a **File System ID** or click the icon in the Actions column corresponding to a file system and choose **Details** from the shortcut menu.

   The **System Details** page shows basic information about the file system, such as file system ID, region, and capacity, as shown in the following figure.



16.4.2 Delete file systems

You can delete a file system in the DFS console.

Prerequisites

Before deleting a file system, you need to complete the steps in *Quick start*, or ensure that at least one file system has been created.

Procedure

1. *Log on to the DFS console*.
2. Click the  icon in the Actions column corresponding to a file system instance and choose **Delete** from the shortcut menu.
3. In the **Delete File System Instance** dialog box, click **OK**.

16.4.3 Change file system information

You can change file system information in the DFS console.

Prerequisites

Before changing the information, you need to complete the steps in *Quick start*, or ensure that at least one file system and permission group have been created.

Procedure

1. *Log on to the DFS console*.
2. Click the  icon in the Actions column corresponding to a file system instance and choose **Change** from the shortcut menu.
3. In the **Change File System** dialog box, configure **Name**, **Description**, and **File System Capacity** for the file system.

   Note:
   The maximum file system capacity is 10 TB.
4. Click **OK**.

16.5 Mount points

16.5.1 View the mount point list

You can view the list of mount points in the DFS console.

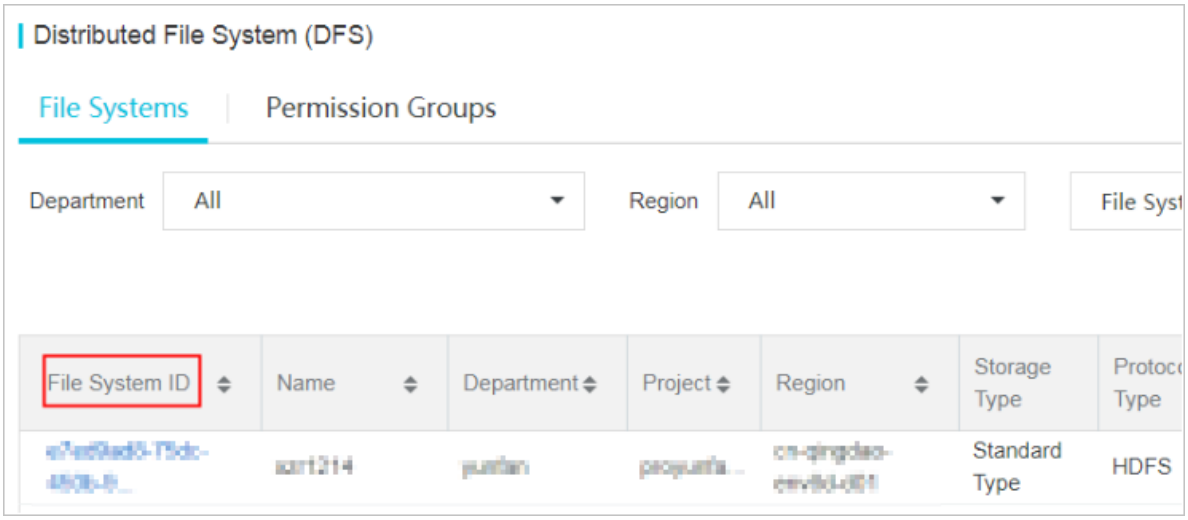
Prerequisites

Before viewing the mount point list, you need to complete the steps in *Quick start*, or ensure that at least one file system and mount point have been created.

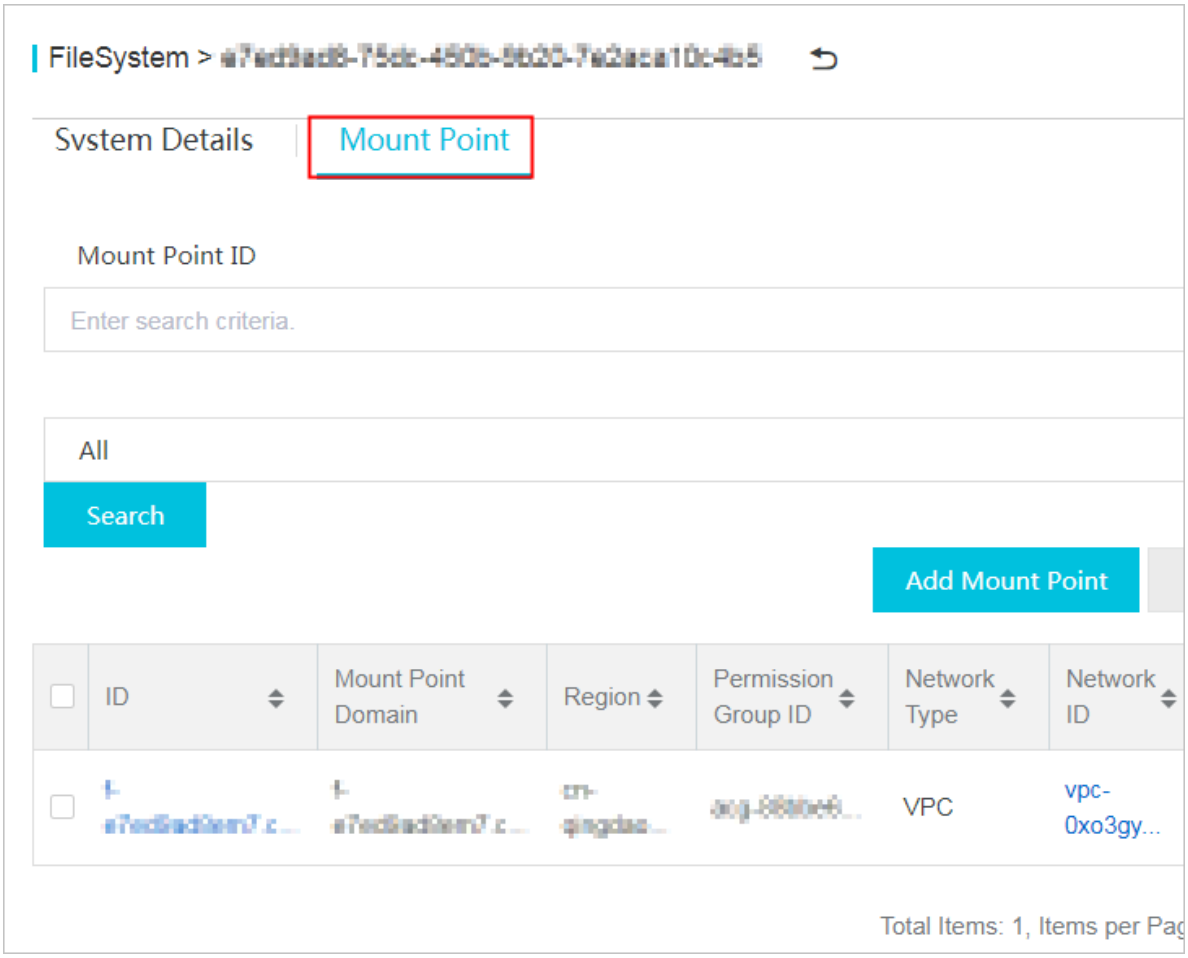
Procedure

1. *Log on to the DFS console*.

2. On the **Distributed File System (DFS)** page, click a **File System ID** to go to the System Details page, as shown in the following figure.



3. Click the **Mount Point** tab to view the list of all mount points in the file system, as shown in the following figure.



16.5.2 Manage mount points

You can change, enable, or delete a mount point in the DFS console.

Prerequisites

Before changing, enabling, or deleting a mount point, you need to complete the steps in *Quick start*, or ensure that at least one file system and one mount point have been created.

Procedure

1. *Log on to the DFS console*.
2. On the **Distributed File System (DFS)** page, click a **File System ID** to go to the file system details page, and click **Mount Point**.
3. You can change, enable, and delete a mount point.
   - Locate the mount point that you want to change and perform the following operations:
     a. Click the  icon in the Actions column corresponding to the mount point and choose **Change** from the shortcut menu.
     b. In the **Change Mount Point** dialog box, you can change the permission group to which the mount point is bound, change the status to **Start** or **Pause**, and enter the description of the mount point.
     c. After you complete the modifications, click **OK**.
   - Locate the mount point that you want to pause and perform the following operations:
     a. Click the  icon in the Actions column corresponding to the mount point and choose **Pause** from the shortcut menu.
     b. In the displayed dialog box that prompts you to confirm the pause operation, click **OK**.
   - Locate the mount point that you want to delete and perform the following operations:
     a. Click the  icon in the Actions column corresponding to the mount point and choose **Delete** from the shortcut menu.
     b. In the displayed dialog box that prompts you to confirm the delete operation, click **OK**.

16.6 Permission groups

16.6.1 View the permission group list

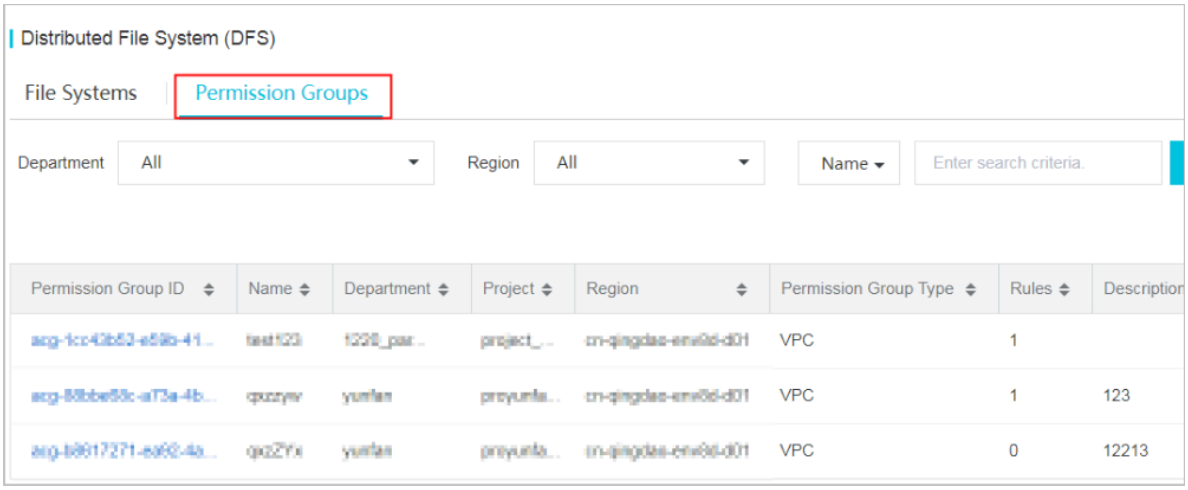
You can view the list of permission groups in the DFS console.

Prerequisites

Before viewing the permission group list, you need to complete the steps in *Quick start*, or ensure that at least one file system and one permission group have been created.

Procedure

1. *Log on to the DFS console*.
2. On the **Distributed File System (DFS)** page, click the **Permission Groups** tab to view the permission group list, as shown in the following figure.



16.6.2 Change permission group information

You can change the information about a permission group in the DFS console, including name and description.

Prerequisites

Before changing the information, you need to complete the steps in *Quick start*, or ensure that at least one file system and permission group have been created.

Context

You must bind a permission group to each of your mount points. Each permission group has a source IP address whitelist used to restrict access to the mount point from ECS instances. You can change the permission group bound to a mount point as required.

Procedure

1. *Log on to the DFS console*.
2. On the **Distributed File System (DFS)** page, click the **Permission Groups** tab.
3. Click the  icon in the Actions column corresponding to a permission group and choose **Change** from the shortcut menu.

4. In the **Change Permission Group** dialog box, configure **Name** and **Description** for the permission group.
5. Click **OK**.

16.6.3 Delete permission groups

You can delete permission groups in the DFS console.

Prerequisites

Before deleting a permission group, you need to complete the steps in *Quick start*, or ensure that at least one file system and one permission group have been created.

Procedure

1. *Log on to the DFS console*.
2. On the **Distributed File System (DFS)** page, click the **Permission Groups** tab.
3. Click the  icon in the Actions column corresponding to a permission group and choose **Delete** from the shortcut menu.
4. In the **Delete Permission Group** dialog box, click **OK**.

16.6.4 Manage permission group rules

You can manage the rules of a permission group in the DFS console, including changing and deleting.

Prerequisites

Before managing a permission group rule, you need to complete the steps in *Quick start*, or ensure that at least one file system and one permission group have been created, and the permission group has at least one rule.

Context

In DFS, the permission group acts as a whitelist. You can add rules to the permission group to allow access or grant different levels of access permissions to IP addresses or CIDR blocks.

Warning:

To secure your data, we strongly recommend that you exercise caution when adding permission group rules and granting permissions to IP addresses.

Procedure

1. *Log on to the DFS console*.

2. On the **Distributed File System (DFS)** page, click the **Permission Groups** tab.
3. Click the  icon in the Actions column corresponding to a permission group and choose **Manage Rules** from the shortcut menu.
4. On the rule list page, you can change or delete a rule.

   To change a rule of the permission group:

   a. Click the  icon in the Actions column corresponding to the rule and choose **Change** from the shortcut menu.
   b. In the **Change Rule** dialog box, you can change **Access Type**, **Authorized IP Address**, and **Priority**.
   c. Click **OK**.

   To delete a rule of the permission group:

   a. Click the  icon in the Actions column corresponding to the rule and choose **Delete** from the shortcut menu.
   b. In the **Delete Rule** dialog box, click **OK**.

17 ApsaraDB for RDS

17.1 What is ApsaraDB for RDS?

Alibaba Cloud ApsaraDB for Relational Database Service (RDS) is a stable, reliable, and auto-scaling online database service.

Based on Alibaba Cloud's distributed file system and high-performance storage, ApsaraDB for RDS provides a complete set of solutions for disaster tolerance, backup, recovery, monitoring, and migration to free you from worries of database operations and maintenance.

ApsaraDB for RDS provides three storage engines: MySQL, PostgreSQL, and PPAS. They help you conveniently and rapidly create database instances suitable for your scenarios.

ApsaraDB for MySQL

Based on Alibaba Cloud's MySQL source code branch, ApsaraDB for MySQL has proven to have excellent performance and throughput. It has withstood the massive data traffic and large number of concurrent users during Double 11. ApsaraDB for MySQL provides basic functions such as instance management, account management, database management, instance whitelist setting, backup, recovery, transparent data encryption, and data migration. It also provides the following advanced functions:

- **Read-only instance:** In scenarios where there are a few write requests but a great number of read requests, you can enable read/write splitting to distribute read pressure of the primary instance. To achieve auto scaling of the reading capability and relieve the database pressure, ApsaraDB for MySQL 5.6 allows you to create read-only instances. You can use read-only instances to read large amounts of data from the database and increase the application throughput.
- **Read/Write splitting:** The read/write splitting function provides an extra read/write splitting address. This address links the primary instance with all its read-only instances to enable an automatic link for read/write splitting. The application can use this method to read and write data by connecting to the same read/write splitting address. Write requests are automatically routed to the primary instance, and read requests are routed to each read-only instance based on their weight. You can add more read-only instances to scale up the processing capacity of the system. No application change is required.
- **CloudDBA database performance optimization:** CloudDBA provides the intelligent diagnostics and optimization features based on the SQL statement performance, CPU

utilization, IOPS, memory usage, disk usage, number of connections, lock information, and hotspot tables. It can discover the existing or potential health issues of databases to the maximum extent. CloudDBA performs diagnostics for a single instance. It provides issue details and solutions to facilitate your instance maintenance.

- **Data compression:** ApsaraDB for MySQL 5.6 allows you to compress data through the TokuDB storage engine. Extensive tests show that the data volume is reduced by 80% to 90% after data tables are transferred from the InnoDB storage engine to the TokuDB storage engine. 2 TB of data can be compressed to 400 GB or less. Aside from data compression, the TokuDB storage engine supports transaction and online DDL operations and is compatible with the applications running on the MyISAM and InnoDB storage engines.

ApsaraDB for PostgreSQL

PostgreSQL is the most advanced open source database in the world. PostgreSQL excels for its full compliance with SQL specifications and robust support for a diverse range of data formats such as JSON, IP, and geometric data. In addition to excellent support for features such as transactions, subqueries, Multi-Version Concurrency Control (MVCC), and data integrity check, ApsaraDB for PostgreSQL integrates a series of important functions including high availability, backup, and recovery that help ease your operations and maintenance burden.

ApsaraDB for PostgreSQL provides basic functions such as instance management, account management, database management, whitelist configuration for instances, backup, recovery, and data migration.

ApsaraDB for PPAS

Postgres Plus Advanced Server (PPAS) is a stable, secure, and scalable enterprise-class relational database. Based on PostgreSQL, the most advanced open source database in the world , PPAS brings enhancements in terms of performance, application solutions, and compatibility. It also provides the capability of directly running Oracle applications. You can run enterprise-class applications on PPAS stably and obtain cost-effective services.

ApsaraDB for PPAS provides basic functions such as instance management, account management, database management, whitelist configuration for instances, backup, recovery, and data migration.

17.2 Limits

17.2.1 Usage limits of ApsaraDB RDS for MySQL

Before you use ApsaraDB RDS for MySQL, you need to understand its limits and take precautions against them.

To guarantee instance stability and security, ApsaraDB RDS for MySQL has some service limits, as listed in *Table 17-1: Limits on ApsaraDB RDS for MySQL*.

Table 17-1: Limits on ApsaraDB RDS for MySQL

Operation	Description
Database parameter modification	Database parameters can only be modified from the RDS console or through APIs. Due to security and stability considerations, only specific parameters can be modified.
Root permission of databases	The root and SA permissions are not provided.
Database backup	• Logical backup can be performed from the command line interface (CLI) or graphical user interface (GUI). • Physical backup can only be performed from the RDS console or through APIs.
Database restoration	• Logical restoration can be performed from the CLI or GUI. • Physical restoration can only be performed from the RDS console or through APIs.
Data import	• Logical import can be performed from the CLI or GUI. • Data can only be immigrated by using the MySQL command-line client.
ApsaraDB RDS for MySQL storage engine	• Only InnoDB and TokuDB are supported. Due to the inherent defects of the MyISAM engine, data may be lost. Only some stock instances are using MyISAM engine. MyISAM engine tables in newly created instances will be automatically converted to InnoDB engine tables. • For safety performance and security considerations, we recommend that you use the InnoDB storage engine. • The Memory engine is not supported. Newly created Memory tables will be automatically converted into InnoDB tables.
Database replication	ApsaraDB RDS for MySQL provides dual-node clusters based on a primary/secondary replication architecture. The secondary instances in this replication architecture are hidden and cannot be accessed directly.

Operation	Description
RDS instance restart	Instances must be restarted through the RDS console or APIs.
Account and database management	ApsaraDB RDS for MySQL uses the RDS console to manage accounts and databases by default. ApsaraDB RDS for MySQL also allows you to create a superuser account to manage users, passwords, and databases.
Standard account	• Custom authorization is not supported. • The account management and database management interfaces are provided in the RDS console. • Instances that support standard accounts also support superuser accounts.
Superuser account	• Custom authorization is supported. • The account management and database management interfaces are not provided in the RDS console. The relevant operations can only be performed through code or DMS. • The superuser account cannot be reverted back into a standard account.

17.2.2 Usage limits of ApsaraDB RDS for PostgreSQL

Before you use ApsaraDB RDS for PostgreSQL, you need to understand its limits and take precautions against them.

To guarantee instance stability and security, ApsaraDB RDS for PostgreSQL has some service limits, as listed in *Table 17-2: Limits on ApsaraDB RDS for PostgreSQL*.

Table 17-2: Limits on ApsaraDB RDS for PostgreSQL

Operation	Description
Database parameter modification	Not supported.
Root permission of databases	Superuser permissions are not provided.
Database backup	Data can only be backed up by using pg_dump.
Data migration	Only PostgreSQL can be used to restore data that was backed up by using pg_dump.

Operation	Description
Database replication	• The system automatically builds HA databases based on PostgreSQL streaming replication without user input. • PostgreSQL standby nodes are hidden and cannot be accessed directly.
RDS instance restart	RDS instances must be restarted from the RDS console or through APIs.
Network settings	For instances that are operating in safe mode, net.ipv4.tcp_timestamps cannot be enabled in SNAT mode.

17.2.3 Usage limits of ApsaraDB RDS for PPAS

Before you use ApsaraDB RDS for PPAS, you must understand its limits and take precautions against them.

To guarantee instance stability and security, ApsaraDB RDS for PPAS has some service limits, as listed in *Table 17-3: Limits on ApsaraDB RDS for PPAS*.

Table 17-3: Limits on ApsaraDB RDS for PPAS

Operation	Description
Database parameter modification	Not supported.
Root permission of databases	Superuser permissions are not provided.
Database backup	Data can only be backed up by using pg_dump.
Data migration	Only PostgreSQL can be used to restore data that was backed up by using pg_dump.
Database replication	• The system automatically builds HA databases based on PPAS streaming replication without user input. • PPAS standby nodes are hidden and cannot be accessed directly.
RDS instance restart	RDS instances must be restarted from the RDS console or through APIs.
Network settings	For instances that are operating in safe mode, net.ipv4.tcp_timestamps cannot be enabled in SNAT mode.

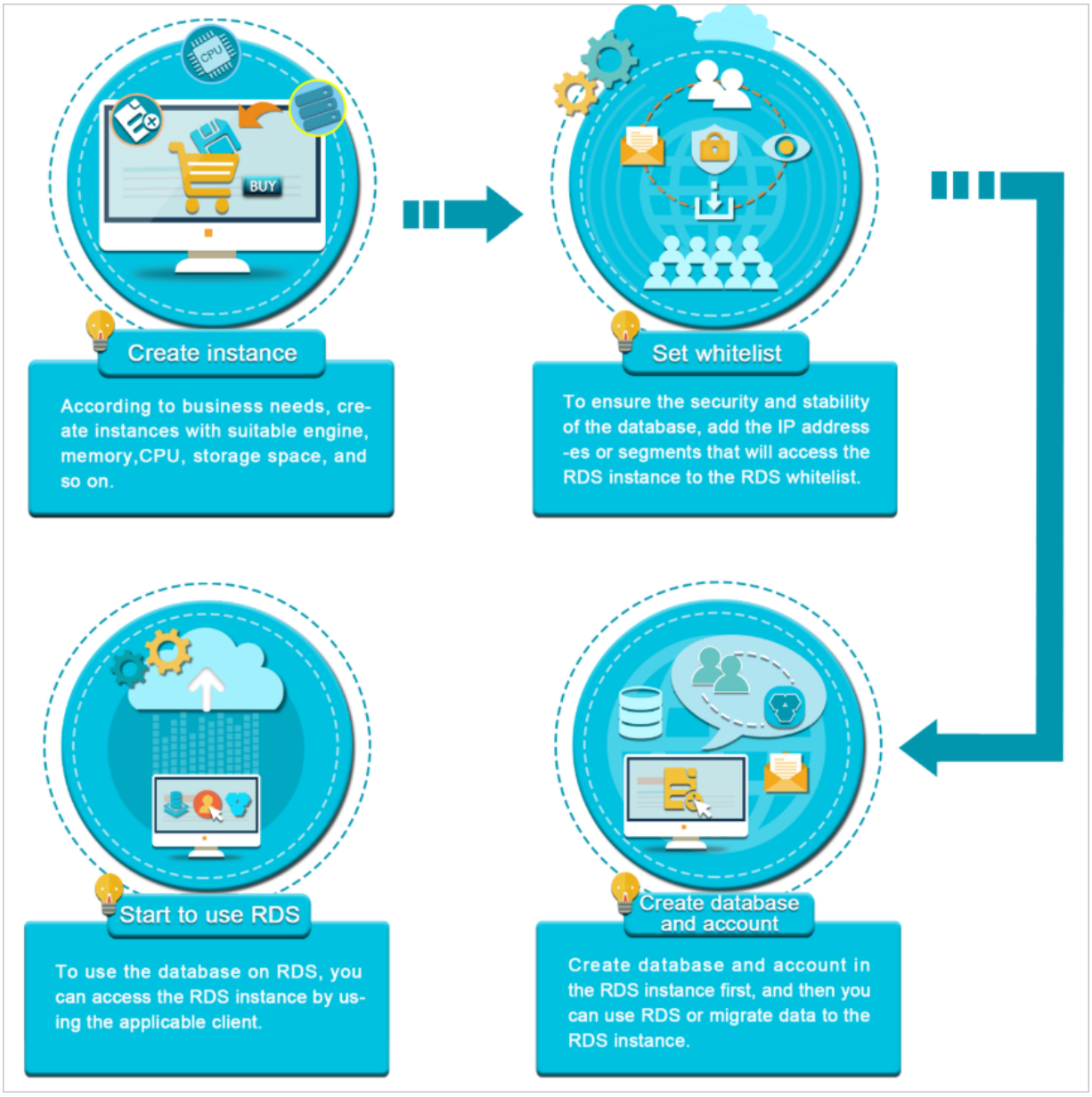
17.3 Quick start

17.3.1 Quick start

ApsaraDB for RDS quick start covers the following topics: ApsaraDB for RDS instance creation, whitelist configuration, database creation, account creation, and instance connection. This topic uses RDS for MySQL as an example to describe how to use ApsaraDB for RDS. It provides all the information you need to build an ApsaraDB for RDS database.

Typically, from instance creation until the instance can be used, the following operations have to be performed, as shown in *Figure 17-1: Quick start flow*.

Figure 17-1: Quick start flow



- *Create an instance*

  An instance is a virtualized database server. You can create and manage multiple databases in an instance.

- *Configure a whitelist*

  After creating an RDS instance, you need to configure its whitelist to allow access from external devices.

  The whitelist provides an RDS instance with high-level security. We recommend that you maintain the whitelist periodically. Configuring the whitelist does not affect the normal operations of the RDS instance.

- *Create a database* and an *account*

  Before you use a database, you need to create the database and an account in the RDS instance. Different engines support different account modes. For more information, see the console UI and documentation.

- *Log on to an instance through DMS* or *Connect to a MySQL instance from a client*

  After creating an instance, configuring a whitelist, and creating a database and an account, you can use Data Management Service (DMS) or a general database client to connect to the RDS instance.

17.3.2 Log on to the RDS console

This topic describes how to log on to the RDS console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.

- The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
- You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click **LOGIN** to go to the **Dashboard** page.
5. In the top navigation bar, choose **Console** > **Database** > **Relational Database Service**.

17.3.3 Create an instance

This topic describes how to create an instance in the RDS console.

Prerequisites

Before you create an RDS instance, you need to apply for an Apsara Stack Management Console account.

Procedure

1. *Log on to the RDS console*.
2. On the **Relational Database Service (RDS)** page, click **Create Instance** in the upper-right corner. On the **Create Instance** page, configure parameters as promoted.

   Table 17-4: Instance creation parameters describes the parameter configurations.

   Table 17-4: Instance creation parameters

Category	Parameter	Description
Basic Configuration	**Department**	The department to which the instance belongs.
	Project	The project to which the instance belongs.
	Region	The region where the instance is located.
	Zone	The zone of the instance. Common RDS instances adopt the hot standby architecture. A single zone means that the primary and secondary nodes are in the same zone.
Network Type	**Instance Type**	The type of the instance.

Category	Parameter	Description
	Network Type	The network types supported by RDS instances : • `Classic Network`: Cloud services in a classic network are not isolated. Unauthorized access to a cloud service is blocked only by the security group or whitelist policy of the service. • VPC: Virtual Private Cloud (VPC) helps you build an isolated network environment in Alibaba Cloud. You can customize route tables, IP address ranges, and gateways in a VPC. We recommend that you choose a VPC to improve security. You must create a VPC instance in advance. Alternativ ely, you can change the network type after you create an instance.
Access Mode	**Access Mode**	The access modes supported by RDS instances, including: • `Standard Mode`: ApsaraDB for RDS uses Server Load Balancer to eliminate the impact from HA switching of the database engine on the application layer. This shortens the response time, but slightly increases the probability of transient disconnections. • `Safe Mode`: This mode prevents 90% of transient disconnections. However, the response time is increased by 20% or more, and performance is affected.
Specification Configuration	**Instance Name**	The name of the RDS instance. It is a string of 2 to 256 characters including letters, numbers, and underscores (_). It must start with a letter.
	Database Type	Database types vary according to regions. The available database types are displayed on the page.
	Database Version	The version of the database.
	CPU/ Memory	The specification of the instance, which includes: • Common: for example, 1 core and 1 GB memory. • Dedicated: The specification is suffixed with "Dedicated." • Dedicated Host: The specification is suffixed with " Dedicated Host." • Financial Version Single Data Center: The specification is suffixed with "Financial Version Single Data Center."

Category	Parameter	Description
		Memory size determines the maximum number of connections and IOPS. The actual values are displayed on the console interface.
	Storage Size	The storage size of the instance, which includes data, system files, binlog files, and transaction files.
Quantity	**Instances**	The number of RDS instances that can be created simultaneously. The maximum number is 20.

3. Click **Create**.

17.3.4 Initiate the configuration

17.3.4.1 RDS for MySQL

17.3.4.1.1 Configure a whitelist

To guarantee database security and reliability, you need to modify the whitelist of the RDS instance before you enable the instance. You need to add the IP addresses or IP address segments used for database access to the whitelist of the RDS instance.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. For each newly created RDS instance, IP address 0.0.0.0/0 is added to the **default** whitelist group by default. 0.0.0.0/0 allows all IP addresses to access the instance, which greatly reduces database security. Delete 0.0.0.0/0 from the whitelist.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Security Control** > **Configure Whitelist**.
4. You can use two methods to add an IP address or IP address segment.
   - Add an IP address or IP address segment directly to the default whitelist group.
   a) Click the  icon corresponding to the **default** whitelist group, and add an IP address or IP address segment.
   b) Click **OK**.

- Add a whitelist group and add an IP address or IP address segment to the group.

a) Click **Create Whitelist Group**. In the dialog box that appears, enter a group name and an IP address or IP address segment.

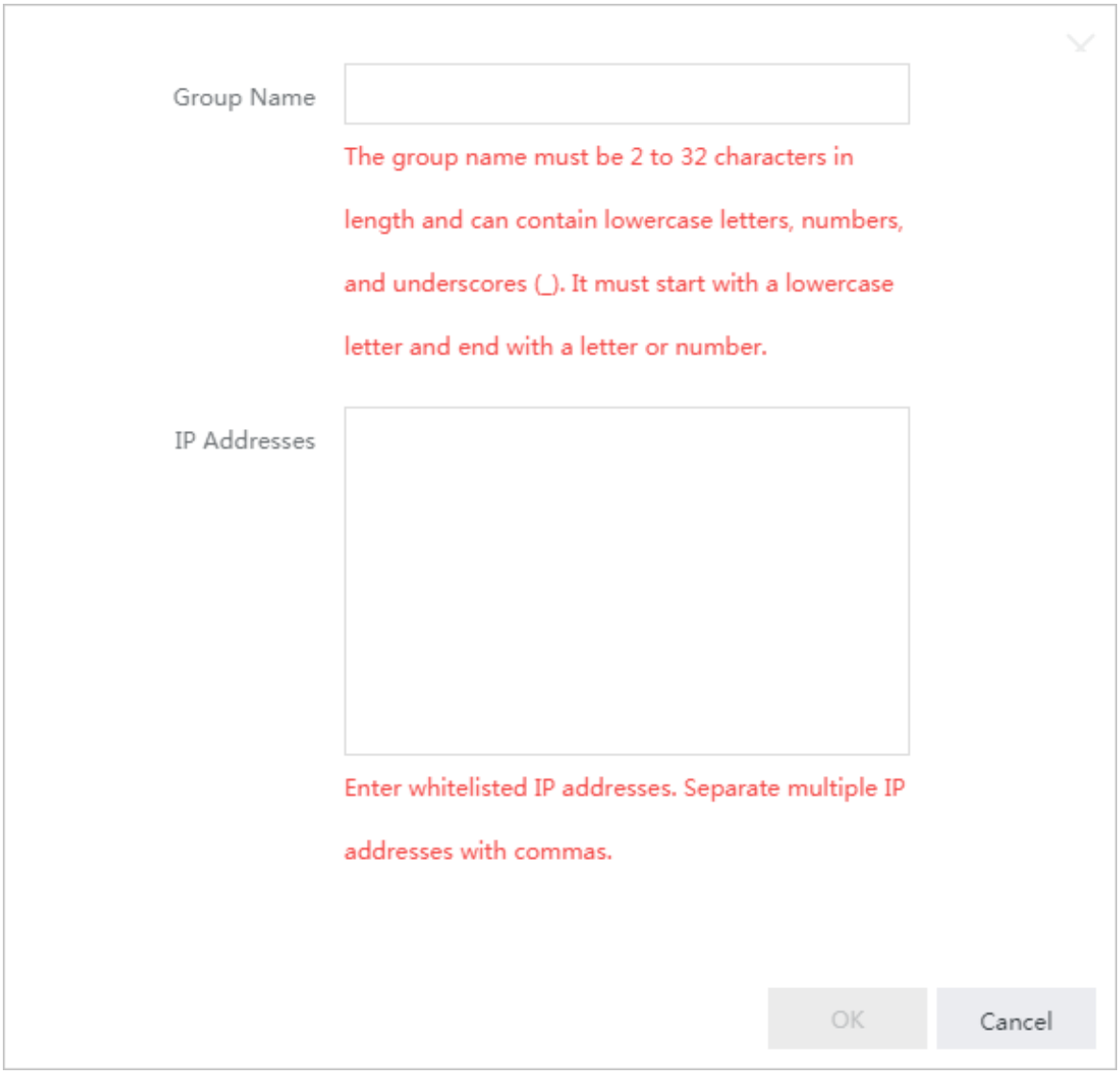
b) Click **OK**.

Table 17-5: Whitelist configuration parameters describes the parameter configurations.

Table 17-5: Whitelist configuration parameters

Parameter	Description
Group Name	The name of the new whitelist group. The naming rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 2 to 32 characters in length. You cannot modify the name of a created whitelist group.
IP Addresses	The IP addresses or IP address segments allowed to access the RDS instance. **Note:** • If you enter an IP address segment, such as 10.10.10.0/24, any IP address in the format of 10.10.10.X can access the RDS instance . • If multiple IP addresses are entered, use commas (,) to separate the addresses and do not add spaces between the addresses and commas, such as 192.168.0.1,172.16.213.9. • 127.0.0.1 indicates that no IP address is allowed to access the RDS instance. • 0.0.0.0/0 indicates that all IP addresses are allowed to access the RDS instance.

Figure 17-2: Create a whitelist group



What's next

Correct use of the whitelist can improve access security for your RDS instance. We recommend that you maintain the whitelist periodically. After you configure the whitelist, you can perform the following operations:

- Click the icon to modify the whitelist group.
- Click the icon to clear the default whitelist group or delete a custom whitelist group.

17.3.4.1.2 Create a premier account

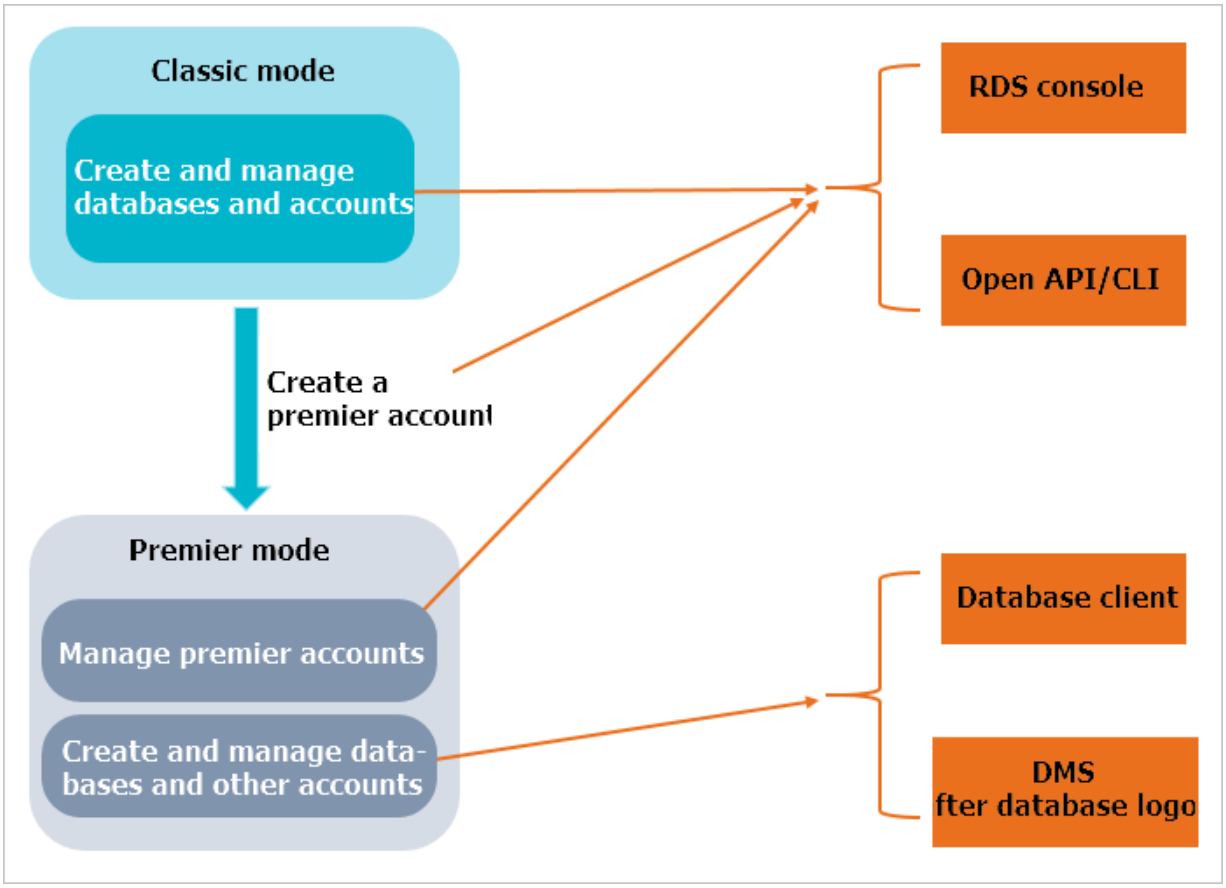
ApsaraDB for RDS supports two account management modes: classic and premier. For MySQL 5.6 instances, you can create a premier account to upgrade the account management mode from classic to premier.

Context

Compared with the classic mode, the premier mode enables more permissions. In premier mode , you can use SQL to directly manage databases and accounts. Therefore, we recommend that you use the premier mode. After a premier account is created for a primary instance, the premier account is synchronized to read-only instances.

Figure 17-3: Comparison between account management modes shows the differences in creating and managing databases and accounts between the two modes.

Figure 17-3: Comparison between account management modes



Note:

- On a MySQL 5.6 instance, the account management mode can be upgraded only from classic to premier, but cannot be downgraded from premier to classic.
- When a premier account is created, the instance restarts once, which causes a transient disconnection of less than 30 seconds. Choose an appropriate time and make sure that your application can be automatically reconnected when you create a premier account to avoid service impacts from transient disconnections.
- The following changes occur after an instance switches to the premier account mode:
  - You cannot use the RDS console or APIs to manage databases and standard accounts. Instead, you must use SQL commands or Data Management Service (DMS). **Create Account** is not displayed on the **Accounts** page.
  - In MySQL 5.6, you cannot directly access the mysql.user or mysql.db tables. However, you can view the existing account and permissions through mysql.user_view and mysql.db_view.
  - You cannot use the premier account to change the passwords of standard accounts. To change the password of a standard account, you must delete the account and create a new one.
  - You can use the RDS console or APIs to reset the password and permissions of a premier account. The resetting operation does not affect the other accounts that have been created .

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Database Management** > **Accounts**.
4. On the **Accounts** page, click **Create Account**. On the **Create Account** page that appears, configure parameters as prompted.

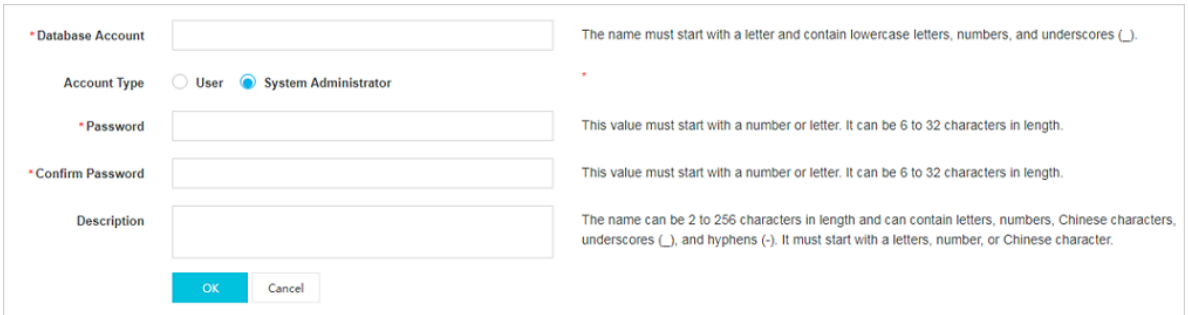
   Table 17-6: Premier account creation parameters describes the parameter configurations.

   Table 17-6: Premier account creation parameters

Parameter	Description
Database Account	The name of the account. The naming rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number.

Parameter	Description
	• It can contain lowercase letters, numbers, and underscores (_). • It can be 2 to 16 characters in length. • Reserved keywords cannot be used. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.
Account Type	The type of the account. The following two types are available: • `User` • `System Administrator`: This option should be selected.
Password	The password for this account. The requirements are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 6 to 32 characters in length.
Confirm Password	The same as the password. The requirements are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 6 to 32 characters in length.
Description	The description of the account. The requirements are as follows: • It must start with a letter or number. • It can contain letters, numbers, underscores (_), and hyphens (-). • It can be 2 to 256 characters in length.

Figure 17-4: Create an account



5. Click **OK**.

17.3.4.1.3 Create a standard account

After you create an RDS instance and configure the whitelist, you need to create a database and an account in the instance. This topic describes how to create a standard account.

Context

Before you migrate data from the on-premises database to ApsaraDB for RDS, you must create a database and an account for the database. Ensure that the database has the same properties as the on-premises database, and the account of the database has the same permissions as the account of the on-premises database. Databases under the same instance share all resources of this instance. You can create up to 500 databases and 500 accounts under each instance in MySQL 5.6.

When assigning database account permissions, follow the least privilege principle. Create accounts by service roles and assign proper read-only and read/write permissions to the accounts. When necessary, you may split database accounts and databases into smaller units so that each database account can only access data for their own services.

Note:

For database security, set strong passwords for the accounts and change the passwords periodically.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Database Management** > **Accounts**.
4. On the **Accounts** page, click **Create Account**. On the **Create Account** page, configure parameters as prompted.

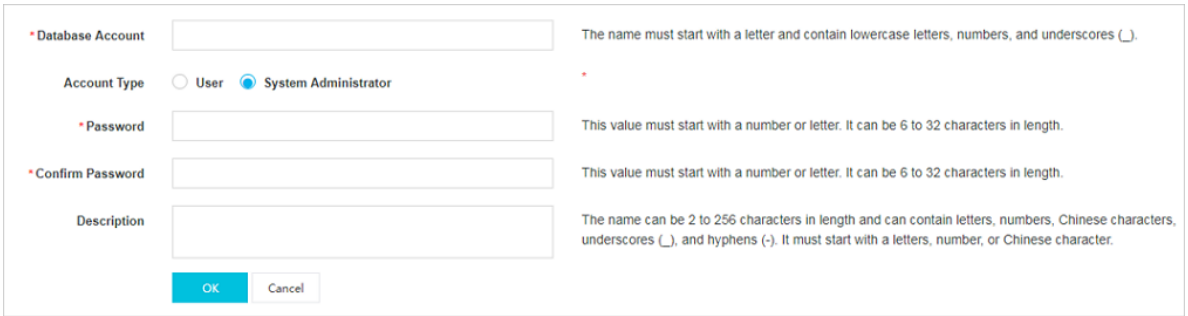
   Table 17-7: Standard account creation parameters describes the parameter configurations.

Table 17-7: Standard account creation parameters

Parameter	Description
Database Account	The name of an account. The naming rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_).

Parameter	Description
	• It can be 2 to 16 characters in length. • Reserved keywords cannot be used. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.
Account Type	The type of an account. The following types are available: • `User`. This option is selected here. • `System Administrator`.
Password	The password for this account. The rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 6 to 32 characters in length.
Confirm Password	The same as the password. The rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 6 to 32 characters in length.
Description	The description for this account. The rules are as follows: • It must start with a letter or number. • It can contain letters, numbers, underscores (_), and hyphens (-). • It can be 2 to 256 characters in length.

Figure 17-5: Create an account



5. Click **OK**.

17.3.4.1.4 Create a database

After you create an RDS instance and configure the whitelist, you need to create a database and an account in the instance.

Context

Before you migrate data from the on-premises database to ApsaraDB for RDS, you must create a database and an account for the database. Ensure that the database has the same properties as the on-premises database, and the account of the database has the same permissions as the account of the on-premises database. Follow the least privilege principle to assign permissions to the database account. Use service roles to create accounts and assign proper read-only and read/write permissions to the roles. When necessary, you may define database accounts and databases in a fine-grained manner so that each database account can only access data for their own services.

Procedure

1. *Log on to the RDS console*.
2. In the left-side navigation pane, choose **Database Management** > **Databases**.
3. On the **Databases** page, click **Create Database**. On the **Create Database** page, configure parameters as prompted.

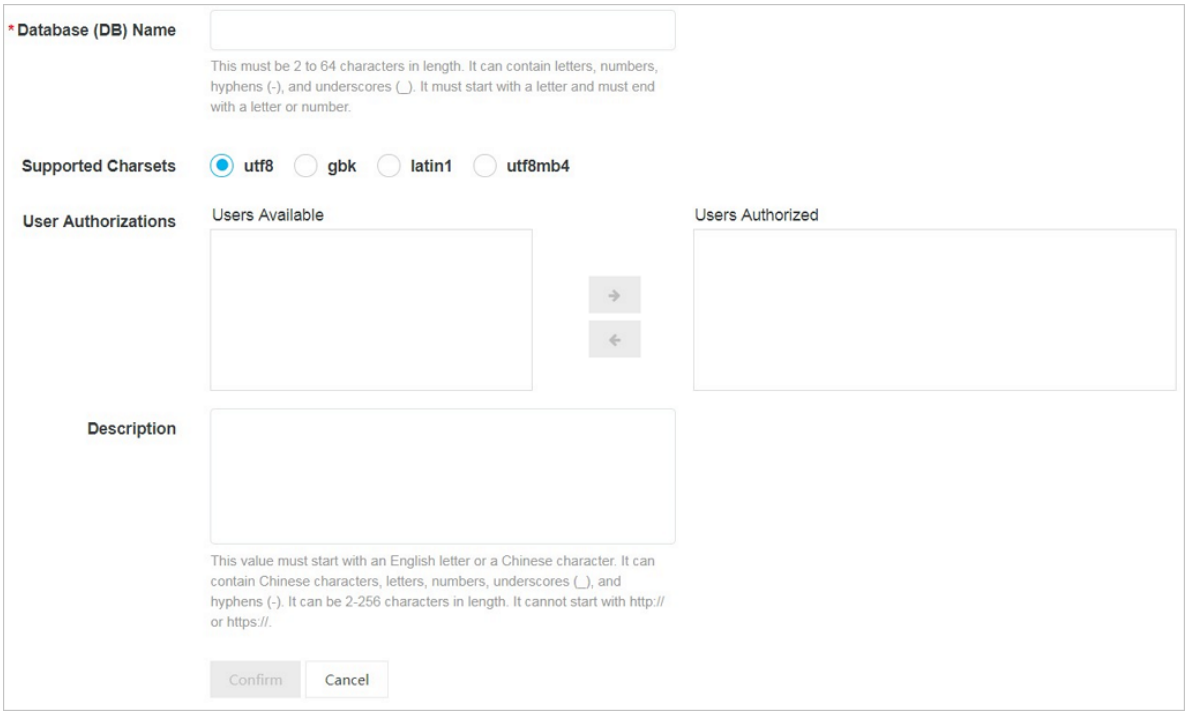
   Table 17-8: Database creation parameters describes the parameter configurations.

   Table 17-8: Database creation parameters

Parameter	Description
Database (DB) Name	The database name. The naming rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, underscores (_), and hyphens (-). • It can be 2 to 64 characters in length. • Reserved keywords cannot be used. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.
Supported Character Set	The character sets supported by the database, including: • `UTF-8` • `gbk` • `latin1`

Parameter	Description
	• `utf8mb4`
User Authorizations	• Select an account that is authorized to use the database. This parameter can be empty if no account is created. • The permissions on the database can be granted only to standard accounts. • The premier account is authorized to use the database by default.
Description	The description of the database. The rules are as follows: • It must start with a lowercase letter. • It can contain lowercase letters, numbers, underscores (_), and hyphens (-). • It can be 2 to 256 characters in length.

Figure 17-6: Create a database



4. Click **OK**.

17.3.4.2 RDS for PostgreSQL

17.3.4.2.1 Configure a whitelist

To guarantee database security and reliability, you need to modify the whitelist of the RDS instance before you enable the instance. You need to add the IP addresses or IP address segments used for database access to the whitelist of the RDS instance.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. For each newly created RDS instance, IP address 0.0.0.0/0 is added to the **default** whitelist group by default. 0.0.0.0/0 allows all IP addresses to access the instance, which greatly reduces database security. Delete 0.0.0.0/0 from the whitelist.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Security Control** > **Configure Whitelist**.
4. You can use two methods to add an IP address or IP address segment.
   - Add an IP address or IP address segment directly to the default whitelist group.
   a) Click the  icon corresponding to the **default** whitelist group, and add an IP address or IP address segment.
   b) Click **OK**.
   - Add a whitelist group and add an IP address or IP address segment to the group.
   a) Click **Create Whitelist Group**. In the dialog box that appears, enter a group name and an IP address or IP address segment.
   b) Click **OK**.

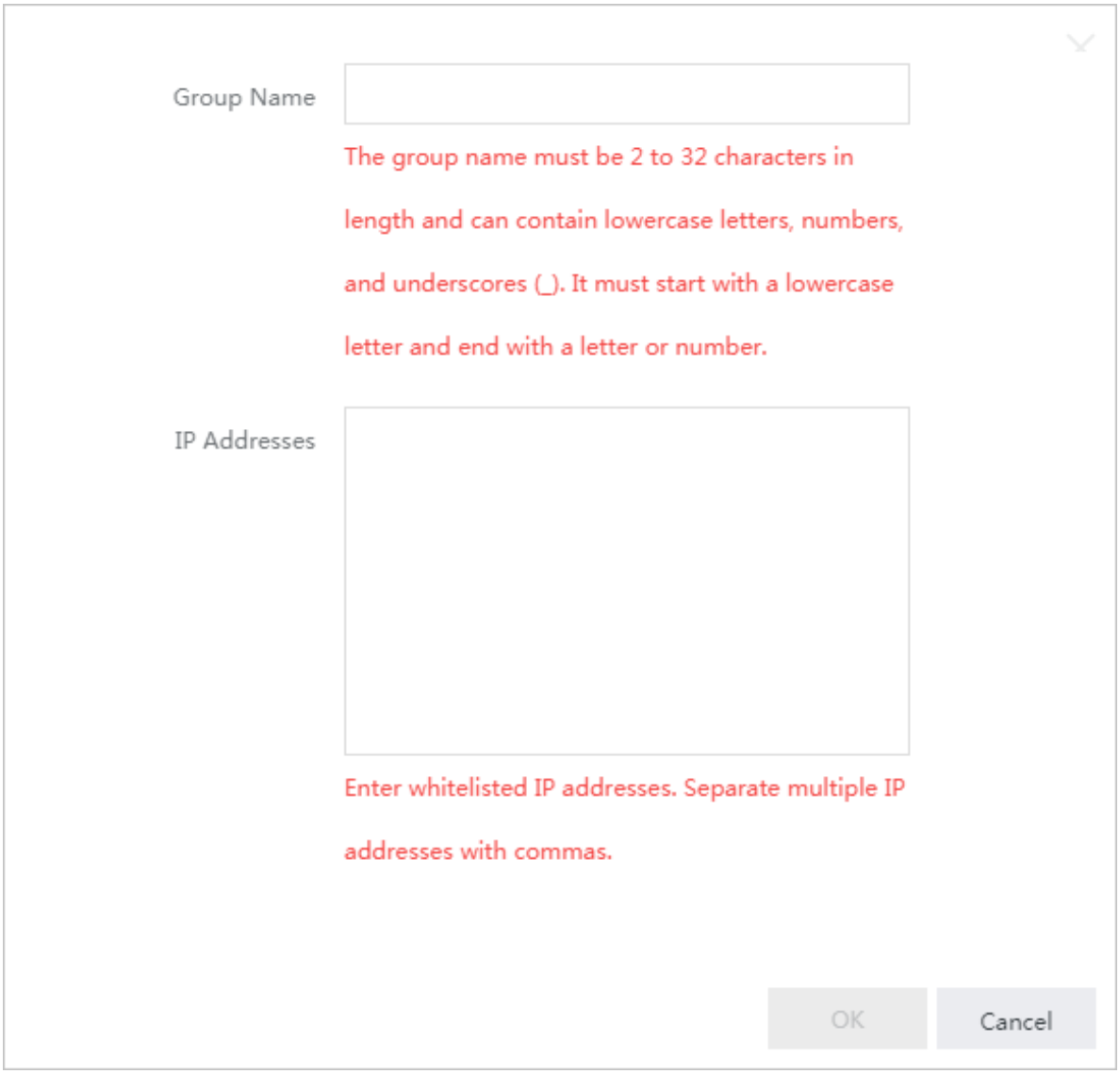
   Table 17-9: Whitelist configuration parameters describes the parameter configurations.

Table 17-9: Whitelist configuration parameters

Parameter	Description
Group Name	The name of the new whitelist group. The naming rules are as follows:

Parameter	Description
	• It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 2 to 32 characters in length. You cannot modify the name of a created whitelist group.
IP Addresses	The IP addresses or IP address segments allowed to access the RDS instance.  **Note:** • If you enter an IP address segment, such as 10.10.10.0/24, any IP address in the format of 10.10.10.X can access the RDS instance . • If multiple IP addresses are entered, use commas (,) to separate the addresses and do not add spaces between the addresses and commas, such as 192.168.0.1,172.16.213.9. • 127.0.0.1 indicates that no IP address is allowed to access the RDS instance. • 0.0.0.0/0 indicates that all IP addresses are allowed to access the RDS instance.

Figure 17-7: Create a whitelist group



What's next

Correct use of the whitelist can improve access security for your RDS instance. We recommend that you maintain the whitelist periodically. After you configure the whitelist, you can perform the following operations:

- Click the icon to modify the whitelist group.
- Click the icon to clear the default whitelist group or delete a custom whitelist group.

17.3.4.2.2 Create a database and an account

After you create an RDS instance and configure the whitelist, you need to create a database and an account in the instance. This topic describes how to create a database and an account.

Context

Before you migrate data from the on-premises database to ApsaraDB for RDS, you must create a database and an account for the database. Ensure that the database has the same properties as the on-premises database, and the account of the database has the same permissions as the account of the on-premises database. When assigning database account permissions, follow the least privilege principle. Create accounts by service roles and assign proper read-only and read/write permissions to the accounts. When necessary, you may define database accounts and databases in a fine-grained manner so that each database account can only access data for their own services.

Note:
For database security, set strong passwords for the accounts and change the passwords periodically.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Database Management** > **Accounts**.
4. On the **Accounts** page, click **Create Account**. On the **Create Account** page that appears, configure parameters as prompted.

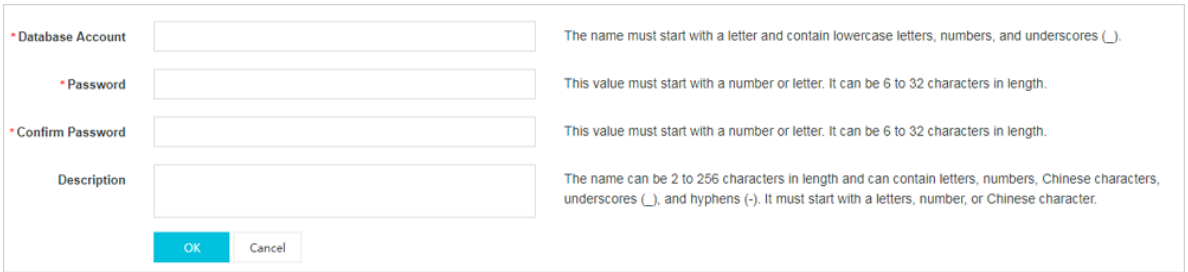
   Table 17-10: Account creation parameters describes the parameter configurations.

   Table 17-10: Account creation parameters

Parameter	Description
Database Account	The name of an account. The naming rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 2 to 16 characters in length.

Parameter	Description
	• Reserved keywords cannot be used. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.
Password	The password for this account. The rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 6 to 32 characters in length.
Confirm Password	The same as the password. The rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 6 to 32 characters in length.
Description	The description of the account. The rules are as follows: • It must start with a letter or number. • It can contain letters, numbers, underscores (_), and hyphens (-). • It can be 2 to 256 characters in length.

Figure 17-8: Create an account



5. Click **OK**.
6. Connect to your RDS instance from the client. For more information, see *Connect to a PostgreSQL instance from a client*.
7. Run the following command to create a database:

   `CREATE DATABASE"databasename"`

   databasename is the name of the database to be created, such as `CREATE DATABASE "mydatabase"`.

17.3.4.3 RDS for PPAS

17.3.4.3.1 Configure a whitelist

To guarantee database security and reliability, you need to modify the whitelist of the RDS instance before you enable the instance. You need to add the IP addresses or IP address segments used for database access to the whitelist of the RDS instance.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. For each newly created RDS instance, IP address 0.0.0.0/0 is added to the **default** whitelist group by default. 0.0.0.0/0 allows all IP addresses to access the instance, which greatly reduces database security. Delete 0.0.0.0/0 from the whitelist.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Security Control** > **Configure Whitelist**.
4. You can use two methods to add an IP address or IP address segment.
   - Add an IP address or IP address segment directly to the default whitelist group.
   a) Click the  icon corresponding to the **default** whitelist group, and add an IP address or IP address segment.
   b) Click **OK**.
   - Add a whitelist group and add an IP address or IP address segment to the group.
   a) Click **Create Whitelist Group**. In the dialog box that appears, enter a group name and an IP address or IP address segment.
   b) Click **OK**.

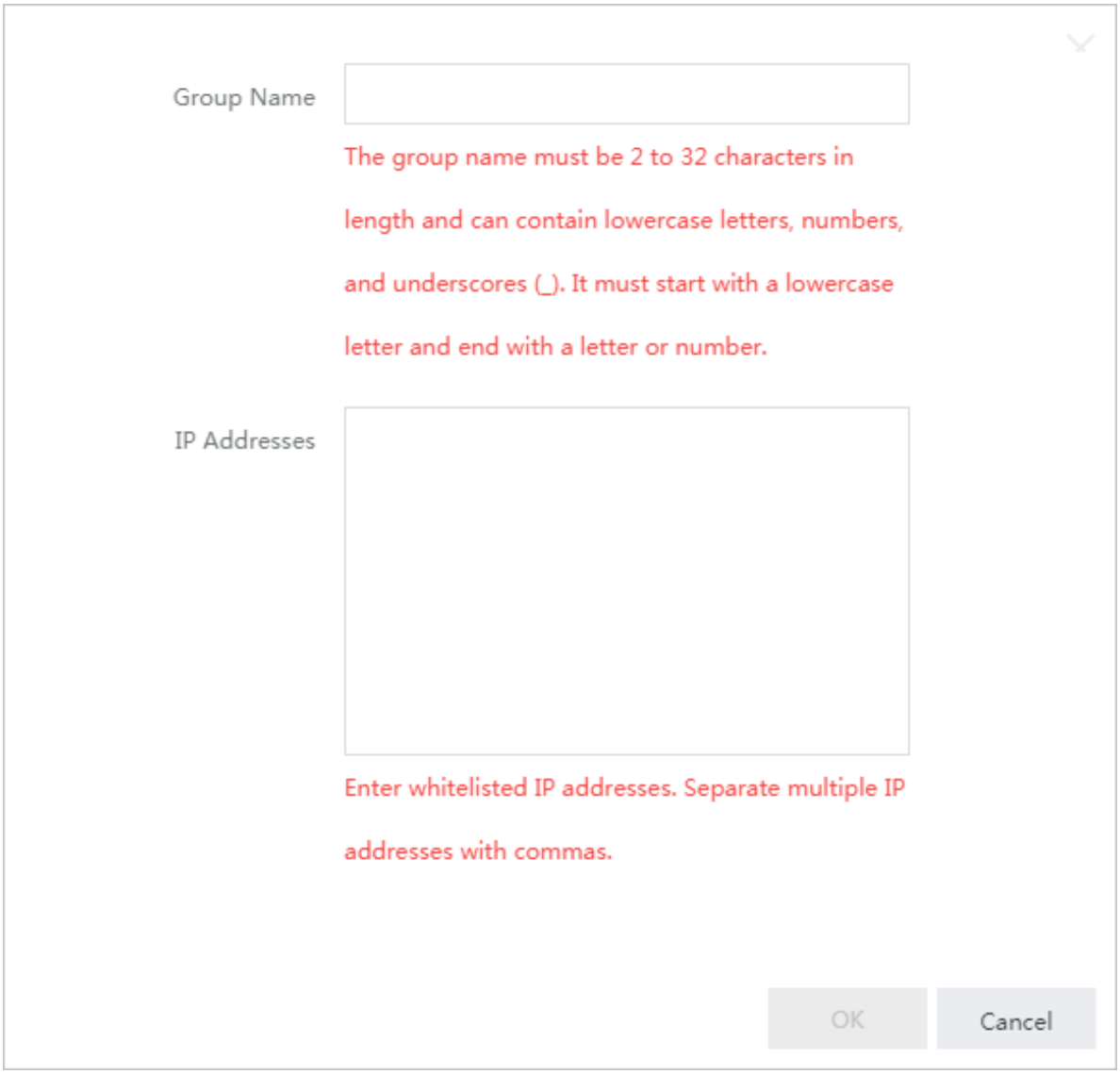
   Table 17-11: Whitelist configuration parameters describes the parameter configurations.

Table 17-11: Whitelist configuration parameters

Parameter	Description
Group Name	The name of the new whitelist group. The naming rules are as follows:

Parameter	Description
	• It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 2 to 32 characters in length. You cannot modify the name of a created whitelist group.
IP Addresses	The IP addresses or IP address segments allowed to access the RDS instance.  **Note:** • If you enter an IP address segment, such as 10.10.10.0/24, any IP address in the format of 10.10.10.X can access the RDS instance . • If multiple IP addresses are entered, use commas (,) to separate the addresses and do not add spaces between the addresses and commas, such as 192.168.0.1,172.16.213.9. • 127.0.0.1 indicates that no IP address is allowed to access the RDS instance. • 0.0.0.0/0 indicates that all IP addresses are allowed to access the RDS instance.

Figure 17-9: Create a whitelist group



What's next

Correct use of the whitelist can improve access security for your RDS instance. We recommend that you maintain the whitelist periodically. After you configure the whitelist, you can perform the following operations:

- Click the icon to modify the whitelist group.
- Click the icon to clear the default whitelist group or delete a custom whitelist group.

17.3.4.3.2 Create a database and an account

After you create an RDS instance and configure the whitelist, you need to create a database and an account in the instance. This topic describes how to create a database and an account.

Context

Before you migrate data from the on-premises database to ApsaraDB for RDS, you must create a database and an account for the database. Ensure that the database has the same properties as the on-premises database, and the account of the database has the same permissions as the account of the on-premises database. When assigning database account permissions, follow the least privilege principle. Create accounts by service roles and assign proper read-only and read/write permissions to the accounts. When necessary, you may define database accounts and databases in a fine-grained manner so that each database account can only access data for their own services.

Note:

For database security, set strong passwords for the accounts and change the passwords periodically.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Database Management** > **Accounts**.
4. On the **Accounts** page, click **Create Account**. On the **Create Account** page that appears, configure parameters as prompted.

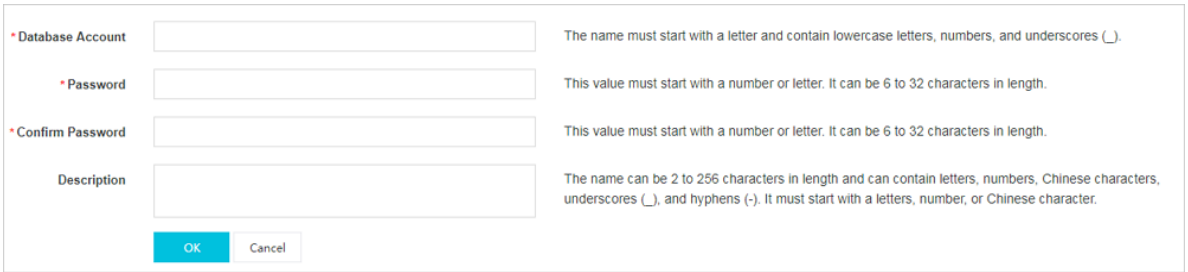
   Table 17-12: Account creation parameters describes the parameter configurations.

   Table 17-12: Account creation parameters

Parameter	Description
Database Account	The name of an account. The naming rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 2 to 16 characters in length.

Parameter	Description
	• Reserved keywords cannot be used. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.
Password	The password for this account. The rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 6 to 32 characters in length.
Confirm Password	The same as the password. The rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 6 to 32 characters in length.
Description	The description of the account. The rules are as follows: • It must start with a letter or number. • It can contain letters, numbers, underscores (_), and hyphens (-). • It can be 2 to 256 characters in length.

Figure 17-10: Create an account



5. Click **OK**.
6. Connect to your RDS instance from the client. For more information, see *Connect to a PostgreSQL instance from a client*.
7. Run the following command to create a database:

   `CREATE DATABASE"databasename"`

   databasename is the name of the database to be created, such as `CREATE DATABASE "mydatabase"`.

17.3.5 Connect to an instance

17.3.5.1 Log on to an instance through DMS

This topic describes how to connect to an RDS instance through Data Management Service (DMS).

Context

On the RDS console, you can connect to an RDS instance through DMS. DMS offers an integrated solution for data management, structure management, access security, BI charts, data trends, data tracking, performance and optimization, and server management. It can manage relational databases (such as MySQL and PostgreSQL) and OLAP databases.

Procedure

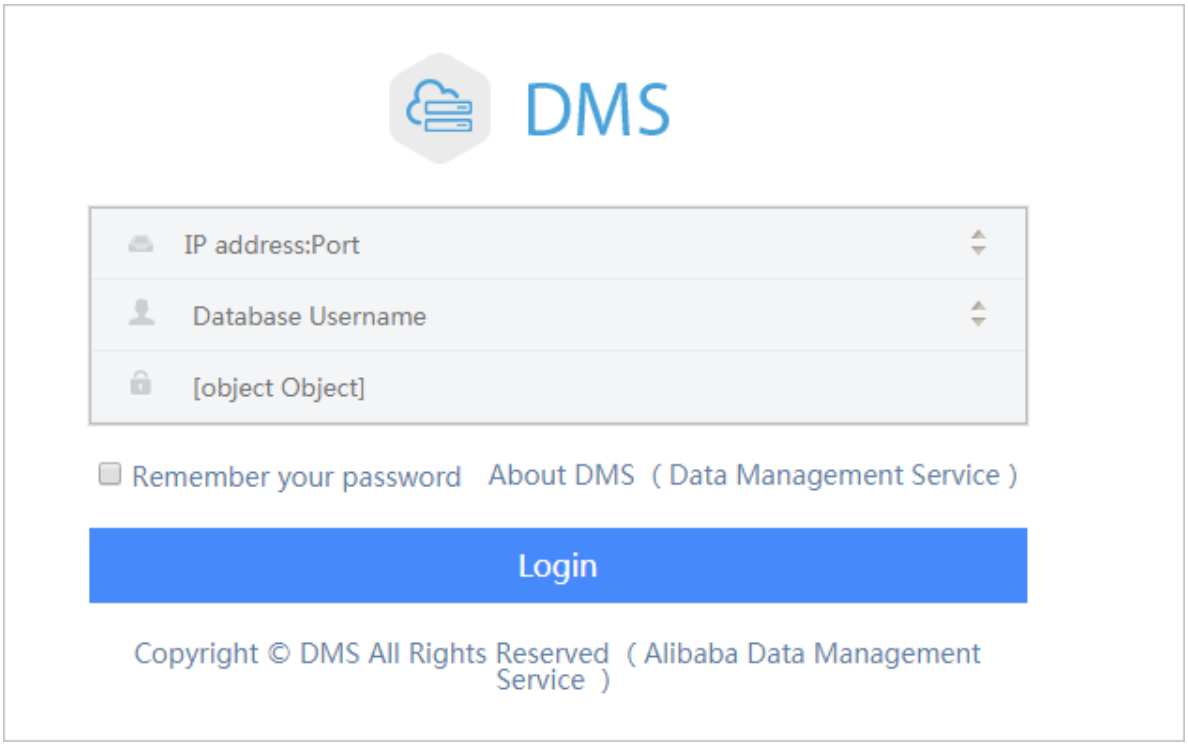
1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. On the **Basic Information** page, click **Log On to DMS**. On the logon page of the DMS console, enter the correct logon information as prompted.

   Table 17-13: DMS logon parameters describes the logon information.

   Table 17-13: DMS logon parameters

Parameter	Description
Network Address : Port	The internal network IP address and port number used to connect to the instance, such as rm-test00000k012.mysql.aliyun-inc.com:3306. You can view the IP address and port number as follows: **a.** *Log on to the RDS console*. **b.** Click the ID of the instance. **c.** In **Internal Network Connection Information** of the **Basic Information** page, view the internal network IP address and port number of the instance.
Username	The account used to connect to the RDS instance (the account you created in the instance). For more information about how to create an account in the MySQL instance, see *Create a standard account* or *Create a premier account*.
Password	The password for the account used to connect to the RDS instance (the password you specified for the account created in the instance).
Database Type	The type of the database to be connected.

Figure 17-11: DMS logon



4. Click **Log On**.

> **Note:**
> If you want the Web browser to remember the password, select **`Remember Password`** and click **Log On**.

17.3.5.2 Connect to a MySQL instance from a client

This topic describes how to connect to an RDS for MySQL instance from the MySQL-Front client.

Prerequisites

- You have installed the MySQL-Front client.
- Your client is deployed in the same VPC as the RDS instance.
- You have added the IP address used to access the RDS instance to the RDS whitelist. For more information about how to configure the whitelist, see *Configure a whitelist*.

Context

RDS for MySQL is fully compatible with the native database service. You can connect to RDS in the same way you connect to an on-premises MySQL server. This topic describes how to connect

to an RDS for MySQL instance through the MySQL-Front client. You can refer to this topic as an example when you connect through other clients.

Procedure

1. Start the MySQL-Front client on your PC.
2. In the **Open Connection** window, click **New**. Configure parameters as prompted in the **Create Account** dialog box that appears.

   Table 17-14: MySQL-Front logon parameters describes the parameter configurations.

   Table 17-14: MySQL-Front logon parameters

Parameter	Description
Name	The name of the database connection task. If this parameter is left blank, the system assumes it is the same as the Host filed by default.
Host	The internal network IP address used to connect to the RDS instance. You can view the IP address and port number as follows: **a.** *Log on to the RDS console*. **b.** Click the ID of the instance. **c.** In **Internal Network Connection Information** of the **Basic Information** page, view the internal network IP address and port number of the instance.
Port	The port of the internal network used to connect to the RDS instance.
User	The account used to connect to the RDS instance (the account you created in the instance).
Password	The password for the account used to connect to the RDS instance (the password you specified for the account created in the instance).

Figure 17-12: Create a connection

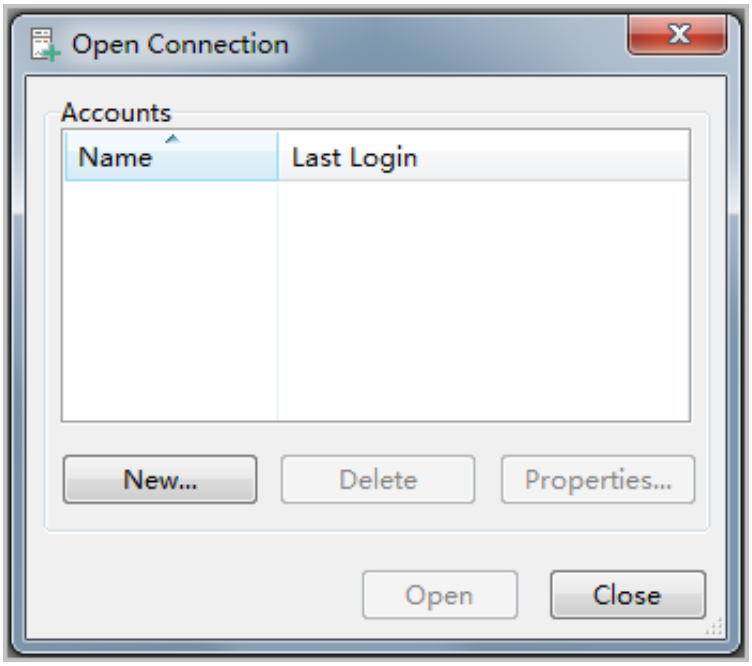
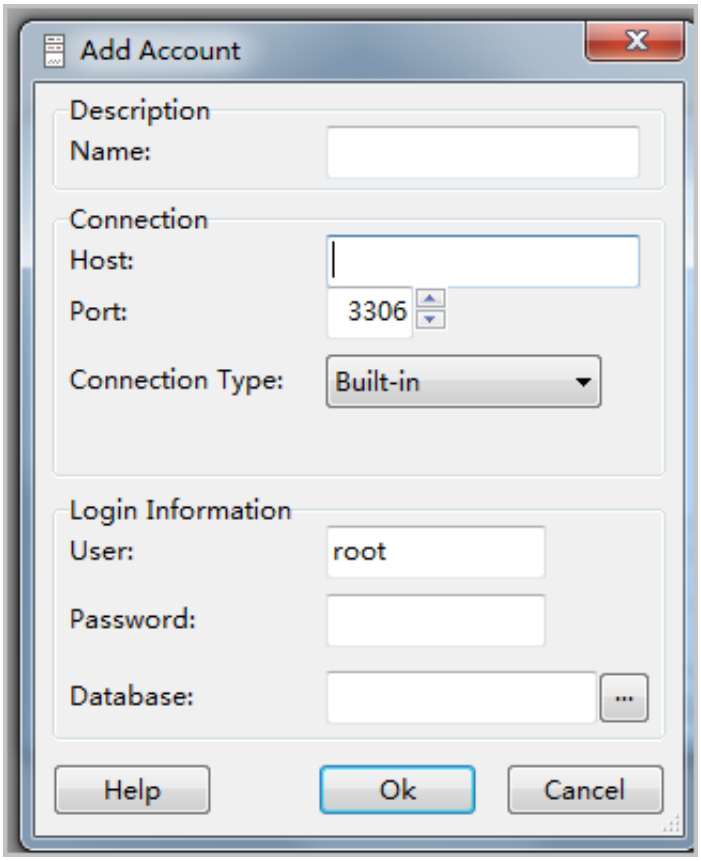


Figure 17-13: Enter connection information

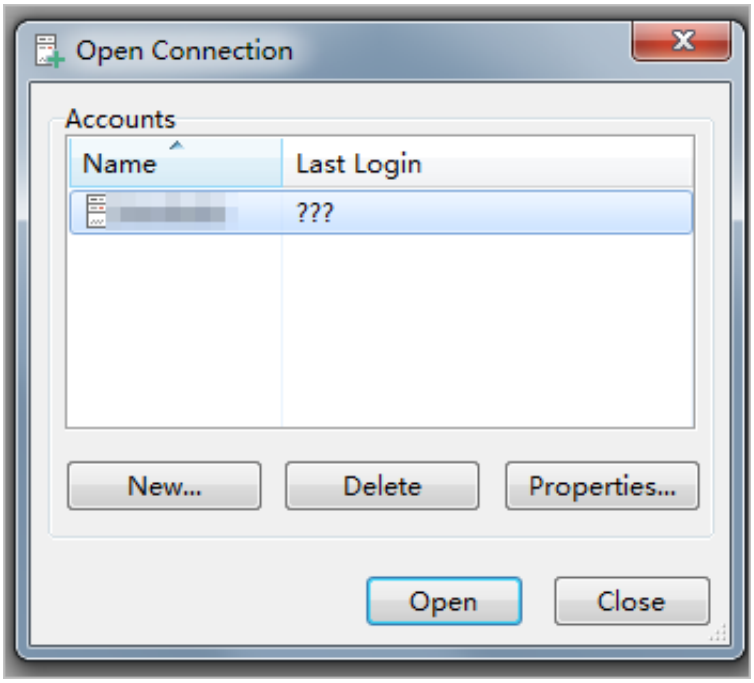


3. Click **OK**.

4. In the **Open Connection** window, select the created connection and click **Open**, as shown in *Figure 17-14: Connect to an instance*.

> **Note:**
> If the connection information is correct, you can connect to the RDS instance successfully.

Figure 17-14: Connect to an instance



17.3.5.3 Connect to a PostgreSQL instance from a client

This topic describes how to connect to an RDS for PostgreSQL instance from a pgAdmind 4 client.

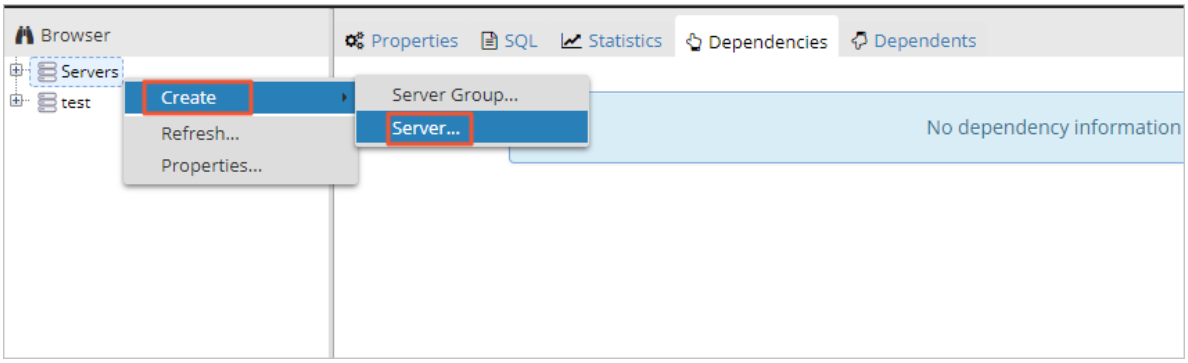
Prerequisites

- You have installed the pgAdmind 4 client.
- Your client is deployed in the same VPC as the RDS instance.
- You have added the IP address used to access the RDS instance to the RDS whitelist. For more information about how to configure the whitelist, see *Configure a whitelist*.

Procedure

1. Open the pgAdmind 4 client on your PC.
2. Right-click **Servers** and choose **Create** > **Server** from the shortcut menu.

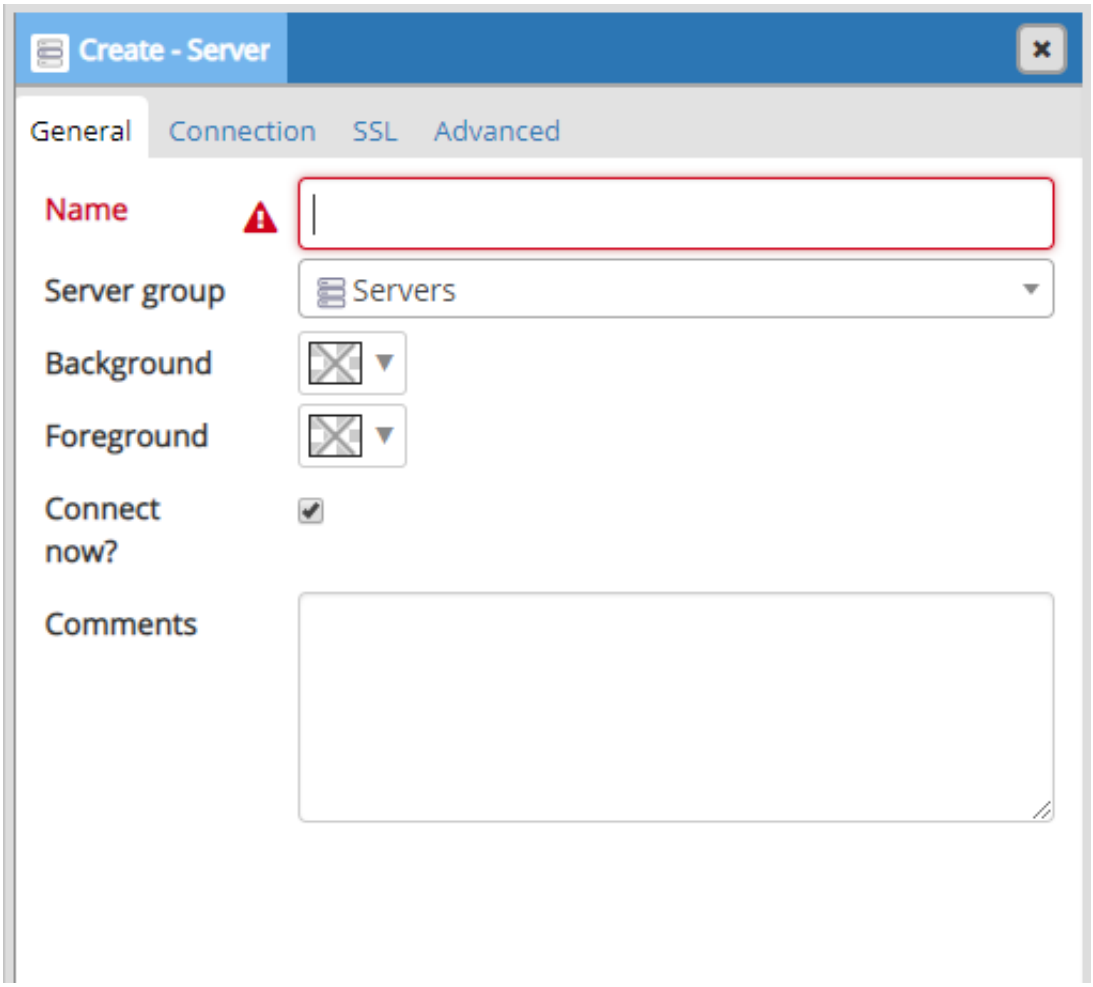
Figure 17-15: Create a server



3. On the **Create - Server** page, click the **General** tab and configure parameters as prompted.

   - **Name**: specifies the name of the server to be created. Select an appropriate name to ease future searches.
   - **Comments**: specifies remarks of the server to be created.

Figure 17-16: Enter a server name



4. Click the **Connection** tab and configure parameters as prompted.

Figure 17-17: Configure instance connection information

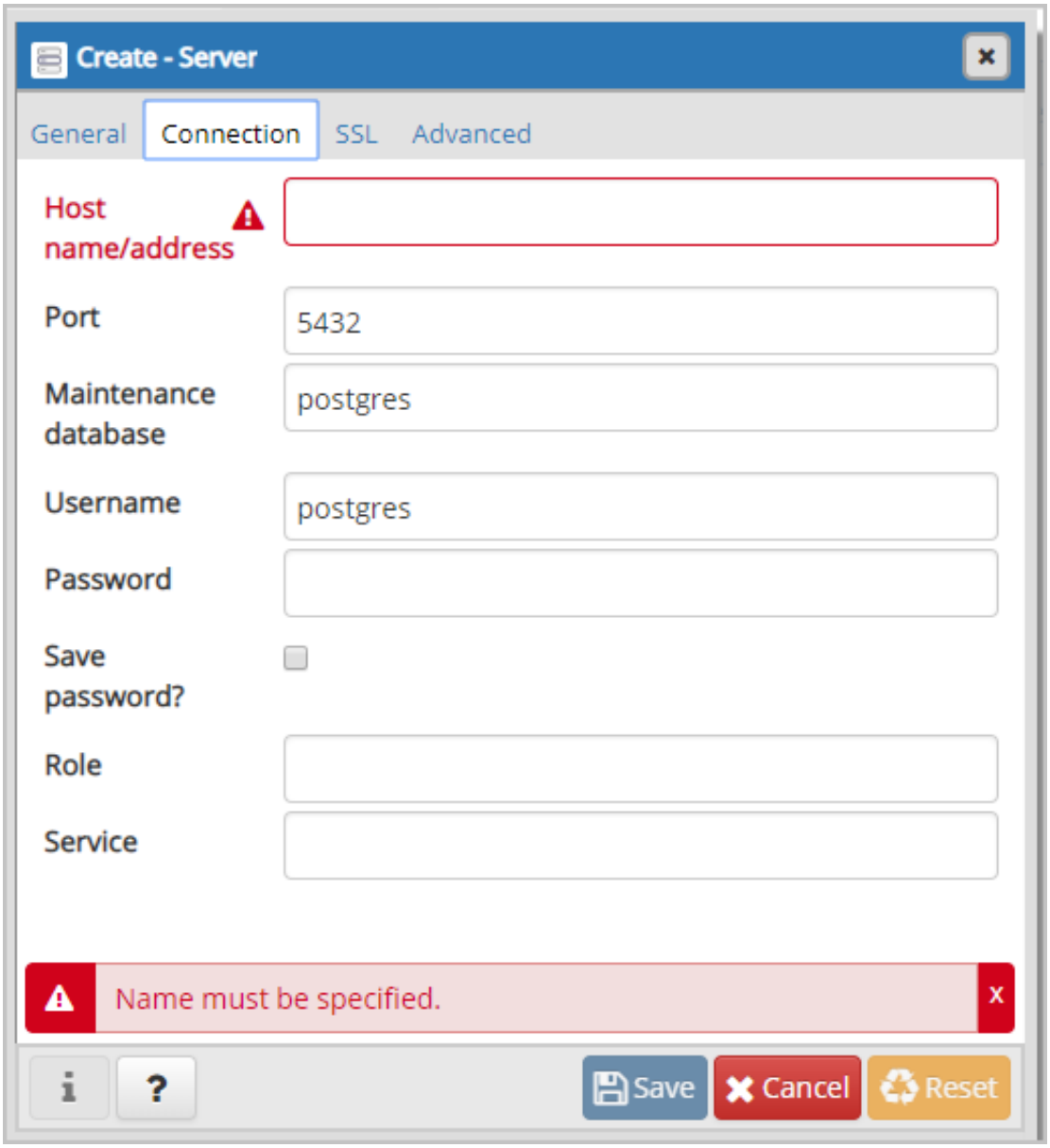


Table 17-15: pgAdmind 4 logon parameters describes the parameter configurations.

Table 17-15: pgAdmind 4 logon parameters

Parameter	Description
Host Name/ Address	The internal network IP address used to connect to the RDS instance. You can view the IP address and port number as follows: **a.** *Log on to the RDS console*. **b.** Click the ID of the instance.

Parameter	Description
	c. In **Internal Network Connection Information** of the **Basic Information** page, view the internal network IP address and port number of the instance.
Port	The port of the internal network used to connect to the RDS instance.
Username	The account used to connect to the RDS instance (the account you created in the instance).
Password	The password for the account used to connect to the RDS instance (the password you specified for the account created in the instance).

5. Click **Save**.
6. If the connection information is correct, choose **Servers** > **Server Name** > **Database** > **postgres**.

17.3.5.4 Connect to a PPAS instance from a client

This topic describes how to connect to an RDS for PPAS instance from a pgAdmind 4 client.

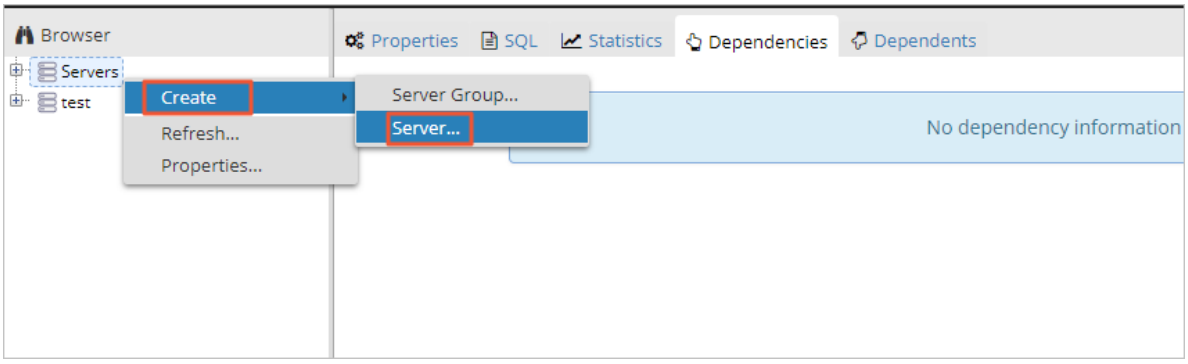
Prerequisites

- You have installed the pgAdmind 4 client.
- Your client is deployed in the same VPC as the RDS instance.
- You have added the IP address used to access the RDS instance to the RDS whitelist. For more information about how to configure the whitelist, see *Configure a whitelist*.

Procedure

1. Open the pgAdmind 4 client on your PC.
2. Right-click **Servers** and choose **Create** > **Server** from the shortcut menu.

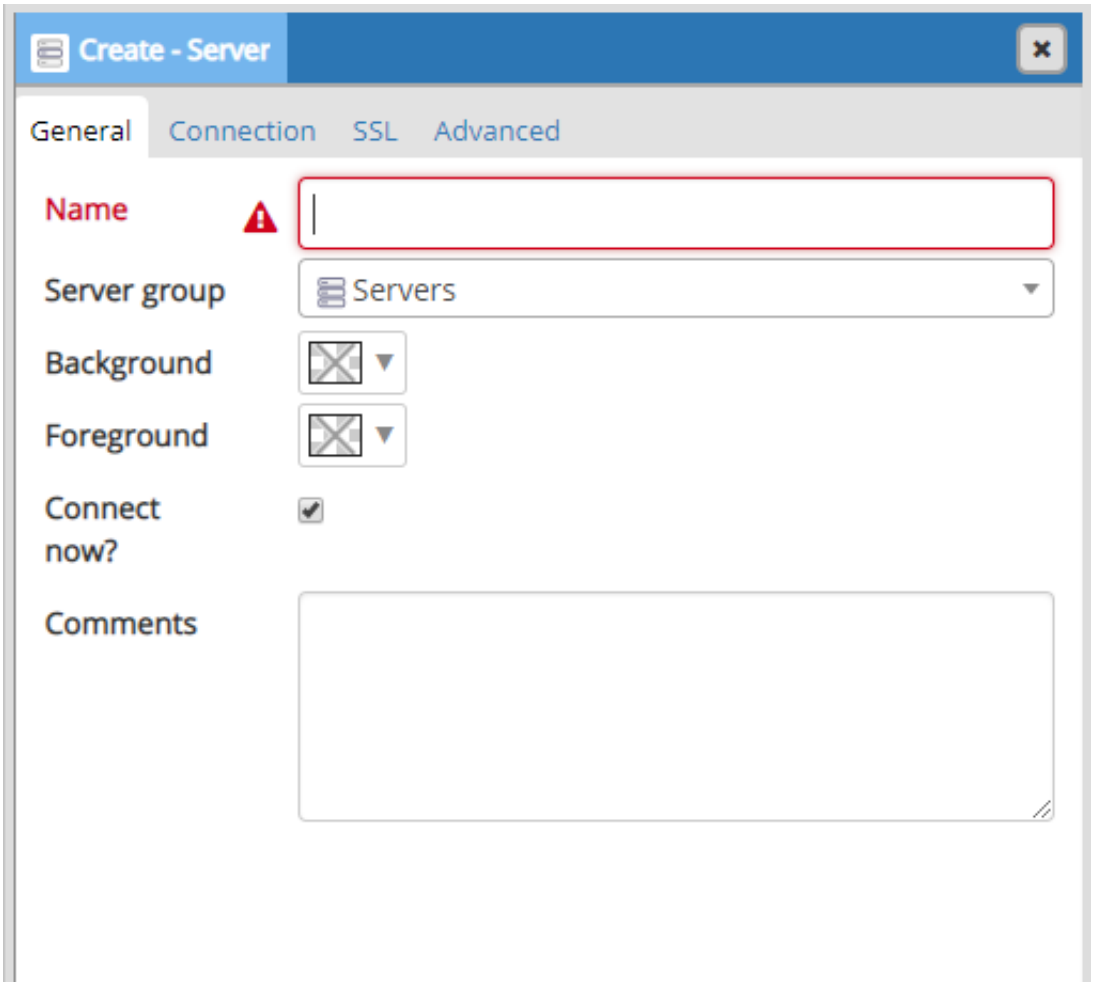
Figure 17-18: Create a server



3. On the **Create - Server** page, click the **General** tab and configure parameters as prompted.

- **Name**: specifies the name of the server to be created. Select an appropriate name to ease future searches.
- **Comments**: specifies remarks of the server to be created.

Figure 17-19: Enter a server name



4. Click the **Connection** tab and configure parameters as prompted.

Figure 17-20: Configure instance connection information

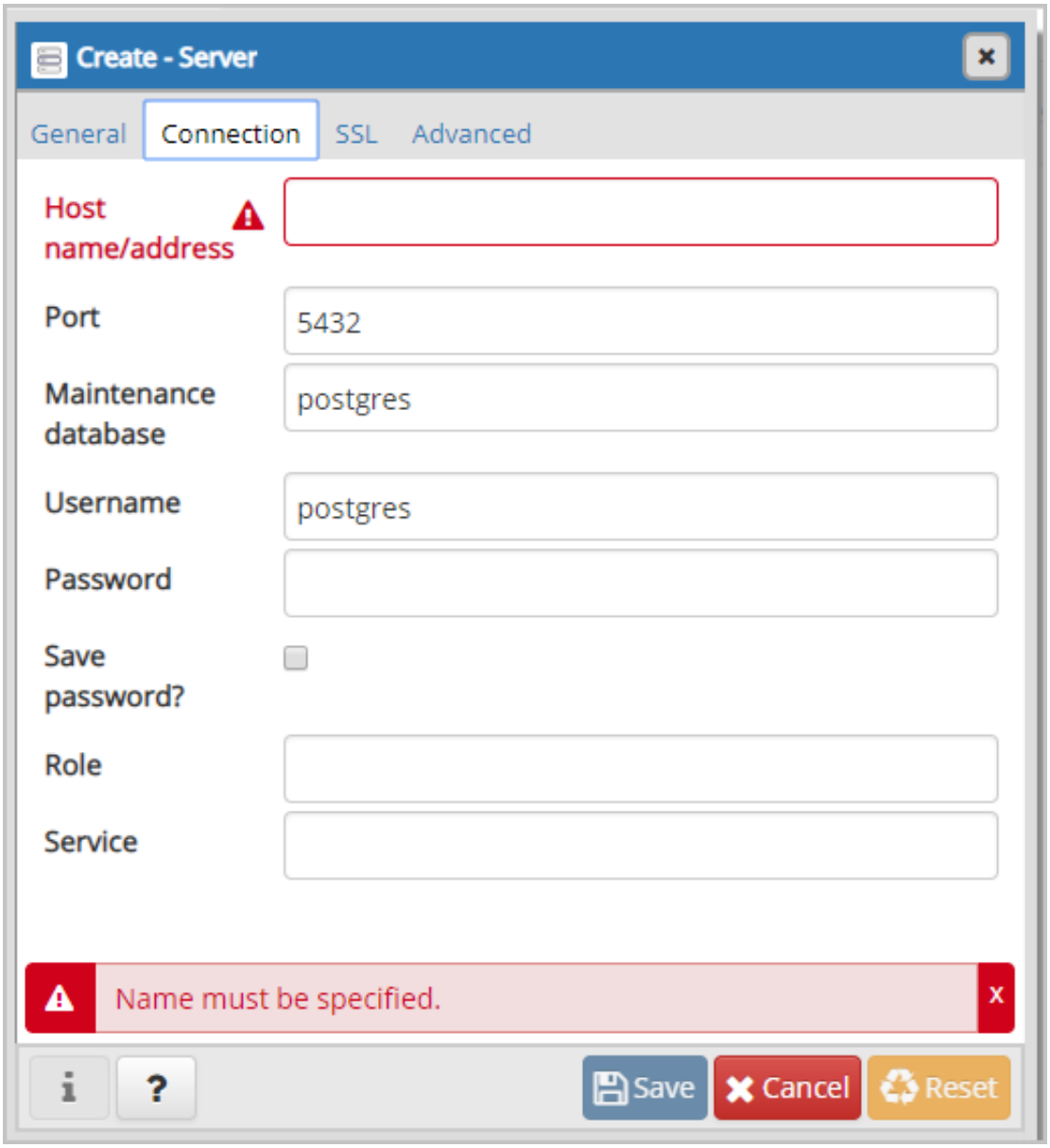


Table 17-16: pgAdmind 4 logon parameters describes the parameter configurations.

Table 17-16: pgAdmind 4 logon parameters

Parameter	Description
Host Name/ Address	The internal network IP address used to connect to the RDS instance. You can view the IP address and port number as follows: **a.** *Log on to the RDS console*. **b.** Click the ID of the instance.

Parameter	Description
	c. In **Internal Network Connection Information** of the **Basic Information** page, view the internal network IP address and port number of the instance.
Port	The port of the internal network used to connect to the RDS instance.
Username	The account used to connect to the RDS instance (the account you created in the instance).
Password	The password for the account used to connect to the RDS instance (the password you specified for the account created in the instance).

5. Click **Save**.
6. If the connection information is correct, choose **Servers** > **Server Name** > **Database** > **postgres**.

17.4 Instances

17.4.1 Create an instance

This topic describes how to create an instance in the RDS console.

Prerequisites

Before you create an RDS instance, you need to apply for an Apsara Stack Management Console account.

Procedure

1. *Log on to the RDS console*.
2. On the **Relational Database Service (RDS)** page, click **Create Instance** in the upper-right corner. On the **Create Instance** page, configure parameters as promoted.

   Table 17-17: Instance creation parameters describes the parameter configurations.

   Table 17-17: Instance creation parameters

Category	Parameter	Description
Basic Configuration	**Department**	The department to which the instance belongs.
	Project	The project to which the instance belongs.
	Region	The region where the instance is located.

Category	Parameter	Description
	Zone	The zone of the instance. Common RDS instances adopt the hot standby architecture. A single zone means that the primary and secondary nodes are in the same zone.
Network Type	**Instance Type**	The type of the instance.
	Network Type	The network types supported by RDS instances : • `Classic Network`: Cloud services in a classic network are not isolated. Unauthorized access to a cloud service is blocked only by the security group or whitelist policy of the service. • `VPC`: Virtual Private Cloud (VPC) helps you build an isolated network environment in Alibaba Cloud. You can customize route tables, IP address ranges, and gateways in a VPC. We recommend that you choose a VPC to improve security. You must create a VPC instance in advance. Alternativ ely, you can change the network type after you create an instance.
Access Mode	**Access Mode**	The access modes supported by RDS instances, including: • `Standard Mode`: ApsaraDB for RDS uses Server Load Balancer to eliminate the impact from HA switching of the database engine on the application layer. This shortens the response time, but slightly increases the probability of transient disconnections. • `Safe Mode`: This mode prevents 90% of transient disconnections. However, the response time is increased by 20% or more, and performance is affected.
Specification Configuration	**Instance Name**	The name of the RDS instance. It is a string of 2 to 256 characters including letters, numbers, and underscores (_). It must start with a letter.
	Database Type	Database types vary according to regions. The available database types are displayed on the page.
	Database Version	The version of the database.
	CPU/ Memory	The specification of the instance, which includes: • Common: for example, 1 core and 1 GB memory.

Category	Parameter	Description
		• Dedicated: The specification is suffixed with "Dedicated." • Dedicated Host: The specification is suffixed with " Dedicated Host." • Financial Version Single Data Center: The specification is suffixed with "Financial Version Single Data Center." Memory size determines the maximum number of connection s and IOPS. The actual values are displayed on the console interface.
	Storage Size	The storage size of the instance, which includes data, system files, binlog files, and transaction files.
Quantity	**Instances**	The number of RDS instances that can be created simultaneously. The maximum number is 20.

3. Click **Create**.

17.4.2 View details

You can view the details of an instance, such as the basic information, internal network connection information, running status, and configurations. This topic describes how to view the details of an instance.

Procedure

1. *Log on to the RDS console*.
2. You can use either of the following ways to go to the instance details page:
   - Click the ID of the instance to go to the **Basic Information** page.
   - In the Actions column corresponding to the instance, click the icon and select **View Details**. The **Basic Information** page for the instance is displayed.

17.4.3 Restart an instance

You can manually restart an instance when the number of connections exceeds the threshold or any performance issue occurs on the instance.

Context

Note:
Restarting an instance will cause service interruptions. Perform this operation only when necessary and make sure that the restart does not affect other services.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. On the **Basic Information** page, click the  icon and select **Restart Instance** in the Actions column corresponding to the instance. In the **Restart Instance** dialog box that appears, click **OK** to restart the instance.

17.4.4 Modify configurations

You can modify the configurations of your instance, such as memory and storage space, if the instance configurations are too high, too low, or the instance is unable to meet application requirements.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. On the **Basic Information** page, click **Change Configuration**.
4. On the **Change Configuration** page, select the **Specifications** and **Storage Size (GB)** for the instance.
5. Click **OK**.

17.4.5 Release an instance

You can manually release an instance when necessary.

Context

Note:

- You can only manually release instances in the running status.
- If read/write splitting is enabled for the primary instance, you must *Disable read/write splitting* first.

Procedure

1. *Log on to the RDS console*.
2. In the Actions column corresponding to the instance, click the  icon and select **Delete Instance**.

3. In the **Delete Instance** dialog box that appears, click **OK**.

17.4.6 Configure parameters

ApsaraDB for RDS allows you to define some instance parameters. For more information about the parameters that can be modified, see **Parameter Settings** in the RDS console.

Context

ApsaraDB for RDS is fully compatible with the native database service. The parameter configurat ion methods for both services are similar. This example uses the RDS console to modify the parameters. You can also use APIs to execute commands to modify the parameters.

Note:

- RDS for PostgreSQL and RDS for PPAS instances do not currently support custom parameters.
- Configure parameters on the **Parameter Settings** page based on the specified **Parameter Range**.
- Modifying some parameters requires you to restart the instance. Go to the **Parameter Settings** page and check **Requires Restart** to determine whether a restart is required. Before you restart the instance, make sure that the instance restart does not affect other services.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Performance Optimization** > **Parameter Settings**.
4. On the **Parameter Settings** page, click the icon and select **Edit** in the Actions column. On the **Change Parameters** page, configure parameters as prompted.

   For more information about parameter setting rules, see *Modifiable MySQL instance parameters*.
5. Click **OK**.

17.4.7 Change ownership

You can change the ownership (department and project) of an instance based on your service requirements.

Procedure

1. *Log on to the RDS console*.
2. In the Actions column corresponding to the instance, click the icon and select **Change Ownership**. In the **Change Ownership** dialog box that appears, configure parameters as prompted.

   Table 17-18: Ownership changing parameters describes the parameter configurations.

   Table 17-18: Ownership changing parameters

Parameter	Description
Instance Name	The name of the instance to which the ownership is to be transferred. The instance name is specified by the system and is unmodifiable. For more information about modifying an instance name, see *Modify an instance name*.
Department	The department to which the instance belongs. Select a department.
Project	The project to which the instance belongs. Select a department.

3. Click **OK**.

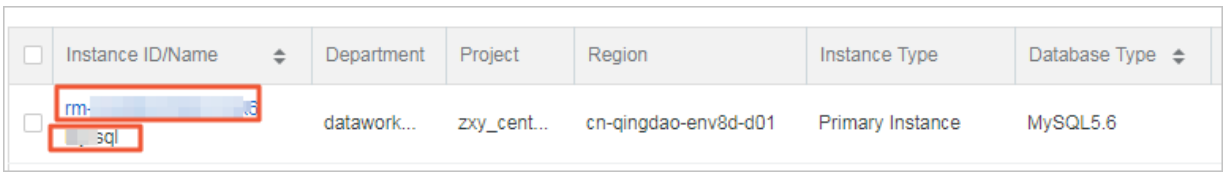
17.4.8 Modify an instance name

You can modify instance names to assist in management.

Context

In the instance list, the **`Instance ID/Name`** column shows instance IDs in the upper part and instance names in the lower part, as shown in *Figure 17-21: Instance ID/Name*. You can modify instance names but cannot modify instance IDs.

Figure 17-21: Instance ID/Name



Procedure

1. *Log on to the RDS console*.
2. In the Actions column corresponding to the instance, click the icon and select **Change Instance Name**. In the **Change Instance Name** dialog box that appears, modify the instance name as prompted.

   Note:

   - The instance name must begin with a letter.
   - The instance name can contain letters, underscores (_), and numbers.
   - The instance name can be 2 to 64 characters in length.
3. Click **OK**.

17.4.9 Typical parameter configuration

17.4.9.1 Modifiable MySQL instance parameters

This topic describes modifiable RDS for MySQL parameters.

Table 17-19: Modifiable MySQL instance parameters lists the modifiable MySQL instance parameters. For more information about the parameters, see the MySQL official documentation at *https://dev.mysql.com/doc/*.

Table 17-19: Modifiable MySQL instance parameters

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
auto_increment_increment	1	1	No	[1-65535]	auto_increment_increment and auto_increment_offset are intended for use with master-to-master replication, and can be used to control the operation of AUTO_INCREMENT columns. Both variables

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					have global and session values, and each can assume an integer value between 1 and 65,535 inclusive. Setting the value of either of these two variables to 0 causes its value to be set to 1 instead. Attempting to set the value of either of these two variables to an integer greater than 65,535 or less than 0 causes its value to be set to 65,535 instead. Attempting to set the value of auto_increment_increment or auto_increment_offset to a noninteger value produces an error, and the actual value of the variable remains unchanged.
auto_increment_offset	1	1	No	[1-65535]	determines the starting point for the AUTO_INCREMENT column value.
back_log	3000	3000	Yes	[0-65535]	The number of outstanding connection requests

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					that MySQL can have.
binlog_cache_size	2097152	128 KB	No	[4096-16777216]	The size of the cache to hold changes to the binary log during a transaction.
binlog_checksum	CRC32	CRC32	Yes	[CRC32|NONE]	The master to write a checksum for each event in the binary log.
Set binlog_row_image	Full	Full	No	[full|minimal]	Binlog save every column or actually required column in binlog images.
binlog_stmt_cache_size	32768	32768	No	[4096-16777216]	The size of the statement cache for updates to non-transactional engines for the binary log.
character_set_server	utf8	utf8	Yes	[utf8|latin1|gbk|utf8mb4]	The server's default character set.
concurrent_insert	1	1	No	[0|1|2]	NEVER-Disables concurrent inserts ; AUTO-(Default) Enables concurrent insert for MyISAM tables that do not have holes; ALWAYS-Enables concurrent inserts for all MyISAM tables, even those that have holes. For a table with a hole, new rows are

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					inserted at the end of the table if it is in use by another thread. Otherwise, MySQL acquires a normal write lock and inserts the row into the hole.
connect_timeout	10	10	No	[1-3600]	The number of seconds that the mysqld server waits for a connect packet before responding with Bad handshake. The default value is 10 seconds as of MySQL 5.1.23 and 5 seconds before that. Increasing the connect_timeout value might help if clients frequently encounter errors of the form Lost connection to MySQL server at 'XXX', system error: errno.
default_storage_engine	InnoDB	TokuDB	Yes	[InnoDB|TokuDB|innodb|tokudb]	The default storage engine for new tables.
default_time_zone	SYSTEM	SYSTEM	Yes	[SYSTEM|-12:00|-11:00|-10:00|-9:00|-8:00|-7:00|-6:00|-5:00|-4:00|-3:00|-2:00|-1:00|+0:00|+1:00|+2:00|+3:00|+4:00|+5:00|+5:30|+6	The default time zone for the database.

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
				:00|+6:30|+7:00|+8:00|+9:00|+10:00|+11:00|+12:00|+13:00]	
default_week_format	0	0	No	[0-7]	The default mode value to use for the WEEK() function.
delayed_insert_limit	100	100	No	[1-4294967295]	After inserting delayed_insert_limit delayed rows, the INSERT DELAYED handler thread checks whether there are any SELECT statements pending. If so, it permits them to execute before continuing to insert delayed rows.
delayed_insert_timeout	300	300	No	[1-3600]	How many seconds an INSERT DELAYED handler thread should wait for INSERT statements before terminating.
delayed_queue_size	1000	1000	No	[1-4294967295]	This is a per-table limit on the number of rows to queue when handling INSERT DELAYED statements. If the queue becomes full, any client that issues an INSERT DELAYED statement waits until

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					there is room in the queue again.
delay_key_write	ON	ON	No	[ON|OFF|ALL]	This option applies only to MyISAM tables. It can have one of the following values to affect handling of the DELAY_KEY_WRITE table option that can be used in CREATE TABLE statements.
div_precision_increment	4	4	No	[0-30]	This variable indicates the number of digits by which to increase the scale of the result of division operations performed with the / operator. The default value is 4. The minimum and maximum values are 0 and 30, respectively. The following example illustrates the effect of increasing the default value.
eq_range_index_dive_limit	10	10	No	[1-200]	This variable indicates the number of equality ranges in an equality comparison condition when the optimizer should

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					switch from using index dives to index statistics in estimating the number of qualifying rows.
explicit_defaults_for_timestamp	false	false	Yes	true|false	As indicated by the warning, to turn off the nonstandard behaviors, enable the explicit_defaults_for_timestamp system variable at server startup.
ft_min_word_len	4	4	Yes	[1-3600]	The minimum length of the word to be included in a FULLTEXT index.
ft_query_expansion_limit	20	20	Yes	[0-1000]	The number of top matches to use for full-text searches performed using WITH QUERY EXPANSION.
group_concat_max_len	1024	1024	No	[4-18446744073709547520]	The maximum permitted result length in bytes for the GROUP_CONCAT() function. The default is 1024, Unit :Byte.
innodb_adaptive_hash_index	ON	ON	No	[ON|OFF]	It may be desirable, depending on your workload, to dynamically enable or disable adaptive

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					hash indexing to improve query performance. The size in bytes of a memory pool InnoDB uses to store data dictionary information and other internal data structures. The more tables you have in your application, the more memory you allocate here. If InnoDB runs out of memory in this pool, it starts to allocate memory from the operating system and writes warning messages to the MySQL error log. The default value is 8MB.
innodb_additional_mem_pool_size	2097152	2097152	Yes	[2097152-104857600]	The locking mode to use for generating auto-increment values. The permissible values are 0, 1, or 2.
innodb_autoinc_lock_mode	1	1	Yes	[0|1|2]	The number of threads that can enter InnoDB concurrently is determined by the innodb_thread_concurrency variable.

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
innodb_concurrency_tickets	5000	5000	No	[1-4294967295]	The number of threads that can enter InnoDB concurrently is determined by the innodb_thread_concurrency variable.
innodb_ft_max_token_size	84	84	Yes	[10-84]	Maximum length of words that are stored in an InnoDB FULLTEXT index.
innodb_ft_min_token_size	3	3	Yes	[0-16]	Minimum length of words that are stored in an InnoDB FULLTEXT index.
innodb_large_prefix	OFF	OFF	No	[ON|OFF]	Enable this option to allow index key prefixes longer than 767 bytes (up to 3072 bytes), for InnoDB tables that use the DYNAMIC and COMPRESSED row formats.
innodb_lock_wait_timeout	50	50	No	[1-1073741824]	The timeout in seconds an InnoDB transaction may wait for a row lock before giving up. The default value is 50 seconds. Unit: Second.
innodb_max_dirty_pages_pct	75	75	No	[50-90]	This is an integer in the range from 0 to 100. The default value is 90 for the

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					built-in InnoDB, 75 for InnoDB Plugin . The main thread in InnoDB tries to write pages from the buffer pool so that the percentage of dirty (not yet written) pages will not exceed this value.
innodb_old_blocks_pct	37	37	No	[5-95]	(InnoDB Plugin only) Specifies the approximate percentage of the InnoDB buffer pool used for the old block sublist. The range of values is 5 to 95. The default value is 37 (that is, 3/8 of the pool).
innodb_old_blocks_time	1000	1000	No	[0-1024]	(InnoDB Plugin only) Specifies how long in millisecon ds (ms) a block inserted into the old sublist must stay there after its first access before it can be moved to the new sublist. The default value is 0 : A block inserted into the old sublist moves immediatel y to the new sublist the first time it is accessed, no matter how soon after

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					insertion the access occurs. If the value is greater than 0, blocks remain in the old sublist until an access occurs at least that many ms after the first access . Unit: ms.
innodb_online_alter_log_max_size	134217728	134217728	No	[134217728-2147483647]	Specifies an upper limit on the size of the temporary log files used during online DDL operations for InnoDB tables.
innodb_open_files	3000	3000	Yes	[1-8192]	This variable is relevant only if you use multiple InnoDB tablespaces. It specifies the maximum number of .ibd files that MySQL can keep open at one time. The minimum value is 10. The default value is 300.
innodb_print_all_deadlocks	OFF	OFF	No	[OFF|ON]	When this option is enabled, information about all deadlocks in InnoDB user transactions is recorded in the mysqld error log.
innodb_purge_batch_size	300	300	Yes	[1-5000]	The granularity of changes, expressed in units of redo

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					log records, that trigger a purge operation, flushing the changed buffer pool blocks to disk.
innodb_purge_threads	1	1	Yes	[1-32]	The number of background threads devoted to the InnoDB purge operation.
innodb_read_ahead_threshold	56	56	No	[0-64]	(InnoDB Plugin only) Controls the sensitivity of linear read-ahead that InnoDB uses to prefetch pages into the buffer pool. If InnoDB reads at least innodb_read_ahead_threshold pages sequentially from an extent (64 pages), it initiates an asynchronous read for the entire following extent.
innodb_read_io_threads	4	4	Yes	[1-64]	(InnoDB Plugin only) The number of I/O threads for read operations in InnoDB. The default value is 4.
innodb_rollback_on_timeout	OFF	OFF	Yes	[OFF|ON]	InnoDB rolls back only the last statement on a transaction timeout by default. If --innodb_rol

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					lback_on_timeout is specified, a transaction timeout causes InnoDB to abort and roll back the entire transaction (the same behavior as in MySQL 4.1). This variable was added in MySQL 5.1.15
innodb_stats_method	nulls_equal	nulls_equal	No	[nulls_equal| nulls_unequal| nulls_ignored]	How the server treats NULL values when collecting statistics about the distribution of index values for InnoDB tables. This variable has three possible values, nulls_equal, nulls_unequal, and nulls_ignored. For nulls_equal, all NULL index values are considered equal and form a single value group that has a size equal to the number of NULL values. For nulls_unequal, NULL values are considered unequal, and each NULL forms a distinct value group of size 1. For nulls_ignored, NULL values are ignored.

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
innodb_stats_on_metadata	OFF	OFF	No	[ON|OFF]	When this variable is enabled (which is the default, as before the variable was created), InnoDB updates statistics during metadata statements such as SHOW TABLE STATUS or SHOW INDEX, or when accessing the INFORMATION_SCHEMA tables TABLES or STATISTICS. (These updates are similar to what happens for ANALYZE TABLE.) When disabled, InnoDB does not update statistics during these operations. Disabling this variable can improve access speed for schemas that have a large number of tables or indexes. It can also improve the stability of execution plans for queries that involve InnoDB tables.

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
innodb_stats_sample_pages	8	8	No	[1-4294967296]	(InnoDB Plugin only) The number of index pages to sample for index distribution statistics such as are calculated by ANALYZE TABLE. The default value is 8.
innodb_strict_mode	OFF	OFF	No	[ON|OFF]	(InnoDB Plugin only) Whether InnoDB returns errors rather than warnings for certain conditions. This is analogous to strict SQL mode. The default value is OFF. See InnoDB Strict Mode for a list of the conditions that are affected.
innodb_table_locks	ON	ON	No	[ON|OFF]	If autocommit = 0, InnoDB honors LOCK TABLES; MySQL does not return from LOCK TABLES ... WRITE until all other threads have released all their locks to the table. The default value of innodb_table_locks is 1, which means that LOCK TABLES causes InnoDB to lock a

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					table internally if autocommit = 0.
innodb_thread_concurrency	0	0	No	[0-128]	InnoDB tries to keep the number of operating system threads concurrently inside InnoDB less than or equal to the limit given by this variable. Once the number of threads reaches this limit, additional threads are placed into a wait state within a FIFO queue for execution.
innodb_thread_sleep_delay	10000	10000	No	[1-3600000]	How long InnoDB threads sleep before joining the InnoDB queue, in microseconds. The default value is 10,000. A value of 0 disables sleep. Unit: ms
innodb_write_io_threads	4	4	Yes	[1-64]	(InnoDB Plugin only) The number of I/O threads for write operations in InnoDB. The default value is 4.
interactive_timeout	7200	7200	No	[10-86400]	The number of seconds the server waits for activity on an interactive connection before closing it. An

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					interactive client is defined as a client that uses the CLIENT_INTERACTIVE option to mysql_real_connect(). Unit: second.
key_cache_age_threshold	300	300	No	[100-4294967295]	This value controls the demotion of buffers from the hot sublist of a key cache to the warm sublist. Lower values cause demotion to happen more quickly. The minimum value is 100. The default value is 300. Unit: Second.
key_cache_block_size	1024	1024	No	[512-16384]	The size in bytes of blocks in the key cache. The default value is 1024. Unit: Byte.
key_cache_division_limit	100	100	No	[1-100]	The division point between the hot and warm sublists of the key cache buffer list. The value is the percentage of the buffer list to use for the warm sublist. Permissible values range from 1 to 100. The default value is 100.

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
log_queries_not_using_indexes	OFF	OFF	No	[ON|OFF]	If a query takes longer than this many seconds, the server increments the Slow_queries status variable. If the slow query log is enabled, the query is logged to the slow query log file; Unit: Second.
long_query_time	1	1	No	[0.03-10]	If a query takes longer than this many seconds, the server increments the Slow_queries status variable. If the slow query log is enabled, the query is logged to the slow query log file;Unit: Second.
loose_max_statement_time	0	0	No	[0-4294967295]	statement be interrupted if the executing time exceeds this value.
loose_rds_indexstat	OFF	OFF	No	[ON|OFF]	If ON, start to collect index information.
loose_rds_max_tmp_disk_space	10737418240	10737418240	No	[10737418240-10737418240]	RDS maximum temp disk space.
loose_rds_tablestat	OFF	OFF	No	[ON|OFF]	RDS table statistics.
loose_rds_threads_running_high_watermark	50000	50000	No	[0-50000]	Max concurrency allowed for SELECT .

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
loose_tokudb_buffer_pool_ratio	0	0	Yes	[0-100]	TokuDB buffer pool size ratio.
low_priority_updates	0	0	No	[0|1]	If set to 1, all INSERT, UPDATE, DELETE, and LOCK TABLE WRITE statements wait until there is no pending SELECT or LOCK TABLE READ on the affected table. This affects only storage engines that use only table-level locking (such as MyISAM, MEMORY, and MERGE). This variable previously was named sql_low_priority_updates.
max_allowed_packet	1073741824	1024M	No	[16384-1073741824]	The maximum size of one packet or any generated/ intermediate string. Unit: Byte.
max_connect_errors	100	100	No	[1-4294967295]	If more than this many successive connection requests from a host are interrupted without a successful connection, the server blocks that host from further connections. You can unblock blocked

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					hosts by flushing the host cache.
max_length_for_sort_data	1024	1024	No	[0-838860]	The cutoff on the size of index values that determines which filesort algorithm to use.
max_prepared_stmt_count	16382	16382	No	[0-1048576]	This variable limits the total number of prepared statements in the server.
max_write_lock_count	102400	102400	No	[1-102400]	After this many write locks, permit some pending read lock requests to be processed in between.
myisam_sort_buffer_size	262144	262144	No	[262144-16777216]	The size of the buffer that is allocated when sorting MyISAM indexes during a REPAIR TABLE or when creating indexes with CREATE INDEX or ALTER TABLE.
net_read_timeout	30	30	No	[1-31536000]	The number of seconds to wait for more data from a connection before aborting the read.
net_retry_count	10	10	No	[1-4294967295]	If a read or write on a communication port is interrupted, retry this many

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					times before giving up.
net_write_timeout	60	60	No	[1-31536000]	The number of seconds to wait for a block to be written to a connection before aborting the write.
open_files_limit	65535	65535	Yes	[4000-65535]	The number of files that the operating system permits mysqld to open. The value of this variable at runtime is the real value permitted by the system and might be different from the value you specify at server startup. The value is 0 on systems where MySQL cannot change the number of open files.
performance_schema	OFF	OFF	Yes	[ON|OFF]	Enable performance_schema or not.
query_alloc_block_size	8192	8192	No	[1024-16384]	The allocation size of memory blocks that are allocated for objects created during statement parsing and execution. Unit: Byte.
query_cache_limit	1048576	1048576	No	[1-1048576]	Do not cache results that are larger than

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					this number of bytes . The default value is 1MB.
query_cache_size	3145728	3145728	No	[0-104857600]	The amount of memory allocated for caching query results. The default value is 0, which disables the query cache. The permissible values are multiples of 1024; other values are rounded down to the nearest multiple. Unit: Byte.
query_cache_type	0	0	Yes	[0|1|2]	Set the query cache type. Setting the GLOBAL value sets the type for all clients that connect thereafter. Individual clients can set the SESSION value to affect their own use of the query cache. Possible values are shown in the following table. • 0: Do not cache results in or retrieve results from the query cache. Note that this does not deallocate the query cache buffer. To do

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					that, you should set query_cache_size to 0. • 1: Cache all cacheable query results except for those that begin with SELECT SQL_NO_CACHE. • 2: Cache results.
query_cache_wlock_invalidate	OFF	OFF	No	[ON|OFF]	Normally, when one client acquires a WRITE lock on a MyISAM table, other clients are not blocked from issuing statements that read from the table if the query results are present in the query cache. Setting this variable to 1 causes acquisition of a WRITE lock for a table to invalidate any queries in the query cache that refer to the table. This forces other clients that attempt to access the table to wait while the lock is in effect.
query_prealloc_size	8192	8192	No	[8192-1048576]	The size of the persistent buffer used for statement parsing and

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
					execution. This buffer is not freed between statements. If you are running complex queries, a larger query_prealloc_size value might be helpful in improving performance, because it can reduce the need for the server to perform memory allocation during query execution operations. Unit: Byte.
rds_reset_all_filter	0	0	No	[0|1]	rds_reset_all_filter=1 ,means reset the rule of filter.
slow_launch_time	2	2	No	[1-1024]	If creating a thread takes longer than this many seconds, the server increments the Slow_launch_threads status variable.
sql_mode	\s	\s	No	(Support space and |REAL_AS_FLOAT|PIPES_AS_CONCAT|ANSI_QUOTES|IGNORE_SPACE|ONLY_FULL_GROUP_BY|NO_UNSIGNE	Modes define what SQL syntax MySQL should support and what kind of data validation checks it should perform.

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
				D_SUBTRACTION|NO_DIR_IN_CREATE|POSTGRESQL|ORACLE|MSSQL|DB2|MAXDB|NO_KEY_OPTIONS|NO_TABLE_OPTIONS|NO_FIELD_OPTIONS|MYSQL323|MYSQL40|ANSI|NO_AUTO_VALUE_ON_ZERO|NO_BACKSLASH_ESCAPES|STRICT_TRANS_TABLES|STRICT_ALL_TABLES|NO_ZERO_IN_DATE|NO_ZERO_DATE|ALLOW_INVALID_DATES|ERROR_FOR_DIVISION_BY_ZERO|TRADITIONAL|HIGH_NOT_PRECEDENCE|NO_ENGINE_SUBSTITUTION|PAD_CHAR_TO_FULL_LENGTH)(,REAL_AS_FLOAT|,PIPES_AS_C	

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
				ONCAT|, ANSI_QUOTES |,IGNORE_SPA CE|,ONLY_FULL_ GROUP_BY|, NO_UNSIGNE D_SUBTRACT ION|,NO_DIR_IN_ CREATE|, POSTGRESQL|, ORACLE|,MSSQL |,DB2|,MAXDB|, NO_KEY_OPT IONS|, NO_TABLE_O PTIONS|, NO_FIELD_O PTIONS|, MYSQL323|, MYSQL40|,ANSI |,NO_AUTO_VA LUE_ON_ZERO |,NO_BACKSLA SH_ESCAPES |,STRICT_TRA NS_TABLES|, STRICT_ALL _TABLES|, NO_ZERO_IN _DATE|, NO_ZERO_DATE |,ALLOW_INVA LID_DATES|, ERROR_FOR_ DIVISION_B Y_ZERO|, TRADITIONAL |,HIGH_NOT_P RECEDENCE|, NO_ENGINE_	

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
				SUBSTITUTION|,PAD_CHAR_TO_FULL_LENGTH)*	
table_definition_cache	512	512	No	[400-80480]	The number of table definitions (from .frm files) that can be stored in the definition cache. If you use a large number of tables, you can create a large table definition cache to speed up opening of tables. The table definition cache takes less space and does not use file descriptors, unlike the normal table cache. The minimum and default values are both 400.
table_open_cache	2000	2000	No	[1-524288]	The stack size of each thread.
thread_stack	262144	262144	Yes	[131072-18446744073709551615]	The stack size of each thread.
tmp_table_size	2097152	2097152	No	[262144-67108864]	The maximum size of internal in-memory temporary tables.
transaction_isolation	READ-COMMITTED	READ-COMMITTED	Yes	[READ-COMMITTED|REPEATABLE-READ]	The default transaction isolation level.

Parameter	Default value	Running parameter value	Restart required	Parameter range	Parameter description
wait_timeout	86400	86400	No	[60-259200]	The number of seconds the server waits for activity on a noninteractive connection before closing it.

17.4.9.2 Best practices for MySQL instance parameter optimization

17.4.9.2.1 Overview

You can optimize MySQL parameters for RDS instances. This topic describes the best practices for modifiable and unmodifiable MySQL parameters. It also describes how to optimize modifiable parameters to improve instance performance.

17.4.9.2.2 Unmodifiable MySQL instance parameters

This topic describes the unmodifiable RDS for MySQL parameters.

Different types of RDS for MySQL instances have different maximum number of connections and memory sizes. Therefore, parameters related to the instance type, such as connections and memory, are restricted and cannot be modified. You can resolve connection or memory bottlenecks through the following methods:

- Memory bottleneck: An out of memory (OOM) error occurs in your instance, which results in a primary/secondary failover.
- Connection bottleneck: If applications cannot establish new connections to the database, you need to optimize the applications or slow SQL statements, or upgrade the instance type.

For the sake of data security of the primary and secondary instances, the following parameters related to data security are unmodifiable: innodb_flush_log_at_trx_commit, sync_binlog, gtid_mode, semi_sync, and binlog_format.

17.4.9.2.3 Modifiable MySQL instance parameters

This topic describes modifiable RDS for MySQL parameters.

Except for the unmodifiable parameters described in *Unmodifiable MySQL instance parameters*, most of the parameters of your RDS for MySQL have been optimized by DBA and source code

teams. This helps you run your database without the need to adjust any parameters. For more information about the modifiable RDS for MySQL instance parameters, see *Modifiable MySQL instance parameters*. These parameters are applicable in most scenarios. You need to adjust some parameters in some special cases. For example:

- If you use the TokuDB storage engine, you need to use tokudb_buffer_pool_ratio to adjust the percentage of the memory available for the engine.
- If your applications require a relatively long lock timeout period, you need to adjust innodb_lock_wait_timeout.

17.4.9.2.4 How to configure parameters

This topic describes how to configure important parameters in the RDS for MySQL console. If these parameters are incorrectly configured, your instances may encounter performance problems or applications may report errors.

open_files_limit

Function: This parameter controls the number of file handles that can be simultaneously enabled by each MySQL instance.

Cause: Opening database tables consumes file handles allocated to each instance. RDS for MySQL sets open_files_limit to 8192 when initializing an instance. When the number of opened tables exceeds this value, errors are returned for all database requests.

Note:

Access to MyISAM engine tables consumes file descriptors. The InnoDB storage engine uses table_open_cache to manage opened tables.

Symptom: If this parameter is set to an excessively small value, applications may report the following error:

```
[ERROR] /mysqld: Can't open file: './mysql/user.frm' (errno: 24 -Too
many open files);
```

Suggestion: Increase the value of open_files_limit. Currently, RDS for MySQL allows you to set the maximum value to 65535 for this parameter. We also recommend that you replace the MyISAM storage engine with the InnoDB storage engine.

back_log

Function: RDS for MySQL creates a thread for every connection request that it processes. If front-end applications initiate too many transient connection requests to the database when a new

thread is created, RDS for MySQL uses back_log to restrict the number of queued connection requests. RDS for MySQL denies new connection requests when the number of connection requests in the queue exceeds the value of back_log. If you want MySQL to process a large number of transient connection requests, increase the value of back_log.

Symptom: If this parameter is set to an excessively small value, applications may report the following error:

```
SQLSTATE[HY000] [2002] Connection timed out;
```

Suggestion: Increase the value of back_log. The default value of this parameter has been increased from 50 to 3000.

Note:

You must restart your instances after you change the parameter value.

innodb_autoinc_lock_mode

Function: In RDS for MySQL 5.1.22 and later versions, innodb_autoinc_lock_mode is introduced to InnoDB to control auto-increment locks. This parameter can be set to 0, 1, or 2. The default value is 1 in RDS for MySQL. This indicates that InnoDB uses the lightweight mutex lock to obtain auto-increment locks in place of table-level locks. However, the load data command (including INSERT … SELECT statement and REPLACE … SELECT statement) uses auto-increment table locks. A deadlock may occur when multiple applications use this command to load data at the same time.

Symptom: A deadlock occurs when multiple applications use the load data command (including INSERT … SELECT statement and REPLACE … SELECT statement) to load data at the same time. The following error is reported:

```
RECORD LOCKS space id xx page no xx n bits xx index PRIMARY of table
 xx.xx trx id xxx lock_mode X insert intention waiting. TABLE LOCK
table xxx.xxx trx id xxxx lock mode AUTO-INC waiting;
```

Suggestion: We recommend that you change the value of innodb_autoinc_lock_mode to 2 to enable the use of the lightweight mutex lock (only in row mode) for all INSERT statements. This avoids auto_inc deadlocks and greatly improves the performance of the INSERT ... SELECT statement.

Note:

If you set the parameter value to 2, you must set the format of binlog to row.

query_cache_size

Function: This parameter controls the memory size of the MySQL query cache. If the query cache is enabled, MySQL locks the query cache when it executes a query. Then MySQL determines whether the query cache contains the queried data. If so, MySQL returns results directly. Otherwise, MySQL proceeds to perform other operations such as engine query. The INSERT, UPDATE, and DELETE statements can invalidate the query cache and any changes made to schemas and indexes. The cost of maintaining the invalid query cache is relatively high, which puts a lot of pressure on MySQL. The query cache helps improve instance performance when the database is not frequently updated. However, when data is frequently written to several tables in the database, the query cache lock results in frequent lock conflicts. The write and read operations of a specific table have to wait for the query cache lock to unlock. This reduces the query efficiency of the SELECT statement.

Symptom: The database goes through several different statuses, which are checking query cache for query, waiting for query cache lock, and storing result in query cache.

Suggestion: ApsaraDB for RDS disables the query cache by default. If you have enabled query cache for your instances, you can disable it when the preceding problem occurs. However, you can enable the query cache to solve database performance problems in some cases.

net_write_timeout

Function: This parameter determines the timeout period a block has to wait before it is sent to a client.

Symptom: If the parameter is set to an excessively small value, the client may report the following error:

```
the last packet successfully received from the server was milliseconds
 ago, the last packet sent successfully to the server was milliseconds
 ago.
```

Suggestion: The default value is 60 seconds in RDS for MySQL. A small value of net_write_timeout may result in frequent disconnections in poor network conditions or it takes a long time for the client to process each block. In cases such as these, we recommend that you increase the value of this parameter.

tmp_table_size

Function: This parameter determines the maximum size of the internal temporary memory table. It is assigned to each thread. (The minimum value of tmp_table_size and max_heap_table_size

decides the actual value.) If the size of a temporary memory table exceeds the value of this parameter, MySQL automatically converts the table to a disk-based MyISAM table. Avoid using temporary tables when you optimize query statements. If you need to use a temporary table, make sure that the temporary table is in the memory.

Symptom: If you use a temporary table for complex SQL statements that contain GROUP BY or DISTINCT clauses, which cannot be optimized through indexes, SQL execution takes a longer time.

Suggestion: If an application involves many GROUP BY or DISTINCT clauses and the database has enough memory, you can increase the values of tmp_table_size and max_heap_table_size to improve query performance.

17.4.9.2.5 New MySQL parameters

This topic describes the new parameters of RDS for MySQL.

oose_rds_max_tmp_disk_space

Function: This parameter controls the temporary file size available for MySQL. The default value is 10 GB in RDS for MySQL.

Symptom: If the temporary file size exceeds the limit indicated by this parameter, applications may report the following error:

```
The table ‘/home/mysql/dataxxx/tmp/#sql_2db3_1’ is full.
```

Suggestion: Evaluate whether you can optimize the SQL statements that cause additional temporary files through indexing or other means. If your instance has enough space, you can increase the value of this parameter to guarantee normal execution of SQL statements.

Note:

You must restart your instances after you change the value of this parameter.

loose_tokudb_buffer_pool_ratio

Function: This parameter controls the buffer size available for the TokuDB storage engine. For example, if you set innodb_buffer_pool_size to 1000 MB and tokudb_buffer_pool_ratio to 50 (which indicates 50% of the buffer size), the TokuDB storage engine can use up to 500 MB of the buffer space.

Suggestion: The default value is 0 in RDS for MySQL. If you use the TokuDB storage engine in your RDS for MySQL instance, we recommend that you increase the value of this parameter to improve the access performance of the TokuDB engine table.

> **Note:**
> You must restart your instances after you change the value of this parameter.

loose_max_statement_time

Function: This parameter controls the maximum query time in MySQL.

Symptom: The query time is not limited by default. If the query time exceeds the limit indicated by this parameter, the query fails as follows:

```
ERROR 3006 (HY000): Query execution was interrupted, max_statem
ent_time exceeded
```

Suggestion: Modify this parameter if you want to control the SQL execution time (in milliseconds) for your database.

loose_rds_threads_running_high_watermark

Function: This parameter controls the maximum number of concurrent MySQL queries. For example, if you set rds_threads_running_high_watermark to 100, 100 MySQL queries can be initiated concurrently. Additional queries are denied. This parameter is used with rds_threads_running_ctl_mode (default value: select).

Suggestion: This parameter is typically used to handle peak-hour or burst requests, which provides effective database protection.

17.5 Accounts

17.5.1 Create an account

This topic describes the functions and features of accounts in classic and premier modes, as well as how to create accounts in different modes.

You need to create an account in an RDS instance before you can use the database. ApsaraDB for RDS supports two account management modes: classic and premier. Classic mode is an earlier management mode. You cannot use SQL to manage databases and accounts in classic mode. Premier mode is a later management mode that enables more permissions. You can use SQL to manage databases and accounts in premier mode. In the long run, we recommend

that you use premier mode if you need personalized and fine-grained control over database permissions.

Account modes

In classic mode, all accounts are created through the RDS console or APIs, instead of through SQL. All accounts are created equally. The RDS console is used to create and manage all accounts and databases.

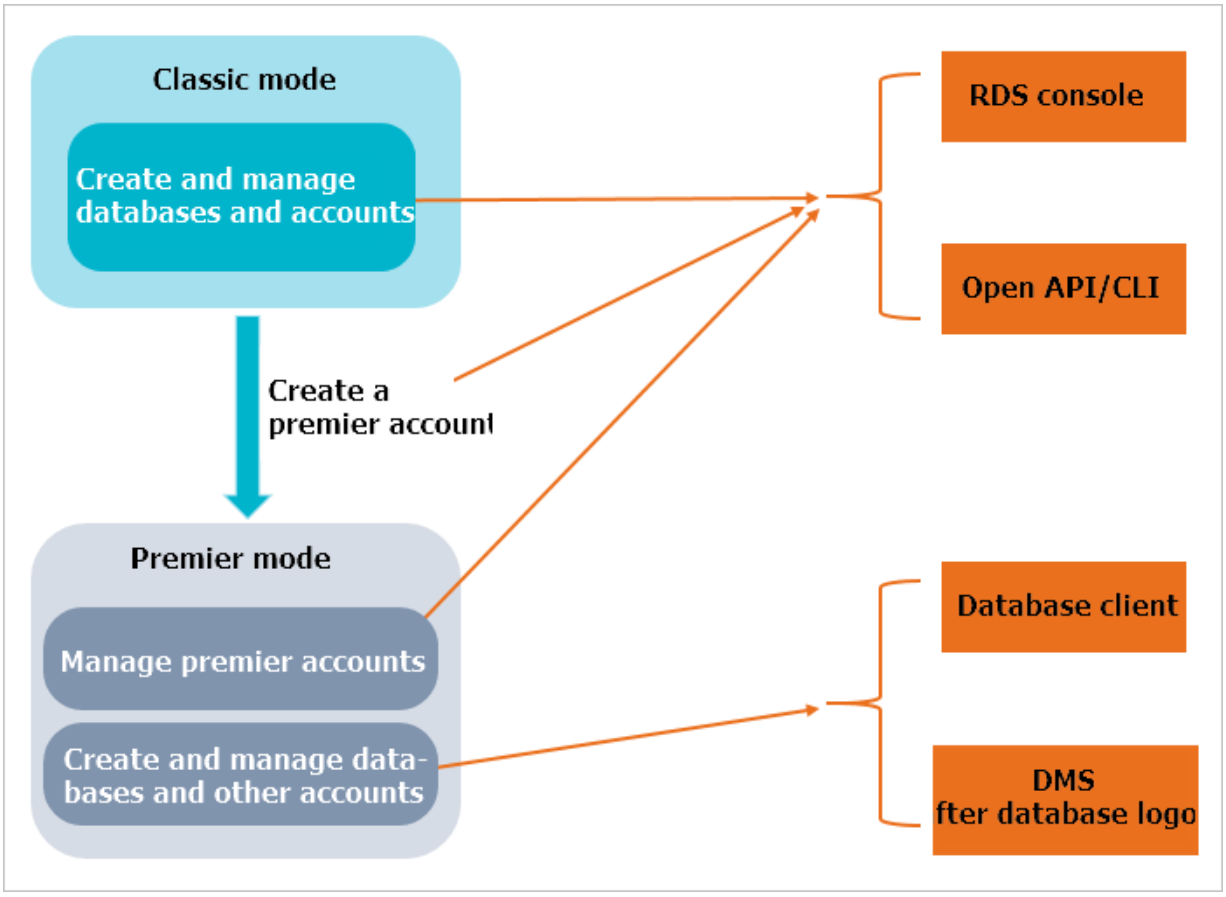
In premier mode, the first account you create is the initial account. You must use the RDS console or APIs to create and manage it. Log on to a database with your initial account. You can then create and manage other standard accounts through SQL commands. However, you cannot use your initial account to change the passwords of other standard accounts. To change the password of a standard account, you must delete the standard account and create a new one. In the following example, the initial account "root" is used to log on to the database. Then, a standard account "jeffrey" is created.

```
mysql -hxxxxxxxxx.mysql.rds.aliyuncs.com -uroot -pxxxxxx -e "
    CREATE USER 'jeffrey'@'%' IDENTIFIED BY 'password';
    CREATE DATABASE DB001;
    "
```

In premier mode, the database management page is not available in the RDS console. APIs such as CreateDatabase cannot be used to manage databases. You must use SQL commands to create and manage databases.

Figure 17-22: Difference between standard and premier accounts shows how to create and manage databases and accounts in classic and premier modes.

Figure 17-22: Difference between standard and premier accounts



How to create an account

Note:

- When assigning database account permissions, follow the least privilege principle. Use service roles to create accounts and assign proper read-only and read/write permissions. When necessary, you may define database accounts and databases in a fine-grained manner so that each database account can only access data for their own services. If the account does not need to write data to a database, assign read-only permissions.
- Use strong passwords for database accounts and change the passwords on a regular basis.

Procedure

- For more information about how to create a standard account for MySQL, see *Create a standard account*.
- For more information about how to create a premier account for MySQL, see *Create a premier account*.

- For more information about how to create an account for PostgreSQL, see *Create a database and an account*.
- For more information about how to create an account for PPAS, see *Create a database and an account*.

17.5.2 Reset your password

You can use the RDS console to reset the password of your database account if you forget the password.

Context

Note:
For data security, we recommend that you change your password periodically.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Database Management** > **Accounts**.
4. On the **Accounts** page, click the icon and select **Reset Password** in the Actions column corresponding to the instance. In the **Reset Password** dialog box that appears, configure parameters as prompted.

   Table 17-20: Password reset parameters describes the parameter configurations.

   Table 17-20: Password reset parameters

Parameter	Description
New Password	It can be 6 to 32 characters in length and can contain letters, numbers, and underscores (_).
Retype Password	It can be 6 to 32 characters in length and can contain letters, numbers, and underscores (_).

5. Click **OK**.

17.5.3 Modify account permissions

When using ApsaraDB for RDS, you can modify the account permissions of your instance at any time.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Database Management** > **Accounts**.
4. On the **Accounts** page, click the icon and select **Modify Permissions** in the Actions column corresponding to the account. Modify the account permission as prompted on the page that appears.

   Table 17-21: Account permission modification parameters describes the parameter configurations.

   Table 17-21: Account permission modification parameters

Parameter	Description
Available Databases	The list of databases that have been created.
Selected Databases	The databases whose permissions are to be authorized to the account. You can select one of the following permissions: • **Read-Only**: authorizes the database read-only permission to the account. • **Read and Write**: authorizes the database read/write permission to the account.

5. Click **OK**.

17.5.4 Delete an account

You can use the console to delete standard accounts that you do not need any more.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Database Management** > **Accounts**.

4. On the **Accounts** page, click the icon and select **Delete** in the Actions column corresponding to the account. In the **Delete User** dialog box that appears, click **OK** to delete the account.

17.5.5 Modify descriptions

You can add a description when you create an account to assist in account management. You can also modify the description after the account is created.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Database Management** > **Accounts**.
4. On the **Accounts** page, click the icon and select **Edit Description** in the Actions column corresponding to the account. In the **Edit Account Description** dialog box that appears, add a description in the **Description** text box.
5. Click **OK**.

17.6 Databases

17.6.1 Create a database

After you create an RDS instance and configure the whitelist, you need to create a database and an account in the instance.

Context

Before you migrate data from the on-premises database to ApsaraDB for RDS, you must create a database and an account for the database. Ensure that the database has the same properties as the on-premises database, and the account of the database has the same permissions as the account of the on-premises database. Follow the least privilege principle to assign permissions to the database account. Use service roles to create accounts and assign proper read-only and read/write permissions to the roles. When necessary, you may define database accounts and databases in a fine-grained manner so that each database account can only access data for their own services.

Procedure

1. *Log on to the RDS console*.
2. In the left-side navigation pane, choose **Database Management** > **Databases**.

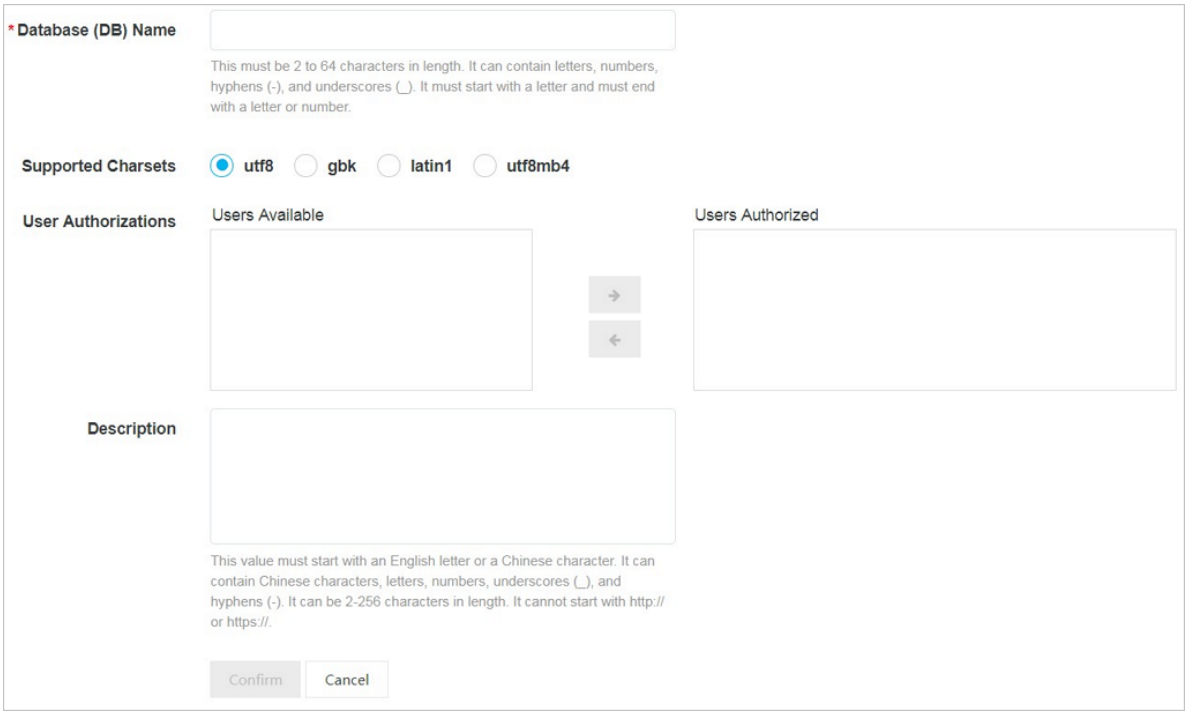
3. On the **Databases** page, click **Create Database**. On the **Create Database** page, configure parameters as prompted.

   Table 17-22: Database creation parameters describes the parameter configurations.

Table 17-22: Database creation parameters

Parameter	Description
Database (DB) Name	The database name. The naming rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, underscores (_), and hyphens (-). • It can be 2 to 64 characters in length. • Reserved keywords cannot be used. For more information about the reserved keywords, see the relevant document in the official website of the corresponding engine.
Supported Character Set	The character sets supported by the database, including: • `UTF-8` • `gbk` • `latin1` • `utf8mb4`
User Authorizations	• Select an account that is authorized to use the database. This parameter can be empty if no account is created. • The permissions on the database can be granted only to standard accounts. • The premier account is authorized to use the database by default.
Description	The description of the database. The rules are as follows: • It must start with a lowercase letter. • It can contain lowercase letters, numbers, underscores (_), and hyphens (-). • It can be 2 to 256 characters in length.

Figure 17-23: Create a database



4. Click **OK**.

17.6.2 Modify database description

You can modify the description of a database for easy management. This topic describes how to modify the description of a database.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Database Management** > **Databases**.
4. On the **Databases** page, click the icon and select **Edit** in the Actions column corresponding to the database.
5. In the **Edit Database** dialog box that appears, set **Database Description**.

   The database description must meet the following requirements:

   - It must start with a letter. It should not start with "http://" or "https://."
   - It can contain letters, numbers, underscores (_), and hyphens (-).
   - It can be 2 to 256 characters in length.

6. Click **OK**.

17.6.3 Delete a database

You can delete out-of-date databases from the RDS console.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Database Management** > **Databases**.
4. On the **Databases** page, click the icon and select **Delete** in the Actions column corresponding to the database. In the **Delete Database** dialog box that appears, click **OK** to delete the database.

17.7 Access mode

ApsaraDB for RDS supports two access modes: **`Standard Mode`** and **`Safe Mode`**. This topic describes the differences between the two access modes and their configuration methods.

Prerequisites

Set the network type of the instance to **Classic Network**.

Context

`Standard Mode` and **`Safe Mode`** have the following differences:

- Standard mode: ApsaraDB for RDS uses Server Load Balancer (SLB) to eliminate the impact from HA switching of the database engine on the application layer. This shortens the response time, but slightly increases the probability of transient disconnections and disables SQL interception.
- Safe mode: This mode prevents 90% of transient disconnections and supports SQL intercepti on (SQL injection attacks are prevented based on SQL semantic analysis). However, it increases the response time by 20% or more.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. On the **Basic Information** page, click **Switch Connection Mode**.

Note:

When the access mode change is in progress, **Status** of the instance changes to **Switching the access mode**. When **Status** changes to **Running**, the access mode is successfully changed.

17.8 Backup and recovery

17.8.1 RDS data backup

17.8.1.1 Automatic backup

ApsaraDB for RDS automatic backup supports full physical backups. ApsaraDB for RDS automatically backs up data based on the pre-configured policies. This topic describes how to configure a policy for automatic backup.

Procedure

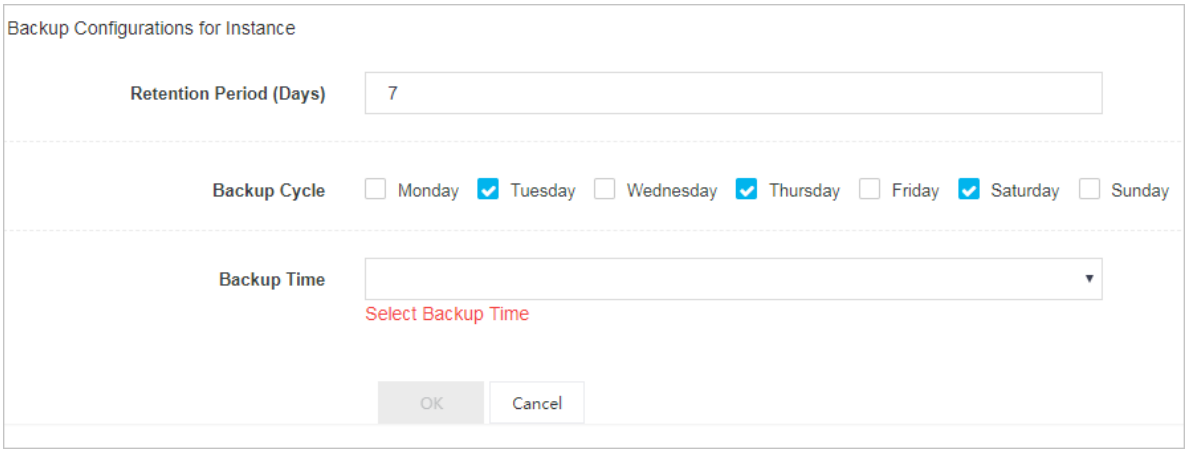
1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Backup and Restore** > **Backups**.
4. On the **Backups** page, click the **Backup Settings** tab and click **Settings**. On the **Backup Configurations for Instance** page, configure parameters as prompted.

   Table 17-23: Backup policy configuration parameters describes the parameter configurations.

Table 17-23: Backup policy configuration parameters

Parameter	Description
Retention Period (Days)	The number of days for which backup files are retained. The default value is 7 days and the value range is from 1 to 30 days.
Backup Cycle	One or multiple days in a week.
Backup Time	Any time range of a day, in hours.

Figure 17-24: Configure a backup policy



5. Click **OK**.

17.8.1.2 Manual backup

Manual backup supports full physical backups and full logical backups. This topic describes how to manually back up RDS data.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. On the **Basic Information** page, click **Back Up Instance**. On the **Back Up Instance** page, set **Backup Mode** and **Backup Type**.

   Instance backup parameters describes the parameter configurations.

Table 17-24: Instance backup parameters

Category	Configuration	Description
Backup Mode	**Physical Backup**	This mode dumps the physical files of the RDS database (such as data files, control files, and log files). In case the database fails, these files can be used to restore data.
	Logical Backup	This mode stores all schema definition statements and data insertion statements of the RDS database. You can execute these SQL statements to restore data. A database that is exactly the same as the original database is created.

Category	Configuration	Description
Backup Type	**Automatic Backup**	This type automatically backs up data based on the pre-configured backup policy. For more information, see *Automatic backup*.
	Full Backup	This type backs up all the files in the database.

Figure 17-25: Configure manual backup

4. Click **OK**.

17.8.2 RDS data recovery

17.8.2.1 Clone an instance

A cloned instance is a new instance with the same content as the primary instance, including data and settings. This feature allows you to restore data of the primary instance or create multiple instances that are the same as the primary instance.

Prerequisites

To clone an instance, make sure that the primary instance meets the following conditions:

- The instance is in running state.
- It does not have an ongoing migration task.
- Data backup and log backup are enabled.
- To clone an instance by using a backup set, ensure that the primary instance must have at least one completed backup set.

Context

You can specify a backup set or any point in time within the backup retention period to clone an instance.

Note:

- A cloned instance only copies the data of the primary instance, but not the content of read-only or disaster recovery instances under the primary instance. The copied data includes database data, account information, and instance configurations (such as whitelist configurations, backup configurations, parameter configurations, and alarm thresholds configurations).
- The database type of a cloned instance must be the same as that of the primary instance. Other settings (such as the instance series, zone, network type, instance type, and storage space) can be different. If you want to clone an instance to restore data of the primary instance, we recommend that you select a higher instance type and larger storage space than the primary instance. Otherwise, data recovery takes a longer time to complete.
- The account type of a cloned instance must be the same as that of the primary instance. The account password of the cloned instance can be modified.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Backup and Restore** > **Backups**.
4. On the **Backups** page, click the **Backups** tab. Click **Clone Instance**. On the **Clone Instance** page that appears, configure parameters as prompted.

   Table 17-25: Instance cloning parameters describes the parameter configurations.

   Table 17-25: Instance cloning parameters

Parameter	Description
Restore Mode	The restore mode of data in the primary instance, including: • `By Time` • `By Backup Set`
Restore Point Time	The time by which the data in the primary instance is to be restored when the restore mode is `By Time`.
Backup Set	The backup set by which the data in the primary instance is to be restored when the restore mode is `By Backup Set`.
Specifications	The specifications of the cloned instance.
Storage Size	The storage size of the cloned instance.

5. Click **Create**.

17.8.3 Binary log (binlog)

This topic describes how to check and download the binlogs of an RDS instance.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Backup and Restore** > **BinLogs**.
4. On the **BinLogs** page, select a time range and click **Search** to search for the binlogs generated within the selected time range.
5. To download a binlog, click the icon and select **Download**.

17.9 Security

17.9.1 Configure a whitelist

To guarantee database security and reliability, you need to modify the whitelist of the RDS instance before you enable the instance. You need to add the IP addresses or IP address segments used for database access to the whitelist of the RDS instance.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. For each newly created RDS instance, IP address 0.0.0.0/0 is added to the **default** whitelist group by default. 0.0.0.0/0 allows all IP addresses to access the instance, which greatly reduces database security. Delete 0.0.0.0/0 from the whitelist.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Security Control** > **Configure Whitelist**.
4. You can use two methods to add an IP address or IP address segment.
   - Add an IP address or IP address segment directly to the default whitelist group.

     a) Click the icon corresponding to the **default** whitelist group, and add an IP address or IP address segment.

b) Click **OK**.

- Add a whitelist group and add an IP address or IP address segment to the group.

a) Click **Create Whitelist Group**. In the dialog box that appears, enter a group name and an IP address or IP address segment.

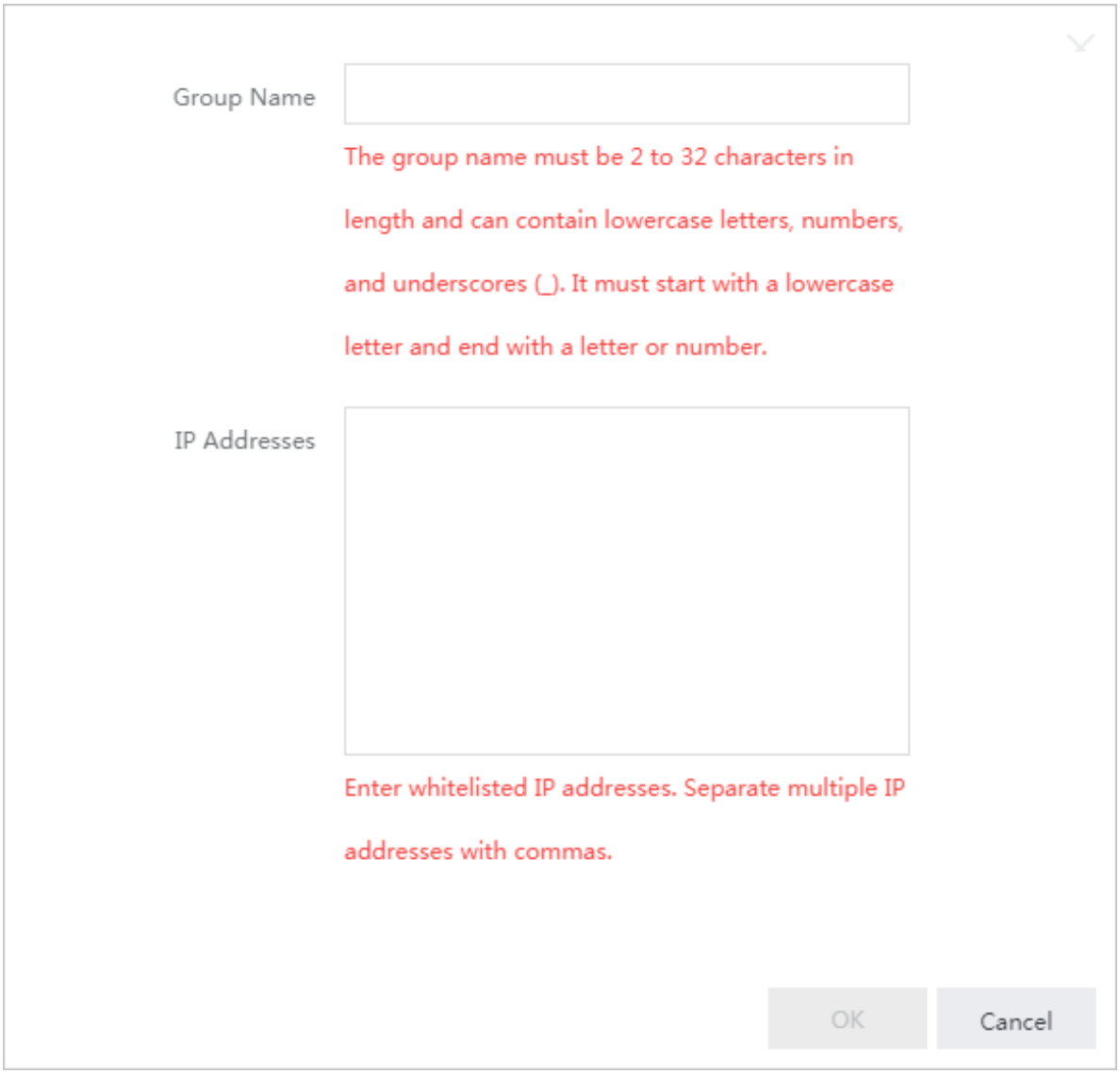
b) Click **OK**.

Table 17-26: Whitelist configuration parameters describes the parameter configurations.

Table 17-26: Whitelist configuration parameters

Parameter	Description
Group Name	The name of the new whitelist group. The naming rules are as follows: • It must start with a lowercase letter and end with a lowercase letter or number. • It can contain lowercase letters, numbers, and underscores (_). • It can be 2 to 32 characters in length. You cannot modify the name of a created whitelist group.
IP Addresses	The IP addresses or IP address segments allowed to access the RDS instance. **Note:** • If you enter an IP address segment, such as 10.10.10.0/24, any IP address in the format of 10.10.10.X can access the RDS instance . • If multiple IP addresses are entered, use commas (,) to separate the addresses and do not add spaces between the addresses and commas, such as 192.168.0.1,172.16.213.9. • 127.0.0.1 indicates that no IP address is allowed to access the RDS instance. • 0.0.0.0/0 indicates that all IP addresses are allowed to access the RDS instance.

Figure 17-26: Create a whitelist group



What's next

Correct use of the whitelist can improve access security for your RDS instance. We recommend that you maintain the whitelist periodically. After you configure the whitelist, you can perform the following operations:

- Click the icon to modify the whitelist group.
- Click the icon to clear the default whitelist group or delete a custom whitelist group.

17.9.2 Audit logs

You can query the SQL logs, operation logs, and error logs of an instance in the RDS console to locate and analyze faults.

Context

> **Note:**
> Audit logs are stored for seven days. They are automatically cleared after seven days.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Security Control** > **Audit Logs**.
4. On the **Audit Logs** page, click the **SQL Logs**, **Error Logs**, **Operation Logs**, or **Audit Log Files** tab.

   Table 17-27: Differences among various kinds of audit logs describes the differences among SQL logs, error logs, and operation logs.

   Table 17-27: Differences among various kinds of audit logs

Log Type	Description
SQL Logs	It records the modification operations on all databases.
Error Logs	It records the SQL statements that failed to be executed on databases in the past month.
Operation Logs	It records all operations performed on the console.

5. Select a time range and click **OK**.

What's next

On the **Operation Logs** tab page, you can click **Export** to export operation logs for offline analysis.

17.9.3 Configure SSL encryption

To enhance link security, you can enable Secure Sockets Layer (SSL) encryption and install SSL CA certificates on the necessary application services.

Context

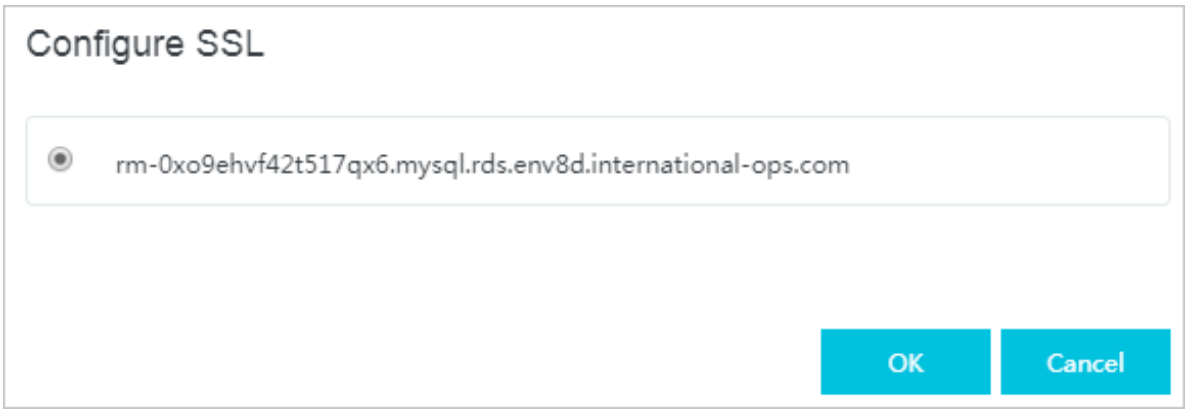
Note:

- SSL is used on the transport layer to encrypt network connections. It increases the security and integrity of communication data, but it also increases the network connection response time.
- Exercise caution when enabling SSL encryption, because it cannot be disabled after it is enabled.

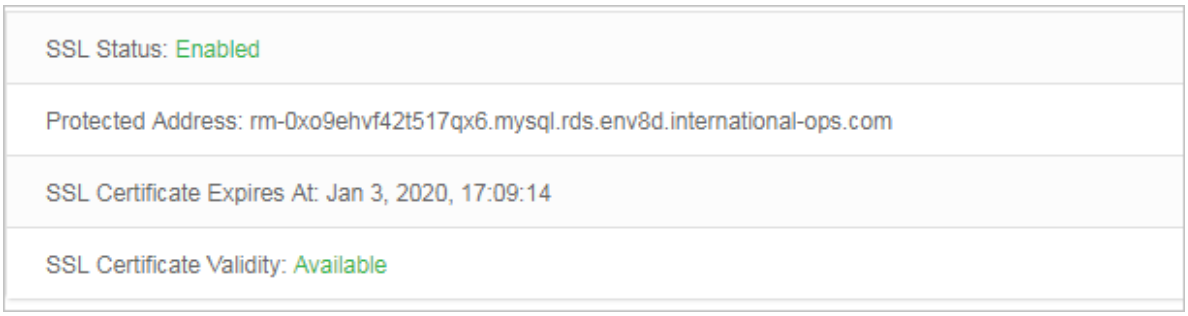
Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Security Control** > **Configure SSL Encryption**. The **Configure SSL Encryption** page displays the SSL details of the instance.
4. Click **Enable SSL**.

5. In the **Configure SSL** dialog box that appears, select an instance ID.



6. Click **OK** to enable SSL encryption, as shown in the following figure.



17.9.4 Download SSL CA certificates

To increase link security, you can enable Secure Sockets Layer (SSL) encryption and install SSL CA certificates on the necessary application services.

Context

Note:
SSL is used on the transport layer to encrypt network connections. It increases the security and integrity of communication data, but it also increases the network connection response time.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Security Control** > **Configure SSL Encryption.** The **Configure SSL Encryption** page displays the SSL details of the instance.
4. Click **Download Certificate**. In the **Download Certificate** dialog box that appears, click **OK** to download SSL CA certificates.

   The downloaded package includes three files:

   - P7b file: used to import CA certificates to the Windows system.
   - PEM file: used to import CA certificates to other operating systems or applications.
   - JKS file: stores truststore certificates in Java. The password is apsaradb. It is used to import the CA certificate chain to Java programs.

   Note:

When the JKS file is used in Java, you need to modify the default JDK security configuration in JDK7 and JDK8. Open the `jre/lib/security/java.security` file on the computer where the database that needs SSL access resides, and modify two configurations as follows:

```
jdk.tls.disabledAlgorithms=SSLv3, RC4, DH keySize < 224
jdk.certpath.disabledAlgorithms=MD2, RSA keySize < 1024
```

If you do not modify the JDK security configuration, the following error will be reported. Other similar errors are also caused by Java security configurations.

```
javax.net.ssl.SSLHandshakeException: DHPublicKey does not comply to
 algorithm constraints
```

17.9.5 Configure transparent data encryption

Transparent data encryption (TDE) implements real-time I/O encryption and decryption for data files. It encrypts data before writing it to a disk and decrypts data before reading from the disk. TDE does not increase the size of data files. Developers can directly use the TDE function without changing any application.

Prerequisites

Before using TDE, make sure that your instance and account meet the following requirements:

- The instance version is RDS for MySQL 5.6.
- You must log on by using an Alibaba Cloud account (not RAM user) to view and modify TDE configurations.
- Before enabling TDE, you need to enable Key Management Service (KMS). If you have not enabled KMS, you can enable it based on the guidance during TDE activation.

Context

Note:

- Once TDE is enabled, it cannot be disabled.
- Keys produced and managed by the KMS are used for encryption. RDS does not provide the keys and certificates required for encryption. After activating TDE, to restore the data to the local device, you must use RDS to decrypt the data first.
- When TDE is enabled, the CPU usage will increase significantly.

Procedure

1. *Log on to the RDS console*.

2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Security Control** > **Configure TDE**. You can view the TDE details of the instance.
4. On the **Configure TDE** page, click **Enable TDE**. In the **Configure TDE** dialog box that appears, click **OK**.
5. Log on to the database and run the following command to encrypt the relevant tables:

   ```
   alter table <tablename> engine=innodb,block_format=encrypted;
   ```

   If you want to decrypt a table encrypted with TDE, run the following command:

   ```
   alter table <tablename> engine=innodb,block_format=default;
   ```

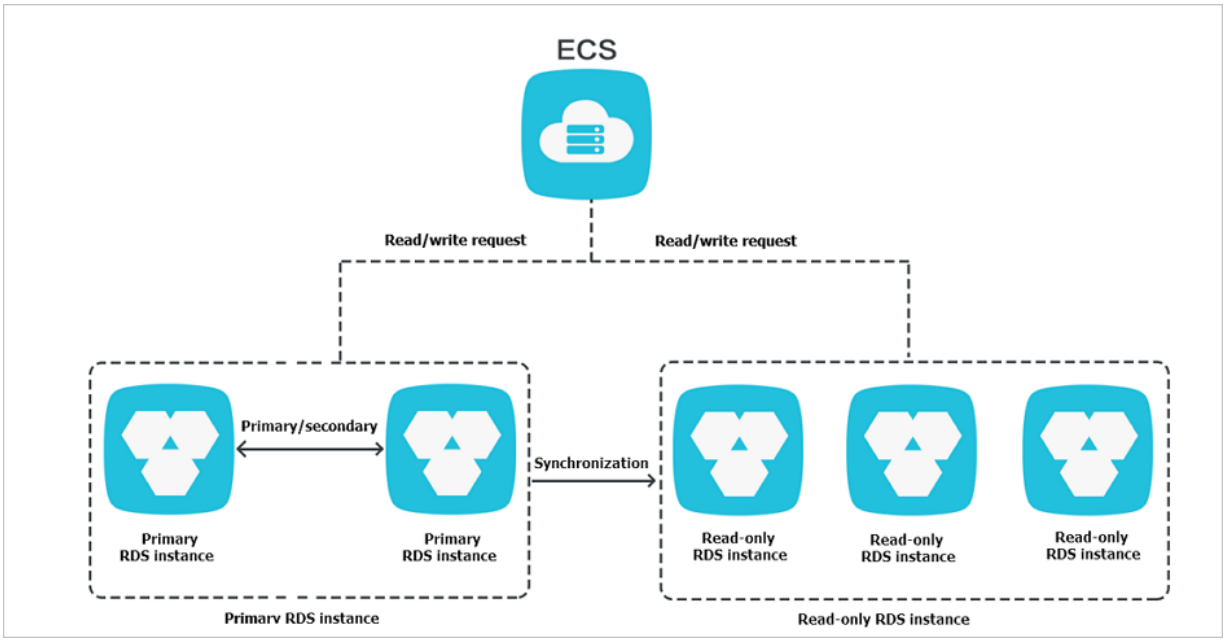
17.10 Read-only instances

17.10.1 Overview

RDS for MySQL 5.6 allows you to create read-only instances. In application scenarios where there are only a few write requests, but a large number of read requests, read-only instances can be created to relieve the database pressure on the primary instance. This topic describes the features and restrictions of read-only instances.

To achieve auto scaling of the read capability and relieve the database pressure, you can create one or more read-only instances in a region. In this way, a large amount of data can be read from the database, and the application throughput can be increased.

A read-only instance with a single physical node and no backup node uses the native replicatio n capability of MySQL to synchronize changes in the primary instance to all relevant read-only instances. Real-only instances are in the same region as the primary instance, but not necessaril y in the same zone as the primary instance. The topology of a read-only instance is shown as follows.



Read-only instances have the following features:

- Specifications of a read-only instance can be different from those of the primary instance and can be changed at any time, which facilitates elastic scaling.
- No account or database maintenance is required for a read-only instance. The account and database information is synchronized from the primary instance.
- Read-only instances support the independent whitelist configuration.
- System performance monitoring is provided.

  ApsaraDB for RDS provides nearly 20 system performance monitoring views, including those for disk capacity, IOPS, connections, CPU utilization, and network traffic. You can view the load of instances with ease.
- Optimization recommendations are provided: ApsaraDB for RDS provides a variety of optimization recommendations, such as storage engine check, primary key check, large table check, and check for excessive indexes and missing indexes. You can optimize your databases based on the optimization recommendations and specific applications.

17.10.2 Create a read-only instance

This topic describes how to create a read-only instance.

Context

Read-only instances have the following usage restrictions:

- A maximum of five read-only instances can be created for a primary instance.

- Backup setting and temporary backup are not supported.
- Instance recovery is not supported.
- Data migration to read-only instances is not supported.
- Database creation and deletion are not supported
- Account creation and deletion are not supported. Account authorization and account password change are not supported.
- After a read-only instance is created, the primary instance does not support data recovery by directly overwriting instances using backup sets.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Performance Optimization** > **Read-Only Instance**.
4. On the **Read-Only Instance** page, click **Create Read-Only Instance**. On the **Create Read-Only Instance** page, configure parameters as prompted.

   Table 17-28: Read-only instance creation parameters describes the parameter configurations.

   Table 17-28: Read-only instance creation parameters

Parameter	Description
Destination Region	The region to which the read-only instance belongs. This parameter is the same as that of the primary instance.
Instance Specification	The specifications of the read-only instance . It can be different from that of the primary instance. The specifications of a read-only instance can be modified at any time to facilitate flexible upgrade and downgrade.
Storage Size (GB)	The storage size of the read-only instance. To guarantee sufficient I/O for data synchroniz ation, we recommend that the memory of read-only instances is not less than that of the primary instance.
Network Type	The network type of the read-only instance. You can choose from the following two network types:

Parameter	Description
	• `Classic Network` • VPC: If you use a VPC, we recommend that you choose the same VPC as that of the primary instance.

5. Click **Create**.

17.10.3 View read-only instance details

17.10.3.1 View instance details through a read-only instance

You can go to the read-only instance management page from the instances page or the read-only instance list page of the primary instance. Read-only instances are managed in the same way as ordinary instances. The page shows the management operations that can be performed. This topic describes how to go to the read-only instance management page from the instances page.

Procedure

1. *Log on to the RDS console*.
2. On the **RDS Instances** page, click the ID of the read-only instance. The **Basic Information** page that appears allows you to manage the read-only instance.

   In the instance list, the **Instance Type** of read-only instances is displayed as **`Read-Only Instance`**, as shown in *Figure 17-27: View read-only instances*.

   Figure 17-27: View read-only instances

17.10.3.2 View instance details through the primary instance

You can go to the read-only instance management page from the instances page or the read-only instance list page of the primary instance. Read-only instances are managed in the same way as ordinary instances. The page shows the management operations that can be performed. This topic describes how to go to the read-only instance management page from the read-only instances page of the primary instance.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.

3. In the left-side navigation pane of the **Basic Information** page, choose **Performance Optimization** > **Read-Only Instance**.
4. On the **Read-Only Instance** page, click the ID of the read-only instance. The **Basic Information** page that appears allows you to manage the read-only instance.

17.11 Read/write splitting

17.11.1 Overview

This topic describes the principles, benefits, and usage of the read/write splitting function.

Both the primary RDS instance and the read-only RDS instance have an independent connection address. The connection address is configured by an application for data read/write splitting.

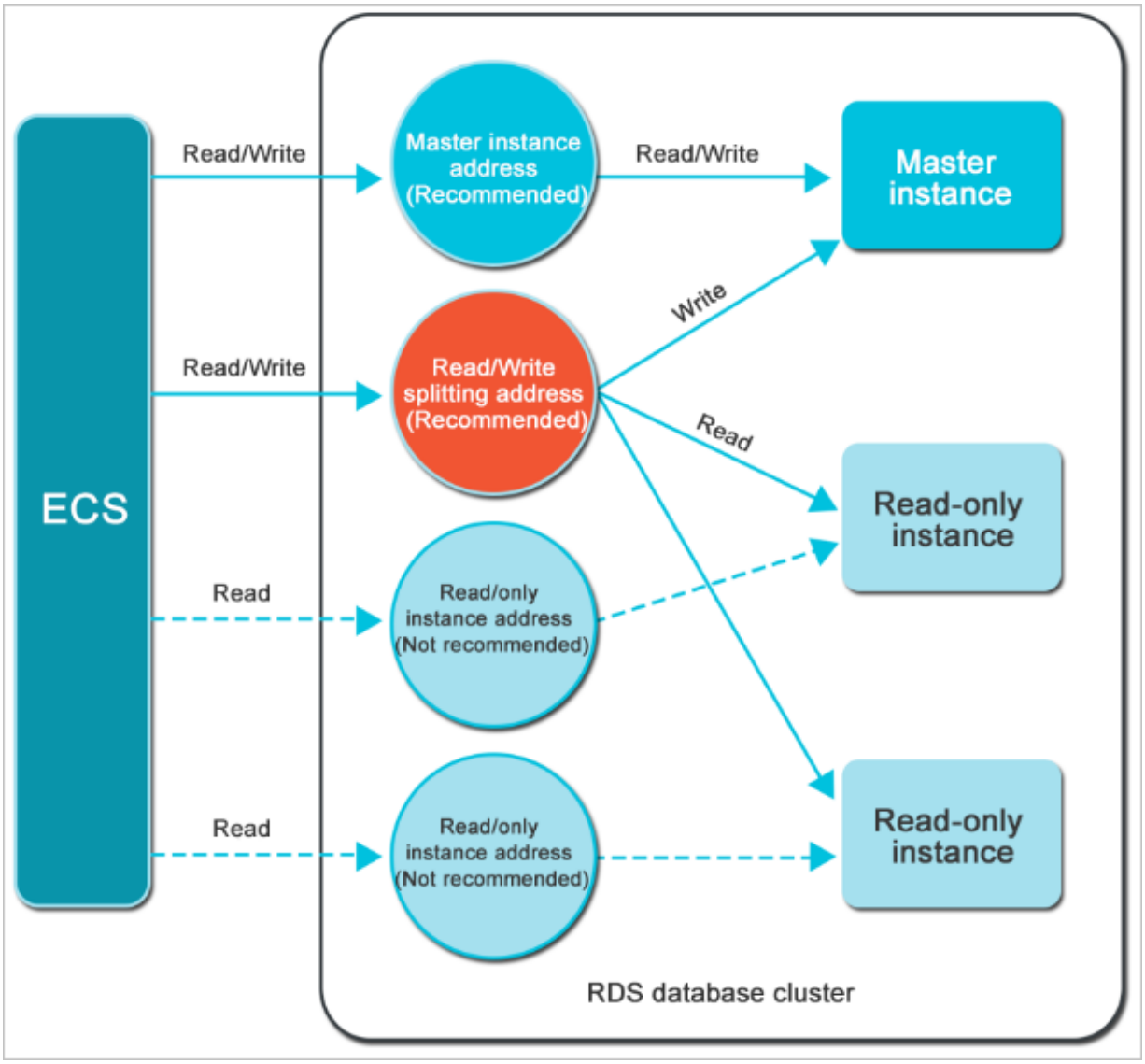
The read/write splitting function provides an extra read/write splitting address. This address links the primary instance with all its read-only instances to enable a link for automatic read/write splitting. An application only need to connect to the same read/write splitting address for data reading and writing. Write requests are automatically routed to the primary instance, and read requests are routed to each read-only instance based on their weight. You can scale out the processing capability of the system by adding more read-only instances. No application change is required.

Read/write splitting is supported in RDS for MySQL 5.6 only. When read/write splitting is enabled, there will be three kinds of addresses for the instances:

- Connection address of the primary instance
- Connection address of the read-only instance
- Connection address of read/write splitting

Figure 17-28: Principle of read/write splitting shows how an application uses different types of connection addresses to access the database.

Figure 17-28: Principle of read/write splitting



Read/write splitting has the following benefits:

- Facilitates maintenance with a single read/write splitting address.

  In the current read-only account mode, both the primary instance and each read-only instance have an independent connection address. You need to configure each of the addresses in your application to have write requests sent to the primary instance and read requests sent to read-only instances.

  RDS read/write splitting provides a read/write splitting address in addition to the existing instance connection addresses. After an application is connected to the read/write splitting address, it can perform read-only and read/write operations on the corresponding primary and

read-only instances. The forwarding logic of read-only and read/write statements is transparent to the user. This reduces the maintenance cost.

- Improves performance with RDS support for the highly secure link.

  For users who build a proxy layer to implement read/write splitting on the cloud, data has to go through multiple components for statement parsing and forwarding before it reaches the database, impacting the response latency significantly. RDS read/write splitting is built on the existing highly secure link without the need of any additional component. This greatly reduces the latency and improves the processing efficiency.

- Applies to various use cases with configurable weights and thresholds.

  RDS read/write splitting allows you to set the read request weight for the primary and read-only instances, and the latency threshold for read-only instances.

- Enhances database availability with instance health check.

  RDS read/write splitting performs health check automatically for all instances in the distribution system. If any instance fails or its latency exceeds the threshold, RDS automatically removes the instance from the distribution system (while marking it as unavailable and stopping allocating read requests to it) and allocates write requests to the remaining instances based on the predefined weight. This ensures normal application access in the case of a read-only instance failure. After the instance is restored, RDS automatically reclaims it into the request distribution system.

  Note:
  We recommend that you create at least two read-only instances for the primary instance when using read/write splitting. This ensures normal database access in case of a single-point failure.

17.11.2 Enable read/write splitting

In scenarios where there are a few write requests but a large number of read requests for the database, you can enable read/write splitting to distribute the read load on the primary instance. This topic describes how to enable the read/write splitting function.

Prerequisites

- The instance is a high-availability RDS for MySQL 5.6 primary instance.

- A read-only instance has been created under the primary instance. If there is no read-only instance, create one. For more information about creating a read-only instance, see *Create a read-only instance*.
- The primary instance has been switched to safe mode.

When you enable the read/write splitting function for the first time, the system automatically upgrades the backend control system of the primary instance and all associated read-only instances to the latest version to guarantee normal service operations. When the function is enabled, the primary and read-only instances automatically restart once. During the restart process, the primary instance is subject to a transient disconnection of 30 seconds or less, and the read-only instances are inaccessible. To avoid service impacts from transient disconnections, we recommend that you enable read/write splitting during off-peak hours and make sure that your application can be automatically reconnected.

- Currently, the following commands or functions cannot be forwarded to a read-only instance:
  - The `stmt prepare sql` command is automatically executed on the primary instance.
  - The `stmt prepare` command cannot be forwarded to a read-only instance before the execution of the `stmt close` command.
  - `set global`, `set user`, and `set once` are automatically executed on the primary instance.
- The execution result is random for the following commands:

  `show processlist`, `show master status`, and `com_process_info` return results based on the instance connected during command execution.
- All transactions are routed to the primary database.
- Read/write splitting does not guarantee consistency of non-transactional reads. If you require such consistency, add a hint to route query requests to the primary database or encapsulate query requests in transactions.
- The following commands or functions are not currently supported:
  - SSL encryption
  - Compression protocols
  - com_dump_table and com_change_user protocols
  - `kill connection [query]`
  - `change user`

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Performance Optimization** > **Read/Write Splitting**.
4. On the **Read/Write Splitting** page, click **Enable**. On the **Configure Read/Write Splitting** page, configure parameters as prompted.

   Table 17-29: Read/write splitting parameters describes the parameter configurations.

Table 17-29: Read/write splitting parameters

Parameter	Description
SLB Type	The read/write splitting address. It must be an internal network address. The internal network type is automatically synchronized with the primary instance.
Latency Threshold	The maximum allowed latency when read-only instances synchronize data from the primary instance. The value range is 0 to 7,200 seconds. If the latency of a read-only instance exceeds this threshold, read requests are not forwarded to this instance regardless of the instance weight. Depending on the running of SQL statements, latencies may occur to read-only instances. We recommend that you set the value to no less than 30 seconds.
Weights of Read Requests	The read request weight of each instance. An instance with a higher weight can process more read requests. For example, a read-write splitting address has a primary instance and three read-only instances. The read weights of the instances are 0, 100, 200, and 200 respectively. This means that the primary instance does not process read requests (write requests are automatically sent to the primary instance). The three read-only instances process read requests in the proportion of 1:2:2. The read request weight can be customized or automatically distributed by the system. • **Default**: The system automatically distributes weights to instances based on their configurations. The new read-only instances under the primary instance are automatically added to the read/write splitting link based on the preset weight without manual configuration. For more information, see *Rules of system weight distribution*. • **Customized**: You can customize the read request weight of instances in the range of 0 to 10,000. If you select this mode, the weight of new read-only instances added to the primary instance defaults to 0. You have to set this parameter manually.

5. Click **OK**.

17.11.3 Modify the latency threshold and weights of read requests

After read/write splitting is enabled, you can modify the latency threshold and weights of read requests.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Performance Optimization** > **Read/Write Splitting**.
4. On the **Read/Write Splitting** page, click **Enable**. On the **Configure Read/Write Splitting** page, configure parameters as prompted.

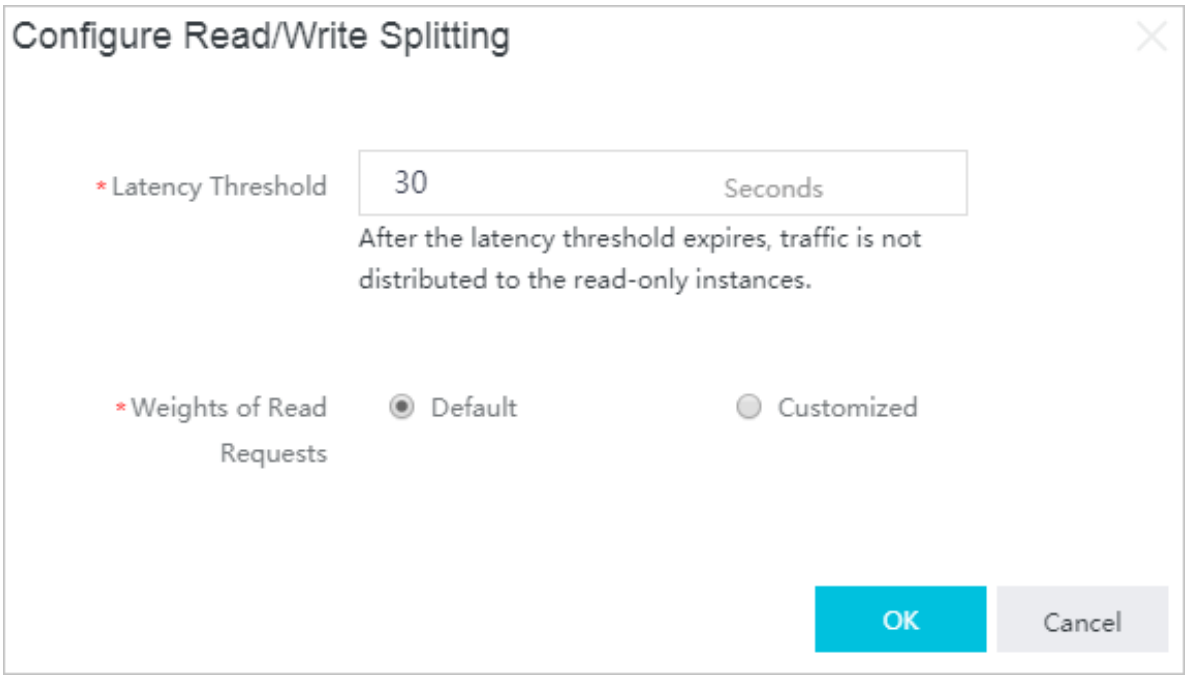
   Table 17-30: Read/write splitting parameters describes the parameter configurations.

   Table 17-30: Read/write splitting parameters

Parameter	Description
Latency Threshold	The maximum allowed latency when read-only instances synchronize data from the primary instance. The value range is 0 to 7,200 seconds. If the latency of a read-only instance exceeds this threshold, read requests are not forwarded to this instance regardless of the instance weight. Depending on the running status of SQL statements, latencies may occur to read-only instances. We recommend that you set the value to no less than 30s.
Weights of Read Requests	The read request weight of each instance. An instance with a higher weight can process more read requests. For example, a read-write splitting address has a primary instance and three read-only instances. The read weights of the instances are 0, 100, 200, and 200 respectively. This means that the primary instance does not process read requests (write requests are automatically sent to the primary instance). The three read-only instances process read requests in the proportion of 1:2:2. The read request weight can be customized or automatically distributed by the system. • **Default**: The system automatically distributes weights to instances based on their configurations. The new read-only instances under the primary instance are automatically added to the read/write splitting link

Parameter	Description
	based on the preset weight without manual configuration. For more information, see *Rules of system weight distribution*. • **Customized**: You can customize the read request weight of instances in the range of 0 to 10,000. If you select this mode, the weight of new read-only instances added to the primary instance defaults to 0. You have to set this parameter manually.

Figure 17-29: Modify read/write splitting parameters



5. Click **OK**.

17.11.4 Disable read/write splitting

You can disable read/write splitting if this function is no longer required. This topic describes how to disable the read/write splitting function.

Context

Note:

- The read/write splitting function can only be used when at least one read-only instance is available. Therefore, you must disable the read/write splitting function before you delete the last read-only instance. Otherwise, you are unable to delete the last read-only instance.

- After read/write splitting is disabled, your application can no longer connect to the read/write splitting address. Make sure that your database connection configuration does not include this connection address.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Performance Optimization** > **Read/Write Splitting**.
4. On the **Read/Write Splitting** page, click **Disable**.
5. Click **OK**.

17.11.5 Monitor read/write splitting performance

You can view the read/write splitting performance on the monitoring page of the RDS console.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **System Resource Monitoring** > **Database Performance**.
4. On the **Database Performance** page, click the **QPS/TPS** tab to view the transaction per second (TPS) and query per second (QPS). You can view the number of reads and writes of all databases including the primary database and read-only databases involved in read/write splitting.

17.11.6 Rules of system weight distribution

ApsaraDB for RDS automatically distributes weights to instances based on the configurations of the instances. This topic describes the rules for the system to distribute read weights and how to use a hint to specify whether an SQL statement is sent to the primary instance or read-only instances.

Weight values list

The system automatically configures fixed read weight values for instances, as listed in *Table 17-31: Weight values*.

Table 17-31: Weight values

Specification code	Specification type	Memory	CPU	Weight
rds.mysql.t1.small	Common instance	1 GB	1	100
rds.mysql.s1.small	Common instance	2 GB	1	100
rds.mysql.s2.large	Common instance	4 GB	2	200
rds.mysql.s2.xlarge	Common instance	8 GB	2	200
rds.mysql.s3.large	Common instance	8 GB	4	400
rds.mysql.m1.medium	Common instance	16 GB	4	400
rds.mysql.c1.large	Common instance	16 GB	8	800
rds.mysql.c1.xlarge	Common instance	32 GB	8	800
rds.mysql.c2.xlarge	Common instance	64 GB	16	1600
rds.mysql.c2.xlp2	Common instance	96 GB	16	1600
mysql.x8.medium. 2	Dedicated instance	16 GB	2	200
mysql.x8.large. 2	Dedicated instance	32 GB	4	400
mysql.x8.xlarge. 2	Dedicated instance	64 GB	8	800
mysql.x8.2xlarge. 2	Dedicated instance	128 GB	16	1600
rds.mysql.st.d13	Dedicated host	220 GB	30	3000

Use a hint to specify whether an SQL statement is sent to the primary instance or read-only instances

In addition to the weight distribution system for read/write splitting, hints are used as the supplementary SQL syntax to force SQL statement execution on the primary instance or read-only instances.

The hint formats supported by RDS read/write splitting are as follows:

- `/*FORCE_MASTER*/`: specifies that the following SQL statements are executed on the primary instance.
- `/*FORCE_SLAVE*/`: specifies that the following SQL statements are executed on the read-only instances.

For example, after a hint is prefixed to the following statement, the statement is always routed to and executed on the primary instance regardless of the preset weight.

```
/*FORCE_MASTER*/ SELECT * FROM table_name;
```

17.12 Performance optimization

17.12.1 Slow SQL statistics

You can use the RDS console to query slow SQL statistics to locate and analyze faults.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation bar of the **Basic Information** page, choose **Performance Optimization** > **Slow SQL Statistics**.
4. Select a time range and click **Search**.

> **Note:**
> The system does not list slow logs from the past two hours. These logs are contained in the slow_log table of the MySQL database.

17.12.2 Missing index

Based on the SQL statement execution status and performance of your RDS instance, the system prompts you about database tables with missing indexes, and provides you with a statement to add indexes.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, choose **Performance Optimization** > **Indexes**. This page allows you to query all missing indexes.

17.13 Monitor system resources

The RDS console provides a variety of performance metrics for you to monitor the status of your instance.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance.
3. In the left-side navigation pane of the **Basic Information** page, click **System Resource Monitoring**.
4. Select the monitored data you wish to view, such as **System Resources**, **Database Performance**, **InnoDB Engine**, and **MyISAM Engine**. *Table 17-32: Metric list* lists the monitored data.

Table 17-32: Metric list

Page	Metric	Description	Monitoring frequency	Monitoring cycle
System Resources	Disk Space	The disk space usage of the instance, such as overall usage of the disk space, data space, log space, temporary file space, and system file space. Unit: MB	60s 300s	30 days
	IOPS	The number of I/O requests of the instance per second. Unit: Times/second	60s 300s	30 days

Page	Metric	Description	Monitoring frequency	Monitoring cycle
	CPU Utilization	The CPU usage of the instance (excluding the CPU resources used by the operating system).	60s 300s	30 days
	Network Traffic	The inbound and outbound traffic of the instance per second. Unit: KB	60s 300s	30 days
Database performance	QPS/TPS	The number of SQL statements executed and transactions processed per second.	60s 300s	30 days
	Temporary Tables	The number of temporary tables automatically created on the hard disk when the database executes SQL statements.	60s 300s	30 days
	COMDML	The number of SQL statements executed by the database per second. The statements include INSERT , DELETE , INSERT_SEL ECT, REPLACE , REPLACE_SE LECT, SELECT, and UPDATE.	60s 300s	30 days
	ROWDML	The number of operations performed on InnoDB per second, such as the number of physical writes to the log file, and the number of InnoDB table rows that are read, updated , deleted, and inserted.	60s 300s	30 days
InnoDB Engine	InnoDB Buffer Pool	The read hit rate, utilization, and dirty data block percentage of the InnoDB buffer pool.	60s 300s	30 days
	InnoDB Read/ Write	The volume of InnoDB data that is read and written per second. Unit: KB	60s 300s	30 days
	InnoDB Reads and Writes	The number of InnoDB reads and writes per second.	60s 300s	30 days

Page	Metric	Description	Monitoring frequency	Monitoring cycle
	InnoDB Log	The number of InnoDB physical writes to the log file, log write requests, and FSYNC writes to the log file.	60s 300s	30 days
MyISAM Engine	MyISAM Key Buffer	The read hit rate, write hit rate, and usage of MyISAM key buffers per second.	60s 300s	30 days
	MyISAM Reads and Writes	The number of MyISAM reads and writes from and to the buffer pool and hard disk per second.	60s 300s	30 days

17.14 Data migration from the on-premises database to RDS

17.14.1 Compress data

RDS for MySQL 5.6 allows you to compress data with the TokuDB storage engine. This topic describes how to compress data.

Context

Extensive tests show that the data volume is reduced by 80% to 90% after data tables are transferred from the InnoDB storage engine to the TokuDB storage engine. 2 TB of data can be compressed to 400 GB or less. Aside from data compression, the TokuDB storage engine supports transaction and online DDL operations. It is also compatible with the applications running on the MyISAM and InnoDB storage engines.

Restrictions on TokuDB:

- The TokuDB storage engine does not support foreign keys.
- The TokuDB storage engine is not applicable to scenarios which require large amounts of data to be read.

Procedure

1. Run the following command to check the MySQL version:

```
SELECT version();
```

2. Run the following command and set **loose_tokudb_buffer_pool_ratio** to indicate the proportion of cache that TokuDB occupies in the shared cache of TokuDB and InnoDB:

```
select sum(data_length) into @all_size from information_schema.
tables where engine='innodb';
select sum(data_length) into @change_size from information_schema
.tables where engine='innodb' and concat(table_schema, '.',
table_name) in ('XX.XXXX', 'XX.XXXX', 'XX.XXXX');
select round(@change_size/@all_size*100);
```

> **Note:**
> In the preceding command, **XX.XXXX** indicates the name of the database or table to be transferred to the TokuDB storage engine.

3. Restart the instance. For more information, see *Restart an instance*.
4. Run the following command to modify the storage engine:

```
ALTER TABLE XX.XXXX ENGINE=TokuDB
```

> **Note:**
> In the preceding command, **XX.XXXX**indicates the name of the database or table to be transferred to the TokuDB storage engine.

17.14.2 Migrate MySQL data

17.14.2.1 Use mysqldump to migrate MySQL data

This topic describes how to use mysqldump to migrate on-premises data to RDS for MySQL.

Prerequisites

An ECS instance must be activated.

Context

mysqldump is easy to use but has a long downtime. The tool is suitable for cases with small data volumes or where a long downtime is allowed.

RDS for MySQL is fully compatible with the native database service. The procedure for migrating the original database to an RDS for MySQL instance is similar to the procedure for migrating data from one MySQL server to another.

Before you perform migration, create a migration account in the on-premises database, and grant the read and write permissions of the database to the migration account.

Procedure

1. Run the following command to create a migration account in the on-premises database:

   ```
   CREATE USER 'username'@'host' IDENTIFIED BY 'password';
   ```

   Parameter description:

   - username: specifies the account name to be created.
   - host: specifies the host of the database which the account logs on to. As an on-premises user, you can use **localhost** to log on to the database. To log on from any host, you can use wildcard %.
   - password: specifies the logon password for this account.

   The following example creates an account named **William** with password **Changme123** that is allowed to log on to the on-premises database from any host.

   ```
   CREATE USER 'William'@'%' IDENTIFIED BY 'Changme123';
   ```

2. Run the following command to grant permissions to the migration account in the on-premises database:

   ```
   GRANT SELECT ON databasename.tablename TO 'username'@'host' WITH
    GRANT OPTION; GRANT REPLICATION SLAVE ON databasename.tablename
   TO 'username'@'host' WITH GRANT OPTION;GRANT REPLICATION SLAVE ON
   databasename.tablename TO 'username'@'host' WITH GRANT OPTION;
   ```

   Parameter description:

   - privileges: specifies the operation permissions of the account, such as SELECT, INSERT, and UPDATE. To grant all permissions to the account, use **ALL**.
   - databasename: specifies the name of the database. To grant all database permissions to the account, use wildcard *****.
   - tablename: specifies the name of the table. To grant all table permissions to the account, use wildcard *****.
   - username: specifies the name of the account to be granted permissions.
   - host: specifies the host from which the account is authorized to log on to the database. As an on-premises user, you can use **localhost** to log on to the database. To log on from any host, you can use wildcard %.

- WITH GRANT OPTION: specifies an optional parameter that enables the account to use the GRANT command.

In the following command, the account named **William** is granted all database and table permissions, and allowed to log on to the on-premises database from any host:

```
GRANT ALL ON *. * TO 'William'@'%';
```

3. Use the data export tool of mysqldump to export data in the database as data files.

Note:

Do not update data during data export. This step exports data only. It does no export stored procedures, triggers, or functions.

```
mysqldump -h localIp -u userName -p --opt --default-character-set=utf8
 --hex-blob dbName --skip-triggers > /tmp/dbName.sql
```

Parameter description:

- localIp: specifies the IP address of the on-premises database server.
- userName: specifies the migration account of the on-premises database.
- dbName: specifies the name of the database to be migrated.
- /tmp/dbName.sql: specifies the name of the backup file.

4. Use mysqldump to export stored procedures, triggers, and functions.

Note:

Skip this step if no stored procedures, triggers, or functions are used in the database. When you export stored procedures, trigger, and functions, you must remove "definer" to be compatible with RDS.

```
mysqldump -h localIp -u userName -p --opt --default-character-set=utf8
 --hex-blob dbName -R | sed -e 's/DEFINER[ ]*=[ ]*[^*]*\*/\*/' > /tmp/
triggerProcedure.sql
```

Parameter description:

- localIp: specifies the IP address of the on-premises database server.
- userName: specifies the migration account of the on-premises database.
- dbName: specifies the name of the database to be migrated.
- /tmp/triggerProcedure.sql: specifies the name of the backup file.

5. Upload the data files and stored procedure files to ECS.

   The example in this topic describes how to upload files to the following path.

   ```
   /tmp/dbName.sql
   ```

   ```
   /tmp/triggerProcedure.sql
   ```

6. Log on to ECS and import the data files and stored procedure files to the target RDS for MySQL instance.

   ```
   mysql -h intranet4example.mysql.rds.aliyuncs.com –u userName -p dbName
    < /tmp/dbName.sql
   ```

   ```
   mysql -h intranet4example.mysql.rds.aliyuncs.com -u userName -p dbName
    < /tmp/triggerProcedure.sql
   ```

   Parameter description:

   - intranet4example.mysql.rds.aliyuncs.com: specifies the IP address used to connect to the RDS for MySQL instance. An internal network IP address is used as an example.
   - userName: specifies the migration account of the RDS for MySQL database.
   - dbName: specifies the name of the database to be imported.
   - /tmp/dbName.sql: specifies the name of the data file to be imported.
   - /tmp/triggerProcedure.sql: specifies the name of the stored procedure file to be imported.

17.15 Typical applications

17.15.1 Store multi-structure data

Alibaba Cloud Object Storage Service (OSS) is a cloud-based storage service that features large capacity, security, low costs, and high reliability. ApsaraDB for RDS and OSS can work together to form various data storage solutions.

Context

For example, when ApsaraDB for RDS and OSS are used in a forum, resources such as the images of registered users and posts on the forum can be stored in OSS, reducing the storage pressure on ApsaraDB for RDS.

The following sample code enables combined use of ApsaraDB for RDS and OSS.

Procedure

1. Run the following command to initialize OssAPI:

```
from oss.oss_api import * endpoint=" oss-cn-hangzhou.aliyuncs.com"
accessKeyId, accessKeySecret="your id","your secret" oss = OssAPI(
endpoint, accessKeyId, accessKeySecret)
```

2. Run the following command to create a bucket:

```
#Set the bucket ACL to Private: res = oss.create_bucket(bucket,"
private") print "%s\n%s" % (res.status, res.read())
```

3. Run the following command to upload an object:

```
res = oss.put_object_from_file(bucket, object, "test.txt") print "%s
\n%s" % (res.status, res.getheaders())
```

4. Run the following command to obtain the corresponding object:

```
res = oss.get_object_to_file(bucket, object, "/filepath/test.txt")
print "%s\n%s" % (res.status, res.getheaders())
```

In the Elastic Compute Service (ECS) application code, the ID of each user is stored in ApsaraDB for RDS, and the avatar resources are stored in OSS. The Python code is as follows :

```
#! /usr/bin/env python from oss.oss_api import *
endpoint" oss-cn-hangzhou.aliyuncs.com"  accessKeyId, accessKeyS
ecret="your id", "your secret" oss = OssAPI(endpoint, accessKeyId,
accessKeySecret)
user_id = mysql_client.fetch_one(sql)#Search for user_id in RDS
#Obtain the user avatar and download it to the corresponding path.
oss.get_object_to_file(bucket, object, your_path/user_id+'.png’)
#Process the uploaded user avatar.
oss.put_object_from_file(bucket, object, your_path/user_id+’.png’)
```

18 KVStore for Redis

18.1 What is KVStore for Redis

KVStore for Redis is an online storage service compatible with the Redis protocol. It supports multiple data types, such as the string, list, set, sorted set, and hash. It also supports advanced features such as transactions and subscribe-publish (Sub/Pub). KVStore for Redis meets persistent storage requirements and provides fast read/write capabilities by using a combined flash memory and hard disk storage architecture.

You can easily deploy and manage KVStore for Redis on the console.

- You can create an instance to initialize a database.
- Before using the KVStore for Redis instance, you must add IP addresses or CIDR blocks used for database access to the whitelist of the target instance.
- You can manage instances on the console.
- You can perform a regularly scheduled backup and recovery or a randomly backup and recovery on the console to ensure data security of the database.
- You can log on to Redis databases through the client tool and manipulate the database by using SQL statements.

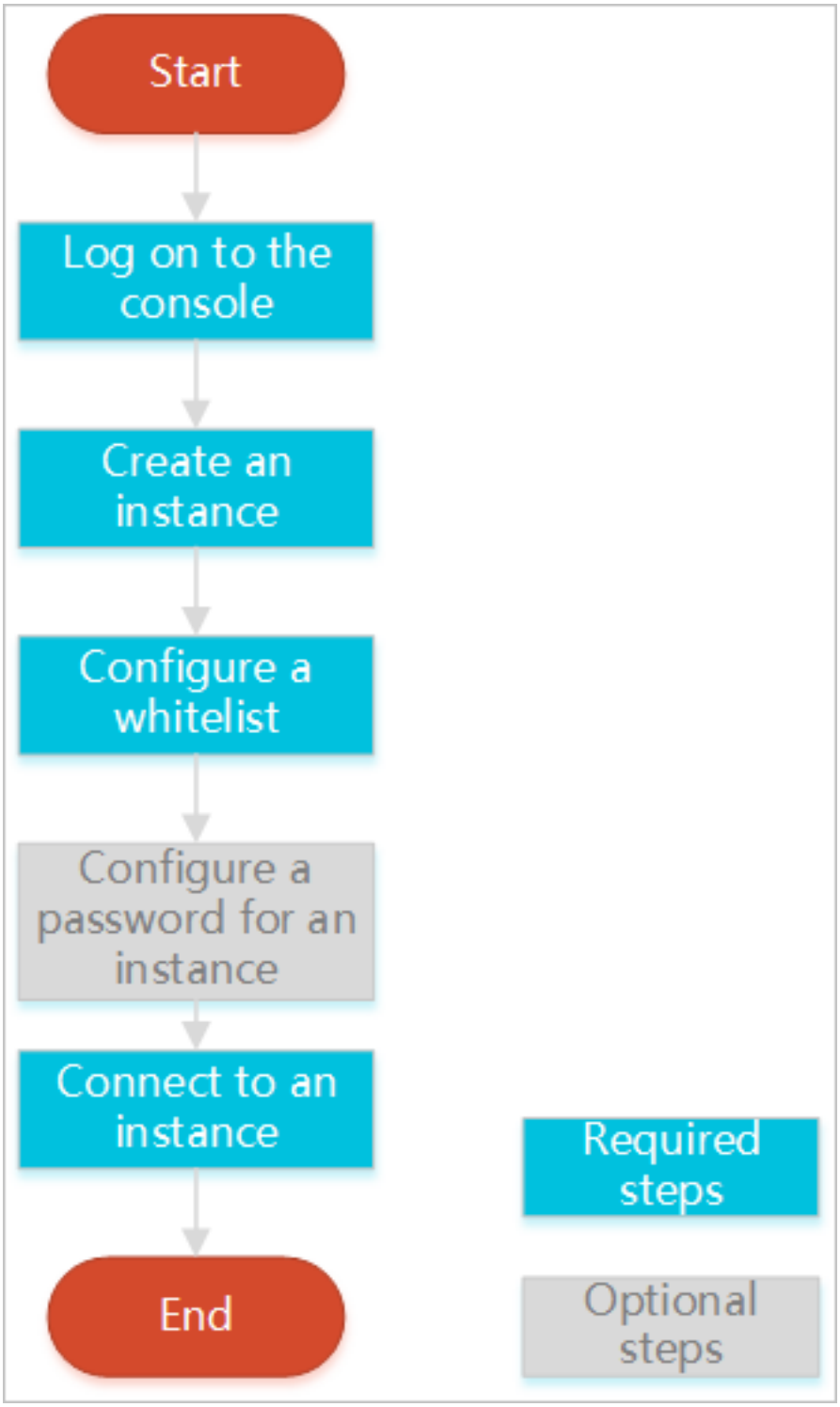
18.2 Quick start

18.2.1 Get started with KVStore for Redis

This topic describes all operations that you can perform on a KVStore for Redis instance, from creating to logging on to a KVStore for Redis instance.

For more information about the procedure, see *Figure 18-1: KVStore for Redis instance operation flow*.

Figure 18-1: KVStore for Redis instance operation flow



- *Log on to the KVStore for Redis console*

  Describes how to log on to the KVStore for Redis console.

- *Create an instance*

Alibaba Cloud databases support two types of networks: classic network and Virtual Private Cloud (VPC). You can create KVStore for Redis instances on different networks.

- *Configure a whitelist*

  Before using a KVStore for Redis instance, you must add the IP addresses or CIDR blocks used for database access to the whitelist of the target instance. This ensures database security and stability.

- *Configure a password for an instance*

  If you do not configure a password for the instance when creating the instance, configure a password on the **Instance Information** tab page.

- *Connect to an instance*

  You can use a client that supports the Redis protocol or use redis-cli to connect to an instance.

18.2.2 Log on to the KVStore for Redis console

Take the Chrome browser as an example to describe how to log on to the KVStore for Redis console through the Apsara Stack console as cloud product users.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*`IP address or domain name address of the Apsara Stack console`*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*`IP address or domain name address of the Apsara Stack console`*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the

minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click **LOGIN** to go to the **Dashboard** page.
5. Choose **Console** > **Database** > **KVStore**.

18.2.3 Create an instance

Alibaba Cloud databases support two types of networks: classic network and Virtual Private Cloud (VPC). You can create KVStore for Redis instances on different networks. This topic describes how to use the KVStore for Redis console to create an instance.

Prerequisites

Make sure that you have obtained an account to log on to the KVStore for Redis console.

To use a VPC Redis instance, you must create a VPC in the same region as the KVStore for Redis instance.

Context

Note:

The network type is specified when the instance is created. It cannot be modified.

Procedure

1. *Log on to the KVStore for Redis console*.
2. In the upper-right corner of the KVStore for Redis tab page, click **Create Instance**. Follow the on-screen tips on the **Create Redis Instance** page to configure instance parameters.

   If you select a VPC Redis instance, you must first create a VPC. For more information about how to create a VPC, see **Create a VPC and a VSwitch** in *VPC User Guide*.

   Table 18-1: KVStore for Redis instance parameters

Area	Parameter	Description
Region	Region	Select a region for the KVStore for Redis instance.
	Zone	Select a zone for the KVStore for Redis instance.
Basic Settings	Department	Select a department for the KVStore for Redis instance.

Area	Parameter	Description
	Project	Select a project for the KVStore for Redis instance.  **Note:** After you select a project, the KVStore for Redis instance is accessible only to the members of the selected project. For more information, see **View project members** in *Apsara Stack Management Console User Guide*.
Instance Specification	Architecture	Select an architecture type for the KVStore for Redis instance. KVStore for Redis provides cluster and standard architectures. The cluster architecture is for high performance business requirements of KVStore for Redis. Running under a single thread mechanism, the standard architecture is recommended for businesses that require a performance capacity of lower than 100,000 QPS. For higher performance, select the cluster architecture.
	Node Type	Select a node type for the KVStore for Redis instance. KVStore for Redis supports the primary/secondary structure.
	Service Plan	Select the standard or premium package. The premium package provides instances with advanced configurations.
	Instance Specification	Select an instance type. The maximum number of connections and maximum intranet bandwidth vary according to instance types.
Network	Network Type	On the Alibaba Cloud platform, a classic network and VPC have the following differences: • Classic network: The cloud services on the classic networks are not isolated. You can configure a security group or whitelist policy for the cloud service to block unauthorized access. • VPC: VPCs can help you build an isolated network environment on Alibaba Cloud. You can customize the routing table, IP address range, and gateway of a VPC. To achieve smooth migration of applications, you can use a leased line or VPN to integrate the on-premises data center and cloud resources on Apsara Stack into a virtual data center.  **Note:** If you select a VPC Redis instance, first you need to create a VPC. For more information about how to create a VPC, see **Create a VPC and a VSwitch** in *VPC User Guide*.

Area	Parameter	Description
Password	Set Password	Configure a password for the instance. You can select **Now** or **Later**. Or you can set a password when you *Reset a password*. The password complexity rules are as follows: • A password must be 8 to 30 characters in length and contain uppercase and lowercase letters, and numbers. • It cannot contain any special character.
Instance Name	Instance Name	Enter a name for the instance. A name must be 2 to 128 characters in length and contain letters , numbers, underscores (_), and hyphens (-). It must start with a letter.

3. Click **Create**.

   After creating the instance, wait for the status of the instance to become **Normal**.

18.2.4 Configure a whitelist

Before using the KVStore for Redis instance, you must add IP addresses or CIDR blocks used for database access to the whitelist of the target instance. This ensures database security and stability.

Context

Correct use of the whitelist improves access security for KVStore for Redis instances. We recommend that you maintain the whitelist on a regular basis.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the Instance ID, or click the  icon in the Actions column corresponding to the instance and choose **View Details** from the shortcut menu.
3. On the **Instance Information** tab page, click **Change Whitelist**.
4. In the **Change Whitelist** dialog box that appears, modify relevant information as prompted.

   IP addresses or CIDR blocks To allow all IP addresses to access the database, set the whitelist to 0.0.0.0/0. To disable database access from all IP addresses, set the whitelist to 127.0.0.1. We recommend that you delete the default IP address 127.0.0.1. Otherwise, the IP addresses added will be invalid.

> **Note:**
> If you enter multiple IP addresses, separate them with commas (do not add spaces before or after each comma), such as 192.168.0.1,172.16.213.9. A maximum of 1000 IP addresses can be added.

5. After setting these parameters, click **OK**.

18.2.5 Connect to an instance

18.2.5.1 Connect to KVStore for Redis instances from a Redis client

18.2.5.1.1 Overview

This topic describes how to connect to KVStore for Redis from a Redis client.

Since the database service provided by KVStore for Redis is entirely compatible with the built-in Redis database service, the connection method is essentially similar. Any Redis-compliant client can access KVStore for Redis services. You can select any Redis client based on your specific application needs.

> **Note:**
> KVStore for RedisKVStore for Redis allows intranet access from Apsara Stack only. You can access the KVStore for Redis instance from the Redis client installed on the ECS instances that share the same node.

For more information about the Redis client, see *http://redis.io/clients*.

18.2.5.1.2 Jedis client

This topic describes how to connect to an instance through a Jedis client.

Download Jedis

Click *Reference Address*.

Example of Jedis single-connection

```
import redis.clients.jedis.Jedis;
public class jedistest {
public static void main(String[] args) {
    try {
        String host = "xx.kvstore.aliyuncs.com";//The IP address for
accessing the host is displayed on the Apsara Stack Management Console
.
        int port = 6379;
```

```
        Jedis jedis = new Jedis(host, port);
        //Authentication information
        jedis.auth("password");//password
        String key = "redis";
        String value = "aliyun-redis";
        //Select a database (The default value is 0).
        jedis.select(1);
        //Configure a key.
        jedis.set(key, value);
        System.out.println("Set Key " + key + " Value: " + value);
        //Obtain the configured key value.
        String getvalue = jedis.get(key);
        System.out.println("Get Key " + key + " ReturnValue: " +
getvalue);
        jedis.quit();
        jedis.close();
    } catch (Exception e) {
        e.printStackTrace();
    }
}
}
```

Example of JedisPool

Configuration file

You can configure the pom configuration file based on the selected client version. Configurations are as follows:

```
<dependency>
<groupId>redis.clients</groupId>
<artifactId>jedis</artifactId>
<version>2.7.2</version>
<type>jar</type>
<scope>compile</scope>
</dependency>
```

Files to be added:

```
import org.apache.commons.pool2. PooledObject;
import org.apache.commons.pool2. PooledObjectFactory;
import org.apache.commons.pool2.impl.DefaultPooledObject;
import org.apache.commons.pool2.impl.GenericObjectPoolConfig;
import redis.clients.jedis.HostAndPort;
import redis.clients.jedis.Jedis;
import redis.clients.jedis.JedisPool;
import redis.clients.jedis.JedisPoolConfig;
```

Example of Jedis-2.7.2

```
JedisPoolConfig config = new JedisPoolConfig();
                //Maximum number of idle connections. You can
customize this parameter. Ensure that the maximum number of idle
 connections does not exceed the maximum connections of each
KVStore for Redis instance.
config.setMaxIdle(200);
```

```
//Maximum number of connections. You can customize this
parameter. Ensure that the maximum number of connections does
 not exceed the maximum connections of each KVStore for Redis
instance.
config.setMaxTotal(300);
config.setTestOnBorrow(false);
config.setTestOnReturn(false);
String host = "*.aliyuncs.com";
String password = "password";
JedisPool pool = new JedisPool(config, host, 6379, 3000,
password);
Jedis jedis = null;
Try {
    jedis = pool.getResource();
    /// ... do stuff here ... for example
    jedis.set("foo", "bar");
    String foobar = jedis.get("foo");
    jedis.zadd("sose", 0, "car");
    jedis.zadd("sose", 0, "bike");
    Set<String> sose = jedis.zrange("sose", 0, -1);
} finally {
    if (jedis ! = null) {
        jedis.close();
    }
}
/// ... when closing your application:
pool.destroy();
```

Examples of Jedis-2.6 and Jedis-2.5

```
JedisPoolConfig config = new JedisPoolConfig();
                //Maximum number of idle connections. You can
customize this parameter. Ensure that the maximum number of idle
 connections does not exceed the maximum connections of each
KVStore for Redis instance.
config.setMaxIdle(200);
//Maximum number of connections. You can customize this
parameter. Ensure that the maximum number of connections does
 not exceed the maximum connections of each KVStore for Redis
instance.
config.setMaxTotal(300);
config.setTestOnBorrow(false);
config.setTestOnReturn(false);
String host = "*.aliyuncs.com";
String password = "password";
JedisPool pool = new JedisPool(config, host, 6379, 3000,
password);
Jedis jedis = null;
boolean broken = false;
try {
    jedis = pool.getResource();
    /// ... do stuff here ... for example
    jedis.set("foo", "bar");
    String foobar = jedis.get("foo");
    jedis.zadd("sose", 0, "car");
    jedis.zadd("sose", 0, "bike");
    Set<String> sose = jedis.zrange("sose", 0, -1);
} catch(Exception e) {
    broken = true;
} finally {
```

```
    if (broken) {
        pool.returnBrokenResource(jedis);
    } else if (jedis ! = null) {
        pool.returnResource(jedis);
    }
}
```

18.2.5.1.3 phpredis client

Use a phpredis client to connect to an instance.

Download phpredis

Click *Reference Address*.

Sample connection code

```
<? php
  /* Replace the following parameter values with the name of the host
and port number used to connect to the instance. */
  $host = "localhost";
  $port = 6379;
  /* Replace the following parameter values with the instance ID and
password used to connect to the instance. */
  $user = "test_username";
  $pwd = "test_password";
  $redis = new Redis();
  if ($redis->connect($host, $port) == false) {
    die($redis->getLastError());
  }
  if ($redis->auth($pwd) == false) {
    die($redis->getLastError());
  }
  /* You can perform operations on the database after authentication.
For more information, see https://github.com/phpredis/phpredis. */
  if ($redis->set("foo", "bar") == false) {
    die($redis->getLastError());
  }
  $value = $redis->get("foo");
  echo $value;
? >
```

18.2.5.1.4 redis-py client

This topic describes how to use the redis-py client to connect to an instance.

Download redis-py

Click *Reference Address*.

Sample connection code

```
#! /usr/bin/env python
#-*-Coding: UTF-8 -*-
import redis
#Replace the following parameter values with the host and the port
number of instance you need to connect to.
```

```
host = 'localhost'
port = 6379
#Replace the following parameter value with the instance password.
pwd = 'test_password'
r = redis.StrictRedis(host=host, port=port, password=pwd)
#You can carry out database operations after you establish a
connection. For more information, see https://github.com/andymccurdy/
redis-py
r.set('foo', 'bar');
print r.get('foo')
```

18.2.5.1.5 C or C++ client

This topic describes how to connect to KVStore for Redis instances through the C or C++ client.

The following example describes how to access the KVStore for Redis instance from the C or C++ client.

Download, compile, and install the C client

```
git clone https://github.com/redis/hiredis.git
cd hiredis
make
sudo make install
```

Compile the test code

```
 #include <stdio.h>
 #include <stdlib.h>
 #include <string.h>
 #include <hiredis.h>
 int main(int argc, char** argv)
     unsigned int j;
     redisContext *c;
     redisReply *reply;
     if (argc < 4) {
         printf("Usage: example xxx.kvstore.aliyuncs.com 6379
instance_id password\n");
         exit(0)
     }
     const char *hostname = argv[1];
     const int port = atoi(argv[2]);
     const char *instance_id = argv[3];
     const char *password = argv[4];
     struct timeval timeout = { 1, 500000 }; // 1.5 seconds
     c = redisConnectWithTimeout(hostname, port, timeout);
     if (c == NULL || c->err) {
         if (c) {
             printf("Connection error: %s\n", c->errstr);
             redisFree(c);
         } else {
             printf("Connection error: can't allocate redis context\n
");
         }
         exit(1);
     }
     /* AUTH */
     reply = redisCommand(c, "AUTH %s", password);
     printf("AUTH: %s\n", reply->str);
```

```
    freeReplyObject(reply);
    /* PING server */
    reply = redisCommand(c,"PING");
    printf("PING: %s\n", reply->str);
    freeReplyObject(reply);
    /* Set a key */
    reply = redisCommand(c,"SET %s %s", "foo", "hello world");
    printf("SET: %s\n", reply->str);
    freeReplyObject(reply);
    /* Set a key using binary safe API */
    reply = redisCommand(c,"SET %b %b", "bar", (size_t) 3, "hello", (
size_t) 5);
    printf("SET (binary API): %s\n", reply->str);
    freeReplyObject(reply);
    /* Try a GET and two INCR */
    reply = redisCommand(c,"GET foo");
    printf("GET foo: %s\n", reply->str);
    freeReplyObject(reply);
    reply = redisCommand(c,"INCR counter");
    printf("INCR counter: %lld\n", reply->integer);
    freeReplyObject(reply);
    /* again ... */
    reply = redisCommand(c,"INCR counter");
    printf("INCR counter: %lld\n", reply->integer);
    freeReplyObject(reply);
    /* Create a list of numbers, from 0 to 9 */
    reply = redisCommand(c,"DEL mylist");
    freeReplyObject(reply);
    for (int j = 0; j < 10; j++) {
        char buf[64];
        snprintf(buf,64,"%d",j);
        reply = redisCommand(c,"LPUSH mylist element-%s", buf);
        freeReplyObject(reply);
    }
    /* Let's check what we have inside the list */
    reply = redisCommand(c,"LRANGE mylist 0 -1");
    if (reply->type == REDIS_REPLY_ARRAY) {
        for (j = 0; j < reply->elements; j++) {
            printf("%u) %s\n", j, reply->element[j]->str);
        }
    }
    freeReplyObject(reply);
    /* Disconnects and frees the context */
    redisFree(c);
    return 0;
```

```
}
```

Compile the code

```
gcc -o example -g example.c -I /usr/local/include/hiredis –lhiredis
```

Test and run the code

```
example xxx.kvstore.aliyuncs.com 6379 instance_id password
```

18.2.5.1.6 .net client

This topic describes how to connect to the KVStore for Redis instance from the .net client.

The following example describes how to use KVStore for Redis through the .net client.

1. Download and use the .net client.

   ```
   git clone https://github.com/ServiceStack/ServiceStack.Redis
   ```

2. Create a *.net* project.
3. Import a file to the client. Read the file on the client and connect to the instance through the .net client. The file is in the library file directory*ServiceStack.Redis/lib/tests*.

Sample test code:

```
using System;
     using System.Collections.Generic;
     using System.Linq;
     using System.Text;
     using System.Threading.Tasks;
     using ServiceStack.Redis;
     namespace ServiceStack.Redis.Tests
     {
         class Program
         {
             public static void RedisClientTest()
             {
                 string host = "127.0.0.1";/*IP address of the access
host*/
                 string password = "password";/*Password*/
                 RedisClient redisClient = new RedisClient(host, 6379
, password);
                 string key = "test-aliyun";
                 string value = "test-aliyun-value";
                 redisClient.Set(key, value);
                 string listKey = "test-aliyun-list";
                 System.Console.WriteLine("set key " + key + " value "
 + value);
                 string getValue = System.Text.Encoding.Default.
GetString(redisClient.Get(key));
                 System.Console.WriteLine("get key " + getValue);
                 System.Console.Read();
             }
             public static void RedisPoolClientTest()
             {
```

```
                string[] testReadWriteHosts = new[] {
                "redis://password@127.0.0.1:6379"/*redis://[password
]@[IP address to access the instance]:[port number]*/
                };
                RedisConfig.VerifyMasterConnections = false;//You
need to set the parameter to
                PooledRedisClientManager redisPoolManager = new
 PooledRedisClientManager(10/*Number of connection pools*/, 10/*
Connection pool timeout time*/, testReadWriteHosts);
                for (int i = 0; i < n; i++) {
                {
                    IRedisClient redisClient = redisPoolManager.
GetClient();//Obtain the connection.
                    RedisNativeClient redisNativeClient = (RedisNativ
eClient)redisClient;
                    redisNativeClient.Client = null;  //You cannot
use client setname for KVStore for Redis. Set Client to null.
                    try
                    {
                        string key = "test-aliyun1111";
                        string value = "test-aliyun-value1111";
                        redisClient.Set(key, value);
                        string listKey = "test-aliyun-list";
                        redisClient.AddItemToList(listKey, value);
                        System.Console.WriteLine("set key " + key +
 " value " + value);
                        string getValue = redisClient.GetValue(key);
                        System.Console.WriteLine("get key " +
getValue);
                        redisClient.Dispose();//
                    }
                    catch (Exception e)
                    {
                        System.Console.WriteLine(e.Message);
                    }
                }
                System.Console.Read();
            }
            static void Main(string[] args)
            {
                //Single connection mode
                RedisClientTest();
                //Connection pool mode
                RedisPoolClientTest();
            }
        }
    }
```

For more information about port use, see *https://github.com/ServiceStack/ServiceStack.Redis*.

18.2.5.1.7 node-redis client

This topic describes how to connect to KVStore for Redis instances using the node-redis client.

1. Install node-redis.

```
npm install hiredis redis
```

2. Connect the node-redis client to KVStore for Redis instances.

```
 var redis = require("redis"),
     client = redis.createClient({detect_buffers: true});
     client.auth("password", redis.print)
```

3. Use KVStore for Redis.

```
// Write the data.
     client.set("key", "OK");
     // Obtain data. A string is returned.
     client.get("key", function (err, reply) {
         console.log(reply.toString()); // print `OK`
     });
     // If a buffer is imported, a buffer is returned.
     client.get(new Buffer("key"), function (err, reply) {
         console.log(reply.toString()); // print `<Buffer 4f 4b>`
     });
     client.quit();
```

18.2.5.2 Connect to KVStore for Redis through redis-cli

You can use the Redis built-in command line interface (CLI) redis-cli to connect to an instance.

Context

Note:

You can access redis-cli from the Intranet only. It does not support internet access. You can access and operate cloud databases after you have installed redis-cli on the ECS instances that share the same node.

- To connect redis-cli to KVStore for Redis, run the following commands:

```
redis-cli -h [instance connection address] -a [password]
```

18.3 Manage instances

18.3.1 Create an instance

Alibaba Cloud databases support two types of networks: classic network and Virtual Private Cloud (VPC). You can create KVStore for Redis instances on different networks. This topic describes how to use the KVStore for Redis console to create an instance.

Prerequisites

Make sure that you have obtained an account to log on to the KVStore for Redis console.

To use a VPC Redis instance, you must create a VPC in the same region as the KVStore for Redis instance.

Context

Note:

The network type is specified when the instance is created. It cannot be modified.

Procedure

1. *Log on to the KVStore for Redis console*.
2. In the upper-right corner of the KVStore for Redis tab page, click **Create Instance**. Follow the on-screen tips on the **Create Redis Instance** page to configure instance parameters.

   If you select a VPC Redis instance, you must first create a VPC. For more information about how to create a VPC, see **Create a VPC and a VSwitch** in *VPC User Guide*.

   Table 18-2: KVStore for Redis instance parameters

Area	Parameter	Description
Region	Region	Select a region for the KVStore for Redis instance.
	Zone	Select a zone for the KVStore for Redis instance.
Basic Settings	Department	Select a department for the KVStore for Redis instance.
	Project	Select a project for the KVStore for Redis instance. **Note:**

Area	Parameter	Description
		After you select a project, the KVStore for Redis instance is accessible only to the members of the selected project. For more information, see **View project members** in *Apsara Stack Management Console User Guide*.
Instance Specification	Architecture	Select an architecture type for the KVStore for Redis instance. KVStore for Redis provides cluster and standard architectures. The cluster architecture is for high performance business requirements of KVStore for Redis. Running under a single thread mechanism, the standard architecture is recommended for businesses that require a performance capacity of lower than 100,000 QPS. For higher performance, select the cluster architecture.
	Node Type	Select a node type for the KVStore for Redis instance. KVStore for Redis supports the primary/secondary structure.
	Service Plan	Select the standard or premium package. The premium package provides instances with advanced configurations.
	Instance Specification	Select an instance type. The maximum number of connections and maximum intranet bandwidth vary according to instance types.
Network	Network Type	On the Alibaba Cloud platform, a classic network and VPC have the following differences: • Classic network: The cloud services on the classic networks are not isolated. You can configure a security group or whitelist policy for the cloud service to block unauthorized access. • VPC: VPCs can help you build an isolated network environment on Alibaba Cloud. You can customize the routing table, IP address range, and gateway of a VPC. To achieve smooth migration of applications, you can use a leased line or VPN to integrate the on-premises data center and cloud resources on Apsara Stack into a virtual data center. **Note:** If you select a VPC Redis instance, first you need to create a VPC. For more information about how to create a VPC, see **Create a VPC and a VSwitch** in *VPC User Guide*.
Password	Set Password	Configure a password for the instance. You can select **Now** or **Later**. Or you can set a password when you *Reset a password*. The password complexity rules are as follows:

Area	Parameter	Description
		• A password must be 8 to 30 characters in length and contain uppercase and lowercase letters, and numbers. • It cannot contain any special character.
Instance Name	Instance Name	Enter a name for the instance. A name must be 2 to 128 characters in length and contain letters , numbers, underscores (_), and hyphens (-). It must start with a letter.

3. Click **Create**.

   After creating the instance, wait for the status of the instance to become **Normal**.

18.3.2 View instance details

After you create an instance, you can view the instance details in Apsara Stack Management Console.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the instance ID, or click the  icon in the Actions column corresponding to the instance and choose **View Details** from the shortcut menu. On the **Instance Information** tab page, you can view the instance details.

   The **Instance Information** tab page contains **Basic Information**, **Configuration**, and **Connection Information**. The information contained in each area is displayed as follows.

Table 18-3: Instance Information tab page

Area	Item
Basic Information	• Instance ID • Name • Status • Region • Department • Project • Created at • Zone • Network Type

Area	Item
	• VPC (only if the instance type is VPC)
Configuration	• Instance Specification • Maximum Connections • Maximum Bandwidth • Maintenance Window • Whitelist
Connection Information	• Address • Port Number • SSL Status (not supported on cluster instances) • SSL Expires at (not supported on cluster instances)

18.3.3 Change the instance name

After you create an instance, you can change the instance name in Apsara Stack Management Console. You can quickly locate the instance through the instance name.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click  in the Actions column corresponding to the instance and choose **Change** from the shortcut menu.
3. In the **Edit Instance Information** dialog box that appears, enter a new instance name and click **OK**.

Note:

- The name must be 2 to 128 characters in length.
- It must start with a letter.
- It can contain letters, numbers, underlines (_), and hyphens (-).

18.3.4 Change instance configurations

KVStore for Redis allows you to change the configurations of an instance.

Context

Note:
Transient disconnections may occur and several errors may be reported when you change instance configurations. We recommend that you perform this operation during off-peak hours.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the target instance ID, or click the  icon in the Actions column corresponding to the instance and choose **Change Instance** from the shortcut menu. On the **Instance Information** tab page, click **Change Instance**.
3. In the **Change Instance** dialog box that appears, select the desired **Instance Specification** and click **OK**.

   `Instance changed` is displayed. You can use the instance only when the status of the instance becomes **Normal**. This process takes a moment to complete.

18.3.5 Configure a whitelist

Before using an instance, you must enable the IP address whitelist. For more information, see *Configure a whitelist*.

18.3.6 Set the O&M time

To ensure the stability of KVStore for Redis instances, the backend system irregularly maintains instances and machines.

Context

Instances enter the **Being Maintained** state before the preset O&M time on the day of maintenance. This guarantees instance stability during the maintenance process. When an instance is in this state, the authorized access to data in the database is not affected. However, change-related functions in the console (for example, configuration change) are temporarily unavailable for this instance, whereas query functions such as performance monitoring are still available.

Note:
After the preset maintenance window time is reached, instances may experience intermittent interruption during maintenance. We recommend that instances be maintained during off-peak hours if possible.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the instance ID, or click the  icon in the Actions column corresponding to the instance and choose **View Details** from the shortcut menu.
3. On the **Instance Information** tab page, click **Change O&M Time**.
4. In the **Change O&M Time** dialog box that appears, select a Maintenance Time and click **OK**.

18.3.7 Enable data transmission encryption

To ensure the security of your instance, you can enable the Secure Sockets Layer (SSL)-encrypted connection after you create the instance.

Context

To enhance link security, you can enable SSL encryption and install SSL CA certificates to necessary application services. SSL encrypts network connection requests at the transport layer, which enhances data security but increases the connection response time.

Note:
You cannot use SSL encryption for cluster instances.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the instance ID, or click the  icon in the Actions column corresponding to the instance and choose **View Details** from the shortcut menu. The **Instance Information** tab page is displayed.
3. Click **Enable SSL**. A message is displayed, indicating that the operation succeeds.
   Wait for a while and refresh the Instance Information tab page. **Disable SSL** and **SSL Certificate Download** will appear on the page, indicating the operation succeeds.

18.3.8 Clear instance data

This topic describes how to clear all data for a KVStore for Redis instance in Apsara Stack Management Console with a single click.

Prerequisites

Note:
This operation clears all data of the instance. The data cannot be restored after being cleared. Use caution when performing the operation.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the  icon in the Actions column corresponding to the instance and choose **Clear** from the shortcut menu.
3. In the **Clear Instance** message that appears, click **OK**.

18.3.9 Reset a password

If you forget your password or do not configure a password when creating an instance, you can reset the password of the instance.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the instance ID, or click the  icon in the Actions column corresponding to the instance and choose **View Details** from the shortcut menu. The **Instance Information** tab page is displayed.
3. Click **Reset Password**. In the **Reset Password** dialogue box that appears, enter the logon password, confirm it, and click **Submit**.

18.3.10 Release an instance

You can release instances through the KVStore for Redis console with a single click. You cannot recover released instances.

Prerequisites

Note:
Exercise caution when you release an instance. Instances are completely deleted if you release them.

Procedure

1. *Log on to the KVStore for Redis console*.

2. Locate the target instance in the instance list, and click  > **Release**.
3. Click **OK** in the **Delete Instance** window.

18.3.11 Set parameters

KVStore for RedisAllows you to set some instance parameters. For more information about the parameters that can be modified, see **Parameter settings** on the KVStore for Redis instance.

Context

KVStore for Redis is completely compatible with the native database service. The setting method of the cloud database parameters is similar to that of the local database parameters. You can modify parameters on the KVStore for Redis console by referring to this example or using other methods such as redis-cli.

For more information about the database parameter descriptions, see the official documentations for the corresponding database version by clicking the following links.

- *redis.conf for Redis 3.0*
- *redis.conf for Redis 2.8*

Procedure

1. *Log on to the KVStore for Redis console*.
2. In the instance list, locate the target instance and click the instance ID or choose  > **Details** to go to the **Instance Information** page.
3. Click the **Parameters** tab.
4. Select the parameter to be modified and choose  > **Edit**.
5. Modify the parameter value and click **Confirm**.

18.4 Backup and recovery

As an increasing number of businesses use KVStore for Redis as their primary persistent storage engine, users need higher reliability for data. KVStore for Redis backup and recovery solutions have increased data reliability in all respects.

18.4.1 Configure automatic backup policies

You can configure an automatic backup policy in the KVStore for Redis console.

Context

An increasing number of applications use KVStore for Redis for persistent storage. Therefore, regular backup mechanisms are required to quickly restore data in the event of misoperation. Apsara Stack uses secondary nodes to back up RDB snapshots to protect the performance of your instance during the backup process. Apsara Stack also provides Apsara Stack Management Console for convenient custom backup configurations.

Note:
You cannot back up or restore data for cluster instances.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the instance ID, or click the  icon in the Actions column corresponding to the instance and choose **View Details** from the shortcut menu.
3. On the **Instance Information** tab page, click the **Backup and Restore** tab.
4. Click the **Backup Settings** tab.
5. Click **Change Settings**. In the **Backup Settings** dialog box that appears, configure the automatic backup cycle and time.

   Note:
   Backup data is retained for seven days by default and cannot be modified.
6. Click **OK**.

18.4.2 Manual backup

In addition to regularly scheduled backup, you can also execute manual backup in the console.

Context

Note:
You cannot back up or restore data for cluster instances.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the instance ID, or click the  icon in the Actions column corresponding to the instance and choose **View Details** from the shortcut menu.
3. On the **Instance Information** tab page, click the **Backup and Restore** tab.
4. Click the **Backups** tab.
5. Click **Create Backup** in the upper-right corner.
6. In the **Back up Instance** message that appears, click **OK**.

> **Note:**
> On the **Backups** tab page, you can select the time range to query historical backup data. Backup data is retained for seven days by default. You can query the historical backup data of the last seven days.

18.4.3 Archive backup data

You can download the backup file of an instance within the last seven days in the console.

Context

Industry regulations and company systems require KVStore for Redis data to be regularly backed up and archived. KVStore for Redis supports archival of backup data: it automatically stores automatic or manual backup files on OSS. KVStore for Redis stores backup files on OSS for seven days. Seven days later, the backup files are automatically deleted.

To archive these backup files for a longer period, you can copy the link in the console and manually download the database backup files to your local machine.

> **Note:**
> You cannot back up or restore data for cluster instances.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the instance ID, or click  in the Actions column corresponding to the instance and choose **View Details** from the shortcut menu.

3. On the **Instance Information** tab page, click the **Backup and Restore** tab.
4. On the **Backups** tab page, locate the backup file you want to download, click the  icon in the Actions column corresponding to the backup file, and choose **Download** from the shortcut menu.
5. In the message that appears, click **OK** to download the file to the default local download path.

18.4.4 Restore data

You can use the data restoration function to minimize damage caused by database misoperation. KVStore for Redis support data restoration from a backup file.

Prerequisites

Note:

- Data restoration is a risky operation. Verify the accuracy of the data before you perform data restoration.
- You cannot back up or restore data for cluster instances.

Ensure that data is backed up before data restoration. After executing a backup operation, KVStore for Redis retains the backup data for seven days by default.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Locate the target instance in the instance list. Click the instance ID, or click the  icon in the Actions column corresponding to the instance and choose **View Details** from the shortcut menu.
3. On the **Instance Information** tab page, click the **Backup and Restore** tab. The **Backups** tab page is displayed by default.
4. Select the time range you want to restore data. Click **Search**. The backup files within that period are listed.

   In order to retrieve backup data, you must have performed a backup operation within the selected time range.

5. Select the target backup file. Click the  icon in the Actions column corresponding to the backup file and choose **Restore** from the shortcut menu.
6. In the Restore message that appears, click **OK**.

18.5 Import data

You can use the redis-cli tool to import the append-only file (AOF) file that contains the KVStore for Redis data to KVStore for Redis.

Context

Redis-cli is the Redis built-in command line interface (CLI). KVStore for Redis allows you to use redis-cli to import existing data to KVStore for Redis for seamless data migration.
Note:

- KVStore for Redis only allows intranet access from Apsara Stack. This solution applies to ECS in Apsara Stack only. If your KVStore for Redis is not deployed on the ECS instance in Apsara Stack, you need to copy the original AOF file to the ECS instance. Then perform the following operations on the ECS.
- Redis-cli is the built-in Redis CLI. If you cannot use redis-cli on ECS, you need to download and install Redis to use redis-cli.

Perform the following steps if you have created a KVStore for Redis instance on ECS in Apsara Stack:

Procedure

1. Enable the AOF function for the current KVStore for Redis instance. Skip this step if this function is already enabled for the instance.

```
# redis-cli -h old_instance_ip -p old_instance_port config set
 appendonly yes
```

2. Use the AOF file to import data to the new KVStore for Redis instance (assume that the generated AOF file is named append.aof).

```
# redis-cli -h aliyun_redis_instance_ip -p 6379 -a password --pipe <
 appendonly.aof
```

Note:

If the AOF function is not required for the original KVStore for Redis instance, run the following command to disable the AOF function after data is imported:

```
# redis-cli -h old_instance_ip -p old_instance_port config set
 appendonly no
```

18.6 Commands supported by KVStore for Redis

KVStore for RedisKVStore for Redis is compatible with Redis 3.0 and can run GEO commands if Redis 3.0 is used. For more information about KVStore for Redis commands, see *http://redis.io/commands*.

Supported command operations

Key	String	Hash	List	Set	SortedSet
DEL	APPEND	HDEL	BLPOP	SADD	ZADD
DUMP	BITCOUNT	HEXISTS	BRPOP	SCARD	ZCARD
EXISTS	BITOP	HGET	BRPOPLPUSH	SDIFF	ZCOUNT
EXPIRE	BITPOS	HGETALL	LINDEX	SDIFFSTORE	ZINCRBY
EXPIREAT	DECR	HINCRBY	LINSERT	SINTER	ZRANGE
MOVE	DECRBY	HINCRBYFLOAT	LLEN	SINTERSTORE	ZRANGEBYSCORE
PERSIST	GET	HKEYS	LPOP	SISMEMBER	ZRANK
PEXPIRE	GETBIT	HLEN	LPUSH	SMEMBERS	ZREM
PEXPTREAT	GETRANGE	HMGET	LPUSHX	SMOVE	ZREMRANGEBYRANK
PTTL	GETSET	HMSET	LRANGE	SPOP	ZREMRANGEBYSCORE
RANDOMKEY	INCR	HSET	LREM	SRANDMEMBER	ZREVRANGE
RENAME	INCRBY	HSETNX	LSET	SREM	ZREVRANGEBYSCORE
RENAMENX	INCRBYFLOAT	HVALS	LTRIM	SUNION	ZREVRANK
RESTORE	MGET	HSCAN	RPOP	SUNIONSTORE	ZSCORE

Key	String	Hash	List	Set	SortedSet
SORT	MSET	-	RPOPLPUSH	SSCAN	ZUNIONSTORE
TTL	MSETNX	-	RPUSH	-	ZINTERSTORE
TYPE	PSETEX	-	RPUSHX	-	ZSCAN
SCAN	SET	-	-	-	ZRANGEBYLEX
OBJECT	SETBIT	-	-	-	ZLEXCOUNT
-	SETEX	-	-	-	ZREMRANGEBYLEX
-	SETNX	-	-	-	-
-	SETRANGE	-	-	-	-
-	STRLEN	-	-	-	-

HyperLogLog	Pub/Sub	Transaction	Connection	Server	Scripting	Geo
PFADD	PSUBSCRIBE	DISCARD	AUTH	FLUSHALL	EVAL	GEOADD
PFCOUNT	PUBLISH	EXEC	ECHO	FLUSHDB	EVALSHA	GEOHASH
PFMERGE	PUBSUB	MULTI	PING	DBSIZE	SCRIPT EXISTS	GEOPOS
-	PUNSUBSCRIBE	UNWATCH	QUIT	TIME	SCRIPT FLUSH	GEODIST
-	SUBSCRIBE	WATCH	SELECT	INFO	SCRIPT KILL	GEORADIUS
-	UNSUBSCRIBE	-	-	KEYS	SCRIPT LOAD	GEORADIUSBYMEMBER
-	-	-	-	CLIENT KILL	-	-
-	-	-	-	CLIENT LIST	-	-
-	-	-	-	CLIENT GETNAME	-	-

HyperLogLog	Pub/Sub	Transaction	Connection	Server	Scripting	Geo
-	-	-	-	CLIENT SETNAME	-	-
-	-	-	-	CONFIG GET	-	-
-	-	-	-	MONITOR	-	-
-	-	-	-	SLOWLOG	-	-

Commands temporarily unavailable

Keys	Server
MIGRATE	BGREWRITEAOF
-	BGSAVE
-	CONFIG REWRITE
-	CONFIG SET
-	CONFIG RESETSTAT
-	COMMAND
-	COMMAND COUNT
-	COMMAND GETKEYS
-	COMMAND INFO
-	DEBUG OBJECT
-	DEBUG SEGFAULT
-	LASTSAVE
-	ROLE
-	SAVE
-	SHUTDOWN
-	SLAVEOF
-	SYNC

Commands not supported by cluster instances

Transaction	Scripting	Connection	Keys	List
DISCARD	EVAL	SELECT	MOVE	BLPOP
EXEC	EVALSHA	-	SCAN	BRPOP
MULTI	SCRIPT EXISTS	-	-	BRPOPLPUSH
UNWATCH	SCRIPT FLUSH	-	-	-
WATCH	SCRIPT KILL	-	-	-
-	SCRIPT LOAD	-	-	-

Commands restricted for cluster instances

Keys	Strings	Lists	Sets	Sorted Sets	HyperLogLog
RENAME	MSETNX	RPOPLPUSH	SINTERSTORE	ZUNIONSTORE	PFMERGE
RENAMENX	-	-	SINTER	ZINTERSTORE	-
-	-	-	SUNIONSTORE	-	-
-	-	-	SUNION	-	-
-	-	-	SDIFFSTORE	-	-
-	-	-	SDIFF	-	-
-	-	-	SMOVE	-	-

Note:

Restricted commands only support scenarios where the operational key is distributed in a single hash slot. You do not have the ability to merge data from multiple hash slots. Because of this, this method ensures that the key for the hash tag you need is distributed to only one hash slot. For example, if there are three keys, key1, aakey, and abkey3, the storage method needs to use {key}1, aa{key}, and ab{key}3 to effectively call the restricted command. . For information how to use the hash tag, see the *Official Redis Documentation*.

19 ApsaraDB for MongoDB

19.1 What is ApsaraDB for MongoDB

ApsaraDB for MongoDB is a stable, reliable, and automatically scalable database service that is fully compatible with MongoDB protocols. The service offers a full range of database solutions, such as disaster recovery, backup, restoration, monitoring, and alarms.

ApsaraDB for MongoDB provides the following features:

- Automatically creates a three-node ApsaraDB for MongoDB replica set and provides ready-to-use advanced functions such as disaster recovery switchover and failover. Users are not aware of the functions.
- Allows you to back up and restore databases with a single click. You can perform conventional database backup and database rollback with a single click on the ApsaraDB for MongoDB console.
- Provides up to 20 performance metrics for monitoring and alarm functions, giving you a full view of database performance.
- Provides visual data management tools for convenient operations and maintenance.

19.2 Instructions

You need to know the limits and guidelines before you use ApsaraDB for MongoDB.

To ensure the stability and security of ApsaraDB for MongoDB instances, pay attention to the limits described in *Table 19-1: ApsaraDB for MongoDB limits*.

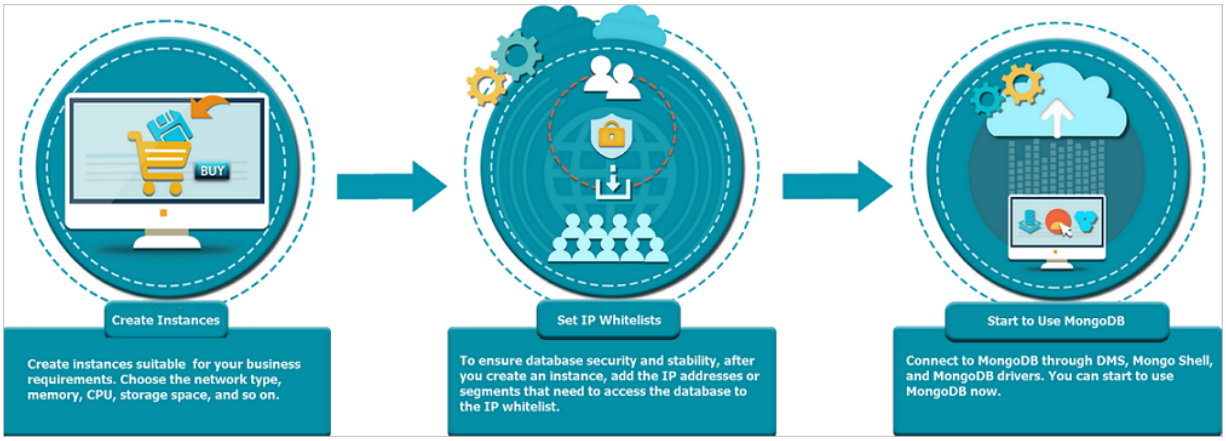
Table 19-1: ApsaraDB for MongoDB limits

Operation	Limit
Create a database copy	• The system automatically creates a three-node replica set. • The primary and secondary nodes are provided to you. The standby node is hidden from you. • Secondary nodes cannot be manually created by users.
Restart a database	Instances must be restarted on the ApsaraDB for MongoDB console.

19.3 Quick start

19.3.1 Procedure

Before you use Alibaba Cloud ApsaraDB for MongoDB for the first time, read *Instructions*. Before using the new instance that you bought, you must complete the following operations:



19.3.2 Log on to the ApsaraDB for MongoDB console

Take the Chrome browser as an example to describe how to log on to the ApsaraDB for MongoDB console through the Apsara Stack console as cloud product users.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super. The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.

- You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click **LOGIN** to go to the **Dashboard** page.
5. In the menu bar, choose **Console** > **Database** > **ApsaraDB for MongoDB**.

19.3.3 Create an instance

This topic describes how to create an instance in the ApsaraDB for MongoDB console.

Prerequisites

Before logging on to the ApsaraDB for MongoDB console, ensure that you have applied for an account.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click **Create Instance** in the upper-right corner. On the **Create MongoDB Instance** page that appears, set the parameters as prompted.

   Table 19-2: Parameter description describes the parameters for creating an instance.

   Table 19-2: Parameter description

Category	Parameter	Description
Basic Settings	**Department**	The department to which the instance belongs.
	Project	The project to which the instance belongs.
	Region	The region where the instance is located.
	Zone	The zone of the instance.
Network Type	**Network Type**	The network type of the instance. An ApsaraDB for MongoDB instance supports the following network types: • `Classic Network`: The cloud services on a classic network are not isolated, and unauthorized access can be blocked only by the security group or whitelist policy of the cloud services.

Category	Parameter	Description
		• VPC: VPCs help you build an isolated network environment on Alibaba Cloud. You can customize the route table, IP address range, and gateway in a VPC. We recommend that you select VPC for enhanced security. To set Network Type to VPC, you must have created a VPC. Alternatively, you can set Network Type to Classic Network and change it to VPC after creating the instance.
Specifications	**Node Specifications**	The specification of the instance. An ApsaraDB for MongoDB instance supports the following specifications: • Three-Member Replica Set: The instance uses dedicated memory and I/O resources while sharing CPU and storage resources with other general instances on the same physical machine. • Exclusive Specifications: The instance uses dedicated CPU, memory, storage, and I/O resources. The performance of this instance can remain stable for a long period of time and will not be affected by the behaviors of other instances on the same physical machine. The highest level of exclusive specifications is **Dedicated Hosts**. This specification type allows an instance to use all resources of a physical machine exclusively. • Dedicated Hosts
	Storage Space	The storage space of the instance. The storage space contains the space for data, system files, binlog files, and transaction files.
Password Settings	**Set Password**	The password for logging on to the ApsaraDB for MongoDB instance. You can select **Now** to set the password immediately or **Later** to set the password later by using the password reset feature. For more information, see *Reset a password*. The password must meet the following requirements:

Category	Parameter	Description
		• Consists of English letters, digits, or underscores (_). • Contains 6 to 32 characters.
Instance Name	**Instance Name**	The instance name. It must be 2 to 256 characters in length and start with an English letter. It can contain English letters, underscores (_), hyphens (-), and digits.

3. After the configurations are complete, click **Create**.

19.3.4 Set a whitelist

Before you use an ApsaraDB for MongoDB instance, you need to add IP addresses or IP segments used for database access to the whitelist of the instance. Configuring the whitelist improves database security and stability. Correct use of the whitelist can enhance access security protection for ApsaraDB for MongoDB. We recommend that you maintain the whitelist regularly.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. After a new instance is created, the system automatically adds the IP address 0.0.0.0/0 to the **default** whitelist group. When the IP address 0.0.0.0/0 is on the whitelist, the instance is accessible from any IP address. To secure your database, delete the IP address 0.0.0.0/0 from the whitelist.

When the IP address 127.0.0.1 is on the whitelist, no IP addresses or IP address segments are allowed to access the instance. Check that the IP address 127.0.0.1 is not on the whitelist before you add any IP address of IP address segment for accessing the instance.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Security Control** > **Whitelist Settings** to go to the **Whitelist** page.
4. On the **Whitelist**page, click **Modify Whitelist** in the upper-right area. In the displayed **Allow Access to IP Addresses** dialog box, follow the on-screen tips to configure parameters.

   Enter IP addresses that are allowed to access the instance. Separate IP addresses with commas (,).

> **Note:**
> When you enter multiple IP addresses, separate them with commas (no space before or after each comma), for example, 192.168.0.1,172.16.213.9.

5. After you configure the parameters, click **Confirm**.

19.3.5 Obtain the seven elements required to connect to an instance

ApsaraDB for MongoDB provides connection addresses for two nodes in a three-node replica set. You can use the addresses to access an ApsaraDB for MongoDB instance. This section describes how to obtain the elements to connect to an instance.

Context

To access an ApsaraDB for MongoDB instance, obtain the following seven elements:

- Instance username
- Password
- Replica set name
- Domain name addresses and port numbers of the two nodes

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. In the left-side navigation pane, click **Database Connections**. On the **Network Information** page, six elements are displayed, including username, replica set name, and domain name addresses and port numbers of the two nodes. *Figure 19-1: Network information* shows the six elements.

   Table 19-3: Element description describes the elements that are used to connect to an instance.

Figure 19-1: Network information

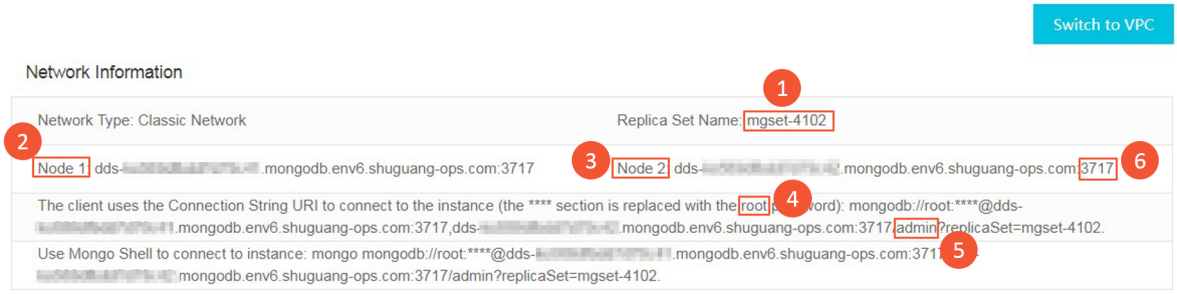


Table 19-3: Element description

Element	Description
Replica set name	Marked as 1 in the previous figure.
Name of Node 1	Marked as 2 in the previous figure.
Name of Node 2	Marked as 3 in the previous figure.
Default account for initial logon to a database: root	Marked as 4 in the previous figure.
Default account name for database connection: admin	Marked as 5 in the previous figure.
Database connection port: 3717	Marked as 6 in the previous figure.

Note:

The password for database connection is set when you create an instance. For information about how to change this password, see *Reset a password*.

19.3.6 Use Mongo shell to connect to an instance

You can create an instance, configure the whitelist, and obtain the seven elements required for instance connection. This section describes how to use Mongo shell to connect to an ApsaraDB for MongoDB instance.

Prerequisites

- Before you use Mongo shell to connect to an ApsaraDB for MongoDB instance, you need to check that Mongo shell and the instance you want to connect to are deployed on the ECS instances within the same region and use the same type of networks.

- You must use Mongo shell 3.0 or later versions to connect to an ApsaraDB for MongoDB instance. Otherwise, authentication fails.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. In the left-side navigation pane, click **Database Connections**. On the **Network Information** page, six elements are displayed, including username, replica set name, and domain name addresses and port numbers of the two nodes.

   For more information about how to obtain the elements, see *Obtain the seven elements required to connect to an instance*.
4. On the ECS, run the mongo command to connect to an ApsaraDB for MongoDB instance.

   Exampe:

   ```
   mongo --host dds-xxxx.mongodb.rds.aliyuncs.com:3717 -u root -p
   123456 --authenticationDatabase admin
   ```

19.4 Manage instances

19.4.1 Create an instance

This topic describes how to create an instance in the ApsaraDB for MongoDB console.

Prerequisites

Before logging on to the ApsaraDB for MongoDB console, ensure that you have applied for an account.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click **Create Instance** in the upper-right corner. On the **Create MongoDB Instance** page that appears, set the parameters as prompted.

   Table 19-4: Parameter description describes the parameters for creating an instance.

   Table 19-4: Parameter description

Category	Parameter	Description
Basic Settings	**Department**	The department to which the instance belongs.
	Project	The project to which the instance belongs.

Category	Parameter	Description
	Region	The region where the instance is located.
	Zone	The zone of the instance.
Network Type	**Network Type**	The network type of the instance. An ApsaraDB for MongoDB instance supports the following network types: • `Classic Network`: The cloud services on a classic network are not isolated, and unauthorized access can be blocked only by the security group or whitelist policy of the cloud services. • VPC: VPCs help you build an isolated network environment on Alibaba Cloud. You can customize the route table, IP address range, and gateway in a VPC. We recommend that you select VPC for enhanced security. To set Network Type to VPC, you must have created a VPC. Alternatively, you can set Network Type to Classic Network and change it to VPC after creating the instance.
Specifications	**Node Specifications**	The specification of the instance. An ApsaraDB for MongoDB instance supports the following specifications: • Three-Member Replica Set: The instance uses dedicated memory and I/O resources while sharing CPU and storage resources with other general instances on the same physical machine. • Exclusive Specifications: The instance uses dedicated CPU, memory, storage, and I/O resources. The performance of this instance can remain stable for a long period of time and will not be affected by the behaviors of other instances on the same physical machine. The highest level of exclusive specifications is **Dedicated Hosts**. This specification type allows an instance to use all resources of a physical machine exclusively. • Dedicated Hosts

Category	Parameter	Description
	Storage Space	The storage space of the instance. The storage space contains the space for data, system files, binlog files, and transaction files.
Password Settings	**Set Password**	The password for logging on to the ApsaraDB for MongoDB instance. You can select **Now** to set the password immediately or **Later** to set the password later by using the password reset feature. For more information, see *Reset a password*. The password must meet the following requirements: • Consists of English letters, digits, or underscores (_). • Contains 6 to 32 characters.
Instance Name	**Instance Name**	The instance name. It must be 2 to 256 characters in length and start with an English letter. It can contain English letters, underscores (_), hyphens (-), and digits.

3. After the configurations are complete, click **Create**.

19.4.2 View instance details

You can view the details of an instance, such as the basic information, internal network connection information, running status, and configurations. This section describes how to view instance details.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. You can go to the Instance Details page in either of the following ways:
   - Click an instance ID to go the **Basic Information** page.
   - In the Actions column of the target instance, click > **Query Details**. On the **Basic Information** page, view basic information about the instance.

19.4.3 Restart an instance

You can manually restart an instance when the number of connections exceeds the threshold or any performance issue occurs on the instance. This section describes how to restart an instance.

Context

Note:

A restart will disconnect the instance. Make appropriate service arrangements before you restart an instance and take caution when you perform this operation.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information**.
3. In the Actions column of the target instance, click > **Restart Instance**. In **Restart MongoDB**, click **Confirm** to restart the instance.

19.4.4 Change specifications

You can change the specifications of an instance, such as the memory and storage space, if the specifications are too high or cannot meet the performance requirements of an application.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. In the Actions column of the target instance, click > **Change Specifications**. In the displayed **Change MongoDB Specifications** dialog box, configure the parameters.

   When you change specifications, you can set **Node Specifications** and **Storage Space** of an instance.
3. Specify the specifications and click **Confirm**.

19.4.5 Switch to VPC

ApsaraDB for MongoDB supports classic networks and Virtual Private Clouds (VPCs). You can switch between two types of networks as required.

Context

The differences between a classic network and a VPC are outlined as follows:

- Classic network: The cloud service in a classic network is not isolated at the network layer . Unauthorized access is blocked only by the security group or whitelist policy of the cloud service.
- VPC: A VPC helps you to build an isolated network environment on Alibaba Cloud. You can customize the route table, IP address segment, and gateway on a VPC. In addition, you can combine your data center and cloud resources in Apsara Stack into a virtual data center through a leased line or VPN to migrate applications to the cloud smoothly.

> **Note:**
> To use a VPC to create an ApsaraDB for MongoDB instance, make sure that the instance and VPC are in the same region.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. In the left-side navigation pane, click **Database Connection** to go to the **Network Information** page.
4. On the **Network Information** page, click **Switch to VPC**. The **Switch to VPC** page is displayed.
5. Specify a VPC and the associated VSwitch based on the description on the **Switch to VPC** page. Then, click **Confirm**.

19.4.6 Change an instance name

To facilitate management, you can change instance names. This topic describes how to change the name of an instance.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click  in the Actions column corresponding to the target instance and choose **Change Instance Name** from the shortcut menu. The **Change Instance Name** page appears.
3. In the **`Instance Name`** field, enter a new instance name.
4. Click **OK**.

   > **Note:**
   > - The instance name must start with an English letter.
   > - The instance name must be 2 to 256 characters in length. It can contain English letters, underscores (_), hyphens (-),
   > - and digits.

5. After the parameters are set, click **OK**.

19.4.7 Reset a password

This section describes how to reset your password on the ApsaraDB for MongoDB console as needed.

Context

> **Note:**
> For your data security, we recommend that you change your password periodically.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. On the **Basic Information** page, click **Reset Password** in the upper-right area and configure parameters in the displayed **Reset Password** dialog box.

   Table 19-5: Password resetting parameters describes the parameter configurations.

   Table 19-5: Password resetting parameters

Parameter	Description
Logon Password	The password can be 6 to 32 characters in length and can contain letters, digits, and underscores (_).
Confirm Password	The password can be 6 to 32 characters in length and can contain letters, digits, and underscores (_).

4. After you configure the parameters, click **OK**.

19.4.8 Release an instance

You can manually release an instance as needed. This section describes how to manually release an instance.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. In the Actions column of the target instance, click and select **Delete**. In **Delete MongoDB**, click **Confirm**.

19.5 Security

19.5.1 Set a whitelist

Before you use an ApsaraDB for MongoDB instance, you need to add IP addresses or IP segments used for database access to the whitelist of the instance. Configuring the whitelist improves database security and stability. Correct use of the whitelist can enhance access security protection for ApsaraDB for MongoDB. We recommend that you maintain the whitelist regularly.

Context

The system creates a default whitelist group for each instance. This whitelist group can be modified or cleared, but cannot be deleted. After a new instance is created, the system automatically adds the IP address 0.0.0.0/0 to the **default** whitelist group. When the IP address 0.0.0.0/0 is on the whitelist, the instance is accessible from any IP address. To secure your database, delete the IP address 0.0.0.0/0 from the whitelist.

When the IP address 127.0.0.1 is on the whitelist, no IP addresses or IP address segments are allowed to access the instance. Check that the IP address 127.0.0.1 is not on the whitelist before you add any IP address of IP address segment for accessing the instance.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Security Control** > **Whitelist Settings** to go to the **Whitelist** page.
4. On the **Whitelist**page, click **Modify Whitelist** in the upper-right area. In the displayed **Allow Access to IP Addresses** dialog box, follow the on-screen tips to configure parameters.

   Enter IP addresses that are allowed to access the instance. Separate IP addresses with commas (,).

   > **Note:**
   > When you enter multiple IP addresses, separate them with commas (no space before or after each comma), for example, 192.168.0.1,172.16.213.9.

5. After you configure the parameters, click **Confirm**.

19.5.2 Audit logs

Context

Audit logs record all operations that a client performs on the connected database. They provide reference for fault analysis, behavior analysis, and security auditing. You can perform log auditing to obtain information about data execution for analysis purposes. Audit logs are necessary for the monitoring of core businesses such as Finance Cloud.

Note:
Audit logs are stored for seven days, and are automatically cleared after seven days.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose**Security Control** > **Audit Log**to go to the **Audit Log** page.
4. On the **Audit Log** page, query or export files.
   - **Query**: Query audit logs based on the time range, database name, database account name, or executed statement.
   - **File List**: Click this button to list audit log files.
   - **Export File**: Click this button to export audit log files.

19.6 Monitoring information

The ApsaraDB for MongoDB console provides abundant performance metrics for you to conveniently check the running status of instances. You can check instance monitoring data on the ApsaraDB for MongoDB console.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. In the left-side navigation pane, select **Monitoring Information**.

   You can select a time range to query historical metrics. *Table 19-6: Metrics* describes the various metrics.

Table 19-6: Metrics

Metric	Description	Monitoring frequency	Monitoring period
CPU Utilization	The instance CPU utilization	300s	30 days
Memory Utilization	The instance memory utilization	300s	30 days
IOPS Usage	The IOPS used by the instance, including: • Data disk IOPS • Log disk IOPS	300s	30 days
IOPS Utilization	The percentage of the IOPS volume used by the instance to the maximum available IOPS volume	300s	30 days
Disk Space Usage	The total disk space used by the instance, including: • Total used space • Data size • Log size	300s	30 days
Disk Space Utilization	The percentage of the total space used by the instance to the maximum available space permitted by the specifications	300s	30 days
opcounters	Operation QPS count on the instance, including: • The number of insert operations • The number of query operations	300s	30 days

Metric	Description	Monitoring frequency	Monitoring period
	• The number of delete operations • The number of update operations • The number of getmore operations • The number of command operations		
connections	The current number of connections to the instance	300s	30 days
cursors	The number of cursors currently used by the instance, including: • Number of currently opened cursors • Number of expired cursors	300s	30 days
network	The network traffic of the instance, including: • Inbound traffic • Outbound traffic • The number of processed requests	300s	30 days
globalLock	The length of the instance queue waiting for global lock, including: • Length of the instance queue	300s	30 days

Metric	Description	Monitoring frequency	Monitoring period
	waiting for global read lock • Length of the instance queue waiting for global write lock • Length of the instance queue waiting to perform operations on the global lock		
wiredTiger	The cache indicators of the wiredTiger engine of an instance, including: • Volume of data read to the cache • Capacity of the disk with data written from the cache • Maximum available disk capacity configured	300s	30 days

19.7 Backup and recovery

19.7.1 Automatic backup

ApsaraDB for MongoDB allows you to specify backup settings and automatically backs up data based on the settings

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.

3. In the left-side navigation pane, choose Backup and Restore > Backup Management. On the displayed **Backups** page, click the **Backup Settings** tab and click **Set**. On the displayed **Backup Settings** page, specify parameters.

   Table 19-7: Backup policy parameters describes the parameter configurations.

   Table 19-7: Backup policy parameters

Parameter	Description
Retention Period (Days)	Number of days for retaining data backups. The value range is from 1 to 30 days. Default value: 7 days.
Backup Period	One or multiple days in a week.
Backup Time	Any period of time in a day, in hours.

4. After you complete the configuration, click **Confirm**.

19.7.2 Back up an instance manually

This section describes how to manually back up an ApsaraDB for MongodDB instance.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Backup and Restore** > **Backup Management.** The **Backup and Restore** page is displayed.
4. On the **Backup and Restore** page, click the **Backups** tab.
5. Click **Back Up Instance**. In **Back Up Instance**, click **OK**.

19.7.3 Search for backups

This section describes how to search for an instance backup.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Backup and Restore** > **Backup Management.**The **Backup and Restore** page is displayed.

4. On the **Backup and Restore** page, click the **Backups** tab, specify a time range, and click **Search** to search for the backups generated in the specified time range.

   Click  in the leftmost column of a backup list to take the following operations:

   - Download: Download the backup files that are generated in the specified time range. For more information, see *Backup download*.
   - Data Restore: Restore data from the backup files that are generated in the specified time range. For more information, see *Restore data*.
   - Create Instance from Backup Point: Create an instance from a specified backup point. For information, see *Create an instance from a backup point*. For information about how to create an instance, see *Create an instance*.

19.7.4 Restore data

The data restore feature minimizes losses caused by misoperations on the database. Apsara for MongoDB allows you to restore data from a backup set.

Context

> **Note:**
> The Apsara for MongoDB rollback operation will overwrite the data. After a rollback, the data cannot be restored to the time point before the rollback. Please perform a rollback with caution.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Backup and Recovery** > **Backup Management**. The **Backup and Recovery** page appears.
4. Click the **Backup List** tab.
5. In the Actions column corresponding to the target backup list, click  and choose **Data Recovery** from the shortcut menu. In the **Data Recovery** dialog box that appears, click **OK**.

19.7.5 Backup download

ApsaraDB for MongoDB allows you to download backup files on the ApsaraDB for MongoDB console.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. In the left-side navigation pane, select **Backup and Restore** > **Backup Management** to go to the **Backup and Restore** page.
4. Click the **Backups** tab.
5. In the Actions column of the backup list, click **Download**.
6. In the displayed **Download** dialog box, click **Confirm** to download the backup file to a local device.

19.7.6 Create an instance from a backup point

You can use a backup file to create a new instance as needed. The new instance contains all the data in the backup set.

Context

Note:
The storage space of the new instance must be equal to or greater than that of the source instance.

Procedure

1. *Log on to ApsaraDB for MongoDB*.
2. Click an instance ID to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Backup and Restore** > **Backup Management.** The **Backup and Restore** page is displayed.
4. Click the **Backups** tab.
5. In the Actions column of the target backup , click and select **Create Instance from Backup Point**. In **Create Instance from Backup Point**, click **OK** to go to the **Create Instance** page.

   For more information about how to create an instance, see *Create an instance*.

20 KVStore for Memcache

20.1 What is KVStore for Memcache

KVStore for Memcache is a memory-based cache service that supports high-speed access to large volumes of small data. KVStore for Memcache can greatly cut down the load on back-end storage, and speed up the response of websites and applications.

KVStore for Memcache supports the key-value data structure. KVStore for Memcachecan communicate with clients compatible with the Memcached protocol.

KVStore for Memcache supports out-of-the-box deployment. It also relieves the loads of dynamic Web applications on databases through the cache service, thus improving the overall response speed of the website.

Similar to local self-built Memcached databases, KVStore for Memcache is also compatible with the Memcached protocol and user environments. You can use ApsaraDB for Memcache directly. The difference is that the hardware and data of ApsaraDB for Memcache are deployed on the cloud, which provides complete infrastructure, network security, and system maintenance services.

20.2 Limits

Before using KVStore for Memcache, you need to understand the limits listed in the following table .

Limit	Description
Data type	KVStore for Memcache supports only data formatted as key-value pairs and does not support complex data types such as array, map, and list.
Data reliability	KVStore for Memcache stores data in the memory, and does not guarantee that the cached data is never lost. Therefore, KVStore for Memcache is unsuitable for storing data that requires high consistency.
Data size	KVStore for Memcache supports a maximum of 1 KB in key size and 1 MB in value size for a single piece of cached data. KVStore for Memcache is unsuitable for storing sizable data.
Transaction support	KVStore for Memcache does not support transactions. Therefore, KVStore for Memcache is unsuitable for storing transaction data. Such data must be written directly to the database.

Limit	Description
Scenarios	When data access traffic is evenly distributed, and there is no obvious hotspot or less popular data, a large number of access requests cannot hit the cached data in KVStore for Memcache. Therefore, KVStore for Memcache does not effectively function as the database cache. You must give full consideration to the data access requirements of the business model when selecting the database cache.
Data deletion policy	Each key in KVStore for Memcache expires at a user-defined time. After expiration, the key becomes inaccessible. The space occupied by the expired key is not recycled immediately after expiration, but is recycled at 02 :00 every day.
Data expiration policy	Like open-source Memcached, KVStore for Memcache adopts the Least Recently Used (LRU) algorithm to determine whether data expires. Expired data is not deleted and the space occupied by the expired data is not recycled immediately after expiration, but is recycled by a background program periodically.
Connection processing	The KVStore for Memcache server does not automatically close idle client connections.
Data expiration	We recommend that you control and manage the key expiration time.

20.3 Quick start

20.3.1 Start to use KVStore for Memcache

This topic describes how to perform a series of operations from creating an instance to logging on to a database. This helps you understand how to operate a KVStore for Memcache instance.

- *Log on to the KVStore for Memcache console*

  This topic describes how to log on to the KVStore for Memcache console.

- *Create an instance*

  KVStore for Memcache supports two types of networks: classic network and VPC. You can create KVStore for Memcache instances of different network types.

- *Set a whitelist*

  To ensure database security and stability, you need to add IP addresses or IP address segments used for database access to the whitelist of the destination instance before using KVStore for Memcache.

- *Reset a password*

If you did not configure a password for the instance when you created the instance, configure a password on the **Instance Information** page.

- *Connect to your instance from a client*

  You can use a client that supports the Memcached protocol to connect to an instance.

20.3.2 Log on to the KVStore for Memcache console

This topic describes how to log on to the KVStore for Memcache console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.
5. In the top navigation bar, choose **Console** > **Database** > **KVStore**.
6. Click the **KVStore for Memcache** tab.

20.3.3 Create an instance

KVStore for Memcache supports two types of networks: classic network and VPC. You can create KVStore for Memcache instances of different network types. This topic describes how to create an instance in the KVStore for Memcache console.

Prerequisites

- At least one ECS instance is required for activating KVStore for Memcache.
- To create a KVStore for Memcache instance of the VPC type, you must first create a VPC. Then, create the instance in the same region as the VPC.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. Click **Create Instance** in the upper-right corner.
3. In the **Create Memcache Instance** dialog box that appears, select a network type and complete other settings.

Table 20-1: Parameter description

Category	Parameter	Description
Region	Region	The region where the KVStore for Memcache instance is located. KVStore for Memcache allows only intranet access. Make sure the KVStore for Memcache instance and ECS are in the same region.
	Zone	The zone where the KVStore for Memcache instance is located. KVStore for Memcache allows only intranet access. Make sure the KVStore for Memcache instance and ECS are in the same zone.
Basic Settings	Department	The department to which the KVStore for Memcache instance belongs.
	Project	The project to which the KVStore for Memcache instance belongs. **Note:** After a project is selected, the KVStore for Memcache instance is accessible only to the members of the selected project. For more

Category	Parameter	Description
		information, see **View project members** in *Apsara Stack Console User Guide*.
Instance Specification	Instance Specification	The instance specification. The maximum number of connections and maximum intranet bandwidth vary depending on different instance specifications.
Network	Network Type	The network type of the instance. On the Alibaba Cloud platform, a classic network and a VPC have the following differences: • Classic network: The cloud services on a classic network are not isolated. Unauthorized access can be blocked only by the security group or whitelist policy of the cloud services. • VPC: VPCs help you build an isolated network environment in Alibaba Cloud. You can customize the routing table, IP address range, and gateway in a VPC. In addition, you can combine your on-premises IDC with cloud resources in the Alibaba Cloud VPC through a leased line or VPN to migrate applications smoothly to the cloud. If you want to set the network type to VPC, you must first create a VPC. For more information, see **Create a VPC and a VSwitch** in *VPC User Guide*.
Password	Set Password	The password for accessing the instance. You can select **Configure Now** to set the password immediately or **Configure After Creation** to set the password later by using the password reset feature. For more information, see *Reset a password*. The password complexity rules are as follows: • A password must be 8 to 30 characters in length. • It can contain uppercase letters, lowercase letters, and digits at the same time. It cannot contain special characters.
Instance Name	Instance Name	The instance name. The instance name must contain 2 to 128 characters in length. The name cannot contain special characters such as at signs (@), forward slashes (/), colons (:), equal signs (=), double quotation marks

Category	Parameter	Description
		(''), angle brackets (<>), square brackets ([]), curly brackets ({}), or spaces.

4. Click **Create**.

   After creating the instance, wait until the instance status becomes **Normal**.

20.3.4 Set a whitelist

To ensure database security and stability, you need to add IP addresses or IP address segments used for database access to the whitelist of the destination instance before using KVStore for Memcache. This topic describes how to set a whitelist.

Context

Correct use of the whitelist improves access security for KVStore for Memcache. We recommend that you maintain the whitelist on a regular basis.

Note:

- The ECS instance whose IP address is added to the whitelist must be in the same region as the KVStore for Memcache instance.
- To enable applications to access multiple KVStore for Memcache instances from the same ECS instance, bind the IP address of the ECS instance to multiple KVStore for Memcache instances.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click an instance ID or choose > **Details** from the shortcut menu.

   The **Instance Information** page appears.
3. Click the **Security Settings** tab.
4. Click **Add Whitelist Group** or the modify icon next to the default whitelist group.
5. In the **Modify Whitelist** dialog box that appears, set parameters as prompted.

   Enter the IP addresses or IP address segments that can access the KVStore for Memcache instance. To allow all IP addresses to access the instance, set the whitelist to 0.0.0.0/0. To disable instance access from all IP addresses, set the whitelist to 127.0.0.1. We recommend

that you delete the default IP address 127.0.0.1 from the whitelist. If you do not delete this IP address, other IP addresses or IP address segments that you add will not take effect.

Note:

If you enter multiple IP addresses or IP address segments, separate them with commas (no space before or after each comma), such as 192.168.0.1,172.16.213.9. For a single instance, a maximum of 1,000 IP addresses can be added.

6. After the parameters are set, click **OK**.

20.3.5 Connect to an instance from a client

20.3.5.1 Overview

Any client compatible with the Memcached protocol can access KVStore for Memcache. Each Memcached client has its own characteristics. You can select any Memcached client that supports Simple Authentication and Security Layer (SASL) and Memcached Binary Protocol based on your application characteristics.

The Memcached clients described in the following sections can interact smoothly with KVStore for Memcache, and therefore are recommended for use.

Note:

The third-party open-source clients described in the following sections are not provided by Alibaba Cloud and may contain defects. Developers must ensure the quality of the clients on their own. Alibaba Cloud is not held liable for any direct or indirect faults or losses arising from the clients.

20.3.5.2 Java: Spymemcache

Use Java: Spymemcache to connect to an instance.

Context

Download a client

- *Client download URL*
- *About the client*
- *Client version*

Java sample code

Procedure

1. Prepare the Java development environment. Log on to an existing Alibaba Cloud ECS instance and install Java Development Kit (JDK) and a commonly used integrated development environment (IDE) such as Eclipse on the instance.
   - Java JDK *Download URL*
   - Eclipse (*Download URL 1*, *Download URL 2*)
2. The first sample code is as follows. Copy the Java code to the Eclipse project.

   Note:

   You must download a JAR package from a third party to call the ApsaraDB for Memcache cache service. Otherwise, you cannot compile the code. With this JAR package added, the code can be compiled.

   OcsSample1.java sample code (username and password required)

```
import java.io.IOException;
 import java.util.concurrent.ExecutionException;
 import net.spy.memcached.AddrUtil;
 import net.spy.memcached.ConnectionFactoryBuilder;
 import net.spy.memcached.ConnectionFactoryBuilder.Protocol;
 import net.spy.memcached.MemcachedClient;
 import net.spy.memcached.auth.AuthDescriptor;
 import net.spy.memcached.auth.PlainCallbackHandler;
 import net.spy.memcached.internal.OperationFuture;
 public class OcsSample1 {
      public static void main(String[] args) {
                final String host = "xxxxxxxx.m.yyyyyyyyyy.ocs.
aliyuncs.com";//"Internal network address" displayed on the console.
                final String port = "11211"; //Default port: 11211.
No changes required.
                final String username = "xxxxxxxxx";//"Access
account" displayed on the console.
                final String password = "my_password";//"Password"
provided in the e-mail.
                MemcachedClient cache = null;
                try {
                          AuthDescriptor ad = new AuthDescriptor(new
String[]{"PLAIN"}, new PlainCallbackHandler(username, password));
                          cache = new MemcachedClient(
                                                new ConnectionFactoryBui
lder().setProtocol(Protocol.BINARY)
                                    .setAuthDescriptor(ad)
                                    .build(),
                                    AddrUtil.getAddresses(host + ":" +
port));
                          System.out.println("OCS Sample Code");
                          //Save a value with the "ocs" key to
ApsaraDB for Memcache to facilitate data verification and
reading.
```

```
                        String key = "ocs";
                        String value = "Open Cache Service,  from
www.Aliyun.com";
                        int expireTime = 1000; // Expiration time,
in
 seconds. The countdown starts from when data is written. After
expireTime
 elapses, the data expires and cannot be read.
                        OperationFuture<Boolean> future = cache.set
(key, expireTime, value);
                        future.get();  // The spymemcached set()
method is asynchronous. The future.get()
operation starts after the cache.set() operation is completed. You
can also choose to execute both operations at the same time.
                        //Save several
 values to ApsaraDB for Memcache and you can view the statistics
on the ApsaraDB for Memcache console.
                        for(int i=0;i<100;i++){
                              key="key-"+i;
                              value="value-"+i;
                              //Perform the Set operation and save

the value to the cache.
                              expireTime = 1000; // Expiration
time,
in seconds.
                              future = cache.set(key, expireTime,
value);
                              future.get();  //  Make sure that
the previous (cache.set()) operation has
 been completed.
                        }
                        System.out.println("Set operation completed
!") ;
                           //Perform the Get operation and read the
 value with the "ocs" key from
the cache.
                        System.out.println("Get operation: "+cache.
get(key));
                              } catch (IOException e) {
                                    e.printStackTrace();
                              } catch (InterruptedException e) {
                                    e.printStackTrace();
                              } catch (ExecutionException e) {
                                    e.printStackTrace();
                              }
                              if (cache ! = null) {
                                    cache.shutdown();
                              }
      }//eof
 }
```

OcsSample2.java sample code (username and password not required)

```
import java.io.IOException;
 import java.util.concurrent.ExecutionException;
 import net.spy.memcached.AddrUtil;
 import net.spy.memcached.BinaryConnectionFactory;
 import net.spy.memcached.MemcachedClient;
 import net.spy.memcached.internal.OperationFuture;
 public class OcsSample2 {
```

```
 public static void main(String[] args) {
     final String host = "xxxxxxxx.m.yyyyyyyyyy.ocs.aliyuncs.com
"; //"Internal network address" displayed on the console
     final String port = "11211"; //Default port: 11211. No changes
required.
     MemcachedClient cache = null;
     try {
         cache = new MemcachedClient(new BinaryConnectionFactory(),
AddrUtil.getAddresses(host + ":" + port));
         System.out.println("OCS Sample Code");
         //Save a value with the "ocs" key to ApsaraDB
for Memcache to facilitate data verification and reading.
         String key = "ocs";
         String value = "Open Cache Service,  from www.Aliyun.com";
         int expireTime = 1000; // Expiration time, in seconds. The
countdown starts from when data is written. After expireTime elapses
, the data
expires and cannot be read.
         OperationFuture<Boolean> future = cache.set(key, expireTime
, value);
         future.get();
         //Save several values
to ApsaraDB for Memcache and you can view the statistics on the
 ApsaraDB for Memcache console.
         for (int i = 0; i < 100; i++) {
             key = "key-" + i;
             value = "value-" + i;
             //Perform the Set operation
and save the value to the cache.
             expireTime = 1000; //Expiration
 time, in seconds.
             future = cache.set(key, expireTime, value);
             future.get();
         }
         System.out.println("Set operation completed!") ;
         //Perform
the Get operation and read the value with the "ocs" key from the
 cache.
         System.out.println("Get operation: " + cache.get(key));
     } catch (IOException e) {
         e.printStackTrace();
     } catch (InterruptedException e) {
         e.printStackTrace();
     } catch (ExecutionException e) {
         e.printStackTrace();
     }
     if (cache ! = null) {
         cache.shutdown();
     }
 }//eof
 }
```

3. Modify the instance ID and internal network address in OcsSample1.java opened in Eclipse based on your instance information.

4. After the information is modified, you can run your program. Run the main function. The following result is displayed in the console window under Eclipse (ignore the red INFO debugging information that may be displayed).

```
OCS Sample Code
 Set operation completed!
 Get operation: Open Cache Service,  from www.Aliyun.com
```

20.3.5.3 PHP: memcached

Use PHP: memcached to connect to an instance.

Context

Download a client

- *Download a client*
- *About the client*
- *Client version*

System requirements and environment configuration

Note:
If you already have a PHP Memcache environment, pay attention to the tips in the tutorial. Otherwise, your production environment may be overwritten and services may become unavailable. We recommend that you back up your data before upgrading or compiling the environment.

ApsaraDB for Memcache for Windows

If the environment cannot be established using the standard PHP Memcached extensions, you can splice packets manually to access KVStore for Memcache. For connection methods, see the following link. The sample code is simple compared to PHP Memcached. It only supports mainstream interfaces, so you need to perform additional operations to use it with other specific interfaces. For installation and usage methods, click *here*.

On a CentOS or Alibaba Cloud Linux 6 operating system,

Note:
Memcached 2.2.0 extensions must use Libmemcached 1.0.x libraries. Libraries earlier than 1.0 cannot be compiled. To compile Libmemcached, use GCC version 4.2 or later.

1. Check whether gcc-c++ and other components are installed (use the `gcc -v` command to check whether GCC version 4.2 or later is used). If the components are not installed, run the `yum install gcc+ gcc-c++` command.
2. Run the `rpm -qa | grep php` command to check whether the PHP environment is ready. If not, run the `yum install php-devel php-common php-cli` command to install PHP with source code compiling.

   PHP 5.3 or later is recommended. The PHP 5.2 source code contains the zend_parse _parameters_none function, which may cause errors. If you need to use the function, read the official PHP documentation. If you compile the source code, follow the official PHP compiling and upgrading methods.
3. Check whether SASL-related environment packages are installed. If not, run the `yum install cyrus-sasl-plain cyrus-sasl cyrus-sasl-devel cyrus-sasl-lib` command to install SASL-related environments.
4. Check whether the Libmemcached source code package is installed. If not, run the following command to install it (Libmemcached 1.0.18 is recommended):

```
 wget https://launchpad.net/libmemcached/1.0/1.0.18/+download/
libmemcached-1.0.18.tar.gz
 tar zxvf libmemcached-1.0.18.tar.gz
 cd libmemcached-1.0.18
 ./configure --prefix=/usr/local/libmemcached --enable-sasl
 make
 make install
 cd ..
```

5. Run the `yum install zlib-devel` command to install the Memcached source code package (Memcached 2.2.0 is recommended).

   Note:

   - Before installing Memcached, check whether there are any zlib-devel packages to be executed.
   - You must first check whether the Memcached client package (including the source code package) is installed. If the Memcached client package has been installed, recompile it to add the -enable-memcached-sasl extension.

```
wget http://pecl.php.net/get/memcached-2.2.0.tgz
tar zxvf memcached-2.2.0.tgz
cd memcached-2.2.0
phpize (If the system has
two PHP environments, you must call the command by specifying
the absolute path /usr/bin/phpize, which is the PHP environment
path for using KVStore for Memcache.)
```

```
./configure --with-libmemcached-dir=/usr/local/libmemcached --
enable-memcached-sasl (Pay attention to this parameter.)
make
make install
```

6. Run the locate command to find the php.ini file. If the system has two PHP environments, locate the PHP environment path for usingKVStore for Memcache, and add `extension=memcached.so memcached.use_sasl = 1` to the php.ini file in this path.
7. Test whether the production environment is successfully deployed by using the test code provided at the end of the page. Replace the address, port number, username, and password in the test code with actual ones.

On a CentOS or Alibaba Cloud Linux 5 (64-bit) operating system

1. Check whether gcc-c++ and other components are installed. If not, run the `yum install gcc + gcc-c++` command.
2. Run the `rpm –qa | grep php` command to check whether the PHP environment is ready in the system. If not, run the `yum install php53 php53-devel` command to install PHP with source code compiling. If the PHP environment has been prepared, skip this step. PHP 5.3 or later is recommended.

   The PHP 5.2 source code contains the zend_parse_parameters_none function, which may cause errors. If you need to use the function, see the PHP official documentation.
3. Run the `yum install cyrus-sasl-plain cyrus-sasl cyrus-sasl-devel cyrus-sasl-lib` command to install SASL-related environments.
4. Check whether Libmemcached (including the source code package) is installed. If not, run the following command to install Libmemcached (Libmemcached 1.0.2 is recommended):

```
 wget http://launchpad.net/libmemcached/1.0/1.0.2/+download/
libmemcached-1.0.2.tar.gz
 tar -zxvf libmemcached-1.0.2.tar.gz
 cd libmemcached-1.0.2
 ./configure --prefix=/usr/local/libmemcached --enable-sasl
 make
 make install
 cd ..
```

5. Run the `yum install zlib-devel` command to install the Memcached source code package (Memcached 2.0 is recommended).

   Note:

   - Before installing Memcached, check whether there are any zlib-devel packages to be executed.

- You must first check whether the Memcached client package (including the source code package) is installed. If the Memcached client package has been installed, recompile it to add the -enable-memcached-sasl extension.

```
wget http://pecl.php.net/get/memcached-2.0.0.tgz tar -zxvf
memcached-2.0.0.tgz
cd memcached-2.0.0
phpize (If the system has
two PHP environments, you must call the command by specifying
the absolute path /usr/bin/phpize, which is the PHP environment
path for using KVStore for Memcache. Run the phpize command in
the Memcached source
code directory.)
./configure --with-libmemcached-dir=/usr/local/libmemcached --
enable-memcached-sasl (Pay attention to this parameter.)
make
make install
```

6. Run the locate command to find the php.ini file, which is in /etc/php.ini for yum installation. If the system has two PHP environments, you must locate the PHP environment path for using ApsaraDB for Memcache, and add `extension=memcached.so memcached.use_sasl = 1` to the php.ini file in this path.
7. Run the `php –m |grep memcached` command. If the displayed result includes "memcache", ApsaraDB for Memcache is supported in the environment.
8. Test whether the production environment is successfully deployed by using the test code provided at the end of the page. Replace the address, port number, username, and password in the test code with actual one.

On an Ubuntu Debian operating system

1. Change the Ubuntu source.

   Solution 1: Run the `vim /etc/apt/source.list` command and add the following content at the beginning of the file:

```
deb http://mirrors.aliyun.com/ubuntu/ precise main restricted
universe multiverse
 deb http://mirrors.aliyun.com/ubuntu/ precise-security main
restricted universe multiverse
 deb http://mirrors.aliyun.com/ubuntu/ precise-updates main
restricted universe multiverse
 deb http://mirrors.aliyun.com/ubuntu/ precise-proposed main
restricted universe multiverse
 deb http://mirrors.aliyun.com/ubuntu/ precise-backports main
restricted universe multiverse
 deb-src http://mirrors.aliyun.com/ubuntu/ precise main restricted
universe multiverse
 deb-src http://mirrors.aliyun.com/ubuntu/ precise-security main
restricted universe multiverse
 deb-src http://mirrors.aliyun.com/ubuntu/ precise-updates main
restricted universe multiverse
```

```
 deb-src http://mirrors.aliyun.com/ubuntu/ precise-proposed main
restricted universe multiverse
 deb-src http://mirrors.aliyun.com/ubuntu/ precise-backports main
restricted universe multiverse
 apt-get update //Update the list.
```

Solution 2: Run the `wget http://oss.aliyuncs.com/aliyunecs/update_source.zip` command to download and decompress the update_source package. Run the `chmod 777 file name` command to grant the file execution permission, and run the script to automatically change the source.

2. Run the `ape-get` command to configure GCC and G++.

   You must first run the `dpkg -s installation package name` command (for example, `dpkg -s gcc`) to check whether gcc-c++ and other components are installed. If the components are not installed, run the `apt-get build-dep gcc apt-get install build-essential` command.

3. Install php5 and php5-dev.

   You must first run dpkg –s installation package name command (for example, dpkg –s php) to check whether PHP and other components are installed. If the components are not installed, run the `apt-get install php5 php5-dev` command (php5-cli and php5-common are automatically installed at the same time).

4. Install and configure SASL support.

   You must first run the `dpkg -s installation package name` command (for example, `dpkg -s libsasl2`) to check whether libsasl2 cloog-ppl and other components are installed. If they are not installed, run the following command:

   ```
    apt-get install libsasl2-dev cloog-ppl
    cd /usr/local/src
   ```

5. Run the following command to install Libmemcache of the specified version:

   Note:

   Before running the command, check whether the specified package (including the source code package) is installed. If yes, skip this step.

   ```
    wget https://launchpad.net/libmemcached/1.0/1.0.18/+download/
   libmemcached-1.0.18.tar.gz
    tar -zxvf libmemcached-1.0.18.tar.gz
    cd libmemcached-1.0.18
    ./configure --prefix=/usr/local/libmemcached
    make
    make install
   ```

```
 cd ..
```

6. Run the following command to install Memcached of the specified version:

> **Note:**
> Check whether the Memcached client package (including the source code package) has been installed. If the Memcached client package has been installed, recompile it to add the -enable-memcached-sasl extension.

```
 wget
 http://pecl.php.net/get/memcached-2.2.0.tgz
tar zxvf memcached-2.2.0.tgz
 cd memcached-2.2.0 phpize5
 ./configure --with-libmemcached-dir=/usr/local/libmemcached --
enable-memcached-sasl
 make
 make install
```

7. Configure PHP to support Memcached and then test the configuration.

```
 echo "extension=memcached.so" >>/etc/php5/conf.d/pdo.ini
echo "memcached.use_sasl = 1" >>/etc/php5/conf.d/pdo.ini
php -m |grep mem memcached
```

If this component is displayed, the installation and configuration are complete.

PHP sample code

Example 1: Establish a connection with KVStore for Memcache and perform the Set and Get operations

```
<? php
$connect = new Memcached;  //Declare a new Memcached connection.
$connect->setOption(Memcached::OPT_COMPRESSION, false); //Disable the
compression function.
$connect->setOption(Memcached::OPT_BINARY_PROTOCOL, true); //Use the
binary protocol.
$connect->setOption(Memcached::OPT_TCP_NODELAY, true); //Note: PHP
Memcached has
 a bug that causes a fixed latency of 40 ms when there is no Get value
. Set this parameter to prevent this bug.
$connect->addServer('aaaaaaaaaa.m.yyyyyyyyyy.ocs.aliyuncs.com', 11211
); //Add
the address and port number of the ApsaraDB for Memcache instance.
$connect->setSaslAuthData('aaaaaaaaaa', 'password'); //Set the
ApsaraDB for
 Memcache account and password for authentication. Skip this step if
 the password-free feature is enabled.
$connect->set("hello", "world");
echo 'hello: ',$connect->get("hello");
$connect->quit();
```

```
? >
```

Example 2:KVStore for Memcache Cache an array in MEMCACHE

```
<? php
$connect= new Memcached; //Declare a new Memcached connection.
$connect->setOption(Memcached::OPT_COMPRESSION, false); //Disable the
compression function.
$connect->setOption(Memcached::OPT_BINARY_PROTOCOL, true);//Use the
binary protocol.
$connect->setOption(Memcached::OPT_TCP_NODELAY, true); //Note: PHP
Memcached has a bug that causes a fixed latency of 40 ms when there is
 no Get value. Enable this parameter to prevent this bug.
$connect->addServer('xxxxxxxx.m.yyyyyyyy.ocs.aliyuncs.com', 11211);//
Add the address and
port number of the ApsaraDB for Memcache instance.
$connect->setSaslAuthData('xxxxxxxx', 'bbbbbbbb');//Set the ApsaraDB
for Memcache account
and password for authentication. Skip this step if the password-free
feature is enabled.
$user = array(
    "name" => "ocs",
    "age" => 1,
    "sex" => "male"
); //Declare an array.
$expire = 60; //Set an expiration time.
test($connect->set('your_name',$user,$expire), true, 'Set cache failed
');
if($connect->get('your_name')){
$result =$connect->get('your_name');
}else{
echo "Return code:", $connect->getResultCode();
echo "Return Message:", $connect->getResultMessage (); //If an error
is returned, parse the return code.
$result=" ";
}
print_r($result);
$connect->quit();
function test($val, $expect, $msg)
{
    if($val! = $expect) throw new Exception($msg);
}
? >
```

Example 3: use KVStore for Memcache together with a MySQL database

```
<? php
$connect = new Memcached; //Declare a new Memcached connection.
$connect->setOption(Memcached::OPT_COMPRESSION, false);//Disable the
compression function.
$connect->setOption(Memcached::OPT_BINARY_PROTOCOL, true);//Use the
binary protocol.
$connect->setOption(Memcached::OPT_TCP_NODELAY, true); //Note: PHP
Memcached has a bug that causes a fixed latency of 40 ms when there is
 no Get value. Set this parameter to prevent this bug.
$connect->addServer('xxxxxx.m.yyyyyyyy.ocs.aliyuncs.com', 11211);//Add
 the instance address
and port number.
$connect->setSaslAuthData('xxxxxx', 'my_passwd');//Set the
ApsaraDB for Memcache account and password for authentication. Skip
```

```
 this step if the password-free feature is enabled.
$user = array(
   "name" => "ocs",
   "age"  => 1,
   "sex"  => "male"
); //Define an array.
if($connect->get('your_name'))
{
  $result =$connect->get('your_name');
  print_r($result);
  echo "Found in OCS, get data from OCS"; //If the value is
available, the value source is displayed as ApsaraDB for Memcache.
  exit;
}
else
{
  echo "Return code:", $connect->getResultCode();
  echo "Return Message:", $connect->getResultMessage ();//Throw the
return code.
  $db_host='zzzzzz.mysql.rds.aliyuncs.com'; //Database address.
  $db_name='my_db';         //Database name.
  $db_username='db_user';    //Database username.
  $db_password='db_passwd'; //Database password.
  $connection=mysql_connect($db_host,$db_username,$db_password);
  if (! mysql_select_db($db_name, $connection))
  {
    echo 'Could not select database'; //An error is thrown
if the database connection fails.
    exit;
  }
  $sql = "SELECT name,age,sex FROM test1 WHERE name = 'ocs'";
  $result = mysql_query($sql, $connection);
  while ($row = mysql_fetch_assoc($result))
  {
    $user = array(
      "name" => $row["name"],
      "age"  => $row["age"],
      "sex"  => $row["sex"],
    );
    $expire = 5; //Set the value expiration time in the cache.
    test($connect->set('your_name',$user,$expire), true, 'Set cache
failed'); //Write the
 value to the ApsaraDB for Memcache cache.
  }
  mysql_free_result($result);
  mysql_close($connection);
}
print_r($connect->get('your_name')); //Print the value obtained.
echo "Not Found in OCS,get data from MySQL"; //Confirm the value
obtained from the database.
$connect->quit();
function test($val, $expect, $msg)
{
  if($val! = $expect) throw new Exception($msg);
}
```

```
? >
```

20.3.5.4 Python

Use Python to connect to an instance.

Download a client

- *Client download URL*
- *About the client*
- *Client version*

Environment configuration

Dependent on Bmemcached (SASL extensions supported). To download Bmemcached, click *here* .

Python sample code

```
#! /usr/bin/env python
import bmemcached
client = bmemcached.Client(('ip:port'), 'user', 'passwd')
print client.set('key', 'value11111111111')
print client.get('key')
```

20.3.5.5 C#/. NET: EnyimMemcached

Use C#/. NET: EnyimMemcached to connect to an instance.

Download a client

- *Client download URL*
- *About the client*
- *Client version*

C#/. NET sample code

```
using System.Net;
using Enyim.Caching;
using Enyim.Caching.Configuration;
using Enyim.Caching.Memcached;
namespace OCS.Memcached
{
    public sealed class MemCached
    {
        private static MemcachedClient MemClient;
        static readonly object padlock = new object();
        //Thread-safe single instance mode.
        public static MemcachedClient getInstance()
        {
            if (MemClient == null)
            {
                lock (padlock)
```

```
                {
                    if (MemClient == null)
                    {
                        MemClientInit();
                    }
                }
            }
            return MemClient;
        }
        static void MemClientInit()
        {
            //Initialize the cache.
            MemcachedClientConfiguration memConfig = new MemcachedC
lientConfiguration();
            IPAddress newaddress =
   IPAddress.Parse(Dns.GetHostEntry
 ("your_ocs_host"). AddressList[0]. ToString());//Replace your_ocs_h
ost with the ApsaraDB for Memcache internal network address.
            IPEndPoint ipEndPoint = new IPEndPoint(newaddress, 11211);
              // Configuration file - IP address.
            memConfig.Servers.Add(ipEndPoint);
            // Configuration file - protocol.
           memConfig.Protocol = MemcachedProtocol.Binary;
            // Configuration file - permission.
            memConfig.Authentication.Type = typeof(PlainTextA
uthenticator);
            memConfig.Authentication.Parameters["zone"] = "";
            memConfig.Authentication.Parameters["userName"] = "
username";
            memConfig.Authentication.Parameters["password"] = "
password";
      //Complete the following settings based on the maximum number of
 connections of the instance.
            memConfig.SocketPool.MinPoolSize = 5;
            memConfig.SocketPool.MaxPoolSize = 200;
            MemClient=new MemcachedClient(memConfig);
                  }
          }
}
```

Dependency

Code:

```
MemcachedClient MemClient = MemCached.getInstance();
```

20.3.5.6 C++

You can use a C++ program to connect to the ApsaraDB for Memcache instance.

Download a client

- *Client download URL*
- *About the client*
- *Client version*

Environment configuration

1. Download, compile, and install the C++ client.

   https://launchpad.net/libmemcached/1.0/1.0.18/+download/libmemcached-1.0.18.tar.gz

2. Run the following command:

   ```
   tar -xvf libmemcached-1.0.18.tar.gz

   cd libmemcached-1.0.18

   ./configure

   sudo make install
   ```

C++ sample code

1. Download *ocs_test.tar.gz*.
2. Run the following command:

   ```
   tar -xvf ocs_test.tar.gz

   cd ocs_test

   vim ocs_test_sample1.cpp
   ```

3. Set TARGET_HOST to the internal network address of the ApsaraDB for Memcache instance, USERNAME to the username of your instance, and PASSWORD to the password you set.
4. Run the `build.sh` command to generate ocs_test. Run the `./ocs_test` command. A key is written to the ApsaraDB for Memcache instance. Get the key from the ApsaraDB for Memcache instance and delete it from the instance.

   The code of ocs_test_sample1.cpp is as follows:

   ```
   #include <iostream>
    #include <string>
    #include <libmemcached/memcached.h>
    using namespace std;
    #define TARGET_HOST  ""
    #define USERNAME ""
    #define PASSWORD ""
    int main(int argc, char *argv[])
    {
        memcached_st *memc = NULL;
        memcached_return rc;
        memcached_server_st *server;
        memc = memcached_create(NULL);
        server = memcached_server_list_append(NULL, TARGET_HOST, 11211
   ,&rc);
        /* SASL */
        sasl_client_init(NULL);
        rc = memcached_set_sasl_auth_data(memc, USERNAME, PASSWORD);
   ```

```
     if(rc ! = MEMCACHED_SUCCESS) {
         cout<<"Set SASL err:"<< endl;
     }
     rc = memcached_behavior_set(memc,MEMCACHED_BEHAVIOR_BINARY_PROT
OCOL,1);
     if(rc ! = MEMCACHED_SUCCESS) {
         cout<<"Binary Set err:"<<endl;
     }
     /* SASL */
     rc = memcached_server_push(memc,server);
     if(rc ! = MEMCACHED_SUCCESS) {
       cout <<"Connect Mem err:"<< rc << endl;
     }
     memcached_server_list_free(server);
     string key = "TestKey";
     string value = "TestValue";
     size_t value_length = value.length();
     size_t key_length = key.length();
     int expiration = 0;
     uint32_t flags = 0;
     //Save data.
     rc = memcached_set(memc,key.c_str(),key.length(),value.c_str(),
value.length(),expiration,flags);
     if (rc ! = MEMCACHED_SUCCESS){
       cout <<"Save data failed: " << rc << endl;
       return -1;
     }
     cout <<"Save data succeed, key: " << key << " value: " << value
 << endl;
     cout << "Start get key:" << key << endl;
     char* result = memcached_get(memc,key.c_str(),key_length,&
value_length,&flags,&rc);
     cout << "Get value:" << result << endl;
     //Delete data.
     cout << "Start delete key:" << key << endl;
     rc = memcached_delete(memc,key.c_str(),key_length,expiration);
     if (rc ! = MEMCACHED_SUCCESS) {
       cout << "Delete key failed: " << rc << endl;
     }
     cout << "Delete key succeed: " << rc << endl;
     //free
     memcached_free(memc);
     return 0;
 }
```

The following example shows the use of ApsaraDB for Memcache through a different C++ program, where the ApsaraDB for Memcache cache and MySQL database are combined. You can follow the steps in the preceding example to compile and install the C++ client.

1. Create a sample database and table in the MySQL database.

```
mysql -h host -P port -u USER -p PASSWORD

create database testdb;

create table user_info (user_id int, user_name char(32) not null,
password char(32) not null, is_online int, primary key(user_id) );
```

2. Download *ocs_test_2.tar.gz* and run the following command:

```
tar -xvf ocs_test_2.tar.gz

cd ocs_test

vim ocs_test_sample2.cpp
```

> **Note:**
> Set OCS_TARGET_HOST to the internal network address of the ApsaraDB for Memcache instance, OCS_USERNAME to the ApsaraDB for Memcache instance name, OCS_PASSWORD to the password you set, MYSQL_HOST to the MySQL database address, MYSQL_USERNAME to the database username, and MYSQL_PASSWORD to the database password.

3. Run the `build.sh` command to generate ocs_test and run the `./ocs_test` command .

   The code of ocs_test_sample2.cpp is as follows:

```
#include <iostream>
 #include <string>
 #include <sstream>
 #include <libmemcached/memcached.h>
 #include <mysql/mysql.h>
 using namespace std;
 #define OCS_TARGET_HOST  "xxxxxxxxxx.m.yyyyyyyyy.ocs.aliyuncs.com"
 #define OCS_USERNAME "your_user_name"
 #define OCS_PASSWORD "your_password"
 #define MYSQL_HOST        "zzzzzzzzzz.mysql.rds.aliyuncs.com"
 #define MYSQL_USERNAME    "db_user"
 #define MYSQL_PASSWORD    "db_paswd"
 #define MYSQL_DBNAME      "testdb"
 #define TEST_USER_ID      "100"
 MYSQL *mysql = NULL;
 memcached_st *memc = NULL;
 memcached_return rc;
 int InitMysql()
 {
   mysql = mysql_init(0);
   if (mysql_real_connect(mysql, MYSQL_HOST, MYSQL_USERNAME,
MYSQL_PASSWORD, MYSQL_DBNAME, MYSQL_PORT, NULL, CLIENT_FOUND_ROWS)
 == NULL )
   {
     cout << "connect mysql failure!" << endl;
     return EXIT_FAILURE;
   }
   cout << "connect mysql success!" << endl;
   return 0;
 }
 bool InitMemcached()
 {
   memcached_server_st *server;
   memc = memcached_create(NULL);
   server = memcached_server_list_append(NULL, OCS_TARGET_HOST,
11211,&rc);
```

```
   /* SASL */
   sasl_client_init(NULL);
   rc = memcached_set_sasl_auth_data(memc, OCS_USERNAME, OCS_PASSWO
RD);
   if (rc ! = MEMCACHED_SUCCESS)
   {
     cout<<"Set SASL err:"<< endl;
     return false;
   }
   rc = memcached_behavior_set(memc,MEMCACHED_BEHAVIOR_BINARY_PROT
OCOL,1);
   if (rc ! = MEMCACHED_SUCCESS)
   {
     cout<<"Binary Set err:"<<endl;
     return false;
   }
   /* SASL */
   rc = memcached_server_push(memc,server);
   if (rc ! = MEMCACHED_SUCCESS)
   {
     cout <<"Connect Mem err:"<< rc << endl;
     return false;
   }
   memcached_server_list_free(server);
   return true;
 }
 struct UserInfo
 {
   int user_id;
   char user_name[32];
   char password[32];
   int is_online;
 };
 bool SaveToCache(string &key, string &value, int expiration)
 {
   size_t value_length = value.length();
   size_t key_length = key.length();
   uint32_t flags = 0;
   //Save data.
   rc = memcached_set( memc,key.c_str(), key.length(), value.c_str
(), value.length(), expiration, flags);
   if (rc ! = MEMCACHED_SUCCESS){
       cout <<"Save data to cache failed: " << rc << endl;
       return false;
   }
   cout <<"Save data to cache succeed, key: " << key << " value: "
 << value << endl;
   return true;
 }
 UserInfo *GetUserInfo(int user_id)
 {
   UserInfo *user_info = NULL;
   //get from cache.
   string key;
   stringstream out;
   out << user_id;
   key = out.str();
   cout << "Start get key:" << key << endl;
   size_t value_length;
   uint32_t flags;
   char* result = memcached_get(memc, key.c_str(), key.size(), &
value_length, &flags, &rc);
```

```
   if (rc ! = MEMCACHED_SUCCESS)
   {
     cout << "Get Cache Failed, start get from mysql."<< endl;
     int status;
     char select_sql[1024];
     memset(select_sql, 0x0, sizeof(select_sql));
     sprintf(select_sql, "select * from user_info where user_id = %d
", user_id);
     status = mysql_query(mysql, select_sql);
     if (status ! =0 )
     {
       cout << "query from mysql failure!" << endl;
       return NULL;
     }
     cout << "the status is :" << status << endl;
     MYSQL_RES *mysql_result = mysql_store_result(mysql);
     user_info = new UserInfo;
     MYSQL_ROW row;
     while (row = mysql_fetch_row(mysql_result))
     {
       user_info->user_id = atoi(row[0]);
       strncpy(user_info->user_name, row[1], strlen(row[1]));
       strncpy(user_info->password, row[2], strlen(row[2]));
       user_info->is_online = atoi(row[3]);
     }
     mysql_free_result(mysql_result);
     return user_info;
   }
   cout << "Get from cache succeed" << endl;
   user_info = new UserInfo;
   memcpy(user_info, result, value_length);
   return user_info;
 }
 bool DeleteCache(string &key, int expiration)
 {
   rc = memcached_delete(memc, key.c_str(), key.length(), expiration
);
   if (rc ! = MEMCACHED_SUCCESS) {
     cout << "Delete key failed: " << rc << endl;
     return false;
   }
   cout << "Delete key succeed: " << rc << endl;
   return true;
 }
 void PrintUserInfo(UserInfo *user_info)
 {
   cout << "user_id: " << user_info->user_id << " " << " name: " <<
user_info->user_name << endl;
 }
 bool SaveMysql(UserInfo *user_info)
 {
   char insert_sql[1024];
   memset(insert_sql, 0x0, sizeof(insert_sql));
   sprintf(insert_sql, "insert into user_info(user_id, user_name,
password, is_online) values(%d, '%s', '%s', %d)", user_info->user_id
, user_info->user_name, user_info->password, user_info->is_online);
   int status = mysql_query(mysql, insert_sql);
   if (status ! = 0)
   {
     cout << "insert failed" << endl;
     return false;
   }
```

```
  cout << "insert user_info" << endl;
  //insert mysql.
  return true;
}
int main(int argc, char *argv[])
{
  if (InitMysql() ! = 0)
  {
    return -1;
  }
  if (! InitMemcached())
  {
    return -1;
  }
  //generate user_info.
  UserInfo user_info;
  user_info.user_id = atoi(TEST_USER_ID);
  strcpy(user_info.user_name, "James");
  strcpy(user_info.password, "12345678");
  user_info.is_online = 1;
  //save to mysql.
  if (! SaveMysql(&user_info))
  {
    //return -1;
  }
  string user_str;
  user_str.assign((char*)&user_info, sizeof(UserInfo));
  //save to memcached.
  string key_str = TEST_USER_ID;
  SaveToCache(key_str, user_str, 10);
  //start get, exist in memcahced.
  UserInfo *get_user_info = GetUserInfo(user_info.user_id);
  PrintUserInfo(get_user_info);
  //wait 10 seconds.
  sleep(2);
  //delete memcached or expired.
  DeleteCache(key_str, 0);
  //start get, exist in mysql.
  delete get_user_info;
  get_user_info = GetUserInfo(user_info.user_id);
  PrintUserInfo(get_user_info);
  delete get_user_info;
  //free
  memcached_free(memc);
  mysql_close(mysql);
  return 0;
}
```

20.4 Instances

20.4.1 Create an instance

KVStore for Memcache supports two types of networks: classic network and VPC. You can create KVStore for Memcache instances of different network types. This topic describes how to create an instance in the KVStore for Memcache console.

Prerequisites

- At least one ECS instance is required for activating KVStore for Memcache.
- To create a KVStore for Memcache instance of the VPC type, you must first create a VPC. Then, create the instance in the same region as the VPC.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. Click **Create Instance** in the upper-right corner.
3. In the **Create Memcache Instance** dialog box that appears, select a network type and complete other settings.

   Table 20-2: Parameter description

Category	Parameter	Description
Region	Region	The region where the KVStore for Memcache instance is located. KVStore for Memcache allows only intranet access. Make sure the KVStore for Memcache instance and ECS are in the same region.
	Zone	The zone where the KVStore for Memcache instance is located. KVStore for Memcache allows only intranet access. Make sure the KVStore for Memcache instance and ECS are in the same zone.
Basic Settings	Department	The department to which the KVStore for Memcache instance belongs.
	Project	The project to which the KVStore for Memcache instance belongs. **Note:** After a project is selected, the KVStore for Memcache instance is accessible only to the members of the selected project. For more information, see **View project members** in *Apsara Stack Console User Guide*.
Instance Specification	Instance Specification	The instance specification. The maximum number of connections and maximum intranet bandwidth vary depending on different instance specifications.

Category	Parameter	Description
Network	Network Type	The network type of the instance. On the Alibaba Cloud platform, a classic network and a VPC have the following differences: • Classic network: The cloud services on a classic network are not isolated. Unauthorized access can be blocked only by the security group or whitelist policy of the cloud services. • VPC: VPCs help you build an isolated network environment in Alibaba Cloud. You can customize the routing table, IP address range, and gateway in a VPC. In addition, you can combine your on-premises IDC with cloud resources in the Alibaba Cloud VPC through a leased line or VPN to migrate applications smoothly to the cloud. If you want to set the network type to VPC, you must first create a VPC. For more information, see **Create a VPC and a VSwitch** in *VPC User Guide*.
Password	Set Password	The password for accessing the instance. You can select **Configure Now** to set the password immediately or **Configure After Creation** to set the password later by using the password reset feature. For more information, see *Reset a password*. The password complexity rules are as follows: • A password must be 8 to 30 characters in length. • It can contain uppercase letters, lowercase letters , and digits at the same time. It cannot contain special characters.
Instance Name	Instance Name	The instance name. The instance name must contain 2 to 128 characters in length. The name cannot contain special characters such as at signs (@), forward slashes (/), colons (:), equal signs (=), double quotation marks ("), angle brackets (<>), square brackets ([]), curly brackets ({}), or spaces.

4. Click **Create**.

   After creating the instance, wait until the instance status becomes **Normal**.

20.4.2 View instance details

After you create an instance, you can view instance details in the Apsara Stack console.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click an instance ID or choose  > **Details** from the shortcut menu.

   On the **Instance Information** page that appears, view the instance details.

   The **Instance Information** page contains the **Basic Information**, **Configuration Information**, and **Connection Information** areas. The *Instance information* table lists the configuration items in each area.

   Table 20-3: Instance information

Area	Item
Basic Information	• Instance ID • Name • Status • Region • Department • Project • Created At • Zone • Network Type • VPC (displayed only when the network type is VPC)
Configuration Information	• Instance Specification • Max Connections • Maximum Internal Network Bandwidth • Maintenance Time
Connection Information	• Connection Address • Port Number

20.4.3 Change an instance name

After you create an instance, you can change the instance name in the Apsara Stack console. In this way, you can locate the instance through the instance name.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, locate the relevant instance, choose  > **Edit** from the shortcut menu.
3. In the **Edit Instance Information** dialog box that appears, enter a new instance **name** and click **OK**.

20.4.4 Change the instance specification

KVStore for Memcache allows you to change the specification of an instance.

Context

Note:
The instance will experience intermittent interruption for several seconds during specification change, so change the specification during off-peak hours.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click an instance ID or choose  > **Details** from the shortcut menu.

   The **Instance Information** page appears.
3. Click **Change Instance** in the upper-right corner. In the **Change Instance** dialog box that appears, select the required **instance specification** and click **OK**.

   The message `Instance changed.` is displayed. You need to wait until the instance status becomes **Normal** before using the instance.

20.4.5 Set a whitelist

Before using an instance, you must set an IP address whitelist. For more information, see *Set a whitelist*.

20.4.6 Configure a maintenance time period

You can configure a maintenance time period for an instance in the console. The instance is maintainable in the specified time period.

Context

To ensure the stability of KVStore for Memcache instances, the backend system irregularly maintains instances and machines on the Alibaba Cloud platform.

Before official maintenance, KVStore for Memcache sends SMS messages and emails to contacts configured in your Alibaba Cloud account.

To ensure the stability in the maintenance process, instances enter the **Maintaining** state before the preset maintenance time on the day of maintenance. When an instance is in this state, the normal access to data in the database is not affected. However, change-related features (for example, specification change) are temporarily unavailable for this instance in the console. Query features such as performance monitoring are still available.

Note:
During maintenance, instances may experience intermittent interruption. We recommend that you select a maintenance time period from off-peak hours if possible.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click an **instance ID**, or choose  > **Details** from the shortcut menu.

   The **Instance Information** page appears.
3. Click **Modify Maintenance Time Period** in the upper-right corner.

   The default maintenance time period for KVStore for Memcache is from 02:00 to 06:00.
4. Select a maintenance time period and click **OK**.

   The time period is in China Standard Time (UTC+8).

20.4.7 Clear the instance data

You can clear all data of an instance with a single click in the console. After the data is cleared, it cannot be recovered.

Context

Note:
The instance data clear operation will delete all data of an instance, and the data cannot be recovered. Exercise caution when performing this operation.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, locate the relevant instance, choose  > **Clear** from the shortcut menu.
3. In the **Clear Instance** dialog box that appears, click **OK**.

20.4.8 Reset a password

If you forget the password for logging on to an instance, need to change the password, or have not set a password when creating the instance, you can reset the password.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click an instance ID or choose  > **Details** from the shortcut menu.

   The **Instance Information** page appears.
3. Click **Reset Password** in the upper-right corner.
4. On the **Reset Password** page, enter a new password and click **Submit**.

20.4.9 Set a data eviction policy

KVStore for Memcache supports six data eviction policies. You can modify the **`EvictionPolicy`** parameter in the console to set an eviction policy that meets your business needs.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click an instance ID, or choose  > **Details** from the shortcut menu.

   The **Instance Information** page appears.
3. Click the **Parameters** tab.
4. Click  next to **`EvictionPolicy`** and select **Modify**.

5. Select a data eviction policy and click **OK**.

20.5 Backup and recovery

20.5.1 Automatic backup

You can configure an automatic backup policy in the console.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, locate the relevant instance, and click the instance ID, or choose  > **Details** from the shortcut menu. The **Instance Information** page appears.
3. Click the **Backup and Recovery** tab.
4. Click **Backup Settings**.
5. Click **Configure**. In the **Backup Settings** dialog box that appears, configure the automatic backup cycle and time.

   Backup data is retained for seven days by default. You cannot modify this configuration.
6. Click **OK**.

20.5.2 Manual backup

In addition to automatic backups, you can also initiate manual backups in the console at any time.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click an **instance ID**, or choose  > **Details** from the shortcut menu.

   The **Instance Information** page appears.
3. Click the **Backup and Recovery** tab.
4. On the **Backups** tab page, click **Create Backup** in the upper-right corner.
5. Click **OK** to back up the instance immediately.

   On the **Backups** page, you can select the time range to query historical backup data. Backup data is retained for seven days by default, so you can query historical backup data from the past seven days.

20.5.3 Data restore

The data restore feature minimizes losses caused by misoperations on the database. Currently, KVStore for Memcache allows you to restore data from backup sets.

Procedure

1. *Log on to the KVStore for Memcache console*.
2. In the instance list, click an instance ID or choose  > **Details** from the shortcut menu.

   The **Instance Information** page appears.
3. Click the **Backup and Recovery** tab.
4. On the **Backup and Recovery** page, click the **Backups** tab.
5. Locate the backup file for restoring data, and choose  > **Restore** from the shortcut menu.
6. In the **Recovery** dialog box that appears, click **OK** to restore instance data from the backup file.

   The data restore operation is highly risky. Fully verify the correctness of the data to be restored before performing this operation.

20.6 Supported protocols and commands

Any clients compatible with the Memcached protocol can access KVStore for Memcache. You can select any Memcached client supporting SASL or Memcached Binary Protocol based on the application features.

Protocol

- Memcached Binary Protocol (binary)
- SASL authentication protocol

Operation

KVStore for Memcache supports the following command operations.

Operation code	Operation command	Remarks
0x00	Get	-
0x01	Set	-
0x02	Add	-

Operation code	Operation command	Remarks
0x03	Replace	-
0x04	Delete	-
0x05	Increment	-
0x06	Decrement	-
0x07	Quit	-
0x08	Flush	ApsaraDB for Memcache supports the second-level time accuracy.
0x09	GetQ	-
0x0a	No-op	-
0x0b	Version	-
0x0c	GetK	-
0x0d	GetKQ	-
0x0e	Append	-
0x0f	Prepend	-
0x10	Stat	Not supported
0x11	SetQ	-
0x12	AddQ	-
0x13	ReplaceQ	-
0x14	DeleteQ	-
0x15	IncrementQ	-
0x16	DecrementQ	-
0x17	QuitQ	-
0x18	FlushQ	-
0x19	AppendQ	-
0x1a	PrependQ	-
0x1b	Verbosity	Not supported
0x1c	Touch	-
0x1d	GAT	-
0x1e	GATQ	-

Operation code	Operation command	Remarks
0x20	SASL list mechs	-
0x21	SASL Auth	-
0x22	SASL Auth	-

21 Data Management Service (DMS)

21.1 What is DMS?

Data Management Service (DMS) offers an integrated solution for data management, structure management, access security, BI charts, data trends, data tracking, performance optimization, and server management. It supports the management of relational databases such as MySQL, SQL Server, PostgreSQL, MongoDB, and Redis, as well as NoSQL databases and Linux servers.

Function interface

Relational databases shows the interface of ApsaraDB for RDS.

Figure 21-1: Relational databases

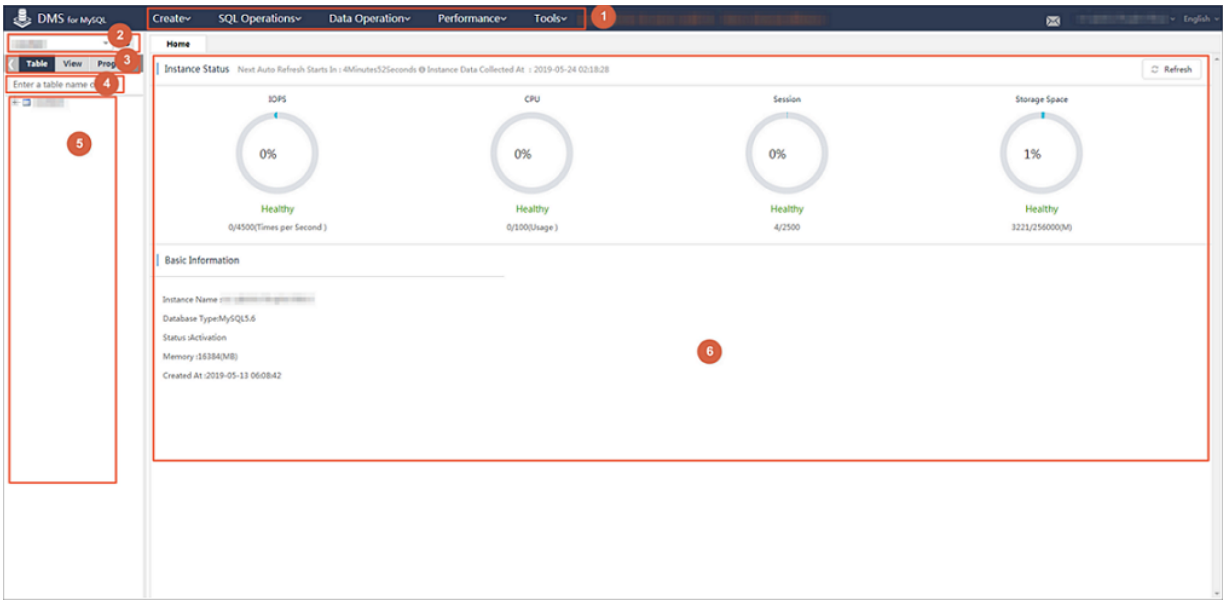


Table 21-1: Description of function modules shows the function modules.

Table 21-1: Description of function modules

No.	Name	Description
1	Top navigation bar	You can access the main function modules of DMS.
2	Drop-down list for database switching	You can select a database from the list to access the tables and data objects in that database.

No.	Name	Description
3	Navigation buttons for database objects	You can locate tables, views, and programmable objects such as functions, stored procedures, triggers, and events.
4	Table search box	You can search for a table through fuzzy match.
5	DMS object list	You can view the details of database objects such as tables.
6	Instance health status report	The report shows the current health status of the database service.

Supported database types

- DMS for MySQL
- DMS for SQLServer
- DMS for PostgreSQL/PPAS

Supported database operations

- SQL operations
  - SQL window
  - SQL command window
  - Save work environments
  - SQL execution
  - SQL optimization
  - SQL formatting (SQL statement improvement)
  - View execution plans
  - SQL input prompt
- Operations on database objects
  - Operations on data tables
  - Operations on table structures: Add table, modify structure, and delete table
  - Change table data: Insert, update, and delete data
  - Query and edit table data visually
- Operations on views and programmable objects such as functions, stored procedures, triggers, and events

  - Add
  - Edit
  - Delete
  - Enable or Disable
- Data processing
  - Import data
  - Export data
- Performance and diagnosis
  - Real-time performance
  - Real-time session
  - Lock wait analysis
- Data processing tools
  - E-R diagram
  - Table data volume statistics
  - Batch operation on tables

Effective user interaction

Simple operation. When an error occurs, an improvement scheme is generated to help you complete data operations.

21.2 Log on to an instance through DMS

This topic describes how to use DMS to log on to a database instance.

Context

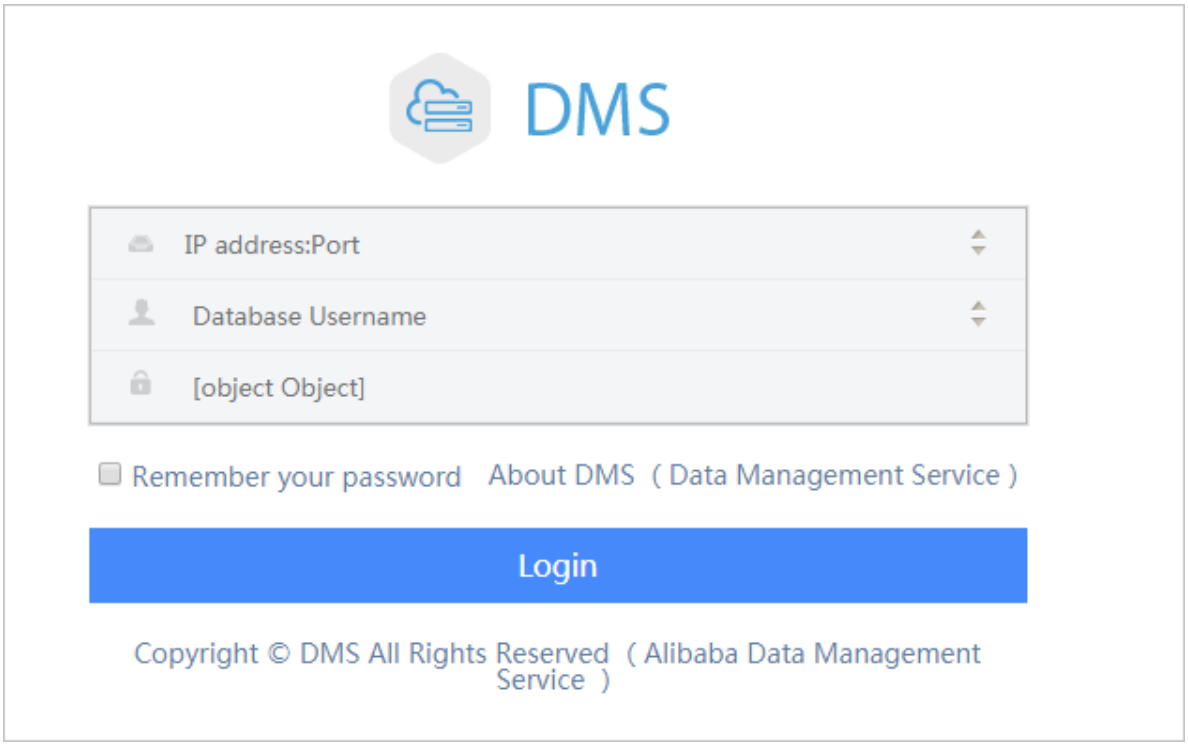
Log on to the Apsara Stack Management Console. Choose Console > Database > Relational Database Service. On the Relational Database Service page, you can use Data Management Service (DMS) to log on to an RDS instance. DMS integrates data management with structure management. It supports the management of relational databases such as MySQL, SQL Server, and PostgreSQL, as well as OLAP databases.

Procedure

1. *Log on to the RDS console*. For details, see *Log on to the ApsaraDB for RDS console* in *ApsaraDB for RDS User Guide*.

2. Click an instance ID, or click the icon in the **Actions** column of the instance, and select **View Details** to go to the **Basic Information** page.
3. Click **Log on to DMS** to go to the logon page of the DMS console.
4. Enter the logon information.

Figure 21-2: DMS logon page



Parameter description

No.	Description
1	The internal or public network address and port number of the instance to be connected, such as rm-test00000k012.mysql.aliyun-inc.com:3306. Follow the following steps to query the internal or public network address and port number of the instance: **a.** *Log on to the RDS console*. **b.** Click an instance ID, or click the icon in the **Actions** column of the instance, and select **View Details** to go to the **Basic Information** page. **c.** You can view the network connection address and port number in the **Private Network Connection Information** area.
2	Account used to connect to the instance. **Note:**

No.	Description
	The account is created in the ApsaraDB for RDS instance. For example, if your instance is a MySQL instance, see *Create a standard account* for how to create an account to connect to the instance.
3	Password of the account used to connect to the instance. **Note:** The password is specified for the account created in the ApsaraDB for RDS instance.
4	Type of the connected database.

5. Click **Log On**.

Note:
If you want the browser to remember the password, select **Remember Password**, and click **Log On**.

21.3 SQL operations

21.3.1 Use the Command Window

This topic describes how to use the DMS Command Window.

Context

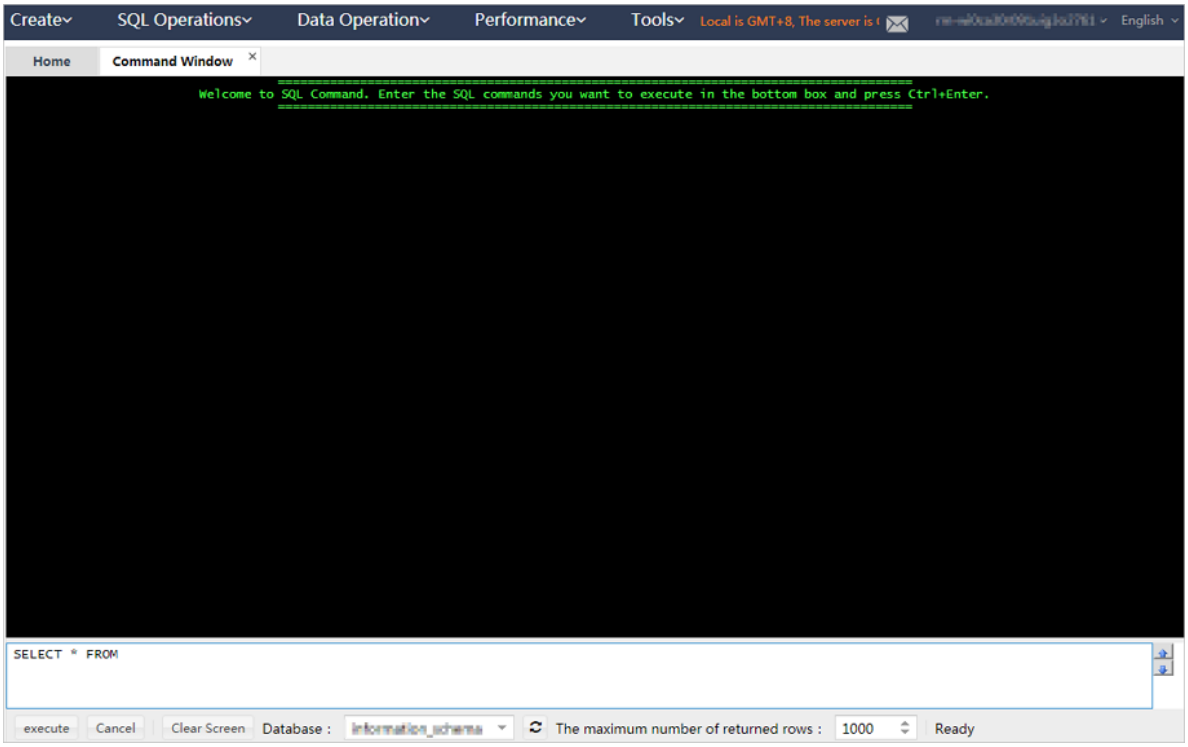
A MySQL database is used in this example.

Procedure

1. *Log on to an instance through DMS*.
2. From the top navigation bar, choose **SQL Operations** > **Command Window**.

   A blank command window is displayed, as shown in *Command Window*.

Figure 21-3: Command Window



3. Enter an SQL statement and click **Execute** in the window, as shown in *Executing an SQL statement*.

Figure 21-4: Executing an SQL statement

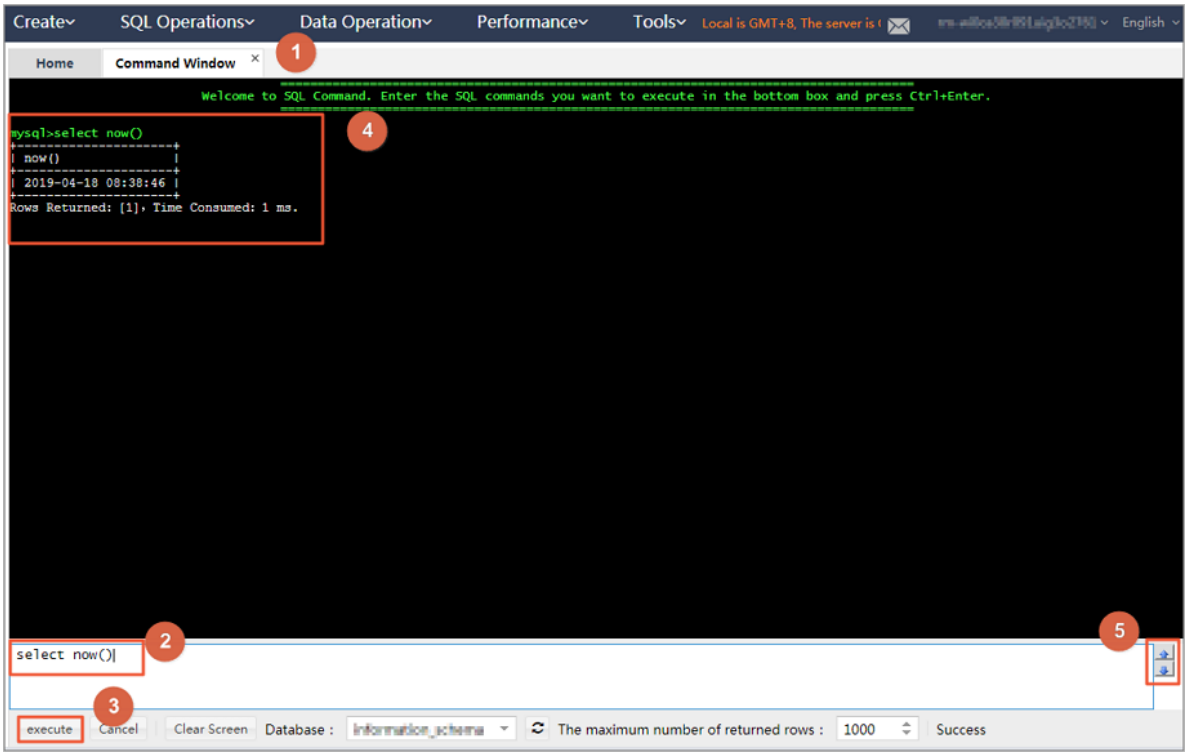


Table 21-2: Area description describes the numbered areas shown in the preceding figure.

Table 21-2: Area description

Number	Area	Description
1	Command Window	Displays the execution results of SQL statements.
2	SQL statement input area	Provides you with an area to enter SQL statements.
3	**Execute** button	Executes the entered SQL statements.
4	Result display area	**Displays the execution results**.
5	**Up and down** arrows	You can click the up or down arrows to view a previously executed SQL statement and execute it again.

4. Optional: If the execution process takes longer time than expected, you can click **Cancel** to cancel the execution.
5. Optional: Click **Clear Screen** to clear the screen.

   To use a different database, select the database from the **Database** drop-down list as needed.

21.3.2 Use the SQL window

21.3.2.1 Open an empty SQL window

This topic describes how to perform relevant operations in an SQL window.

Context

- A MySQL database is used as an example.
- A maximum of 20 SQL windows (including the homepage) can be opened in DMS. We recommend that you open no more than 5 SQL windows.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **SQL Operations** > **SQL Window** to open an SQL window.

   Figure 21-5: Empty SQL window shows the empty **`SQL window`** you have opened.

Figure 21-5: Empty SQL window

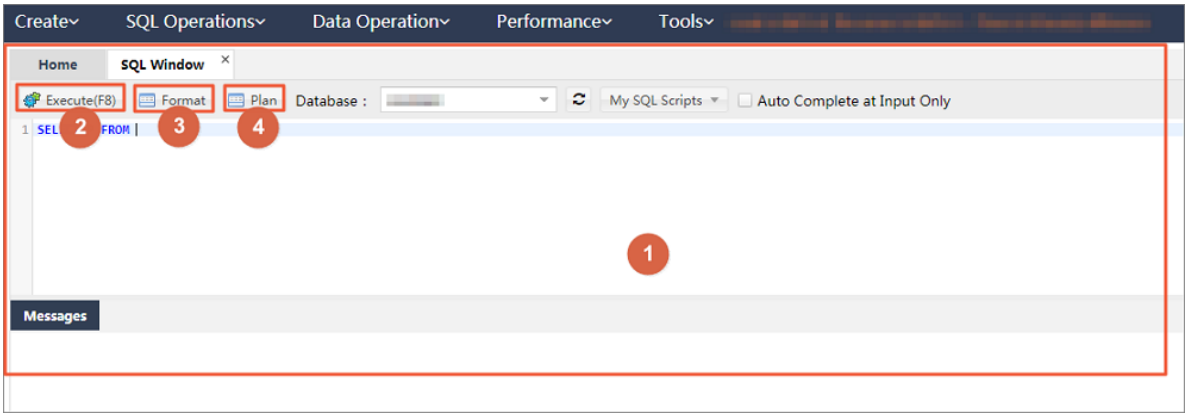


Table 21-3: Numbered items in the SQL window describes the numbered items in the SQL window.

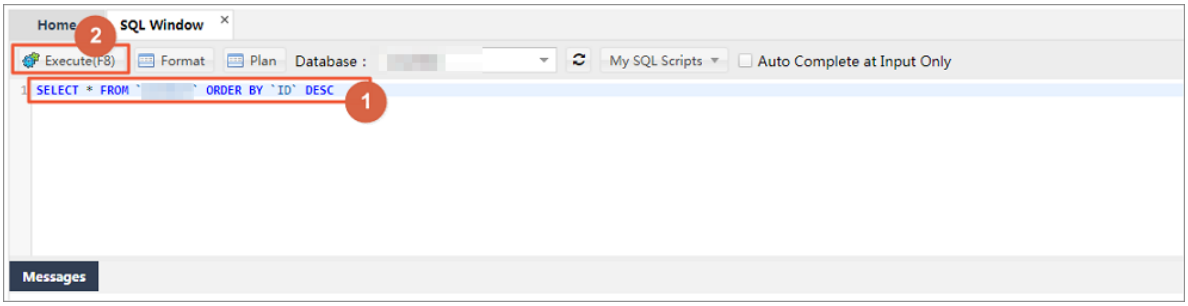
Table 21-3: Numbered items in the SQL window

No.	Name	Description
1	**SQL window**	The area in the green frame is the main area of the SQL window.
2	**Run (F8) button**	Click this button to run the entered SQL statement.
3	**Format button**	Click this button to format the entered SQL statement to make it more readable.

No.	Name	Description
4	**Execution Plan button**	Click this button to display the execution plan of the selected SQL statement. You can optimize the SQL statement and improve SQL processing performance based on the execution plan.

3. Enter the SQL statement you want to execute, click **Run** to complete the SQL query or update, as shown in *Run the SQL statement*.

Figure 21-6: Run the SQL statement



4. You can view the execution result set, as shown in *View the result set*.

Figure 21-7: View the result set

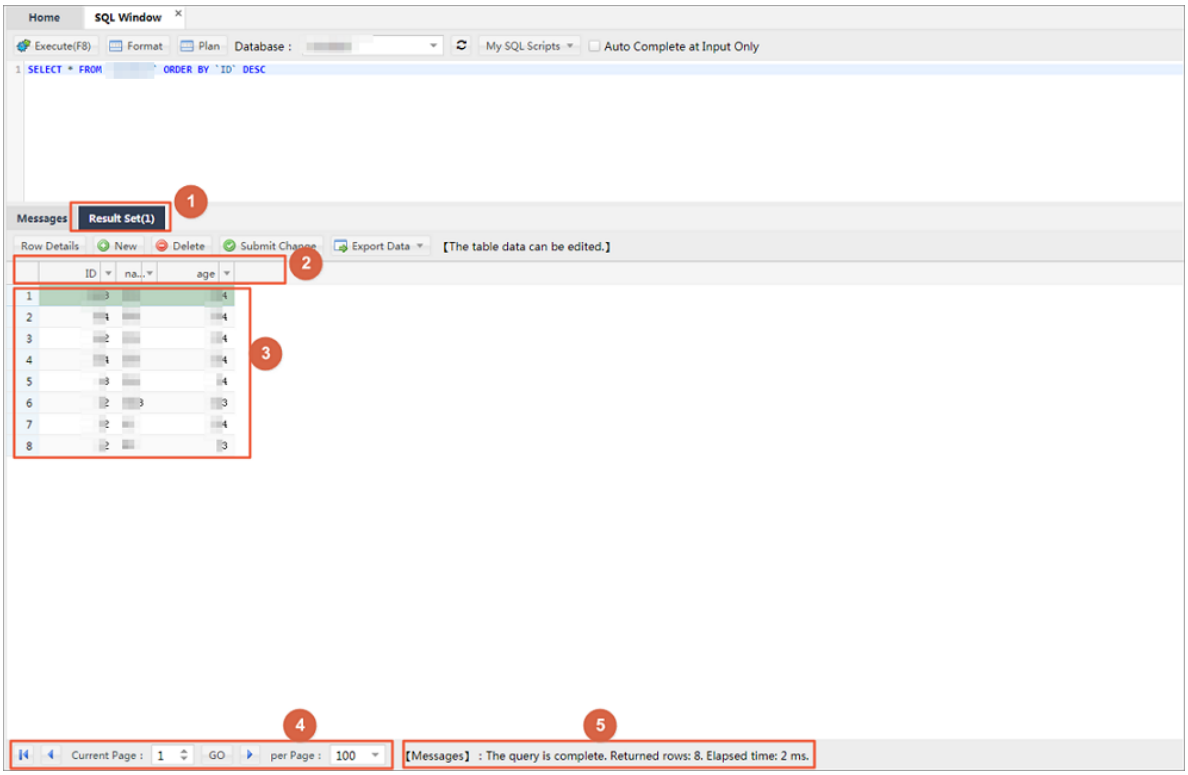


Table 21-4: Numbered items in the result set window

No.	Description
1	The **Result Set** tab shows the results returned for the SQL query statement.
2	The first row of the table shows the field names. If an alias has been specified for a field in SQL, the alias is displayed in this table.
3	The data area of the table shows the query results row by row. If the data area is unable to show the full results, horizontal and vertical scroll bars will appear to help you navigate the results.
4	Click **Show in Pages** or **Next Page** to view the results. • Each page shows 100 query results by default. Go to the next page to view more results. • You can set the number of results displayed **per page** as needed. • The results on the next page are appended to the **table numbered 3** in the figure.
5	Result acquisition progress and time elapsed.

5. View the message about SQL execution.

   Each time a data query (SELECT) or a data correction (INSERT, UPDATE, or DELETE) statement is executed, DMS returns a message that indicates the execution result, including the status and impact.

   Figure 21-8: Data query shows the message returned for data query.

Figure 21-8: Data query

Figure 21-9: Data correction shows the message returned for data correction.

Figure 21-9: Data correction

Table 21-5: Numbered items in the data correction window describes the numbered items in the data correction window.

Table 21-5: Numbered items in the data correction window

No.	Description
1	After you run an SQL statement, you can click the **Message** tab to view the execution status. No result set is returned for data correction. DMS displays a message after data correction is complete.
2	DMS runs the entered SQL statement step by step. • Analyzes the entered SQL statement. • Runs the SQL statement in the database. • Displays the queried data. • Counts statistics. For example, the number of data rows that are queried or affected.
3	DMS displays the SQL execution results. • Whether the execution is successful. • Number of queried rows, or number of rows affected by the Add, Delete, or Modify operation. • Time consumed to run the SQL statement.

6. Run multiple SQL statements in batches.

   DMS allows you to run multiple SQL statements in batches, as shown in *Batch execution*.

Figure 21-10: Batch execution

- 1: Shows the execution results of the first SQL statement.

- 2: Shows the execution results of the second SQL statement.

a) Enter the SQL statements that you want to run in the SQL window. Separate each SQL statement with a semicolon (;) or another separator.

b) If you want to run only some SQL statements, select the SQL statements you want to run. If you want to run all SQL statements, deselect or select all SQL statements, and click **Run**.

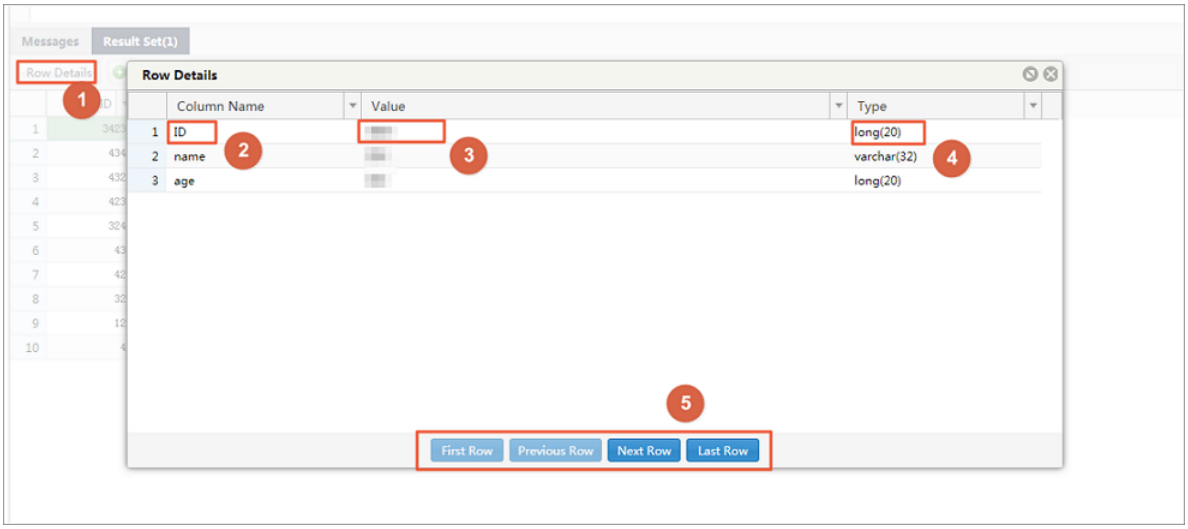
Wait until all SQL statements are completed.

c) View the execution results.

After you run a SELECT statement, DMS displays the result set. If you run other statements , DMS displays the execution results, such as the number of affected rows.

7. Click **Single Row Details** to view the details of a single record in the result set, as shown in *Single row details*.

Figure 21-11: Single row details



The following table describes the numbered items in the Single Row Details window.

Table 21-6: Numbered items in the Single Row Details window

No.	Description
1	Select the single row record you want to display in the **`Result Set`** table, and click **Single Row Details** to view a single data record. The **Single Row Details** dialog box that pops up displays every **`Field name`**, **`Field value`**, and **`Field type`** of the record.
2	**`Field name`**: If you have specified aliases for the fields, the aliases are displayed.

No.	Description
3	**`Field value`**: DMS automatically parses and displays the field values. Data such as time and binary code is formatted as an easy-to-read string and displayed.
4	**`Field type`**: You can view the type and length of each field.
5	Record navigation area. The **Previous**, **Next**, **First**, and **Last** buttons make it easier for you to view single row details of the previous and next records.

8. Optional: Edit the queried data in the result set.

   - Click **Add** to add a data row to the currently queried table.
   - Click **Delete** to delete the selected data row from the result set table.
   - Select the data row that you want to perform operations on.
   - Update the field values in the selected row directly.

   After you modify data, click **Submit Changes** to save the changed results to the database.

   After you click **Submit Changes**, DMS displays the SQL statement required to save your changes. This allows you to confirm the changes and prevent misoperations that cause loss of data.

   Click **OK** to apply the changes to the database as expected.

9. Click **Format** to format the selected SQL statement to make it readable.

   - Only the selected SQL statement is formatted. If you do not select any SQL statements, all the SQL statements that you entered are formatted.
   - Formatting converts your SQL statements to standard and readable ones, without changing the SQL execution logic and semantics, or affecting the execution.

   Example:

   Figure 21-12: Unformatted SQL statement shows an unformatted SQL statement.

   Figure 21-12: Unformatted SQL statement

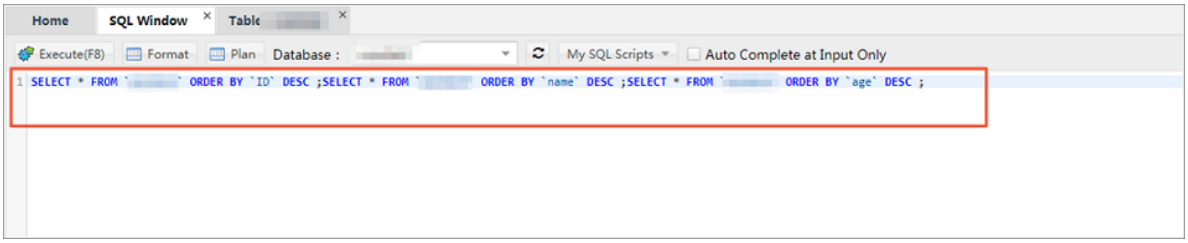
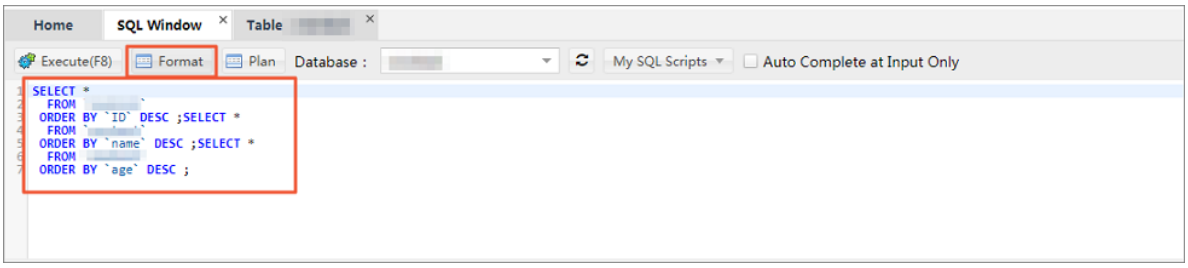


Figure 21-13: Formatted SQL statements shows formatted SQL statements.

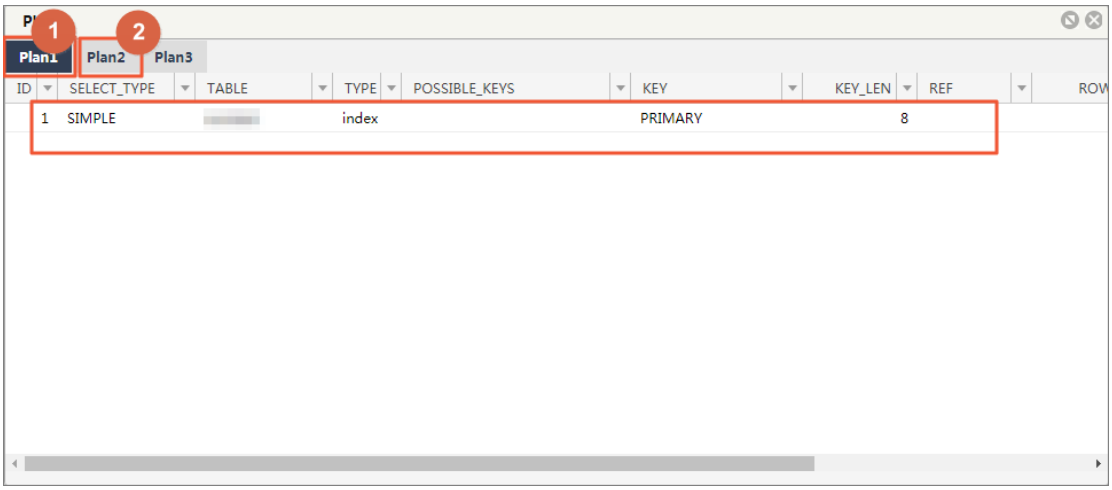
Figure 21-13: Formatted SQL statements



10. Click **Execution Plan** to view the SQL execution plan when you want to troubleshoot SQL-related faults or optimize SQL performance.

- After you click **Execution Plan**, DMS displays the execution plan of the selected SQL statement. If no SQL statement is selected, DMS displays the execution plans of all SQL statements.
- DMS displays an execution plan in detail. You can view information about an execution plan , such as the plan type and possible keys.
  - The display mode varies depending on different databases. The actual displayed content depends on the specific database.
  - When you query the execution plans for multiple SQL statements, DMS displays the execution plan of each SQL statement in detail on different tab pages, as shown in *Execution plan*.

Figure 21-14: Execution plan



- 1: Shows the execution plan of the first SQL statement in detail.

- 2: Shows the execution plan of the second SQL statement in detail.

21.3.2.2 Restore a saved SQL window

This topic describes how to restore a saved SQL window.

Context

- A MySQL database is used as an example.
- A maximum of 20 SQL windows (including the homepage) can be opened in DMS. We recommend that you open no more than five SQL windows.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **SQL Operations** > **SQL Window**.
3. Save the operating environment of the current SQL window.
   - DMS automatically saves the work environment when you close the operation page.
   - When you log on to the DMS console next time, DMS automatically restores the last work environment, including the last used database, the SQL window(s) that you opened, and the SQL statements that you entered in the SQL window(s).
   - When you close an SQL window, DMS prompts you to confirm whether you want to save the window content.
     - 1: Click the Close icon in the upper-right corner of the SQL window to close the window.
     - 2: DMS prompts you to confirm whether you want to save the work content. Click **Close and Save.** DMS then saves the work content of the SQL window, and closes the window after the content is successfully saved.

       If you click **Close**, DMS does not save the work content of the SQL window.
4. Restore the saved SQL window.
   a) Choose **SQL Operations** > **Saved SQL Windows**.

      DMS displays all the saved SQL windows.
   b) Click **New SQL Window** to restore one of the saved SQL windows.
   c) When you log on to the database through DMS, DMS automatically restores the work content of the last saved SQL window.

21.3.2.3 Manage commonly used SQL commands

This topic describes how to use DMS to manage commonly used SQL commands.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **SQL Operations** > **SQL Window** to open an SQL window.
3. Perform the following operations:
   - Add a commonly used SQL command.

     Choose **My SQL** > **Add My SQL** to add a commonly used SQL command.

     - Applicable scope: The custom SQL command is applicable to all application scenarios.
     - All databases: You can access the custom SQL command in any databases that you log on to from DMS.
     - Current instance: You can access the custom SQL command only through the currently connected instance (through an IP address and a port number).
     - Current database: You can access the custom SQL command only through the currently connected database. If you switch to another database, choose **My SQL** > **Select My SQL.** The custom SQL command is not displayed.
   - View saved SQL commands

     Choose **My SQL** > **Select My SQL** to view the commonly used SQL commands that you saved.
   - Manage your SQL commands

     Choose **My SQL** > **Manage My SQL** to manage commonly used SQL commands.

     - On the **Manage My SQL** page, click **Edit** or **Delete** to edit or delete your SQL commands.
     - On the **Manage My SQL** page, click **Add** to add an SQL command.
     - Double-click an SQL command under **My SQL** to insert the command into the **SQL Window**. The command is in the selected state in the SQL window.

21.3.2.4 Use the SQL template

This topic describes how to use the SQL template.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **SQL Operations** > **SQL Window**. An SQL window appears.

   You can view the SQL template in the rightmost part of the SQL window.
3. Double-click an SQL command or drag it to the SQL window. Then you can use or reference the command.

   You can directly modify commands referenced from the template even though you are not familiar with the commands.

21.3.3 Table operations (based on the Table directory tree)

21.3.3.1 Open a table-based SQL window

This topic describes how to open a table-based SQL window.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side directory tree of DMS, right-click a table and choose **SQL Operation Data** from the shortcut menu to open an SQL window.

   DMS automatically runs the SQL statement that queries top 50 data records of the table.

21.3.3.2 Edit table data

This topic describes how to manage common SQL commands.

Context

- A MySQL database is used as an example.
- The edit function is applicable to tables with average data volumes. To edit a table with a large volume of data, locate the data you want to deal with first. Locating data may take some time.

Procedure

1. *Log on to an instance through DMS*.

2. In the left-side directory tree of DMS, right-click a table and choose **Open Table** from the shortcut menu.

   A window showing the data of the selected table appears.

   - 1: In the left-side directory tree, right-click a table and choose **Open Table** from the shortcut menu. The data edit window appears.
   - 2: You can modify the values of the fields in the table.
   - 3: After you modify field values, click **Submit Changes** to submit the modified data.

21.4 Database development

This topic describes how to add, modify, delete, and maintain tables in a database, as well as how to manage objects such as indexes, foreign keys, and stored procedures.

21.4.1 Overview

This topic describes how to add, modify, delete, and manage data tables in a database, and how to manage indexes, foreign keys, and storage objects.

21.4.2 Table

21.4.2.1 Add a table

This topic describes how to use DMS to add a table.

Procedure

1. *Log on to an instance through DMS*.
2. You can create a table by using any of the following methods:

   - On the top navigation bar, choose **Create** > **Table**.
   - In the left-side **Table** directory tree, right-click a table and choose **Add Table** from the shortcut menu.
   - In the **Common Operations** area of the homepage, click **Create Table**.

3. Edit columns.

   Go to the **Create: Table** page, which displays the Column Info tab by default.

   You can edit the basic information and extended information of the fields as needed.

   You can also click **Column Info** to edit the table information.

4. Click the **Index** tab to edit indexes.

- Click **Add** to add an index.
- Click **Delete** to delete an index.
- You can edit the index row to modify index information.

5. Click the **Foreign Key** tab to go to the **Edit Foreign Keys** tab page.
   - Click **Add** to add a foreign key. The new key is editable.
   - Click **Delete** to delete a foreign key.
   - You can edit the index row to modify index information. When you edit a foreign key, you must enter the key name, column name, and information about the referenced database, table, and column.
6. Click the **Partition** tab and enter the SQL information of the partition.
7. Click the **Basic Info** tab to edit the basic information of the table.
   - You can edit the table name, storage engine, character set, and remarks.
   - You can click **More** to edit table parameters.
8. Click **Save**. DMS generates the SQL statement used to create a table.

   Click **OK** after you confirm the SQL statement. DMS then adds the table to your database.

21.4.2.2 Edit a table

This topic describes how to use DMS to edit a table.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Table** directory tree, right-click a table and choose **Edit Table Structure** from the shortcut menu to edit the table structure.
3. The Edit Table window is similar to the Create Table window. DMS automatically loads the table structure to the window.
   - 1: Select a table object type, such as **`Column Info`** or **`Index`**.
   - 2: Click a specific operation on the table object, which is similar to the Create and Edit operations on tables.
   - 3: Click **Open Table Data** to view and modify table data.
   - 4: Click **Create Statement** to view the statement used to create a table.
4. Click **Save**. DMS displays the SQL statement used to modify the table structure.

Click **OK** after you confirm the SQL statement. DMS then saves the modified table structure to your database.

21.4.2.3 Delete a table

This topic describes how to use DMS to delete a table.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Table** directory tree, right-click the table you want to delete and choose **Delete Table** from the shortcut menu.

   Warning:
   Deleting tables is a high-risk operation. Therefore, exercise caution when deleting tables.
3. Click **Yes** to delete the table.

21.4.2.4 Create a similar table

This topic describes how to use DMS to copy a table.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Table** directory tree, right-click the table that you want to copy and choose **Create Similar Table** from the shortcut menu.

   The Create Similar Table window appears.
3. Enter a table name and click **OK**. DMS creates a table similar to the selected table.
4. You can see that the structure of the created table is the same as that of the source table.

   A similar table is successfully created.

21.4.2.5 Generate SQL statement templates

This topic describes how to use DMS to generate SQL templates.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Table** directory tree, right-click the table that you want to copy and choose **Create Template SQL** from the shortcut menu.
3. DMS generates SQL INSERT, UPDATE, SELECT and CREATE TABLE statement templates for this table. Use these templates as a reference when you perform SQL operations.

21.4.2.6 Query table information

This topic describes how to use DMS to query table information.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Table** directory tree, right-click the table that you want to query and choose **Object Info** from the shortcut menu.
3. DMS obtains information about the table object. Click the **Basic Info** tab to view basic information about the table.
4. Click the **Create Statement** tab to view the table creation statement.

21.4.2.7 Clear data

This topic describes how to use DMS to clear table data.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Table** directory tree, right-click the table that you want to clear data from and choose **Clear Table** from the shortcut menu.

   Note:

   Clearing table data is a high-risk operation and may affect your data usage. DMS prompts you to confirm whether you want to clear table data.
3. Click **Yes** if you want to clear table data. DMS then clears the data of the selected table.
4. Open the table to check whether its data is cleared.

21.4.2.8 Perform table operations in batches

This topic describes how to use DMS to perform table operations in batches.

Procedure

1. *Log on to an instance through DMS*.
2. Delete tables in batches.
   a) In the left-side **Table** directory tree, right-click a table and choose **Batch Operate Tables** > **Batch Delete Tables**.

      The Batch Delete Tables window appears.
   b) Select the tables that you want to delete.
   c) Click **OK**.

DMS prompts you to confirm whether you want to delete the selected tables in batches.

d) Click **Yes**.

DMS deletes the selected tables in batches.

3. Perform table operations in batches.

You can clear data, delete tables, maintain tables, and modify table name prefixes in batches.

a) In the left-side Table directory tree, right-click a table and choose **Operate Tables** > **More Batch Operations** from the shortcut menu.

The More Batch Operations window is displayed.

b) Select the tables that you want to operate and click **Clear Data**, **Delete**, **Table Maintenance** or **Table Name Prefix**.

c) Click **OK**.

DMS prompts you to confirm whether you want to perform the batch operation.

d) Click **Yes**.

DMS performs the batch operation.

21.4.2.9 Maintain a table

This topic describes how to use DMS to maintain and optimize a table.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Table** directory tree, right-click the table you want to maintain and choose **Maintain Table** > **Optimize Table**.
3. Click **Yes**.

Click **Yes** if you want to optimize the table. Then DMS starts optimization.

Optimization allows you to reuse the table space in the database and organize file fragments.

Note:

You can check, restore, and analyze tables in a way similar to optimizing tables.

21.4.3 Manage indexes

This topic describes how to use DMS to add, modify, or delete indexes.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Table** directory tree, expand the table you want to modify and choose **Index** > **Add index** .

   The **Add Index** page is displayed.
3. Set index parameters.
   - 1: Enter an index name and select an index type.
   - 2: Click **+** or **-** to add or delete a field to or from the index.
   - 3: Edit the fields of the index. You can set values or select values from the drop-down list. You can set prefix length for variable-length fields (such as varchar) to save index space.
4. Click **Save**.

   DMS generates the SQL statement used to add the index. Confirm the change.
5. Click **Run**.
6. After the index is added, check the indexes of the table to verify that the new index takes effect.

   You can modify or delete the new index as needed.
   - In the left-side Table directory tree, right-click an index and choose **Modify Index** from the shortcut menu. The **Modify Index** window is displayed.

     The method of modifying an index is similar to the method of adding an index. However, the SQL statement deletes the old index and adds a new index to modify the index.
   - In the left-side Table directory tree, right-click an index and choose **Delete Index** from the shortcut menu. The **Delete Index** window is displayed. Click OK to delete the index.

21.4.4 Manage foreign keys

This topic describes how to use DMS to add foreign keys.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Table** directory tree, right-click the table that you want to modify and choose **Edit Table Structure** from the shortcut menu.
3. On the **Edit Table** page that appears, click the **Foreign Key** tab to edit the foreign keys.
4. Enter the foreign key information, set the fields of foreign keys, and specify the referenced table fields.
5. Click **Save**.

21.4.5 Create partitions

This topic describes how to use DMS to create partitions.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Table** directory tree, right-click a table and choose **Add Table** from the shortcut menu.

   The **Create: Table** page is displayed.
3. Enter the basic table information, and set the table fields and partitions.
4. Click **Save** to save the table structure that you created.

   A window is displayed, prompting you to confirm the SQL statement used to create the table.
5. Click **OK**. DMS creates the partition table. The SQL statement creates the table based on the partition fields and the partitioning logic that you have configured.
6. After the SQL statement is completed, check the table structure to verify that the partition table is successfully created.

21.4.6 Create a stored procedure

This topic describes how to use DMS to create and manage stored procedures.

Context

A MySQL database is used as an example.

Stored procedures, functions, triggers, and events are programmable objects in DMS.

Procedure

1. *Log on to an instance through DMS*.
2. Click the left-side **Programmable Object** directory tree, choose **Stored Procedure** > **Create (Stored Procedure)**.

   The Create Stored Procedure page is displayed.
3. Enter a name for the stored procedure.
4. Click **OK**.
5. DMS provides a stored procedure template. You only need to edit the stored procedure on the template.
6. Click **Save** to save the stored procedure to the database.

If a syntax error is found, DMS returns the cause of the error.

7. Click **Run** to run the stored procedure.

   DMS displays a page for you to configure the input parameters for the stored procedure.

   Set the input parameters. In this example, set **`cnt`** to **`80`** so that the records with **`Value=80`** can be displayed.

8. Click **Run**. DMS then runs the stored procedure.

   DMS displays the output parameters or intermediate result set of the stored procedure, if any.

   - The **Message** tab displays the message about the stored procedure execution, such as the output variables and intermediate result set.
   - The **Intermediate Result Set 1** tab displays the result set generated during stored procedure execution. If multiple result sets are available, DMS will generate multiple tabs for these intermediate result sets, such as Intermediate Result Sets 1, 2, and 3.

9. Click the **Intermediate Result Set 1** tab.

   The records with values equal to 80 are displayed.

10. You can set the options when creating the stored procedure. Click **Option Setup** to set options for the stored procedure.

11. After a stored procedure is successfully created, it is added to the Programmable Object directory tree.

    You can use the following menu options to perform operations on the stored procedure:

    - Create
    - Edit
    - Delete
    - Execute

12. You can run the stored procedure in the SQL window.

    - 1: To call a stored procedure, run the call stored_procedure_name command.
    - 2: The SQL window shows the result set of the stored procedure, if any.

21.4.7 Create a function

This topic describes how to use DMS to create a function.

Context

Stored procedures, functions, triggers, and events are programmable objects in DMS.

Procedure

1. *Log on to an instance through DMS*.
2. In the left-side **Programmable Object** directory tree, choose **Function** > **Create (Function)**.

   The **Please Set Basic Information About the New Function** page is displayed.
3. Set basic information about the new function.
4. Click **OK**.

   The Edit Function page appears. DMS generates a function creation template. You only need to enter information in the function part.
5. Enter information in the function part.
6. Click **Save**. DMS then checks whether the function is correctly defined. If the function definition is incorrect, DMS returns a prompt message.

   DMS runs the function definition in your database and returns a message indicating that the function is successfully saved.
7. Click **Run** to run the function.
8. Enter a parameter such as **wednesday** and click **Run**. DMS then runs the function.
9. Click **Option Settings** to set different options for the function.

   You can also run the function in the SQL window.

21.4.8 Create a view

This topic describes how to use DMS to create and manage custom views.

Procedure

Create a view

1. *Log on to an instance through DMS*.
2. Click the **View** directory tree on the left side to list the views of the current database.
3. Right-click the blank space and choose **Add View** from the shortcut menu.

   The **Create: View** page is displayed.
4. Set basic information about the view.

   The following example shows you how to filter the records where values are even numbers in the dmstest table, and how to output the id and name fields.

5. Click **Save Changes**. DMS then generates the SQL statement used to create the view based on your settings.
6. Click **OK** after you confirm the SQL statement. DMS then saves the view that you defined for your database.
7. After the view is saved, it is added to the left-side **View** directory tree. You can click the view to display its definition.

Check the view

8. Right-click the view and choose **Check View** from the shortcut menu to query data through the created view.
9. You can perform view-related operations through the following menu options:
   - View Data
   - Create View
   - Edit View
   - Delete View
   - Refresh Views

21.4.9 Create a trigger

This topic describes how to use DMS to create and manage a trigger.

Context

Stored procedures, functions, triggers, and events are programmable objects in DMS.

Procedure

1. *Log on to an instance through DMS*.
2. Click the left-side **Programmable Object** directory tree and choose **Trigger** > **Create (Trigger)**.

   The **Create: Trigger** page is displayed.

   - 1: Trigger table.
   - 2: Trigger settings.
     - Enter a name for the trigger.
     - Select a trigger table. The dmstest table marked by 1 is selected in this example.
     - Select a trigger time. The After option is selected in this example.
     - Select a trigger event. The Insert event is selected in this example.

- 3: Set a trigger statement.
  - Set the operation that the trigger performs when the specified event occurs.
  - In this example, when data is inserted into the dmstest table, the data is also automatically inserted into the copy_test table by the trigger, and the insertion time is recorded by the copy_test.time field.

3. Click **Save** after you have completed the trigger settings. DMS then generates the SQL statement used to create the trigger based on your settings. Confirm the SQL statement.
4. Click **OK**. DMS then saves the trigger to your database. DMS returns a message indicating that the trigger is successfully saved. In the left-side navigation pane, choose **Programmable Object** > **Trigger**. The trigger you just saved is displayed.
5. You can insert data into the **`dmstest`** table to check whether the data is recorded in the **`copy_test`** table.

   1: Insert data into the dmstest table and query the copy_test table for the inserted data.

   2: The SQL window displays a message about the execution status of the SQL statement. The message indicates that a row is inserted into the dmstest table and the corresponding data item is displayed in the copy_test table.
6. Check the result set in the SQL window to verify that the Insert operation is correctly processed by the trigger.
7. In the left-side navigation pane, choose **Programmable Object** > **Trigger** to perform trigger-related operations through the following menu options:
   - Create (Trigger)
   - Edit (Trigger)
   - Delete (Trigger)

21.4.10 Create event

This topic describes how to use DMS to create and manage events.

Prerequisites

After you log on to a database, make sure that event support has been enabled for the database.

- Execute the `SELECT @@event_scheduler;` statement to check whether the database supports events. If ON is returned, event support is enabled.
- If **`OFF`** is returned, event support is not enabled. You need to modify the configuration file or execute the `SET GLOBAL event_scheduler = ON;` statement to enable event support.

Context

Stored procedures, functions, triggers, and events are programmable objects in DMS.

Procedure

1. *Log on to an instance through DMS*.
2. Click the left-side **Programmable Object** directory tree, choose **Event** > **Create (Event)**.

   The Create Event page is displayed.

   1: In the Event Setup area, set the event name, cyclic execution, cycle, start time, end time, status, and comment.

   2: In the Event Execution Statement area, set the operation to be performed when a scheduled event is triggered.
3. Set an event trigger rule and the SQL statement for event execution.
4. Click **Save**. DMS then generates the SQL statement used to create the event.
5. After you confirm that the SQL statement is correct, click **OK**. DMS then executes the event you just edited in your database.
   - If the event is successfully generated, DMS returns a message indicating that the event is successfully saved.
   - In the left-side navigation pane, choose **Programmable Object** > **Event** to view the event you just edited.
6. Check whether the event is properly executed in the SQL window.

   In this example, the statement inserts a data record to the **`copy_test`** table every minute. Check the copy_test table to verify that the record is properly inserted into the table.
7. You can perform event-related operations through the following menu options:
   - Create (Event)
   - Edit (Event)
   - Delete (Event)

21.5 Data processing

21.5.1 Import Data

This topic describes how to use DMS to import data.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **Data Processing** > **Import**.

   The **Import** page is displayed.

   The Import page contains the import toolbar and import history.

   If you have previously imported data, **Import History List** lists the operations that you performed.
3. Click **Add Task**.

   The Import Task page appears.

   - Select a file type. Currently, only SQL and CSV file types are supported.
   - If the data file to be imported uses a character set, you can manually specify the character set here. DMS detects the character sets of files by default.
   - DMS terminates the import task if an error occurs while running an SQL statement. You can select **`Ignore Error`**, which may affect subsequent operations.
   - You can enter a task description to briefly specify the imported content and reason for later review.
4. Click **Start** to run the import task.

   If the imported data has an error, DMS terminates the import process and returns an error message. You can modify the data file and import it again.

   If the imported data and SQL statement are correct, DMS displays the import progress, imported data volume, and consumed time.

   After data is successfully imported, you can view the import task in **Import History List**.

   Click **Task Number** to view the execution details of the task.

21.5.2 Export data

21.5.2.1 Export a database

This topic describes how to use DMS to export a database.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **Data Processing** > **Export** to go to the **Export** page.
3. On the **Export** page, choose **Add Task** > **Export Database**.
4. On the **Add Export Task** page, select a database, file type (SQL or CSV), and content to be exported (structure and data, data, or structure). Select tables on the right side and additional content in the Additional Content area.
5. Click **OK** to run the export task.

   DMS refreshes the export progress every two seconds.

   You can close the export window. You can go to the Export History List to view the export details and download the exported data.

   After the export is complete, DMS automatically downloads the exported file to your local machine. You can also click **Download File** to download the exported file.

   You can view the export tasks that you submitted in the Export History List. Click a task name to view the task details and download the exported data.

21.5.2.2 Export an SQL result set

This topic describes how to use DMS to export an SQL result set.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **Data Processing** > **Export** to go to the **Export** page.
3. On the **Export** page, choose **Add Task** > **SQL Result Set Export**.
4. On the **Add SQL Export Task** page, complete the settings as needed.

   Select a file type (CSV or SQL_Insert) and a database, set the maximum number of rows of the result set, and enter an SQL statement.
5. Click **OK**. DMS then runs the SQL result set export task in the background.

   After the export task is completed, DMS automatically downloads the exported files to your local computer. You can also click **Download File** to download the exported files.

DMS also summarizes the export results and automatically downloads the exported SQL result set files.

You can view the SQL result set export tasks that you submitted in the Export History List and download SQL result set files.

21.6 Performance

On the DMS console, three menu options are available under the Performance menu: real-time performance, real-time session, and lock wait (InnoDB). This topic describes the operations related to these functions.

21.6.1 Overview

This topic describes how to use the real-time performance, real-time session, and InnoDB lock wait functions in the Performance menu of DMS.

Note:
This function only supports RDS for MySQL.

21.6.2 Lock wait

When an RDS for MySQL session is waiting for an exclusive Innodb row lock held by another session, Innodb lock wait will occur. The following two topics describe how to view and release lock wait.

21.6.2.1 View lock wait

This topic describes how to use DMS to view lock wait.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **Performance** > **InnoDB Lock Wait**.

   The **Lock Wait (InnoDB)** page is displayed.

   If the current instance has transactions waiting for locks, you can view the lock hold and lock wait.

3. Move the cursor to the **Lock Hold** or **Lock Wait** icon. The lock hold or lock wait list and related session ID are displayed.
4. Click  to reload the data.

21.6.2.2 Release lock wait

This topic describes how to use DMS to release lock wait.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **Performance** > **InnoDB Lock Wait**.

   The **Lock Wait (InnoDB)** page is displayed.

   If the current instance has transactions waiting for locks, you can view the lock hold and lock wait.
3. Move the cursor to the **Lock Hold** or **Lock Wait** icon. The lock hold or lock wait list and related session ID are displayed.
4. Click the **Lock Hold** or **Lock Wait** icon.

   The **Delete Session** confirmation box appears.
5. Click **Yes** to release the current session.

21.6.3 Sessions

21.6.3.1 View sessions

This topic describes how to use DMS to view instance sessions.

Context

A MySQL example is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **Performance** > **Instance Sessions**.

   The **Instance Sessions** page is displayed.

The session list is refreshed every 30 seconds by default. You can click **Refresh** to refresh the session list manually.

3. Click the specific SQL text to view the SQL execution details of the current session.

   You can view instance sessions in the instance list by summary, user statistics, access source statistics, or database statistics.

21.6.3.2 End a session

This topic describes how to use DMS to end a specific session.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top menu bar, choose **Performance** > **Instance Sessions**.
3. Select a session, or select multiple sessions by pressing the Shift or Ctrl key.
4. Click **Kill Session**.

   The Kill Session confirmation box appears.

5. Click **OK**.

   The selected session is ended.

   You can end multiple sessions in the instance list grouped by summary, user statistics, access source statistics, and database statistics.

21.6.3.3 Optimize a session

This topic describes how to use DMS to optimize a session.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **Performance** > **Real-time Session**.
3. Select a session and click **Optimize** to run the SQL statement used to optimize the selected session.

21.6.4 View real-time performance

This topic describes how to use DMS to view real-time performance of your database.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top menu bar, choose **Performance** > **Real-time Performance**.

   The Real-time Performance page is displayed.
3. Click **Start** or **Pause** to start or pause real-time performance detection.

   Note:
   Move the cursor to **`Parameter Description`** to view the description of real-time performance parameters.

21.7 Extended tools

21.7.1 Table data volume statistics

This topic describes how to use DMS to view table data volume statistics.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **Tools** > **Table Data Volume Statistics**.

   The **Table Data Volume Statistics** page is displayed.
3. The page shows the information about all user tables of the current instance, including the database, table name, storage engine, number of rows, row length (in bytes), data, index, data and index, creation time, and character set sorting rules.

   You can filter the statistics on table data volumes based on a range of criteria such as the database name, table name, total table size (MB), number of table rows, global sorting, and storage engine. You can also perform the paging, refresh, and reset operations.

21.7.2 E-R diagram

This topic describes how to use DMS to view E-R diagram.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an instance through DMS*.
2. On the top navigation bar, choose **Tools** > **E-R Diagram**.

   The E-R Diagram page is displayed.

   The E-R diagram shows the relationship between the tables of the current database and provides the methods for representing table names, column names, indexes, and relationships.
3. Perform the following operations:
   - Click the **DB:mysql** drop-down list to switch to another database.
   - Click the **Sorting: Sorting Options** drop-down list to sort tables in descending order of relationship count or field count, or in ascending order of table name or field count.
   - Click **Refresh** to refresh the current E-R diagram.
   - Click **View SQL Script** to list the SQL statements used to create all the tables of the current database.
   - Click **Download SQL Script** to download the SQL statements used to create all the tables of the current database.
   - Click **Download XML File** to convert the table creation SQL statements of the current database to XML files.
   - Double-click the name of a table column to view the column definition.
   - Double-click a table name to edit the table on a new page.

21.8 DMS for Redis

21.8.1 Function overview

Homepage shows the areas on the DMS for Redis homepage.

Figure 21-15: Homepage

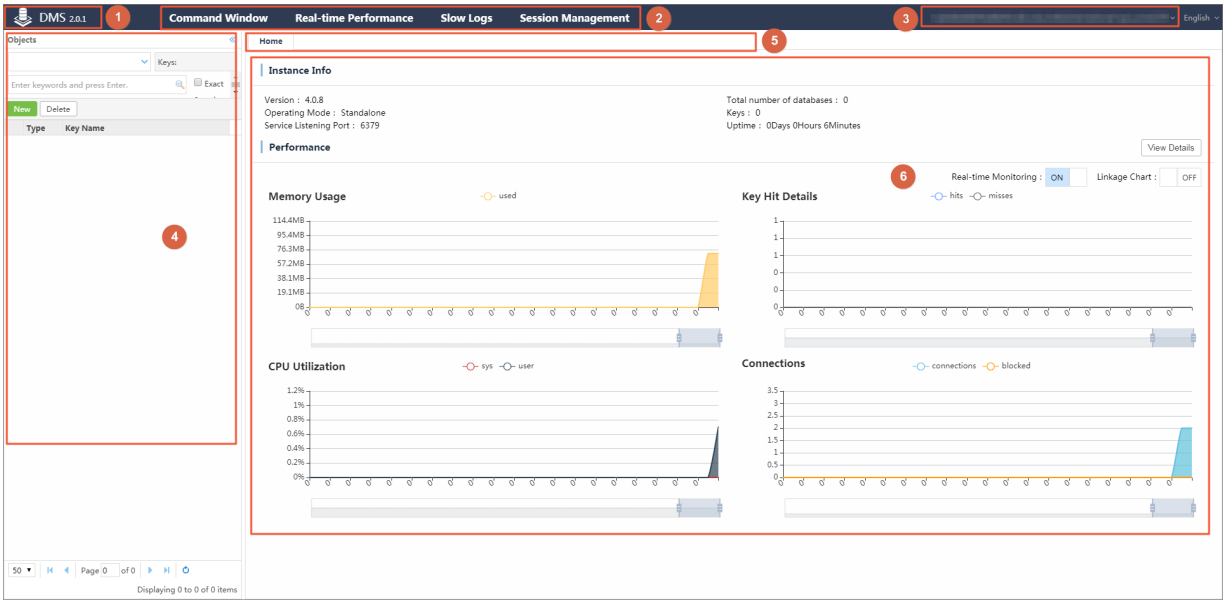


Table 21-7: DMS for Redis homepage description describes the areas of the homepage shown in the figure.

Table 21-7: DMS for Redis homepage description

No.	Area	Description
1	Version	You can hover over this area to view the upgrade records of the current version and go to the upgrade history page.
2	Top navigation bar	This area provides multiple functions, including the command window and real-time performance.
3	Instance display area	This area displays the connection string of the current instance. You can hover over the connection string to show the log out menu.
4	Object list	The object list allows you to perform multiple operations. For example, you can select databases, search for keys

No.	Area	Description
		by keyword, and view search results.
5	Tab	You can click a tab to view the corresponding tab page.
6	Tab page	A tab page displays the basic information and available actions for the corresponding feature.

21.8.2 Data management

21.8.2.1 Create a key

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. After logging on to the instance, select the database for which you want to create a key from the drop-down list.
3. Click **New**.

   The **New Key** dialog box is displayed.
4. Set **Key Name** and select a data type of the value from the **Type** drop-down list.

   A value editing tab varies depending on the data types of the value.
5. Click **OK**. The value editing tab page is displayed.
   - Value type 1: String
     a. Enter a value in the **Value** area.
     b. Click **Commit**. A dialog box is displayed, showing the command that will be executed to add the value.
     c. Click **OK** to create the key.
   - Value type 2: List
     a. Edit the values in the Value list. To add multiple values, click **Add to the Head** to add values to the head of the list or click **Add to the Tail** to add values to the tail of the list.
     b. Click **Commit**. A dialog box is displayed, showing the command that will be executed to add the value.
     c. Click **OK** to create the key.

- Value type 3: Hash
  a. Edit the values in the Value list. To add multiple data entries, click **New**.

     A valid data entry must contain a key and a value. The key must be unique while the same value can be specified for different keys.
  b. Click **Commit**. A dialog box is displayed, showing the command that will be executed to add the value.
  c. Click **OK** to create the key.
- Value type 4: Set
  a. Edit the values in the Value list. To add multiple data entries, click **New**. The values must be unique.
  b. Click **Commit**. A dialog box is displayed, showing the command that will be executed to add the value.
  c. Click **OK** to create the key.
- Value type 5: ZSet (sorted set)
  a. Edit the values in the Value list. To add multiple data entries, click **New**.

     A valid data entry must contain a value and a score. The value must be unique while the same score can be specified for different values. A valid score must be an integer or decimal number.
  b. Click **Commit**. A dialog box is displayed, showing the command that will be executed to add the value.
  c. Click **OK** to create the key.

21.8.2.2 Edit a key

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. After logging on to the instance, select the database that contains the key to be edited from the drop-down list.
3. Enter the key name or part of the key name in the search box, and press Enter or click the search icon.
4. In the search results, double-click the key to be edited. The value editing tab page is displayed.

   A value editing tab varies depending on the data type of the value.

- Example 1: String
  a. Enter a value in the Value list.
  b. Click **Commit**. A dialog box is displayed, showing the command that will be executed to edit the value.
  c. Click **OK** to submit the changes.
- Example 2: List
  a. Add, edit, or delete values in the Value list.
     - To add multiple values, click **Add to the Head** to add values to the head of the list or click **Add to the Tail** to add values to the tail of the list.
     - To edit an existing value, double-click and edit the value.
     - To delete an existing value, select the row containing the value and click **Delete**.
     - Information about the number of data entries per page and the number of pages is displayed at the lower part of each tab page. You can click the buttons to go to different pages, and locate values as needed.
  b. Click **Commit**. A dialog box is displayed, showing the command that will be executed to edit the value.
  c. Click **OK** to submit the changes.
- Example 3: Hash
  a. Edit the values in the Value list. To add multiple data entries, click **New**. To edit an existing data entry, double-click and edit the value. To delete an existing data entry, select the row containing the value and click **Delete**.

     Note:
     A valid data entry must contain a key and a value. The key must be unique while the same value can be specified for different keys. You can enter part of a key in the search box to search for the key as required.
  b. Click **Commit**. A dialog box is displayed, showing the command that will be executed to edit the value.
  c. Click **OK** to submit the changes.
- Example 4: Set

a. Edit the values in the Value list. To add multiple data entries, click **New**. To edit an existing value, double-click and edit the value. To delete an existing value, select the row containing the value and click **Delete**.

> **Note:**
> The values must be unique. You can enter the value or part of the value in the search box to search for the value.

b. Click **Commit**. A dialog box is displayed, showing the command that will be executed to edit the value.

c. Click **OK** to submit the changes.

- Example 5: ZSet (sorted set)

a. Edit the values in the Value list. To add multiple data entries, click **New**. To edit an existing data entry, double-click and edit the value. To delete an existing data entry, select the row containing the value and click **Delete**.

> **Note:**
> A valid data entry must contain a value and a score. The value must be unique while the same score can be specified for different values. A valid score must be an integer or decimal number. You can search for a specified value. Click **Switch search** to switch between search by keyword or search by score range.

b. Click **Commit**. A dialog box is displayed, showing the command that will be executed to edit the value.

c. Click **OK** to submit the changes.

21.8.2.3 Set key timeout

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. After logging on to the instance, select the database that contains the key to be edited from the drop-down list.
3. Enter the key name or part of the key name in the search box, and press Enter or click the search icon.
4. In the search results, locate the key to be edited. Right-click the key and choose **Set Timeout** from the shortcut menu.

5. In the displayed dialog box, enter a value for Set Timeout (unit: second).

   Note:

   - We recommend that you do not manually enter –1 in Set Timeout. A key with a value less than 0 immediately expires and cannot be searched for.
   - If Set Timeout is -1 by default, the timeout of the key has not been set, and the key will not expire. The displayed timeout value is consistent with the output of the TTL command in a Redis database.

6. In the displayed message, click **Yes** to complete the timeout setting.

   Note:

   An expired key cannot be searched for.

21.8.2.4 Delete a key

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. After logging on to the instance, select the database that contains the key to be deleted from the drop-down list.
3. Enter the key name or part of the key name in the search box, and press Enter or click the search icon.
4. In the search results, locate the key to be deleted, right-click the key, and choose **Delete** from the shortcut menu.
5. In the displayed message, click **Yes** to delete the key.

21.8.2.5 Rename a key

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. After logging on to the instance, select the database that contains the target key from the drop-down list.
3. Enter the key name or part of the key name in the search box, and press Enter or click the search icon.
4. In the search results, right-click the key to be renamed and choose **Rename** from the shortcut menu.

5. Enter a new key name.
6. Click **Yes** to rename the key.

21.8.3 Performance monitoring

21.8.3.1 View the homepage

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.

   The homepage is displayed.
2. View the homepage.
   - On the homepage, basic instance information is displayed in the upper part, and performanc e metrics are displayed in the lower part.
   - Real-time monitoring data collection starts when you open the homepage. The data on the page is refreshed every eight seconds. The refresh interval cannot be changed. You can turn the **Real-Time Monitoring** switch on or off to enable or disable monitoring data refresh.
   - You can hover over a chart to view the data collected at a specific point in time.
   - You can turn the **Linkage Chart** switch on or off to enable or disable the linkage among charts. If you turn on the Linkage Chart switch and then hover over one of the charts, all charts display the data collected at the same specific point in time.
   - In most cases, there is a slider indicating time period below a chart. You can adjust both ends of the slider to view the data collected within the adjusted time period.
   - Some charts contain multiple metrics. The metric legends are displayed above each chart. Each metric legend corresponds to a line in a chart in the same color. Click a metric legend to display or hide the metric in the chart.
   - Click **Show Details** to go to the real-time performance page. You can also click **Real-time Performance** in the top navigation bar to go to the page.

21.8.3.2 View real-time performance

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. Click **Real-Time Performance** in the top navigation bar. The real-time performance page is displayed.

- On the real-time performance page, metric data is displayed in rectangular boxes in the upper real-time data area, and metric data trends in different charts are displayed in the lower chart area.
- Real-time monitoring data collection starts when you open the homepage. The data on the page is refreshed every eight seconds. The refresh interval cannot be changed.
- You can turn the **Real-Time Monitoring** switch on or off to enable or disable monitoring data refresh.

In the real-time data area:

- Each metric (displayed in a rectangular box) is correlated to a chart in the lower part of the page. Click a **rectangular box** to display or hide the correlated chart.
- If the color of a rectangular box is blue, the correlated chart is displayed. Otherwise, the chart is hidden.

In the chart area:

- You can hover over a chart to view the data collected at a specific point in time.
- In most cases, there is a slider indicating time period below a chart. You can adjust both ends of the slider to view the data collected within the adjusted time period.
- Some charts contain multiple metrics. The metric legends are displayed above each chart. Each metric legend corresponds to a line in a chart in the same color. Click a metric legend to display or hide the metric in the chart.

21.9 DMS for MongoDB

21.9.1 Function overview

Homepage shows the areas on the DMS for MongoDB homepage.

Figure 21-16: Homepage

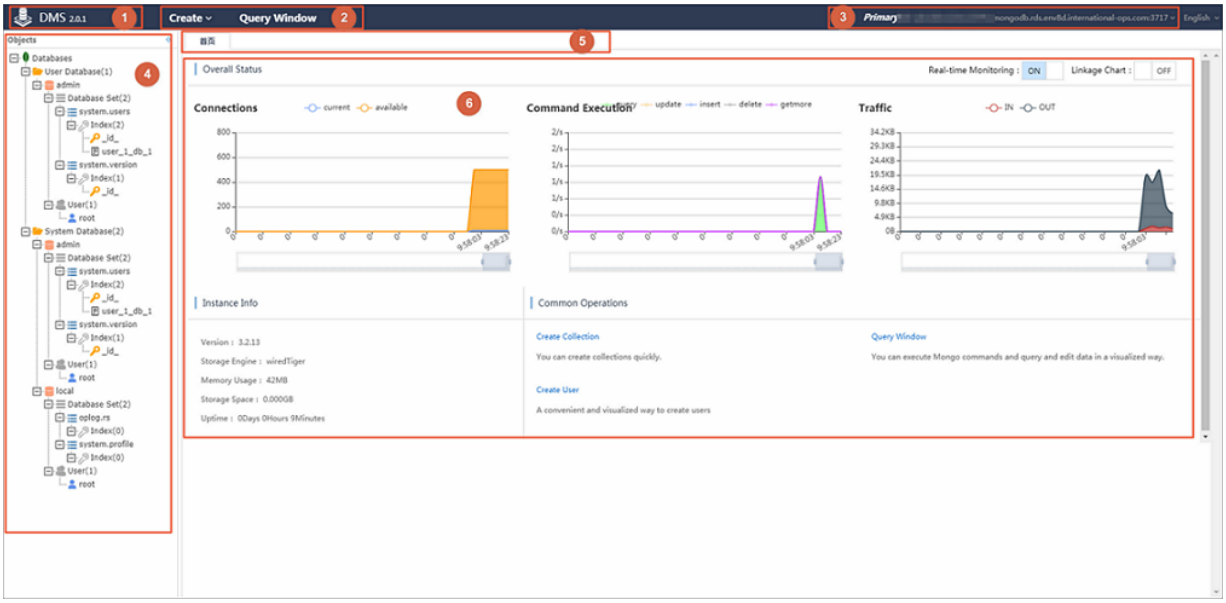


Table 21-8: DMS for MongoDB homepage description describes the areas of the homepage shown in the figure.

Table 21-8: DMS for MongoDB homepage description

No.	Area	Description
1	Version	You can hover over this area to view the upgrade records of the current version and go to the upgrade history page.
2	Top navigation bar	This area provides multiple functions such as database creation, collection creation, user creation and the query window.
3	Instance display area	This area displays the connection string of the current instance. You can hover over the connection string to show the log out menu.
4	Object list	The object list displays the structure of database objects, including databases, collections, users, and indexes. You can right-click these objects to manage them.
5	Tab	You can click a tab to view the corresponding tab page.
6	Tab page	A tab page displays the basic information and available actions for the corresponding feature.

21.9.2 Structure management

21.9.2.1 Create a collection

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. From the top navigation bar, choose **Create** > **Database Set**.

   The **Create Collection** dialog box is displayed.
3. Enter the name of the database for which you want to create a collection, and enter a collection name.
4. Click **Yes** to create the collection.

21.9.2.2 Create a database

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. From the top navigation bar, choose **Create** > **Database**.
3. Enter a database name and collection name.

   Note:
   When you create a database, you must enter a collection name in the Create Database dialog box to create a collection for the database. If you do not enter a collection name, a collection named test is created by default.
4. Click **Yes** to create the database.

21.9.2.3 Create an index

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. Expand the left-side object list and locate the collection for which you want to create an index.
3. Right-click the collection and choose **Create Index** from the shortcut menu. The **Add Index** dialog box is displayed.
4. In **Add Index**, enter an index name, add an index key, and specify a sorting order for the key. You can click **New** to add multiple keys. Make sure you specify a proper sorting order for the index keys.

   Note:
   Some parameters are displayed on the **Advanced Options** tab page. Click the **Advanced Options** tab to configure them.
5. Click **Yes** to complete the index creation.

21.9.2.4 Edit an index

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. Expand the left-side object list and locate the index to be edited.
3. Right-click the index and choose **Edit Index** from the shortcut menu.

   To change the index name, you can enter a new name in the Index Name field. To manage index keys, you can add, delete, or modify the keys in the Index area. Make sure you specify a proper sorting order for the index keys.

   Some parameters are displayed on the **Advanced Options** tab page. Click the **Advanced Options** tab to configure them.
4. Click **Yes** to complete the index configurations.

21.9.2.5 Delete a collection

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. Expand the left-side object list and locate the collection to be deleted. Right-click the collection and choose **Drop Collection** from the shortcut menu.
3. In the displayed **Information** message, click **Yes** to delete the collection.

21.9.2.6 Delete a database

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. Expand the left-side object list and locate the database to be deleted. Right-click the database and choose **Drop Database** from the shortcut menu.
3. In the displayed message, click **Yes** to delete the database.

Related tasks
Delete a collection

21.9.2.7 Delete an index

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. Expand the left-side object list and locate the index to be deleted.
3. Right-click the index and choose **Drop Index** from the shortcut menu.

4. In the displayed message, click **Yes** to delete the index.

21.9.3 User management

21.9.3.1 Create a user

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. From the top navigation bar, choose **Create** > **Users**.

   The **Create User** dialog box is displayed.

   - If you have logged on to the admin database, you can create a user with the highest privilege.
   - You can click the **Privileges on Other Databases** tab, select a database, and assign permissions on the selected database to the user.
3. Click the Privileges on Current Database tab, select the permissions that you want to assign to the user, and click **Yes**.

21.9.3.2 Edit a user

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. Expand the left-side object list and locate the user to be edited. Right-click the user and choose **Edit User** from the shortcut menu.
3. On the **Privileges on Current Database** tab page of the displayed dialog box, select the permissions that you want to assign to the user.

   If the user belongs to the admin database, you can click the **Privileges on Other Databases** tab, select a database, and assign permissions on the selected database to the user.
4. Click **Yes**.

21.9.3.3 Delete a user

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. Expand the left-side object list and locate the user to be deleted. Right-click the user and choose **Drop User** from the shortcut menu.
3. In the displayed message, click **Yes** to delete the user.

21.9.4 Data management

21.9.4.1 Create a document

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. Expand the left-side object list and locate the collection for which you want to create a document. Right-click the collection and choose **View Data** from the shortcut menu.
3. In the displayed Query Window, a query command is executed automatically to query the documents in the collection. Click **Create Document** on the Results tab page.

   A dialog box for creating a document is displayed.

   - Enter the document content in the displayed dialog box and ensure the content is compliant with the mongo shell standards.
   - You can choose whether to enclose an element name in double quotation marks (" "). If the element name contains spaces, you must enclose the name in double quotation marks (" ").
4. After you edit the content, click **Validate Format** to validate the format in the document.

   If a message indicating **the format is valid** is displayed, the document has passed the format validation. Otherwise, you need to modify the content based on the error message.
5. After the format validation succeeds, click **Yes** in the dialog box. The **Confirm Commands** dialog box is displayed. Confirm the entered content and click **Yes** in the **Confirm Commands** dialog box.

21.9.4.2 Edit a document

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. After logging on to the instance, expand the left-side object list and locate the collection that contains the document to be edited. Right-click the collection and choose **View Data** from the shortcut menu.
3. In the displayed Query Window, a query command is executed automatically t to query the documents in the collection. You can modify the query command to add query conditions and find the document to be edited.
4. You can edit the elements in the document or directly edit the document.

   - Edit an element in a document

You can edit elements when the changes do not affect the document structure.

a. Locate the element to be edited and click the value field. If the element can be edited, it enters the edit mode.

The edit mode varies with the data type of the element. For an element of the time type , a calendar is displayed. For an element of the boolean type, a drop-down list that contains boolean values is displayed. For an element of the string type, a text box is displayed.

b. After editing the element, click another field to complete editing the current element. In the displayed Confirm Commands dialog box, confirm the element modification and click **Yes**.

Right-click an element and choose **Delete Document** from the shortcut menu. In the displayed **Confirm Commands** message, click **Yes** to delete the element.

- Directly edit a document

The procedure of directly editing a document is similar to that of creating a document. You can replace the content in a document to edit the document. You can edit a document when you need to make a large number of changes to the document or make changes to the document structure, such as adding or deleting elements.

a. Select the document to be edited in the List view and locate an element in the document. Right-click the element and choose **Edit Document** from the shortcut menu.

b. Edit the document content in the displayed dialog box and ensure the content is compliant with the mongo shell standards.

- You can choose whether to enclose an element name in double quotation marks (" ").
- If the element name contains spaces, you must enclose the name in double quotation marks (" ").
- You can delete an element in the Edit dialog box.
- You can add an element by editing a document or adding the element in the Edit dialog box.

Note:

When you edit an element, follow the format specified for the data type of the element. If you change the format, the data type of the element may also change. For example, **`Value: NumberInt(123)`** indicates that the data type of the value is integer. **`Value`**

: **123** indicates that the data type of the value is double. If you edit the element and confirm the change, the data type of the value changes from integer to double.

c. After you edit the content, click **Validate Format** to validate the format in the document. If a message indicating *the format is valid* is displayed, the document has passed the format validation. Otherwise, you need to modify the content based on the error message.

d. After the format validation succeeds, click **Yes** in the Edit dialog box. The **Confirm Commands** dialog box is displayed. Confirm the changes and click **Yes** in the **Confirm Commands** dialog box.

21.9.4.3 Query a document

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. After logging on to the instance, click **Query Window** in the top navigation bar.
3. Select the database that contains the target document from the Database drop-down list, enter the query command on the command line, and click **Execute (F8)** to execute the command.

Note:

The command output is displayed on the Results tab page in the lower part of the Query Window. If you execute multiple query commands, all command output is displayed on the Results tab page in the execution order of the commands.

The structure of documents is complex and similar to that of a JSON file. Therefore, DMS provides three document views. You can click buttons on the left side to switch between views as needed.

- JSON view

  The JSON view is the default view of a document. You can click − or + to the left of an array or object to collapse or expand the structure of the array or object.
- List view

  The List view displays a document as a tree table. You can click the arrow displayed to the left of an array or object that contains data elements to collapse or expand the structure of the array or object. The List view displays detailed information about each data element, such as data type and value. You can also edit a document in the List view.

- Text view

  The Text view displays a document in the JSON text format with indents, which allows you to quickly copy the content.

21.9.4.4 Delete a document

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. After logging on to the instance, expand the left-side object list and locate the collection that contains the document to be deleted. Right-click the collection and choose **View Data** from the shortcut menu.
3. In the displayed Query Window, a query command is executed automatically to query the documents in the collection. You can modify the query command to add query conditions and find the document to be deleted. Click List on the Results tab page.
4. Select a document in the List view, and click **Delete Document**.
5. In the displayed message, click **Yes** to delete the document.

21.9.5 View the homepage

Procedure

1. On the DMS logon page, enter the database logon information and click **Log On**.
2. After logging on to the instance, the homepage is displayed by default:
   - On the homepage, performance metrics are displayed in the upper part, and basic instance information is displayed in the lower part.
   - Real-time monitoring data collection starts when you open the homepage. The data on the homepage is refreshed every eight seconds. The refresh interval cannot be changed. You can turn the **Real-Time Monitoring** switch on or off to enable or disable monitoring data refresh.
   - You can hover over a chart to view the data collected at a specific point in time.
3. You can turn the **Linkage Chart** switch on or off to enable or disable the linkage among charts. If you turn on the Linkage Chart switch and then hover over one of the charts, all charts display the data collected at the same specific point in time.
4. You can adjust both ends of the slider below each chart to view the data collected within the adjusted time period.

Some charts contain multiple metrics. The metric legends are displayed above each chart. Each metric legend corresponds to a line in a chart in the same color. Click a metric legend to display or hide the metric in the chart.

22 Server Load Balancer (SLB)

22.1 What is Server Load Balancer?

Server Load Balancer (SLB) is a traffic distribution control service that distributes traffic to multiple backend Elastic Compute Service (ECS) instances based on routing algorithms and forwarding rules.

SLB is a complementary service for ECS multi-machine solutions, and must be used in conjunction with ECS. SLB expands application service capabilities by distributing and balancing traffic. It checks the health status of added backend servers and automatically isolates abnormal ECS instances to eliminate the single point of failure (SPOF), improving the overall service capabilities of your applications.

SLB provides the following functions:

- Protocol support: SLB provides both Layer-4 (TCP and UDP) and Layer-7 (HTTP and HTTPS) load balancing services.
- Health check: SLB checks the health status of backend ECS instances. SLB can automatically block abnormal ECS instances and begin distributing requests to these ECS instances only when they become functional again.
- Session persistence: SLB provides the session persistence function. You can configure rules to forward a session request from a client to the same backend ECS instance during the session lifecycle.
- Routing algorithm: SLB supports round-robin and least-connections routing algorithms.
  - Round robin: External requests are sequentially distributed to backend ECS instances based on the number of visits.
  - Least connections: Backend ECS instances with fewer connections will be prioritized and accessed more frequently (probably).
- Domain name-based and URL-based forwarding: For Layer-7 (HTTP and HTTPS) protocols, SLB forwards requests to different VServer groups based on the preset domain names or URLs.
- Certificate management: SLB provides centralized certificate management for applications that use HTTPS. Certificates do not need to be uploaded to backend ECS instances. SLB performs decryption on access addresses, which reduces the CPU overheads of backend ECS instances.

22.2 Planning and preparation

Before creating an SLB instance, you must make a plan for the deployment of backend ECS instances, the SLB instance type (internal or external), and the listener protocols to be configured.

Make the following preparations before creating an SLB instance:

- **Create ECS instances**

  ECS instances will be added as backend servers to the SLB instance to receive and process requests forwarded by the listeners. Before using SLB, you must create ECS instances and deploy applications on them. Make sure that the department of each ECS instance is the same as that of the SLB instance, and the security groups of the ECS instances allow HTTP or HTTPS access over port 80 or 443.

- **Plan the SLB instance type**

  There are two types of SLB instances: external SLB instance and internal SLB instance. Different IP addresses are allocated to each type of SLB instances. You can set the SLB instance type as needed.

  - External: External SLB instances distribute only requests from public networks. After you create an external SLB instance, the system will allocate a public IP address to the instance . You can bind a domain name to the public IP address to provide external services.
  - Internal: Internal SLB instances distribute only requests from internal networks. When configuring an internal SLB instance, you must set Network Type to Classic Network or VPC:
    - If Network Type of an internal SLB instance is set to Classic Network, the IP address of the SLB instance is allocated and managed by Apsara Stack in a unified manner. The SLB instance is accessible to the classic-network ECS instances that are located in the same region as the SLB instance.
    - If Network Type of an internal SLB instance is set to VPC, the IP address of the SLB instance is allocated from the CIDR block of the specified VSwitch. The SLB instance is accessible only to the ECS instances that are located in the same VPC as the SLB instance.

- **Plan listeners**

  Alibaba Cloud Server Load Balancer supports both Layer-4 (TCP and UDP) and Layer-7 (HTTP and HTTPS) load-balancing services. You can configure different listeners as needed.

Compared with Layer-4 listeners, Layer-7 listeners require an additional step of Tengine processing. Therefore, the performance of Layer-7 listeners is inferior to that of Layer-4 listeners. In addition, Layer-7 listener performance may be further deteriorated by factors such as insufficient number of client ports and too many backend server connections. Layer-4 listeners are recommended for high performance purposes.

22.3 Quick start

22.3.1 Overview

This topic describes how to quickly create an external SLB instance and forward requests to two backend ECS instances.

For this example, two ECS instances where the Apache Web application is set up are added as backend servers to the SLB instance, to receive requests forwarded by the listeners.

Note:
To create multiple listeners to forward requests to different ECS instances, you must create VServer groups. For more information about how to create a VServer group, see *Add a VServer group*.

The following operations are involved:

- *Create an SLB instance*

  An SLB instance is a running entity of the SLB service. You can add multiple listeners and backend servers to an SLB instance.
- *Add listeners*

  You must add at least one listener to an SLB instance to forward requests to the backend servers. When configuring a listener, you need to set its basic forwarding rules and health check parameters.
- *Add backend servers*

  After configuring listeners, you must add backend servers to the SLB instance to receive and process requests forwarded by the listeners.

22.3.2 Log on to the Server Load Balancer console

This topic provides an example of how to log on to the Server Load Blanacer console by using the Chrome browser.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.
5. On the top navigation bar, choose**Console** > **Compute, Storage & Networking** > **Server Load Balancer**.

22.3.3 Create an SLB instance

An SLB instance is a running entity of the SLB service. You can add multiple listeners and backend servers to an SLB instance.

Prerequisites

- You have created ECS instances and deployed applications on them.
- Make sure that the department of each ECS instance is the same as that of the SLB instance , and the security groups of the ECS instances allow HTTP or HTTPS access over port 80 or 443.

Procedure

1. *Log on to the Server Load Balancer console*.
2. On the **Instances** tab page, click **Create Instance**.
   - **Department**: Select a department for the SLB instance from the drop-down list.

     Note:
     Make sure that the department of the SLB instance is the same as that of the backend ECS instances.
   - **Project**: Select a project for the SLB instance.
   - **Name**: Enter a name for the SLB instance.
   - **Network Type**: Set the instance type and network type of the SLB instance. For this example, select **External**, and leave **IP Address** empty. The IP address allocated by the system is used.
3. Click **Create**.

What's next

Add listeners

22.3.4 Add listeners

You must add at least one listener to an SLB instance to forward requests to the backend servers. When configuring a listener, you need to set its basic forwarding rules and health check parameters.

Prerequisites

Create an SLB instance

Procedure

1. *Log on to the Server Load Balancer console*.
2. On the Instances tab page, click the ID of the target SLB instance.
3. On the **SLB Instance** page, click the **Listeners** tab.
4. On the **Listeners** tab page, click **Add**.

5. In the **Add Listener** dialog box, configure a listener.

   In this example, the listener is configured as follows:

   - **SLB Protocol [Port]**: HTTP 80
   - **Backend Protocol [Port]**: HTTP 80
   - **Scheduling Algorithm**: Round Robin
   - **Peak Bandwidth**: 100 Mbit/s
   - **Session Persistence**: disabled
   - **Health Check Settings**: default settings

6. Click **OK**.

What's next

Add backend servers

22.3.5 Add backend servers

After configuring listeners, you must add backend servers to the SLB instance to receive and process requests forwarded by the listeners.

Procedure

1. *Log on to the Server Load Balancer console*.
2. On the Instances tab page, click the ID of the target instance.
3. On the **SLB Instance** page, click the **Backend Servers** tab.
4. Click **Add Backend Server**.
5. In the **Add Backend Server** dialog box that appears, select ECS instances and click **Add**.

   An ECS instance with a higher weight receives a greater proportion of access requests. You can set the weights of ECS instances based on their service capabilities. The default weight is used in this example.

22.4 SLB instances

22.4.1 SLB instance overview

An SLB instance is a running entity of the SLB service. To use the SLB service, you must create an SLB instance and add listeners and backend servers to the instance.

In the Server Load Balancer console, you can edit, delete, start, and stop SLB instances.

22.4.2 Create an SLB instance

Before using SLB, you must create an SLB instance. You can add multiple listeners and backend servers to an SLB instance.

Procedure

1. *Log on to the Server Load Balancer console*.
2. On the **Instances** tab page, click **Create Instance**.
3. On the **Create SLB Instance** page, configure an SLB instance.

Table 22-1: Parameters for creating an SLB instance

Parameter	Description
Region	The region to which an SLB instance belongs.
Zone	The zone within the selected region, to which an SLB instance belongs. If Apsara Stack is deployed in two data centers, select both a primary zone and a secondary zone for the SLB instance.
Department	The department to which an SLB instance belongs.
Project	The project to which an SLB instance belongs.
Name	The name of an SLB instance. The name must be 1 to 63 characters in length. It must start with a letter and can contain numbers, letters, hyphens (-), and underscores (_).
Instance Type	The type of an SLB instance. • **Internal**: If your SLB instance distributes only internal traffic, set Instance Type to Internal. For internal SLB instances, you must set the network type to Classic Network or VPC. If you create the SLB instance in a VPC, select a VPC and a VSwitch for the SLB instance. • **External**: If your SLB instance needs to distribute external traffic, set Instance Type to External.
Network Type	The network type of an SLB instance. • **Classic Network**: The instance IP address is allocated by the Apsara Stack platform. • **VPC**: The instance IP address is allocated by the specified VSwitch.
IP Address	The service IP address of an SLB instance.

Parameter	Description
	If the service IP address is left empty, the system automatically allocates an IP address to the instance based on the instance network type.
Instances	The number of instances to be created. The system creates SLB instances with the same configuration in batches.

4. Click **Create**.

22.4.3 Start or stop an instance

You can start or stop an SLB instance as needed.

Procedure

1. *Log on to the Server Load Balancer console*.
2. On the **Instances** tab page, locate the target instance. In the Actions column, click the  icon and choose **Start** or **Stopped** from the shortcut menu.
3. In the message that appears, click **OK**.

22.4.4 View instance details

You can view the details such as departments, projects, and IP addresses of SLB instances in the Server Load Balancer console.

Procedure

1. *Log on to the Server Load Balancer console*.
2. On the **Instances** tab page, locate the SLB instance that you want to view.
3. Click the instance ID, or click the  icon in the Actions column and choose **View Details** from the shortcut menu, to go to the instance details page.

22.4.5 Modify attributes of an SLB instance

You can edit the name and description of an SLB instance.

Procedure

1. *Log on to the Server Load Balancer console*.
2. On the **Instances** tab page, locate the target SLB instance. Click the  icon in the Actions column and choose **Change** from the shortcut menu.

3. In the **Change SLB Instance** dialog box that appears, change the name and description of the SLB instance.
4. Click **OK**.

22.4.6 Modify the ownership of an SLB instance

You can modify the department and project of an SLB instance.

Procedure

1. *Log on to the Server Load Balancer console*.
2. On the **Instances** tab page, locate the target instance. Click the  icon in the Actions column and choose **Change Ownership** from the shortcut menu.
3. In the **Change Ownership** dialog box that appears, change the department and project of the instance.
4. Click **OK**.

22.4.7 Delete an SLB instance

You can delete an existing SLB instance.

Procedure

1. *Log on to the Server Load Balancer console*.
2. On the **Instances** tab page, locate the target instance. Click the  icon in the Actions column and choose **Delete Instance** from the shortcut menu.
3. In the message that appears, click **OK**.

22.5 Listeners

22.5.1 Overview

SLB listeners monitor requests received by SLB instances and forward the requests to backend ECS instances based on the forwarding rules.

Server Load Balancer supports Layer-4 (TCP and UDP) and Layer-7 (HTTP and HTTPS) listeners. The following table describes the features and usage scenarios of each protocol.

Table 22-2: Protocols of Server Load Balancer listeners

Protocol	Feature	Usage scenario
TCP	• A connection-oriented protocol. A reliable connection must be established with the peer before data can be sent and received. • Source address-based session persistence. • Source address available at the network layer. • Fast data transmission.	TCP is applicable to scenarios with high requirements on reliability and data accuracy but with tolerance for low speeds, such as file transmission, email sending or receiving, remote logon, and Web applications with no special requirements.
UDP	• A non-connection-oriented protocol. Before sending data, UDP directly transmits data packets without making three handshakes with the peer, and provides no error recovery or data retransmission. • Fast data transmission and low reliability.	UDP is applicable to scenarios where real-time transmission is more important than reliability, such as video chat and real-time financial market pushes.
HTTP	• An application-layer protocol primarily used to package data. • Cookie-based session persistence. • Get the source address by using X-Forwarded-For.	HTTP is applicable to applications that need to identify data content, such as Web applications and small-size mobile games.
HTTPS	• Similar to HTTP, but with an encrypted connection that prevents unauthorized access. • Centralized certificate management service. You can upload certificates to Server Load Balancer. The decryption operations are completed directly on Server Load Balancer.	HTTPS is applicable to applications that require encrypted transmission.

22.5.2 Configure a Layer-4 listener

Alibaba Cloud provides Layer-4 (TCP and UDP) load balancing services. Layer-4 SLB listeners forward requests directly to backend ECS instances without modifying the packet headers.

Context

For details about the features and usage scenarios of UDP and TCP, see *Listeners*.

Procedure

1. *Log on to the Server Load Balancer console*.

2. Click the ID of the target instance to go to the instance details page. Then, click the **Listeners** tab.
3. On the **Listeners** tab page, click **Add**.
4. In the **Add Listener** dialog box, configure a Layer-4 listener.

Table 22-3: Parameters for configuring a Layer-4 listener

Category	Parameter	Description
Basic Settings	**SLB Protocol [Port]**	The SLB frontend protocol and port that are used to receive requests and forward requests to backend servers. To configure a Layer-4 listener, select **TCP** or **UDP** from the drop-down list.
	Backend Protocol [Port]	The port of applications deployed on backend ECS instances.
	Scheduling Algorithm	The routing algorithm. • **Round Robin**: Requests are sequentially distributed to ECS instances based on the number of visits. • **Least Connections**: Requests are forwarded to the ECS instance with the fewest connections.
	Peak Bandwidth	The bandwidth peak value for the listener, in Mbit/s.
	Session Persistence	Indicates whether to enable session persistence. For Layer-4 (TCP and UDP) listeners, SLB supports IP address-based session persistence. SLB forwards access requests from the same IP address to the same backend ECS instance for processing. If you turn on the Session Persistence switch, you must specify the timeout period of the session in **Timeout**.
	Timeout	The connection timeout period.
	Idle Connection Timeout	The idle connection timeout period, in seconds. Value range: 1—60. If no request is received during the specified timeout period, SLB closes the connection and starts a new connection when the next request comes.

Category	Parameter	Description
	Enable VServer Group	Indicates whether to use a VServer group. If you turn on the Enable VServer Group switch, select a VServer group to bind to the listener. A VServer group consists of multiple ECS instances that provide the same services. Requests are forwarded to the ECS instances in the specified VServer group based on the forwarding rules configured for the listener. If no VServer group is used, client requests are forwarded to the backend ECS instances of the SLB instance based on the forwarding rules configured for the listener. **Note:** A VServer group cannot be changed once it is bound to a listener.
Health Check Settings	**Port**	The port used by the health check service to access backend ECS instances. The backend port configured for the listener is used by default.
	Response Timeout (Seconds)	The maximum response time for a health check before timing out. If a backend ECS instance does not respond to the health check requests within the specified period, the health check fails.
	Health Check Interval (Seconds)	The time interval between two consecutive health checks. All nodes in the LVS cluster perform regular health checks at the specified interval independently and in parallel on backend ECS instances.
	Unhealthy Threshold	The number of consecutive failed health checks that must occur on an ECS instance for an LVS node to declare this ECS instance unhealthy.
	Healthy Threshold	The number of consecutive successful health checks that must occur on an ECS instance for an LVS node to declare this ECS instance healthy.

5. Click **OK**.

22.5.3 Configure a Layer-7 listener

Layer-7 (HTTP or HTTPS) SLB listening is a way to implement reverse proxy. After HTTP requests arrive at an HTTP listener of an SLB instance, the SLB instance establishes TCP connections with its backend servers. Then, the SLB instance sends the HTTP requests to the backend servers over the new TCP connections by using HTTP, instead of forwarding the data packets directly to the backend servers.

Context

For details about the features and usage scenarios of HTTP and HTTPS, see *Listeners*.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target SLB instance to go to the instance details page.
3. Click the **Listeners** tab.
4. On the **Listeners** tab page, click **Add**.
5. In the **Add Listener** dialog box, configure a listener.

Table 22-4: Parameters for configuring a Layer-7 listener

Category	Parameter	Description
Basic Settings	**SLB Protocol [Port]**	The SLB frontend protocol and port that are used to receive requests and forward requests to backend servers. Select **HTTP** or **HTTPS** for Layer-7 listeners.
	Backend Protocol [Port]	The port of applications deployed on backend ECS instances.
	Scheduling Algorithm	The routing algorithm. • **Round Robin**: Requests are sequentially distributed to ECS instances based on the number of visits. • **Least Connections**: Requests are forwarded to the ECS instance with the fewest connections.
	Two-way Authentication	Indicates whether to enable two-way authentication. After two-way authentication is enabled, you must upload both the server and CA certificates. Two-way authentication is disabled and one-way authentica tion is enabled by default.

Category	Parameter	Description
		Click **Upload Certificate** to upload the server and CA certificates. For more information, see *Upload a certificate*.  **Note:** This option is applicable only to HTTPS listeners.
	Select Server Certificate	The server certificate. The server certificate allows your browser to verify whether the server-sent certificate is signed and issued by a trusted center.  **Note:** This option is applicable only to HTTPS listeners.
	Select CA Certificate	The CA certificate. The CA certificate allows a server to verify whether a client certificate sent by your browser is trusted. If the verification fails, the connection request is denied.  **Note:** This option is applicable only to HTTPS listeners with two-way authentication.
	Peak Bandwidth	The bandwidth peak value for the listener, in Mbit/s.
	Session Persistence	Indicates whether to enable session persistence. For Layer-7 (HTTP and HTTPS) listeners, SLB supports cookie-based session persistence.
	Cookie Persistence	The cookie processing method. • **Cookie Insert**: SLB adds a cookie to the first response from a backend server, inserting SERVERID in the HTTP or HTTPS response packet, and records the backend server. The next time the client carries this cookie to access SLB, the listener will forward the request to the recorded backend server. If you use this method, you must specify the cookie timeout period in **Timeout**. • **Cookie Rewrite**: When SLB discovers that a cookie is customized, it overwrites the original cookie in the response from a backend server with the new cookie,

Category	Parameter	Description
		and records the backend server. The next time the client carries the new cookie to access SLB, the listener will forward the request to the recorded backend server. If you use this method, you must customize the cookie to be inserted in the HTTPS or HTTP response in **Cookie Name** and maintain the cookie timeout period in the backend ECS instances. **Note:** You must set the cookie processing method only when you have session persistence enabled.
	Timeout	The connection timeout period.
	Idle Connection Timeout	The idle connection timeout period, in seconds. Value range: 1—60. If no request is received during the specified timeout period , SLB closes the connection and starts a new connection when the next request comes.
	Enable VServer Group	Indicates whether to use a VServer group. If you turn on the Enable VServer Group switch, select a VServer group to bind to the listener. A VServer group consists of multiple ECS instances that provide the same services. Requests are forwarded to the ECS instances in the specified VServer group based on the forwarding rules configured for the listener. If no VServer group is used, client requests are forwarded to the backend ECS instances of the SLB instance based on the forwarding rules configured for the listener. **Note:** A VServer group cannot be changed once it is bound to a listener.
Health Check Settings	**Enable Health Check**	Indicates whether to enable health check. Health check is enabled by default. To ensure service availability, we recommend that you enable health check.
	Domain Name and Health Check URI	The domain name and URI for health check. By default, SLB uses the internal IP address of a backend ECS instance to

Category	Parameter	Description
		initiate an HTTP head request to the default homepage of the application server for health check. • If the page used for health check is not the default homepage of the application server, you must specify the domain name and URI for health check. • If you have defined the host field parameters for the HTTP head requests, you only need to specify the URI for health check.
	Health Status	The HTTP status codes for health check.
	Port	The port used by the health check service to access backend ECS instances. The backend port configured for the listener is used by default.
	Response Timeout (Seconds)	The maximum response time for a health check before timing out. If a backend ECS instance does not respond to the health check requests within the specified period, the health check fails.
	Health Check Interval (Seconds)	The time interval between two consecutive health checks. All nodes in the LVS cluster perform regular health checks at the specified interval independently and in parallel on backend ECS instances.
	Unhealthy Threshold	The number of consecutive failed health checks that must occur on an ECS instance for an LVS node to declare this ECS instance unhealthy.
	Healthy Threshold	The number of consecutive successful health checks that must occur on an ECS instance for an LVS node to declare this ECS instance healthy.

6. Click **OK**.

22.5.4 Configure forwarding rules

You can configure domain name-based or URL-based forwarding rules for an SLB instance that has Layer-7 listeners to distribute requests with different domain names or URLs to different ECS instances.

Context

You can add multiple forwarding rules under a single listener. Each forwarding rule is associated with a different VServer group. (A VServer group consists of multiple ECS instances.) For example

, you can forward all read requests to one VServer group and all write requests to another VServer group to optimize resource usage.

Server Load Balancer has the following judgment rules for request forwarding:

- If a request matches a domain name-based or URL-based forwarding rule configured for a listener, the request is forward to the VServer group based on the rule.
- If a request does not match any domain name-based or URL-based forwarding rules configured for a listener to which a VServer group is bound, the request is forwarded to the VServer group.
- If none of the preceding conditions are met, the requests are forwarded to the backend servers of the SLB instance based on the listener configuration.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target SLB instance to go to the instance details page.
3. Click the **Listeners** tab.
4. On the **Listeners** tab page, locate the target listener.

   Domain name-based or URL-based forwarding rules can be configured only for HTTP and HTTPS listeners.
5. In the **Actions** column, click the  icon and choose **Configure Forwarding Rule** from the shortcut menu.
6. In the **Configure Forwarding Rule** dialog box, click **+ Add Forwarding Rule**.
7. Configure forwarding rules based on the following principles:
   - Configure a domain name-based forwarding rule
     - When configuring a domain name-based forwarding rule, leave the URL field empty (no forward slash is required). The domain name can contain only letters, numbers, hyphens (-), and periods (.).
     - Domain names support both exact match and wildcard match. For example, www.aliyun.com is an exact domain name, while *.aliyun.com and *.market.aliyun.com are wildcard domain names. When a request matches multiple domain name-based forwarding rules simultaneously, an exact match takes precedence over any wildcard match, as described in the following table.

Table 22-5: Domain name matching rule

Type	Request URL	Domain name matching rule (√ indicates that the domain name is matched, while x indicates that the domain name is not matched.)		
		www.aliyun.com	*.aliyun.com	*.market.aliyun.com
Exact match	www.aliyun.com	√	x	x
Wildcard match	market.aliyun.com	x	x	x
	info.market.aliyun.com	x	x	√

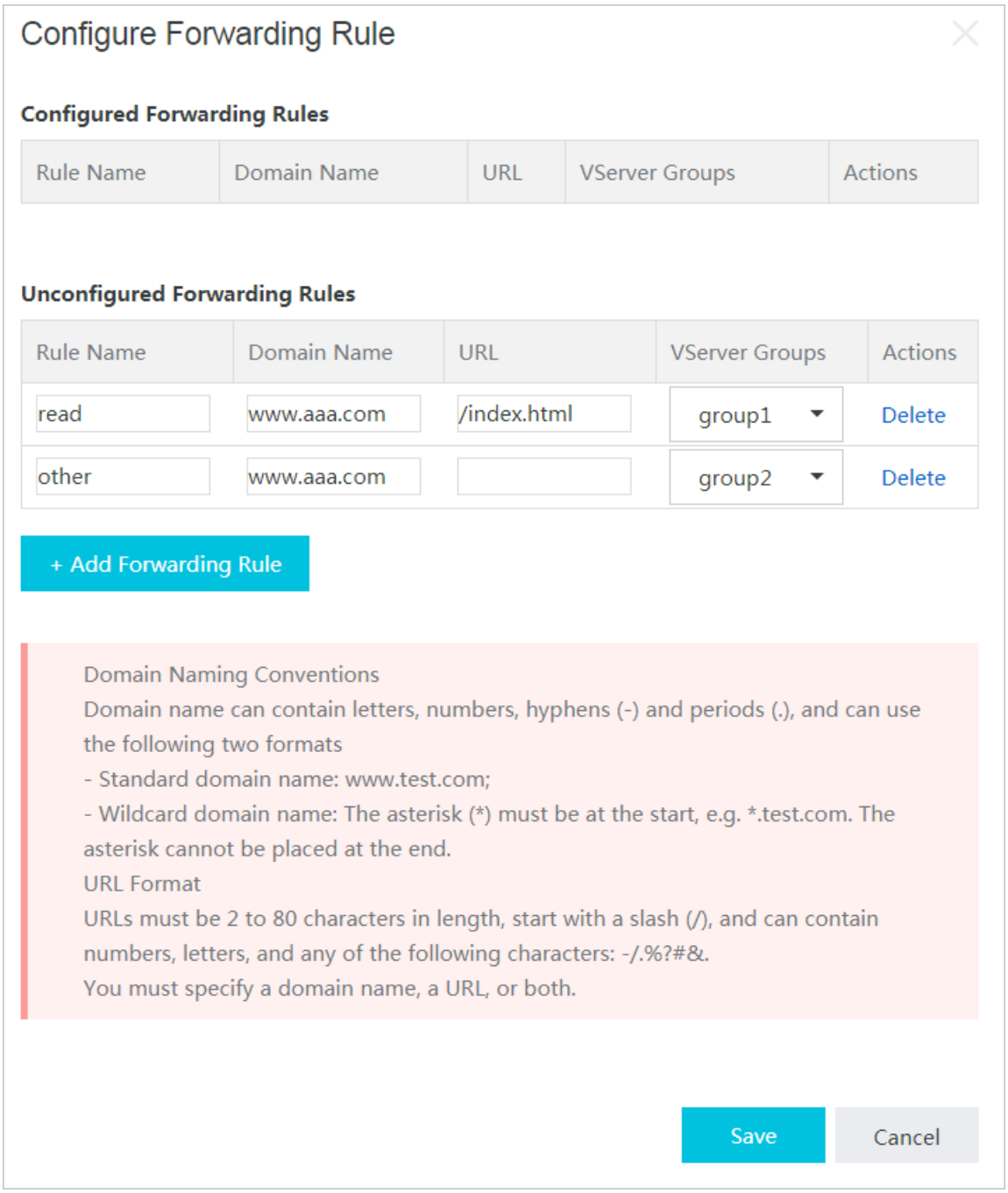
- Configure a URL-based forwarding rule
  - When configuring a URL-based forwarding rule, leave the Domain Name field empty.
  - The URL can contain only letters, numbers, hyphens (-), periods (.), forward slashes (/), percent signs (%), question marks (?), number signs (#), and ampersands (&).
  - The URL must start with a forward slash (/).

    Note:
    If you enter only one forward slash (/) in the URL, the URL-based forwarding rule is invalid.

  - URL-based forwarding rules support string matching and adopt sequential matching. For example, /admin, /bbs_, and /ino_test.
- Configure a domain name- and URL-based forwarding rule

  You can combine domain name- and URL-based forwarding rules to use different URLs under the same domain name for traffic forwarding. We recommend that you configure a default forwarding rule (with the URL field left empty) to prevent failures to access other unmatched URLs.

  Assume that the domain name of a website is *www.aaa.com*, and that VServer group 1 is required to process requests from *www.aaa.com/index.html* while VServer group 2 is required to process other requests. In order to meet these processing requirements, two forwarding rules must be configured, as shown in the following figure. Otherwise, the 404 response code will be returned if no forwarding rule matches domain name *www.aaa.com*.



8. Click **Save**.

22.5.5 Configure access control

You can add a whitelist to allow specific IP addresses to access Server Load Balancer (SLB).

Context

When configuring a whitelist, note that:

- Once a whitelist is configured, only IP addresses in the whitelist can access the SLB listener, which can lead to certain business risks.

- If no whitelist is configured after access control is enabled, no IP address can access the SLB listener.
- During the whitelist configuration process, access to the SLB listener may be interrupted for a short period of time.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target SLB instance to go to the instance details page.
3. Click the **Listeners** tab.
4. On the **Listeners** tab page, locate the target listener.
5. In the **Actions** column, click the  icon and choose **Configure Access Control** from the shortcut menu.
6. In the **Configure Access Control** dialog box that appears, turn on the **Enable Access Control** switch.
7. In Whitelist, enter IP addresses.

   Separate multiple IP addresses with commas (,). You can add up to 300 unique IP addresses and network segments in the form of CIDR blocks, such as 10.23.12.0/24.
8. Click **OK**.

22.5.6 Stop a listener

After a listener is stopped, it no longer forwards traffic.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target instance to go to the instance details page. Then click the **Listeners** tab.
3. On the **Listeners** tab page, locate the target listener.
4. In the **Actions** column, click the  icon and choose **Stopped** from the shortcut menu.

22.5.7 Start a listener

You can restart a stopped listener.

Procedure

1. *Log on to the Server Load Balancer console*.

2. Click the ID of the target instance to go to the instance details page. Then, click the **Listeners** tab.
3. On the **Listeners** tab page, locate the target listener.
4. In the **Actions** column, click the  icon and choose **Start** from the shortcut menu.

22.5.8 Edit listener settings

You can edit the settings of a listener.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target instance to go to the instance details page. Then, click the **Listeners** tab.
3. On the **Listeners** tab page, locate the target listener.
4. In the **Actions** column, click the  icon and choose **Change** from the shortcut menu.
5. Change the listener settings, and click **Save**.

22.5.9 Delete a listener

You can delete listeners which you no long need.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target instance to go to the instance details page. Then, click the **Listeners** tab.
3. On the **Listeners** tab page, locate the target listener.
4. In the **Actions** column, click the  icon and choose **Delete** from the shortcut menu.
5. In the message that appears, click **OK**.

22.6 Backend servers

22.6.1 Backend server overview

Before using SLB, you must add ECS instances as the backend servers of your SLB instance to receive and process requests forwarded by the listeners.

You can use SLB to virtualize multiple ECS instances in the same region into a high-performanc e and high-availability application server pool by configuring virtual IP addresses (VIPs). SLB performs health check on ECS instances in the application server pool, automatically blocks abnormal ECS instances, and begins distributing requests to these ECS instances only when they become functional again. . The health check function improves the overall availability of services and mitigates the impact of exceptions in backend ECS instances on the services.

You can increase or decrease the number of backend ECS instances at any time. Before you perform these operations, make sure that health check is enabled and that there is at least one properly running backend ECS instance to ensure service continuity.

22.6.2 Add backend servers

After creating an SLB instance, you must add ECS instances as backend servers to your SLB instance to process the connection requests forwarded by the listeners.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then, click the **Backend Servers** tab.
3. On the **Backend Servers** tab page, click **Add Backend Server**.
4. In the **Add Backend Server** dialog box that appears, select ECS instances and set their weights.

   An ECS instance with a higher weight receives a greater proportion of access requests. You can set the weights of ECS instances based on their service capabilities.

   Note:

   If the weight of an ECS instance is set to **0**, the ECS instance no longer receives new requests.
5. Click **Add**.

22.6.3 Modify the weight of an ECS instance

You can modify the weight of an added ECS instance. An ECS instance with a higher weight receives a greater proportion of access requests. You can set the weights of ECS instances based on their service capabilities.

Context

> **Note:**
> If a backend ECS instance has been added to a VServer group, you must modify the weight of this ECS instance by editing the VServer group. For more information about how to edit a VServer group, see *Edit a VServer group*.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then, click the **Backend Servers** tab.
3. In the **Actions** column corresponding to the target ECS instance, click the  icon and choose **Change** from the shortcut menu.
4. In the **Change Backend Server** dialog box that appears, change the weight of the ECS instance.

   > **Note:**
   > If the weight is set to **0**, the ECS instance no longer receives new requests.
5. Click **Save**.

22.6.4 Remove a backend ECS instance

Directly removing an ECS instance from an SLB instance may cause intermittent service interruptions. We recommend that before removing an ECS instance from an SLB instance, you change the weight of the ECS instance to 0 so that the SLB no longer forwards traffic to it.

Context

> **Note:**
> For a backend ECS instance that has been added to a VServer group, you must first remove the ECS instance from the VServer group by editing the VServer group. Then, switch to the Backend

Servers tab page to change the weight of the ECS instance to 0 and remove the ECS instance. For more information about how to edit a VServer group, see *Edit a VServer group*.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then, click the **Backend Servers** tab.
3. Locate the ECS instance to be removed. In the **Actions** column, click the  icon and choose **Remove** from the shortcut menu.
4. In the message that appears, click **OK**.

22.7 VServer groups

22.7.1 Add a VServer group

A VServer group is a group of ECS instances. By using VServer groups (different listeners are associated with different VServer groups), SLB forwards requests to different ECS instances. When you need to distribute different requests to different backend servers, or when you want to configure domain name-based or URL-based forwarding rules, you can use VServer groups.

Context

VServer groups have the following restrictions:

- An ECS instance can be added to multiple VServer groups.
- A VServer group can be associated with multiple listeners.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then click the **VServer Groups** tab.
3. Click the **Add VServer Group**.
4. In the **Create VServer Group** dialog box, perform the following operations:
   a) Enter a name for the VServer group.
   b) Select a search condition from the **ECS Instance ID** drop-down list, enter a value in the text box, and click **Search** to search for ECS instances.
   c) In the **Available Servers** list, click the ECS instances to be added one by one.

d) The ECS instances are added to the **Selected Servers** list. Then, set the ports and weights of the added ECS instances.

An ECS instance with a higher weight receives a greater proportion of access requests. You can set the weights of ECS instances based on their service capabilities.

> **Note:**
> If the weight of an ECS instance is set to **0**, the ECS instance no longer receives new requests.

e) Click **OK**.

22.7.2 View a VServer group

You can view information such as status and ports of the ECS instances in a VServer group.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then click the **VServer Groups** tab.
3. Locate the target VServer group. Click the group ID, or click the icon in the **Actions** column and choose **View** from the shortcut menu.

22.7.3 Edit a VServer group

You can modify the name of a VServer group, modify the weights of the ECS instances in the VServer group, and add new ECS instances to or remove ECS instances from the VServer group.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then, click the **VServer Groups** tab.
3. In the **Actions** column corresponding to the target VServer group, click the icon and choose **Change** from the shortcut menu.
4. Change the ECS instances in the VServer group.
5. Click **OK**.

22.7.4 Delete a VServer group

You can delete VServer groups which you no longer need.

Prerequisites

If a VServer group to be deleted is associated with a listener, you must first delete the listener.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the ID of the target SLB instance to go to the instance details page. Then click the **VServer Groups** tab.
3. Locate the target VServer group. In the **Actions** column, click the  icon and choose **Delete** from the shortcut menu.
4. In the message that appears, click **Yes**.
5. Click **OK**.

22.8 Certificates

22.8.1 Certificate overview

Server Load Balancer provides a certificate management function for HTTPS listeners.

To configure HTTPS listeners, you must upload the required certificates.

- For HTTPS two-way authentication, upload both the CA and server certificates.
- For HTTPS one-way authentication, upload only the server certificate.

After the certificates are uploaded to Server Load Balancer, you do not need to deploy the certificates on the backend ECS instances. Private keys uploaded to the certificate management system are encrypted and stored.

- Server certificate: allows your browser to check whether the server-sent certificate is signed and issued by a trusted center. You can purchase a server certificate from Alibaba Cloud Security Certificate Service or other service providers.
- Client certificate: proves your identity when you use a client to communicate with the server. You can sign a client certificate with a self-signed CA certificate.
- CA certificate: allows a server to verify whether a client certificate sent by your browser is trusted. If the verification fails, the connection request is denied.

22.8.2 Certificate format

Only PEM certificates in the Linux environment can be uploaded to Server Load Balancer.

Certificate format requirements

Make sure that the certificate to be uploaded meets the following requirements:

- Certificate issued by the root CA

  If your certificate is issued by the root CA, only this certificate is required for access devices such as browsers to trust your website. Make sure that the certificates comply with the following rules:

  - The certificate content is placed between `-----BEGIN CERTIFICATE-----` and `-----END CERTIFICATE-----`. Ensure that the certificate includes the header and footer when uploading it.
  - Each row contains 64 characters, and the last row can contain less than 64 characters.
  - The certificate content cannot include any spaces.
- Certificate issued by an intermediate CA

  If the certificate has been issued to you by an intermediate CA and the certificate file consists of multiple certificates, you must combine the server certificate and intermediate certificate before uploading them. Combine the certificates based on the following rules:

  - The server certificate must be followed by the intermediate certificate. There cannot be any blank rows between the certificates.
  - The certificate content cannot include any spaces.
  - There cannot be any blank rows between the certificates. Each row contains 64 characters. For more information, see *https://www.ietf.org/rfc/rfc1424.txt*.
  - The certificates must meet the format requirements. Typically, the CA provides a relevant description when issuing a certificate. Pay attention to the rule description.

RSA private key format requirements

When you upload a server certificate, you must also upload an RSA private key for the certificate. Make sure that the RSA private key complies with the following rules:

- The key is placed between `-----BEGIN RSA PRIVATE KEY-----` and `-----END RSA PRIVATE KEY-----`. Ensure that the key includes the header and footer when uploading it.
- There cannot be any blank rows in the content. Each row must contain exactly 64 characters, except for the final row. The final row can contain less than 64 characters.

Note:

If your private key is encrypted (for example, the header and footer of the private key are `-----BEGIN PRIVATE KEY-----` and `-----END PRIVATE KEY-----` or `-----BEGIN ENCRYPTED PRIVATE KEY-----` and `-----END ENCRYPTED PRIVATE KEY-----`) or the private key contains `Proc-Type: 4,ENCRYPTED`, you must first run the following command to convert the private key and upload *new_server_key.pem* together with the server certificate:

```
openssl rsa -in old_server_key.pem -out new_server_key.pem
```

22.8.3 Generate a CA certificate

To configure HTTPS two-way authentication, you must generate and upload a CA certificate after purchasing a server certificate.

Context

This topic describes how to generate a self-signed CA certificate by using OpenSSL.

Procedure

1. Run the following commands to create a *ca* folder under the */root* directory and then four subfolders under the *ca* folder:

```
$ sudo mkdir ca
$ cd ca
$ sudo mkdir newcerts private conf server
```

   - *newcerts* subfolder: is the certificate backup directory that stores the digital certificates signed by the CA.
   - *private* subfolder: stores the private key of the CA.
   - *conf* subfolder: stores the configuration files for simplifying parameters.
   - *server* subfolder: stores the server certificate.

2. Create an *openssl.conf* file that contains the following information under the *conf* directory:

```
[ ca ]
default_ca = foo

[ foo ]
dir = /root/ca
database = /root/ca/index.txt
new_certs_dir = /root/ca/newcerts

certificate = /root/ca/private/ca.crt
serial = /root/ca/serial
private_key = /root/ca/private/ca.key
```

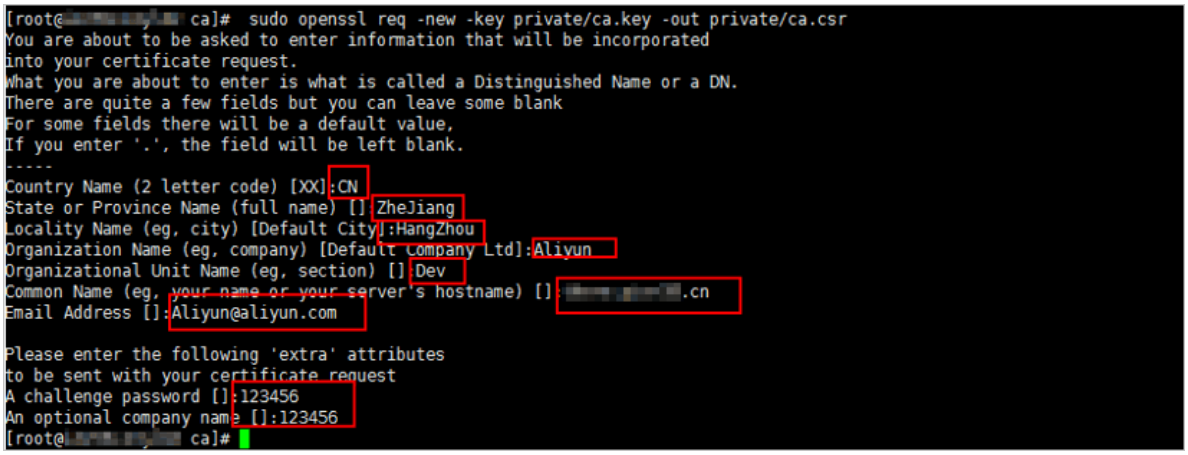
```
RANDFILE = /root/ca/private/.rand

default_days = 365
default_crl_days= 30
default_md = md5
unique_subject = no
policy = policy_any

[ policy_any ]
countryName = match
stateOrProvinceName = match
organizationName = match
organizationalUnitName = match
localityName = optional
commonName      = supplied
emailAddress    = optional
```

3. Run the following commands to generate a private key:

```
$ cd /root/ca
$ sudo openssl genrsa -out private/ca.key
```

4. Run the following command, enter required information as shown in the following figure, and then press Enter to generate a certificate request *.csr* file.

```
$ sudo openssl req -new -key private/ca.key -out private/ca.csr
```



5. Run the following command to generate a *.crt* file:

```
$ sudo openssl x509 -req -days 365 -in private/ca.csr -signkey
private/ca.key -out private/ca.crt
```

6. Run the following command to set the start sequence number for the key, which can be any four characters:

```
$ sudo echo FACE > serial
```

7. Run the following command to create a CA key library:

```
$ sudo touch index.txt
```

8. Run the following command to create a certificate revocation list for removing the client certificate:

```
$ sudo openssl ca -gencrl -out /root/ca/private/ca.crl -crldays 7 -
config "/root/ca/conf/openssl.conf"
```

The command output is as follows:

```
Using configuration from /root/ca/conf/openssl.conf
```

9. Run the following commands to view the generated CA certificate:

```
cd private
ls
```

22.8.4 Generate a client certificate

A client certificate proves your identity when you use a client to communicate with the server.

Prerequisites

A CA certificate is required to sign the client certificate. Make sure that you already *Generate a CA certificate*.

Procedure

1. Run the following command to create the *users* directory under the *ca* directory to store keys:

```
$ sudo mkdir users
```

2. Run the following command to create a key for the client certificate:

```
$ sudo openssl genrsa -des3 -out /root/ca/users/client.key 1024
```

When creating the key, enter the pass phrase as the key password to prevent unauthorized use if the key leaks. Enter the same password twice.

3. Run the following command to create a certificate signature request *.csr* file for the key:

```
$ sudo openssl req -new -key /root/ca/users/client.key -out /root/ca
/users/client.csr
```

Enter the pass phrase stored in the preceding step as prompted, press Enter, and enter the required information as prompted.

Note:
A challenge password is the password of the client certificate (which must be separated from the password of client.key). It can be the same as the password of the server or root certificate.

4. Run the following command to use the CA key to sign the client key:

```
$ sudo openssl ca -in /root/ca/users/client.csr -cert /root/ca/
private/ca.crt -keyfile /root/ca/private/ca.key -out /root/ca/users/
client.crt -config "/root/ca/conf/openssl.conf"
```

Enter y twice when prompted to confirm the operation.

5. Run the following command to convert the certificate to a *PKCS12* file that can be recognized by most browsers:

```
$ sudo openssl pkcs12 -export -clcerts -in /root/ca/users/client.crt
 -inkey /root/ca/users/client.key -out /root/ca/users/client.p12
```

Enter the pass phrase of client.key as prompted and press Enter. Then enter the password used for exporting the client certificate. This password is used to protect the client certificate, which is required when the client certificate is installed.

6. Run the following commands to view the generated client certificate:

```
cd users
ls
```

22.8.5 Upload a certificate

SLB provides the certificate management function to implement data transfer encryption and authentication over HTTPS. You can store certificates in the SLB certificate management system without deploying the certificates on backend ECS instances. Private keys uploaded to the certificate management system are encrypted and stored. A certificate can be applied to one or more listeners.

Note:

Each account can upload up to 100 certificates.

Prerequisites

You have generated a server or CA certificate to be uploaded.

Procedure

1. *Log on to the Server Load Balancer console*.
2. Click the **Certificates** tab to view the certificate list.
3. Click **Upload Certificate**.
4. In the **Upload Certificate** dialog box, set the parameters.

   Table 22-6: Parameters for uploading a certificate

Parameter	Description
Region	The region where a certificate is used. SLB manages certificates by regions. To use a certificate in multiple regions, upload the certificate in each region individually.
Department	The department that uses a certificate.
Project	The project that uses a certificate.
Certificate Type	The type of a certificate to be uploaded. • CA Certificate: A CA certificate allows a server to verify whether a client certificate sent by your browser is trusted. If the verification fails , the connection request is denied. • Server Certificate: A server certificate allows your browser to check whether the server-sent certificate is signed and issued by a trusted center.
Certificate Name	The name of a certificate.
Certificate Contents	The content of a certificate. The certificate must be in the PEM format. You can click **Examples** to view the sample format. For more information, see *Certificate format*.
Private Key	The private key of a server certificate. Private keys must comply with the format requirements in SLB. You can click **Examples** to view the sample format.

5. Click **OK**.

22.8.6 Convert the format of a certificate

Server Load Balancer supports PEM certificates only. Certificates in other formats must be converted to the PEM format before they can be uploaded to Server Load Balancer.

Context

We recommend that you use OpenSSL to convert certificates. This topic describes how to convert popular certificate formats to PEM:

- DER: This format is usually used on Java platforms.
- P7B: This format is usually used in Windows Server and Tomcat.
- PFX: This format is usually used in Windows Server.

Convert the certificate format from DER to PEM

1. Run the following command to convert the format of a certificate:

```
openssl x509 -inform der -in certificate.cer -out certificate.pem
```

2. Run the following command to convert the private key:

```
openssl rsa -inform DER -outform PEM -in privatekey.der -out
privatekey.pem
```

Convert the certificate format from P7B to PEM

1. Run the following command to convert the format of a certificate:

```
openssl pkcs7 -print_certs -in incertificat.p7b -out outcertificate.
cer
```

2. In *outcertificat.cer*, retrieve the `[——-BEGIN CERTIFICATE——-, ——-END CERTIFICATE——-]` content and upload the content as a certificate.

Convert the certificate format from PFX to PEM

1. Run the following command to extract the private key:

```
openssl pkcs12 -in certname.pfx -nocerts -out key.pem -nodes
```

2. Run the following command to extract the certificate:

```
openssl pkcs12 -in certname.pfx -nokeys -out cert.pem
```

22.8.7 Replace a certificate

If your certificate expires or an error is reported when you upload the certificate, you can generate and upload a new certificate, and then delete the existing one.

Procedure

1. Create and upload a new certificate.

   For more information, see *Generate a certificate* and *Upload a certificate*.

2. When you configure an HTTPS listener, you must also configure a new certificate.

   For more information, see *Configure a Layer-7 listener*.

3. On the **Certificates** page, locate the old certificate. In the **Actions** column, click the  icon and choose **Delete Certificate** from the shortcut menu to delete the certificate.

23 Virtual Private Cloud (VPC)

23.1 What is VPC

A Virtual Private Cloud (VPC) is logically isolated from other virtual networks.

You can have full control over your VPC. For example, you can specify its IP address range, and configure routing tables and gateways. You can also use Apsara Stack resources such as ECS, RDS, and SLB in your own VPC. Additionally, you can connect a VPC to another VPC or to an on-premises IDC network to form an on-demand network environment. Consequently, you can smoothly migrate applications to the cloud and expand the on-premises IDCs.

Components

Each VPC consists of a Classless Inter-Domain Routing (CIDR) block, a VRouter, and at least a VSwitch.

- CIDR blocks

  A CIDR block indicates the IP address range of VPC. IP addresses of all cloud resources deployed in the VPC are allocated from the specified CIDR block. When you create a VSwitch or VPC, you must specify the private IP address range in the form of CIDR blocks.

  You can use the standard CIDR blocks and their subnets in the following table as the private IP address range of a VPC.

CIDR blocks	Number of available private IP addresses (Private IP addresses reserved by the system are excluded.)
192.168.0.0/16	65,532
172.16.0.0/12	1,048,572
10.0.0.0/8	16,777,212

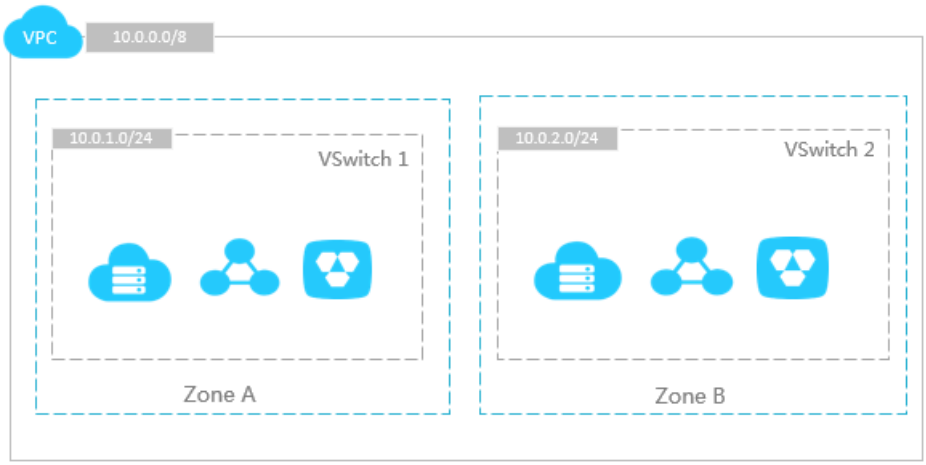
- VRouters

  VRouters are the network hub of VPC. A VRouter is an important component of VPC. It connects VSwitches in VPC and serves as the gateway that connects VPC to another network. A VRouter is automatically created after a VPC is created. Each VRouter has a routing table.

- VSwitches

VSwitch is a basic network device of VPC and is used to connect different cloud service instances. After you create a VPC, you can create VSwitches to divide a VPC into one or more subnets. VSwitches within a VPC are interconnected. You can deploy different applications on the VSwitches that reside in different zones to improve service availability.

Figure 23-1: VPC



23.2 Quick start

23.2.1 Tutorial overview

This topic describes how to quickly create a VPC and deploy an ECS instance in it.

Specific operations are as follows:

- *Log on to the VPC console*

  This topic describes how to log on to the VPC console.
- *Create a VPC and a VSwitch*

  To use cloud services in VPC, you need to create a VPC and VSwitch.
- *Create a security group*

  Before creating an ECS instance in a VPC, you must first create a security group. Security groups are an important means to isolate networks. Security groups allow you to set network access control for one or more ECS instances.
- *Create an ECS instance*

  An ECS instance is a virtual computing environment that consists of the most basic server components such as the CPU, memory, OS, disk, and bandwidth.

23.2.2 Log on to the VPC console

This topic describes how to log on to the VPC console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.
5. Choose **Console** > **Compute, Storage & Networking** > **Virtual Private Cloud** from the top navigation bar.

23.2.3 Create a VPC and a VSwitch

To use a cloud service in VPCs, you must create a VPC and VSwitch.

Context

When creating a VPC, note that:

- Only one CIDR block can be specified for each VPC. For more information, see *Plan a CIDR block*.
- After a VPC is created, a VRouter and a routing table are automatically created. Each VPC can contain only one VRouter and one routing table.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** tab page, click **Create**.
3. Configure the VPC based on the following information.

   Table 23-1: VPC configurations

Name	Configuration method
Name	Enter the VPC name. The name must be 2 to 128 characters in length. It must start with an English letter or Chinese character. It cannot contain special characters such as the at sign (@), backslash (/), colon (:), angle brackets (<>), curly brackets ({}), braces ([]), or space.
Description	Add the VPC description.
Region	Select a region for the VPC.
Department	Select a department for the VPC.
Shared with Subdepartments	Specify whether to allow lower-level department administrators to share VPC resources.
CIDR Block	Select a CIDR block for the VPC. The CIDR block cannot be modified after the VPC is created.

4. Click **OK**. In the Create VPC dialog box that appears, click **Next** to create a VSwitch.
5. In the **Create VSwitch** dialog box that appears, configure the VSwitch based on the following information.

   Table 23-2: VSwitch configurations

Name	Configuration method
Zone	Select a zone for the VSwitch. In a VPC, a VSwitch can be located in only one zone and cannot span across several zones. You can deploy the cloud service instances on VSwitches in different zones to implement cross-zone disaster recovery.

Name	Configuration method
	Note: A cloud service instance can be added to one VSwitch only.
Name	Enter the name of the VSwitch.
CIDR Block	Enter the CIDR block of the VSwitch. • You must specify the network segment of a VSwitch in the form of a CIDR block. The mask length of the VSwitch CIDR block can be between 16 to 29 bits, which can provide 8 to 6,5536 IP addresses. • The CIDR block of a VSwitch must be a subset of the CIDR block of the VPC. **Note:** If the CIDR blocks of your VSwitch and VPC are the same, only this single VSwitch can be created. • The first IP address and the last three IP addresses of each VSwitch are reserved for the system. For example, if the VSwitch CIDR block is 192.168.1.0/24, 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved for the system. • The CIDR block of a VSwitch cannot be the same as the destination CIDR block in the routing entry of the VPC where the VSwitch resides. However, it can be a subset of the destination CIDR block in the current routing entry. • After a VSwitch is created, its CIDR block cannot be modified.
Description	Enter the description of the VSwitch.

6. Click **OK**, and click **Cancel** to close the dialog box.

23.2.4 Create a security group

Before creating an ECS instance in a VPC, you must first create a security group. Security groups are an important means for network security isolation. They are used to set network access control for one or more ECS instances.

Procedure

1. Log on to Apsara Stack Management Console.
2. Choose **Console** > **Compute, Storage & Networking** > **Elastic Compute Service** from the top navigation bar.
3. Click the **Security Groups** tab, and click **Create Security Group**.

4. In the **Create Security Group** dialog box that appears, configure the security group based on the following information.

   Table 23-3: Security group configurations

Name	Configuration method
Security Group Name	Enter a name for the security group.
Region	Select a region for the security group. Make sure the VPC and security group are in the same region.
Department	Select a department for the security group. Make sure that the VPC and security group are in the department.
Description	Add descriptions about the security group.
Project	Select a project for the security group. Make sure that the VPC and security group are in the project.
Network Type	Select **VPC**.
VPC	Select a VPC for the security group.

5. Click **OK**.

23.2.5 Create an ECS instance

An ECS instance is a virtual computing environment that consists of most basic components of a server, such as the CPU, memory, OS, disk, and bandwidth.

Procedure

1. Log on to Apsara Stack Management Console.
2. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Elastic Compute Service**.
3. Click the **Instances** tab, and click **Create Instance**.
4. In the **Create Cloud Server (ECS)** dialog box that appears, configure the ECS instance, and then click **Create**.

   For more information about how to configure ECS instances, see *`Create an instance`* in *`Quick start`* of *ECS User Guide*.

   Note:
   In Network, set Network Type to **VPC**, and select a VPC and VSwitch.

23.3 VPC

23.3.1 Plan a CIDR block

When creating a VPC and VSwitch, you must specify the private IP address range for your VPC as CIDR blocks.

CIDR is a bitwise, prefix-based standard for the representation of IP addresses. It combines multiple IP address blocks into a single routing entry to facilitate routing. IP address blocks with a subnet mask of /25, /26, or /27 can be allocated flexibly. These IP address blocks are called CIDR blocks.

VPC CIDR blocks

When planning the CIDR block for a VPC, note that:

- You can use the standard private network segments (192.168.0.0/16, 10.0.0.0/0, and 172.16.0.0/12) and their subnets as the CIDR blocks of VPCs. Only one CIDR block can be specified for each VPC. When you deploy Apsara Stack, you can use **`vpc_customer_private_cidr`** to specify the CIDR blocks (you can select one from them when you create a VPC) in the global configuration during the delivery planning phase.
- When a VPC is created using APIs, the network mask length of the VPC is from 8 to 24 bits.
- After a VPC has been created, its CIDR block cannot be modified.

VSwitch CIDR blocks

When planning VSwitch CIDR blocks, note that:

- The network mask length of the VSwitch CIDR block can be from 16 to 29 bits, which can provide 8 to 6,5536 IP addresses.
- The CIDR block of a VSwitch must be within the range of the CIDR block of its VPC.

  Note:
  If the CIDR blocks of your VSwitch and VPC are the same, only this single VSwitch can be created.
- The first IP address and the last three IP addresses of each VSwitch are reserved. For example, if the VSwitch CIDR block is 192.168.1.0/24, 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved for the system.

- The CIDR block of a VSwitch cannot be the same as the destination CIDR block in the routing entry of the VPC where the VSwitch resides. However, it can be a subset of the destination CIDR block in the current routing entry.
- After a VSwitch has been created, its CIDR block cannot be modified.

23.3.2 Create a VPC

A VPC is an isolated virtual network environment that is built on Alibaba Cloud. You have full control over your VPC. For example, you can specify its IP address range, and configure routing tables and gateways. You can also use Apsara Stack resources such as ECS, RDS, and SLB in your own VPC. To use a cloud service in VPCs, you must create a VPC and VSwitch.

Context

When creating a VPC, note that:

- Only one CIDR block can be specified for each VPC. For more information, see *Plan a CIDR block*.
- After a VPC is created, a VRouter and a routing table are automatically created. Each VPC can contain only one VRouter and one routing table.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** tab page, click **Create**.
3. Configure the VPC based on the following information.

Table 23-4: VPC configurations

Name	Configuration method
Name	Enter the VPC name. The name must be 2 to 128 characters in length and can contain digits, underscores (_), and hyphens (-). It must start with an English letter or a Chinese character.
Description	Add the VPC description.
Region	Select a region for the VPC.
Department	Select a department for the VPC.
Shared with Subdepartments	Specify whether to allow lower-level department administrators to share VPC resources.
CIDR Block	Select a CIDR block for the VPC.

Name	Configuration method
	The CIDR block cannot be modified after the VPC is created.

4. Click **OK**. In the Create VPC dialog box that appears, click **Next** to create a VSwitch.

   For more information, see *Create a VSwitch*.

23.3.3 View a VPC

On the **VPC** tab page, you can specify filtering conditions to search for the VPC you want to view.

Procedure

1. *Log on to the VPC console*.
2. Locate the relevant VPC, and click the instance ID or click the  icon. Then choose **Details** from the shortcut menu to view VSwitches and VRouters in the VPC.

23.3.4 Modify VPC information

After creating a VPC, you can modify the name and description of this VPC.

Procedure

1. *Log on to the VPC console*.
2. Locate the relevant VPC, click the  icon in the Actions column, and choose **Edit** from the shortcut menu.
3. In the **Modify VPC** dialog box that appears, set the name and description, and click **OK**.

23.3.5 Delete a VPC

You can delete a VPC if you no longer need it.

Prerequisites

Before deleting a VPC, you must release or move all resources, including VSwitches, from the VPC.

After the VPC is deleted:

- The security group in this VPC is deleted as well.
- Data related to the VPC instance cannot be restored.

Procedure

1. *Log on to the VPC console*.

2. Locate the relevant VPC, click the  icon in the Actions column, and choose **Delete** from the shortcut menu.
3. In the message that appears, click **OK**.

23.4 VSwitch

23.4.1 Create a VSwitch

A VSwitch is a basic network device of a VPC. It can connect different cloud service instances. After a VPC is created, you can divide the VPC into several subnets by adding VSwitches.

Context

When creating a VSwitch, note that:

- You can create a maximum of 24 VSwitches for each VPC.
- After a VSwitch is created, the system automatically adds a system routing entry that is in the same CIDR block as the VSwitch.
- In a VPC, a VSwitch can be located in only one zone and cannot span across several zones. You can deploy the cloud service instances on different VSwitches to implement cross-zone disaster recovery.

> **Note:**
> A VSwitch does not support multicasting or broadcasting.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** tab page, locate a relevant VPC and click the VPC ID.
3. Click the **VSwitches** tab, and click **Create**.
4. In the **Create VSwitch** dialog box that appears, configure the VSwitch based on the following information.

Table 23-5: VSwitch configurations

Name	Configuration method
Zone	Select a zone for the VSwitch. In a VPC, a VSwitch can be located in only one zone and cannot span across several zones. You can deploy the cloud service

Name	Configuration method
	instances on VSwitches in different zones to implement cross-zone disaster recovery. **Note:** A cloud service instance can be added to one VSwitch only.
Name	Enter the name of the VSwitch. The name must be 2 to 128 characters in length and can contain digits, underscores (_), and hyphens (-). It must start with an English letter or a Chinese character.
CIDR Block	Enter the CIDR block of the VSwitch. • You must specify the network segment of a VSwitch in the form of a CIDR block. The mask length of the VSwitch CIDR block can be between 16 to 29 bits, which can provide 8 to 6,5536 IP addresses . • The CIDR block of a VSwitch must be a subset of the CIDR block of the VPC. **Note:** If the CIDR blocks of your VSwitch and VPC are the same, only this single VSwitch can be created. • The first IP address and the last three IP addresses of each VSwitch are reserved for the system. For example, if the VSwitch CIDR block is 192.168.1.0/24, 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved for the system. • The CIDR block of a VSwitch cannot be the same as the destination CIDR block in the routing entry of the VPC where the VSwitch resides. However, it can be a subset of the destination CIDR block in the current routing entry. • After a VSwitch is created, its CIDR block cannot be modified.
Description	Enter the description of the VSwitch.

5. Click **OK**.

23.4.2 View VSwitches

On the **VSwitches** tab page, you can view the information of VSwitches that are created.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** tab page, locate the relevant VPC, and click the VPC ID.

3. Click the **VSwitches** tab to view VSwitch information.

23.4.3 Edit VSwitch information

After you create a VSwitch, you can edit the name and description of this VSwitch.

Procedure

1. *Log on to the VPC console*.
2. Locate the relevant VPC, and click the VPC ID.
3. Click the **VSwitch** tab.
4. Locate the relevant VSwitch, click the  icon, and choose **Edit** from the shortcut menu.
5. In the dialog box that appears, edit the name and description of the VSwitch, and click **OK**.

23.4.4 Delete a VSwitch

You can delete a VSwitch that is created.

Prerequisites

Before deleting a VSwitch, you must release or move cloud services from this VSwitch.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** tab page, locate the relevant VPC, and click the VPC ID.
3. Click the **VSwitch** tab.
4. Locate the relevant VSwitch, click the  icon in the Actions column, and choose **Delete** from the shortcut menu.
5. In the message that appears, click **OK**.

23.5 VRouter and route table

23.5.1 Overview

A VRouter is the network hub in a VPC. A VRouter is an important component of VPCs. It connects VSwitches in VPCs and serves as the gateway that connects VPCs to gateways in other networks.

A VRouter is automatically created for a VPC after the VPC is created. When the VPC is deleted , the VRouter is also deleted. A VRouter cannot be created or deleted directly. Each VRouter

maintains a routing table. A VRouter forwards network traffic based on the routing entries in the routing table.

A routing table is a list of routing entries stored in the VRouter. A routing table is automatically created for a VPC after the VPC is created. When the VPC is deleted, the routing table is also deleted. A routing table cannot be created or deleted directly.

Each item in a routing table is a routing entry. The routing entry defines the next-hop IP address for the network traffic to be routed to the specified destination CIDR block. Two types of routes are available: system routes and custom routes.

- System routes

  A system routing entry is automatically created for a VPC after the VPC is created. This routing entry defines the routes for the cloud service instances in the VPC to communicate with each other. A system routing entry is also automatically created for a VSwitch after the VSwitch is created. The CIDR block of this VSwitch is the destination. For more information, see *View VRouters and routing tables*.
- Custom routes

  You can add a custom route to forward the destination traffic to a specified next-hop. For more information, see *Add routing entries*.

23.5.2 View VRouters and routing tables

You can view created VRouters and their routing tables.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** tab page, locate the relevant VPC, and click the VPC ID.
3. Click the **VRouter** tab to view the VRouter and routing table.

23.5.3 Add routing entries

Each item in a routing table is a routing entry. A routing entry defines the next-hop IP address for traffic destined for a specified destination CIDR block. The number of custom routing entries in a routing table cannot exceed 48.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** page, locate the relevant VPC, and click the VPC ID.

3. Click the **VRouter** tab.
4. In **Route Table**, click **Create**.
5. In the dialog box that appears, configure the routing entry based on the following information.

Table 23-6: Routing entry configurations

Name	Configuration method
VPC CIDR Block	Set the destination CIDR block of the routing entry. • The destination CIDR block of a routing entry cannot be the same as the CIDR block of any VSwitch in the VPC, or a subset of the CIDR block of the VSwitch. • The destination CIDR block of a routing entry cannot be 100.64.0.0/10 or a subset of 100.64.0.0/10. • The destination CIDR blocks of routing entries in the same routing table cannot be the same. • If the specified destination CIDR block is an IP address, the default subnet mask 32 is used.
Next Hop Type	Set Next Hop Type. • **ECS Instances**: receives the forwarded traffic.
Next Hop Instance ID	If the next-hop address type is set to **ECS Instances**, select the ECS instance that receives the forwarded traffic. • The next-hop ECS instance specified in a routing entry must belong to the same VPC as the routing table. • Multiple routing entries can be directed to the same ECS instance.

6. Click **OK**.

24 Log Service

24.1 What is Log Service?

As a one-stop service for log data, Log Service (Log for short) provides you with multiple functions such as log data collection, query, analysis, and consumption.

Log Service has been honed by countless big data scenarios at Alibaba Group. Without any development, you can quickly collect, consume, query, and analyze log data by using Log Service . It helps increase the O&M efficiency and build capabilities to process large-volume logs in this data technology (DT) era.

Log Service provides you with the following fuctions:

- **Log collection**: Supports collecting multiple formats of log data such as Event, Binlog, and TextLog in real time through the Logtail client, JS, and other methods.
- **Query and analysis**: Provides real-time query and analysis for collected log data and supports generating visual charts and dashboards based on analysis results.
- **Status alarm**: Supports executing query and analysis statements regularly based on the query and analysis function. When query results meet alarm conditions, real-time alarms are reported based on configured alarm tasks.
- **Real-time consumption**: Provides real-time consumption interfaces for lod data collected to the server.

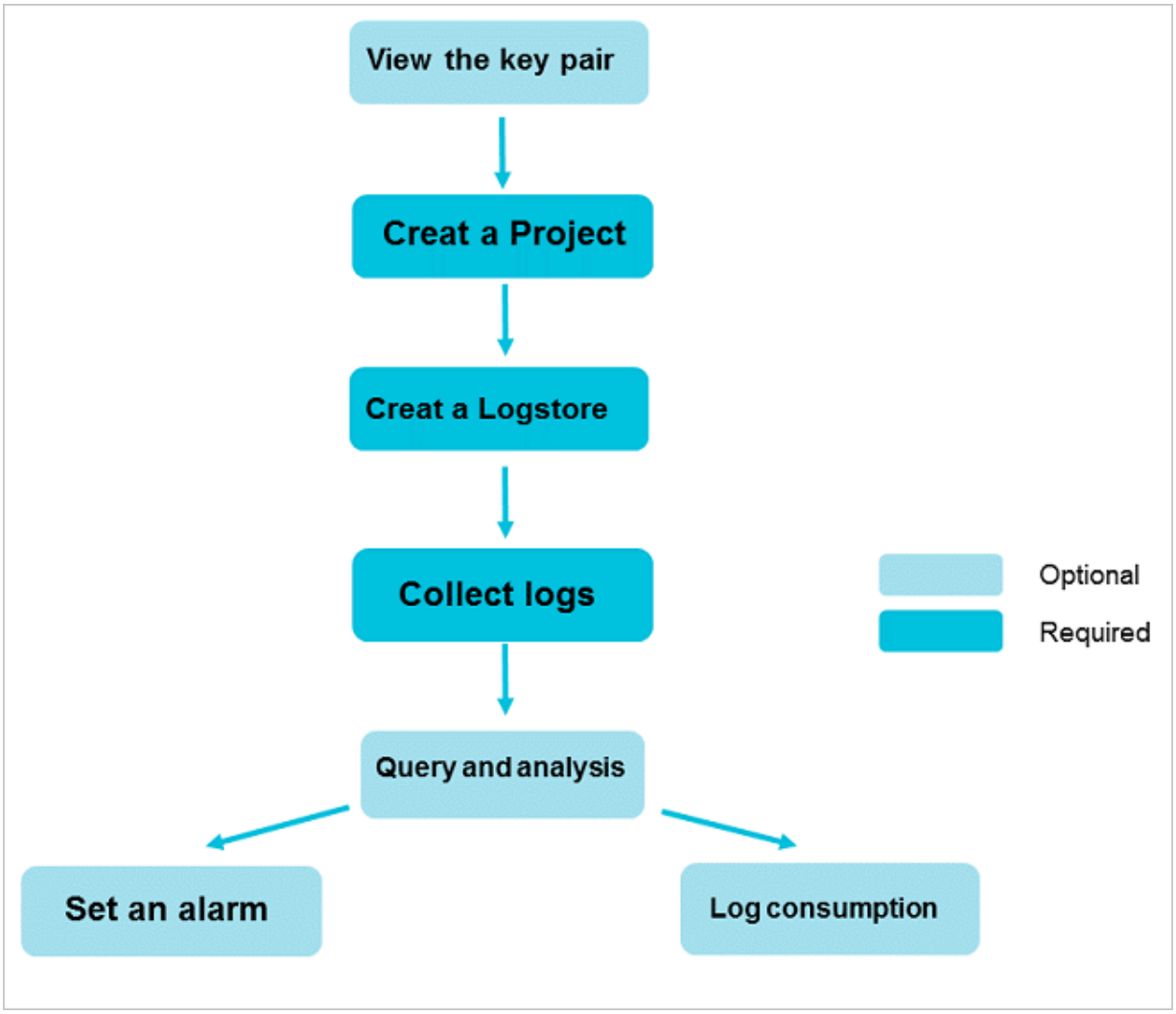
24.2 Quick start

24.2.1 Procedure

This section briefly explains how to use Log Service to quickly create a project and a Logstore, as well as how to collect log data.

For the detailed procedure, see *Figure 24-1: Log Service application flowchart*.

Figure 24-1: Log Service application flowchart



1. (Optional) *View the key pair*.

   AccessKeys are essential for using Log Service through APIs or SDKs. Therefore, ensure that your account has an AccessKey.

2. *Create a Project*.

   Create a project in the specified region and add a comment.

3. *Create a Logstore*.

   Create a Logstore under the created project and specify the number of shards.

4. *Collect text logs*.

   Select a log data collection method. Text log collection is used as an example.

5. *Configure an index*, and query and analyze logs.

Log Service supports *real-time query* and *analytics* for a large number of data. With indexing enabled, you can query logs in real time and configure *Graph description* and *Dashboard*.

6. *Set an alarm*.

   Log Service can trigger alarms based on the log query results, and allows you to configure rules to send alarm content through a custom WebHook method.

7. *Log consumption* logs.

   Log Service supports log consumption through the *Spark Streaming client*, *Storm spout*, and *Flink Connector*.

24.2.2 Log on to the Log Service console

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.

5. From the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Log Service**.
6. Log on to the Apsara Stack console and go to the Log Service homepage. Then, click **Go to Console** in the upper-right corner.
7. Set **Region** and **Department**, and then click **SLS** to log on to the Log Service console.

24.2.3 View the key pair

AccessKey is a requirement for Log Service operations through APIs/SKDs.

Procedure

1. *Log on to the Log Service console*.
2. Click your profile picture in the upper-right corner of the top navigation bar and click **Personal Information**.
3. In the left-side navigation pane, select **AccessKey**.
4. In the displayed dialog box, click **Confirm**.

   Now you can view the AccessKey ID and AccessKey Secret of the current account.

24.2.4 Create a Project

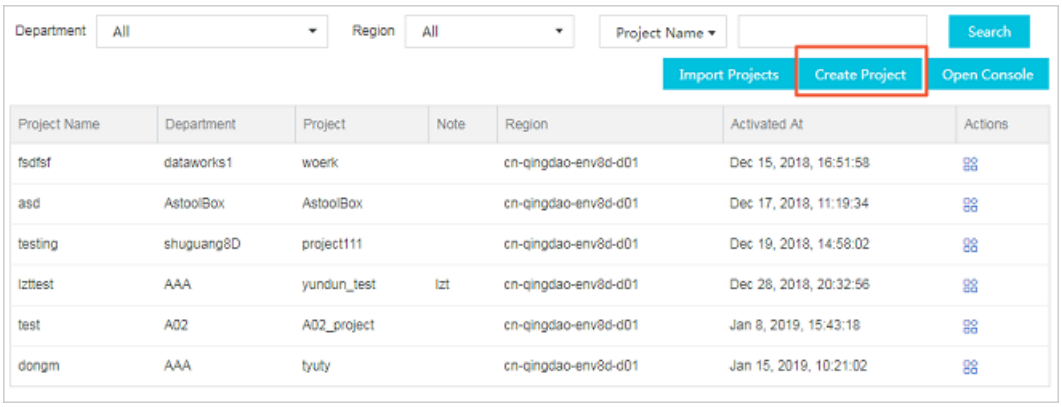
You can create a project on the Log Service homepage.

Context

Each department can have up to 10 projects.

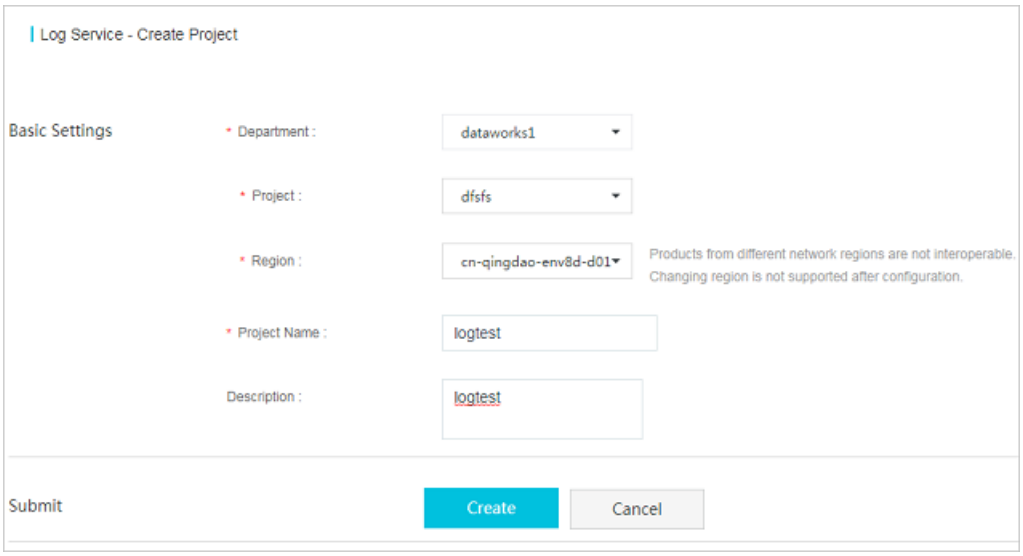
Procedure

1. *Log on to the Log Service console*.
2. Click **Create Project** in the upper-right corner.



3. Complete **Basic Settings**.

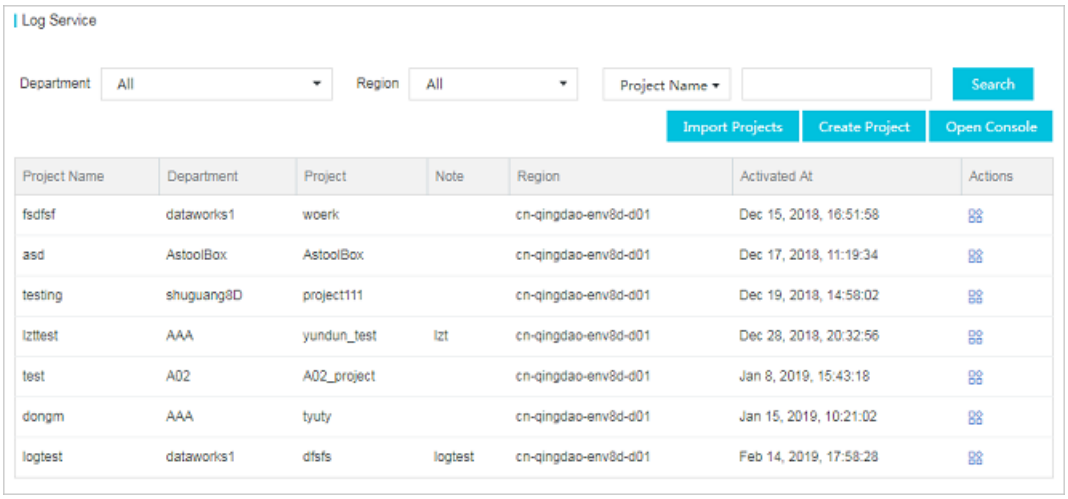
Parameter	Description
Department	The department to which the project belongs.
Project	The project to which the project belongs.
Region	You must specify the region where the project is created. Select a region based on the log sources and other relevant conditions. The products in different regions cannot connect with each other through the internal network. If you want to receive logs from an ECS instance, create a project in the same region as the ECS instance. Once a project is created, its region cannot be changed. Log Service does not support project migration. Therefore, select a region with caution.
Project name	The project name can only contain lowercase letters, numbers, and hyphens (-). It must start and end with a lowercase letter or number and must be 3 to 63 bytes in length. **Note:** • Once a project is created, its name cannot be changed. • The name of the project must be globally unique. The message "**Project XXX already exists**" is displayed if the project name that you entered has already been used by another user. If this is the case, enter another project name and try again.
Comments	Enter a simple description for the project. After the project is created, the description is displayed on the **Projects** page. If you want to modify the description after the project is created, go to the **Projects** page and click **Modify Description**.



4. Click **Create**.

Result

The created project is displayed on the Log Service homepage and the Log Service console.



You can also select a **Department** and **Region**, enter a **project name**, and click **Search** to quickly search for the specified project in the project list.

24.2.5 Create a Logstore

You can use the Log Service console or APIs to create a Logstore. For more information about how to create a Logstore using APIs, see the section CreateLogstore in *Log Service API references*.

Context

A Logstore is a set of resources created under a project. All the data in a Logstore comes from the same source. A Logstore is a unit for querying, analyzing, and shipping the collected log data.

- If your project is created in the Log Service console, it is automatically created under a level-1 department and does not belong to any project. You must import projects to Apsara Stack Management Console and migrate the project to another project by using the ownership change function.
- Up to 100 Logstores can be created under each Log Service project.

Procedure

1. *Log on to the Log Service console*.
2. Click a project name to go to the **Logstores** page, and then click **Create**. Create a Logstore.
3. Configure the parameters of the Logstore. Then, click **OK**.

Parameter	Description
Logstore name	A Logstore name can only contain lowercase letters, numbers, hyphens (-), and underscores (_). It must start and end with a lowercase letter or number and must be 3 to 63 bytes in length. **Note:** • A Logstore name must be unique in the project to which it belongs. • A Logstore name cannot be modified after creation.
Data Retention Time	Specifies the number of days for which the data will be retained in the Logstore. Its value ranges from 1 to 365, in days. After this period expires, logs are deleted. The data retention time can be modified after a Logstore is created. You can go to the **Logstores** page, click **Modify** in the **Actions** column, change **Data Retention Time**, and then click **Modify**.
Shards	Specifies the number of shards in the Logstore. You can create 1 to 10 shards for each Logstore. You can create up to 100 shards for each project.

24.2.6 Configure an index

Log Service supports real-time query and analysis of collected log data. You must enable and configure an index before you use the query and analysis function.

Context

You can use the query and analysis function only after you enable and configure an index.

Note:

- Collected log data can be queried and analyzed only after you enable an index.
- After an index is modified, the new settings take effect only on the logs collected after the modification. The settings are not applicable to history logs before the modification.

Procedure

1. *Log on to the Log Service console*.
2. Click the name of a project.
3. Select a Logstore and click **Search** in the **LogSearch** column.
4. Click **Enable** in the upper-right corner.

   If you have enabled an index, click **Index Attributes** > **Modify**.
5. On the **Search & Analysis** page, configure the index.

   You can configure **Full Text Index** or **Field Index**. If the two types of indexes are both configured, the field index takes effect.

   When you configure a field index, you can click **Disable** on the right of the field to delete the configuration.

Type	Configuration	Description
Full Text Index	**Case Sensitive**	Configure whether the index to be case sensitive or not.
	Include Chinese	Configure whether to include Chinese characters in the index.
	Delimiter	Configure the delimiter used to separate keywords.
Field Search	**Key**	Configure the log filed name.
	Type	Specify the log field type, which includes:

Type	Configuration	Description
		• **text** • **long** • **double** • **json**
	Alias	Configure the alias of a column.
	Delimiter	Configure the delimiter used to separate keywords.
	Case Sensitive	Configure whether the index to be case sensitive or not.
	Include Chinese	Configure whether to include Chinese characters in the index.
	Statistics	Configure whether enable statistics for the field. After statistics is enabled, you can analyze the field by using statistics statements.

6. Click **Confirm**.

   You index configuration takes effect within one minute.

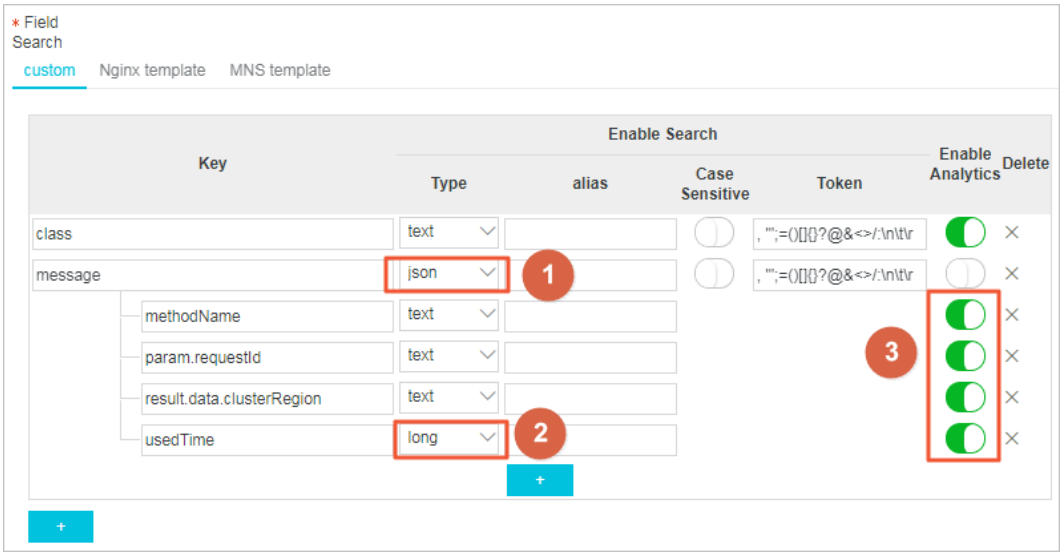
   The following log includes four key values in addition to the time:

NO.	Key	Type
0	time	-
1	class	text
2	status	long
3	latency	double
4	message	json

```
0. time:2018-01-01 12:00:00
  1. class:central-log
  2. status:200
  3. latency:68.75
  4. message:
  {
      "methodName": "getProjectInfo",
      "success": true,
```

```
        "remoteAddress": "1.1.1.1:11111",
        "usedTime": 48,
        "param": {
                "projectName": "ali-log-test-project",
                "requestId": "d3f0c96a-51b0-4166-a850-
f4175dde7323"
        },
        "result": {
            "message": "successful",
            "code": "200",
            "data": {
                "clusterRegion": "ap-southeast-1",
                "ProjectName": "ali-log-test-project",
                "CreateTime": "2017-06-08 20:22:41"
            },
            "success": true
        }
    }
```

Settings are as follows:



where:

- (1) indicates query of all the data of the String and Boolean types in JSON fields.
- (2) indicates query of data of the Long type.
- (3) indicates SQL analysis of configured fields.

What's next

After you configure an index, you can query and analyze collected log data in real time by entering statements in the search box on the search page and clicking **Search**.

24.2.7 Set an alarm

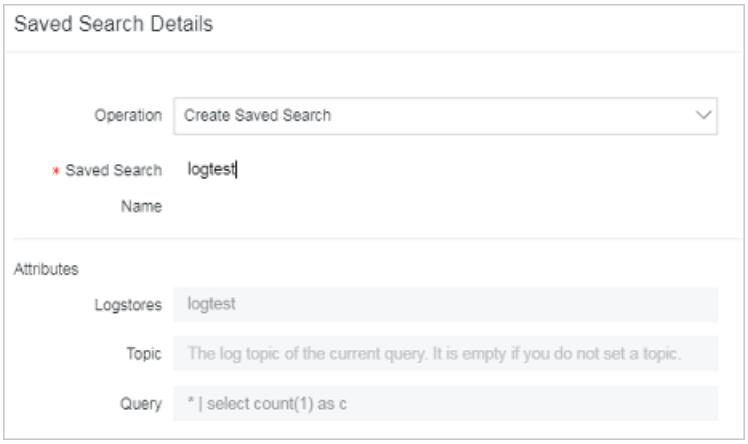
You can save Saved Search as an alarm on the query page so that Log Service performs a scheduled check and sends an alarm when the alarm condition is met.

Prerequisites

- Log data has been collected.
- *Configure an index*.

Procedure

1. *Log on to the Log Service console*.
2. On the **Logstores** page, click **Search** in the **LogSearch** column.
3. Enter a query statement in the search box.
4. Set a time range to be queried and click **Search**.
5. Click **Save As Alarm** in the upper-right corner of the page.



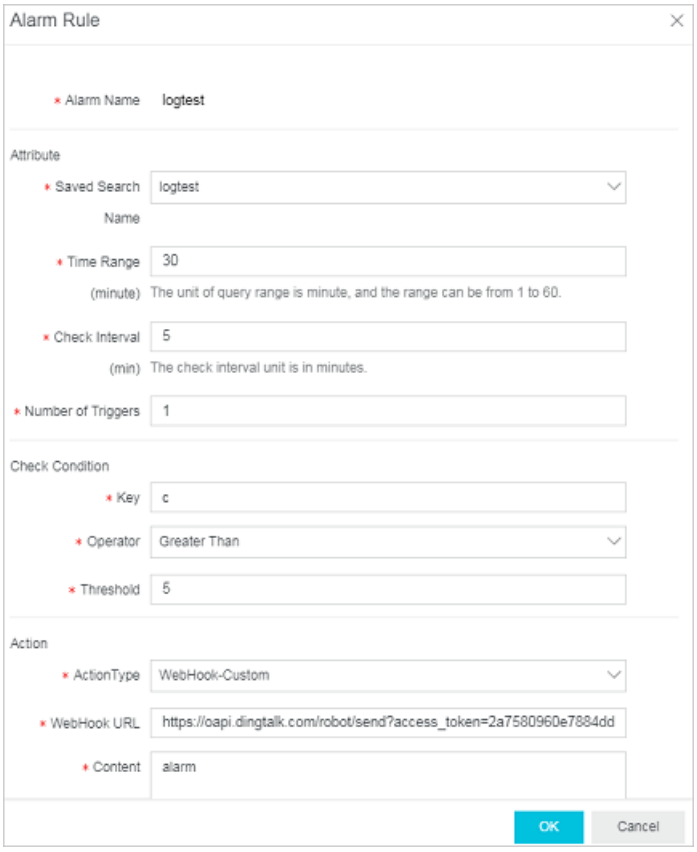
6. Configure alarm rules and click **OK**.

Rule	Description
Alarm rule name	The name of the alarm rule. This name must be a string of 3 to 63 characters.
Alarm rule attributes	
Saved Search name	The name of Saved Search used by the alarm. Saved Search can be set to: • **Current query**, indicating that the current query statement is used as Saved Search; • Another existing Saved Search.

Rule	Description
Data query time (minutes)	The time range of the data read by the server during each alarm check. For example, if it is set to 1, the server queries the data generated within the last 1 minute of the check time. The data query time is in minutes and ranges from 1 to 60.  **Note:** Currently, when performing an alarm check, the server only processes the first 10 data records generated during the time range for sampling purpose.
Check interval (minutes)	The interval at which the server performs an alarm check. The check interval is in minutes and ranges from 1 to 1,440.
Trigger times	The number of consecutive alarm check triggers. An alarm notification is sent when the specified trigger count is reached. Value range: 1 to 10,000. For example, if the **check interval (minutes)** is set to 1 and the **trigger times** is set to 2, the server performs a check every one minute and sends an alarm if the results of two consecutive checks meet the alarm condition. The minimum interval between alarm notifications is two minutes.
Check condition	
Key name	The key name for alarming in the log content.
Comparison operator	The comparison operator in the check condition, which can be of the numeric or character type.
Check threshold	The comparative value in the check condition, which is combined with the **comparison operator** to determine whether the Saved Search results meet the alarm condition.
Alarm action	
Notification type	The method for sending alarm notifications. When the configured alarm rule is triggered, Log Service sends an alarm based on the predefined notification method. Alarms are only sent through the **WebHook-custom** method. That is, notifications are sent to the custom WebHook link through the Post method.
WebHook address	The URL of WebHook.
Notification content	The content of an alarm notification. The maximum length of the content is 500 characters.

Table 24-1: Comparison operators

Operation	Description	Example
>	Checks whether the column value is greater than the specified value.	$count > 0
<	Checks whether the column value is less than a value.	$count<200
>=	Checks whether the column value is greater than or equal to the specified value.	$count>=0
<=	Checks whether the column value is smaller than or equal to the specified value.	$count<=0
like	A matched substring.	$project like "admin"
regex	A string that matches with the regular expression.	$project regex match "^/S+$"



Result

You can view specific alarm results after creating alarm rules.

1. On the **Logstores** page, choose **Search/Analytics** > **Alarm** from the left-side navigation pane.
2. Click **View** for an alarm rule to view the specific alarm records under the rule.

Alarm status:

- **Success**: indicates that the rule is successfully executed and the standard to trigger the alarm is displayed in **Trigger Details**.
- **Failure**: indicates that the rule failed during the query, alarm rule matching, or notification phase. In this case, you can view **Trigger Details** for more information.
  - Query failed: indicates that the query syntax is incorrect.
  - Query call failed: indicates that the query failed to be called. We recommend that you check your network connectivity.
  - Failed to call the rule: indicates that the rule failed to be called We recommend that you check the format consistency between the rule parameters and returned data.

24.2.8 Log consumption

Log Service supports log consumption in a variety of ways.

Logs collected to the LogHub of Log Service can be consumed in the following three methods:

Method	Scenario	Real-time	Storage duration
Real-time consumption (LogHub)	Stream computing and real-time computing	Real-time (< 10 ms)	365 days
Index query (LogSearch)	Online query of recent hot data	Real-time	365 days
Shipping and storage (LogShipper)	Full log storage for offline analysis	5–30 minutes	Dependent on the storage system

Real-time consumption

Consumption process

Logs are consumed after being written. Both log consumption and log query require the capability of reading logs. The logs in a shard are consumed as follows:

1. Obtain a cursor based on conditions such as time, Begin, and End.
2. Read logs by using the cursor and step, and return the next cursor.
3. Move the cursor continuously to consume logs.

Consumption method

Besides the basic APIs, Log Service provides many methods to consume logs, such as SDKs, Storm spout, Spark Streaming client, Flink connector, consumer library, and Web console.

- Use the Spark Streaming client to consume logs.
- Use Storm Spout to consume logs.
- Use the Flink connector to consume logs. The Flink connector consists of the consumer and producer.
- Use LogHub Consumer Library to consume logs. LogHub Consumer Library is an advanced consumption mode for LogHub consumers. It provides a lightweight computing framework to implement automatic shard allocation and sequence guarantee when multiple LogHub consumers consume Logstore simultaneously.
- Use SDKs to consume logs: Log Service provides SDKs in several languages (Java and Python) with support for log consumption APIs. For more information about the SDKs, see *Log Service SDKs*.

Query analysis

- Query logs in the Log Service console.
- Query logs by using the SDKs or APIs of Log Service. Log Service provides HTTP-enabled RESTful APIs. The APIs support full-featured log querying. For more information, see *Log Service API References*.

24.3 Project

A project is the resource management unit in Log Service and is used to isolate and control resources. You can manage all the logs and the related log sources of an application by using projects. Projects manage the information of all your Logstores and the log collection machine configuration, and serve as the portals where you can access the Log Service resources.

The project is the resource management unit in Log Service and is used to isolate and control resources. Specifically, projects provide the following functions:

- Projects help you organize and manage different Logstores. In actual use, you might use Log Service to centrally collect and store the logs of the different projects, products, or environments. You can classify different logs for management in different projects to facilitate subsequent usage, export, or index of logs. In addition, projects are the carriers of the log access permission management.

- Projects serve as the portals where you can access the Log Service resources. Log Service allocates a unique access point for each created project. The access point supports writing, reading, and managing logs by using the network.

You can perform the following operations through the Log Service console:

- *Create a Project*
- *Modify a project comment*
- *Delete a project*

24.3.1 Project

A project is the resource management unit in Log Service and is used to isolate and control resources. You can manage all the logs and the related log sources of an application by using projects. Projects manage the information of all your Logstores and the log collection machine configuration, and serve as the portals where you can access the Log Service resources.

Projects provide the following functions:

- Projects help you organize and manage different Logstores. In actual use, you may have to use Log Service to collect and store different project, product, or environment logs in a centralize d manner. You can classify different logs for management in different projects to facilitate subsequent usage, export, or index of logs. In addition, projects are the carriers of the log access permission management.
- Projects serve as the portals where you can access the Log Service resources. Log Service allocates a unique access point for each created project. The access point supports writing, reading, and managing logs by using the network.

You can use the Log Service console to perform the following project-related operations:

- *Create a Project*
- *Modify a project comment*
- *Delete a project*

24.3.2 Import projects

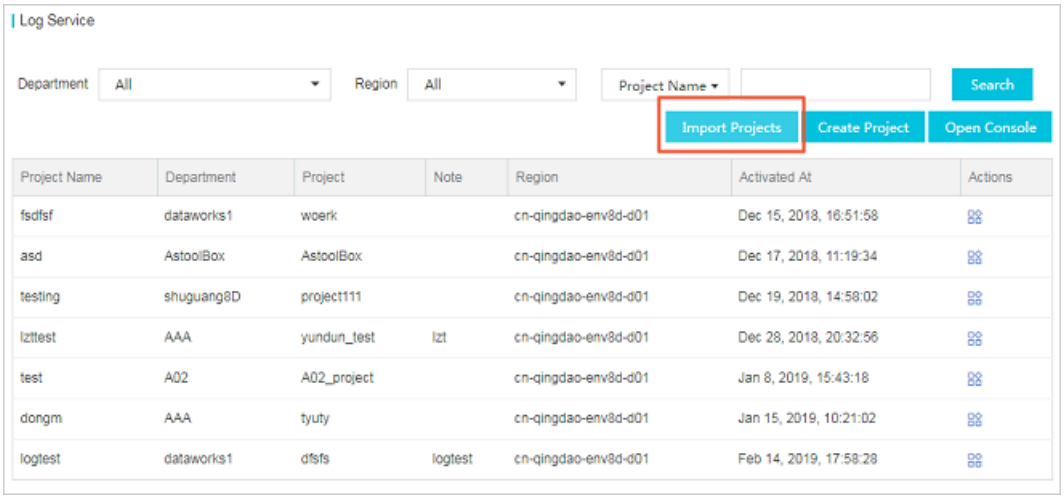
By using the project import function, you can update the project list on the Log Service homepage of Apsara Stack Management Console. After the update, the projects that you have created on the Log Service console are displayed in the project list on the Log Service homepage of Apsara Stack Management Console.

Context

By default, the projects that you have created in the Log Service console are only displayed in the project list of the console. You can use the project import function to import the projects created on the Log Service console to the project list on the Log Service homepage of Apsara Stack Management Console.

Procedure

1. *Log on to the Log Service console*.
2. Click **Import Resources** in the upper-right corner.



3. Select a department and click **OK**.

Result

After the projects are imported, the project list on the Log Service homepage of Apsara Stack Management Console is updated.

24.3.3 Change the project ownership

After creating a project, you can change the department and project to which the project belongs.

Context

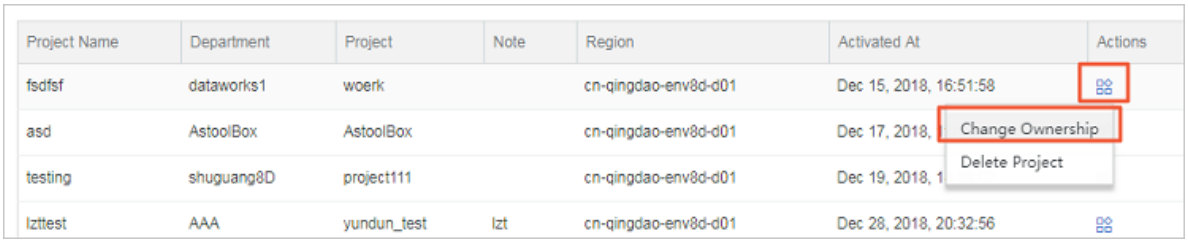
By changing the project ownership, you can change the department and project to which the project belongs, which means migrating the project to another department and project.

Note:

If your project is created in the Log Service console, it is automatically created under a level-1 department and does not belong to any project. In such a case, you must *Import projects* to Apsara Stack Management Console and migrate the project to another project by using the ownership change function.

Procedure

1. *Log on to the Log Service console*.
2. Click the icon in the **Action** column corresponding to the project.
3. Choose **Modify Attributes** from the shortcut menu.
4. Select a new **department** and **project** for the project.



5. Click **OK**.

24.3.4 Modify a project comment

You can add project comments to facilitate project management.

Context

A project is the resource management unit of Log Service, which is used to manage Logstores and machines for log collection. A Logstore is the log storage unit of Log Service, which is used to store a specific type of logs. A project can collect multiple types of logs, such as the access logs of frontend Web servers and the application logs generated by backend applications. You can create separate Logstores for a project and write different types of logs to different Logstores.

Procedure

1. *Log on to the Log Service console*.
2. On the **Projects** page, click **Modify Comment** in the **Action** column cororesponding to the target project.
3. In the displayed dialog box, modify the project comment and click **OK**.

24.3.5 Delete a project

In some cases (for example, to disable a Log Service instance or destroy all the logs in a project), you may want to delete the entire project. Log Service allows you to delete the entire project in the console.

Prerequisites

If the project is created in the Log Service console, the project is not displayed on the Log Service homepage of Apsara Stack Management Console. You must *Import projects* before you can delete a project.

Context

Deleting a project permanently releases all its logs and settings. This action cannot be rolled back. Therefore, use caution when deleting a project to avoid data loss.

Procedure

1. *Log on to the Log Service console*.
2. Click the  icon in the **Action** column corresponding to a specified project name.
3. In the displayed dialog box, click **Delete Project**.
4. A dialog box is displayed, prompting you to confirm whether to delete the project. If yes, click **OK**.

24.4 Logstore

A Logstore is a unit in Log Service to collect, store, and query the log data. Each Logstore belongs to a project, and each project can create multiple Logstores

You can create multiple Logstores for a project according to your actual needs. Typically, an independent Logstore is created for each type of logs in an application. For example, you have a game application **`big-game`**, and three types of logs are on the server: operation_log, application_log, and access_log. You can first create a project named **`big-game`**, and then create three Logstores under this project for these three types of logs to collect, store, and query logs respectively. You must specify the Logstore for writing and querying logs.

Specifically, Logstores provide the following functions:

- Logstores can collect logs and support writing logs in real time.
- Logstores can store logs and support using logs in real time.
- Logstores can create indexes and support querying logs in real time.

You can perform the following operations through the Log Service console:

- *Create a Logstore*
- *Modify Logstore configurations*
- *Delete a Logstore*

24.4.1 Logstore

A Logstore is a unit in Log Service, used for collecting, storing, and querying log data. Each Logstore belongs to a project, and each project can create multiple Logstores.

You can create multiple Logstores for a project as needed. Typically, an independent Logstore is created for each type of logs in an application. For example, you have a game application **`big-game`**, and three types of logs exist on the server: operation_log, application_log, and access_log. In this case, first create a project named **`big-game`** and then create three Logstores under this project to collect, store, and query the three types of logs respectively. Whether writing or querying logs, you must specify a Logstore for the actual operation.

Specifically, Logstores provide the following functions:

- Log collection, with support for real-time logging.
- Log storage, with support for real-time consumption.
- Index creation, with support for real-time log querying.

You can perform the following Logstore operations on the Log Service console:

- *Create a Logstore*
- *Modify Logstore configurations*
- *Delete a Logstore*

24.4.2 Modify Logstore configurations

After a Logstore is created, you can mdofiy the Logstore configurations based on your needs.

Procedure

1. *Log on to the Log Service console*.
2. Click the name of a project.

   You can see Logstores under the project.
3. On the **Logstores** page, select a Logstore and click **Modify** under the actions column.
4. In the displayed dialog box, modify the Logstore configurations and then click **Modify**.

   For information about shard management, see *Split a shard*.

24.4.3 Delete a Logstore

You may need to delete a Logstore in certain cases, for example, to discard the Logstore. Log Service allows you to delete Logstores on the console.

Restrictions

- After a Logstore is deleted, its logs will be lost permanently and cannot be recovered, and a Logstore with the same name cannot be created. Exercise caution when performing this operation.
- Before deleting a Logstore, you must delete all its Logtail configurations.

Procedure

1. *Log on to the Log Service console*.
2. Click the name of a project.
3. On the **LogStores** page, select the LogStore you want to delete and click **Delete** on the right.
4. In the displayed dialog box, click **Confirm**.

24.5 Shard

24.5.1 Shard management

Logstore read/write logs must be saved in shards. Each Logstore is divided into several shards and each shard is composed of MD5 left-closed and right-open intervals. Meanwhile, the range of each interval does not overlap with others and the total range of all the intervals is the entire MD5 value range.

Range

You must specify the number of shards when creating a Logstore. Then, the entire MD5 range is automatically divided evenly based on the specified number of shards. Each shard has a range, which can be expressed in the MD5 mode and must be within the following value range: [00000000000000000000000000000000,ffffffffffffffffffffffffffffffff).

All of the shard ranges are left-closed and right-open intervals, and composed of the following keys:

- BeginKey: indicates the start of a shard. This key is included in the shard range.
- EndKey: indicates the end of the shard. This key is excluded from the shard range.

The shard ranges are used when you write logs by specifying a hash key, split shards, and merge shards. When reading data from a shard, you must specify the corresponding shard. When writing

data, you can use the load balancing mode or the specified hash key mode. By using Server Load Balancer, each data packet is written to an available shard at random. By specifying the hash key, data is written to the shard whose range includes the specified key.

As described in *Table 24-2: Shard example*, a Logstore has four shards and the MD5 value range of this Logstore is [00,FF). The range of each shard is as follows.

Table 24-2: Shard example

Shard No.	Range
Shard0	[00,40)
Shard1	[40,80)
Shard2	[80,C0)
Shard3	[C0,FF)

If you specify the MD5 key as 5F by specifying the hash key when writing logs, the log data is written to Shard1 that contains the MD5 key 5F. If you specify the MD5 key as 8C, the log data is written to Shard2 that contains the MD5 key 8C.

Read/write capacities

Each shard has certain service capacities:

- Write: 5 Mbit/s, 2000 times/s
- Reading: 10 MBit/s, 100 times/s

We recommend that you plan the number of shards based on the actual traffic. If the traffic exceeds the read/write capacities, split the shard in time to increase the number of shards to achieve greater capacities. If the traffic is far less than the maximum read/write capacities of shards, we recommend that you merge the shards to reduce the number of shards to save the rental costs of shards.

For example, assume that you have two shards in read/write status and can write data at 10 Mbit /s at maximum. If you write data at 14 Mbit/s in real time, we recommend that you split a shard to make the number of shards in read/write status reaches three. If you write data at only 3 Mbit/s in real time, we recommend that you merge these two shards because one shard can meet the needs.

Note:

- When the writing API continues to report Error 403 or 500, check whether it is necessary to add shards based on the traffic.
- For read/write operations that exceed the service capacities of shards, the system attempts to provide the needed services, but the service quality cannot be guaranteed.

Statuses of shards

The shard statuses include:

- read/write: Reading and writing data are supported.
- read-only: Only reading data is supported.

When a shard is created, it is in the read/write state. The **split** or **merge** operation changes the shard state to read-only and generates a new shard in the read/write state. The shard state does not affect the performance of reading data. Shards in the read/write state can ensure the normal data writing performance, while shards in the read-only state do not support writing data.

To perform the **split** operation, specify a shard ID in the read/write state and an MD5 value. The MD5 must be greater than the shard BeginKey and smaller than the shard EndKey. Split operations can split two other shards from one, that is, the number of shards is increased by 2 after the split. After the split, the status of the original shard specified to be split is changed from read/write to read-only. Data can still be consumed, while new data cannot be written. The two new shards are in the read/write state and follow the original shard. The MD5 range of the new shards covers that of the original shard.

When **merging** shards, you must specify a shard in read/write status. Make sure the specified shard is not the last (rightmost) shard in read/write status. The server automatically finds the adjacent shard at the right of the specified shard and merges these two shards. After the merge, the specified shard and the following shard enter the read-only state. Data in the two shards can still be consumed, but new data cannot be written to them. A shard in the read/write state is generated, and its MD5 range covers the MD5 ranges of the original two shards.

In the Log Service console, you can perform the following shard-related operations:

- Scale shards in
- Scale shards out

Note:

The life cycle (or data retention period) of a Logstore ranges from 1 to 365 days. Shards and their logs will be automatically deleted when this period expires.

24.5.2 Split a shard

You can set the number of shards according to the quantity of your logs and the log geneatation speed when creating a Logstore, or change the number of shards by merging shards or splitting a shard when modifying a Logstore.

Context

Logstore read/write logs must be saved in a certain shard. Each Logstore has several shards. When creating a Logstore, you must specify its number of shards. After the Logstore is created, you can split or merge shards to increase or reduce shards.

Each shard can write data at 5 MB/s and read data at 10 MB/s. When the data traffic exceeds the service capacity of the shard, we recommend that you increase the number of shards in time by splitting a shard. The expansion partition is completed by split operation.

When splitting a shard, you must specify a ShardId in readwrite status and an MD5. The MD5 must be greater than the shard BeginKey and less than the shard EndKey.

Split operations can split two other shards from one, that is, the number of shards is increased by 2 after the split. After the split, the status of the original shard specified to be split is changed from readwrite to readonly. Data can still be consumed, while new data cannot be written. The two newly generated shards are in readwrite status and arranged behind the original shard. The MD5 range of these two shards covers the range of the original shard.

Procedure

1. *Log on to the Log Service console*.
2. Click the name of a project.
3. On the **Logstores** page, select a Logstore and click **Modify** in the Actions column.
4. Select the shard to be split, and click **Split** on the right.
5. Click **Confirm** and close the dialog box.

   After split, the original shard enters the read-only state, and the MD5 range of the new shards covers that of the original shard.

24.5.3 Merge shards

A shard can be scaled down through the merge operation. The merge operation combines the ranges of a specified shard and its following shard and assigns the combined range to a new shard in the read/write state. The original two shards enter the read-only state.

You must specify a shard (except the last one) in the read/write state. The server finds the shard that follows the specified shard and merges the ranges of the two shards. After merging, the specified shard and the following shard enter the read-only state. Data in the two shards can still be consumed, but new data cannot be written to them. A shard in the read/write state is generated , and its MD5 range covers the MD5 ranges of the original two shards.

Procedure

1. *Log on to the Log Service console*.
2. Click the name of a project or **Manage** on the right of the project.
3. On the **Logstores** page, select the expected Logstore and click **Modify** in the log consumption column.
4. Select shards to be merged and click **Merge** on the right. Then, close the dialog box.

   After the merge operation is completed, the specified shard and the following shard enter the read-only state, and the data can still be consumed. A shard in the read/write state is generated, and its MD5 range covers the MD5 ranges of the original two shards.

24.6 Data collection

LogHub provides multiple log collection methods for your data source, such as by using clients, Web pages, protocols, and SDKs/APIs (mobile devices and games). Select a collection method as you need.

24.6.1 Data collection

LogHub supports lossless log collection through clients, Web pages, protocols, SDKs, and APIs. You can select an appropriate collection mode for your data sources.

24.6.2 Collect Nginx access logs

Log Service supports querying and analyzing real-time logs, and saves the analytical results to Dashboard, which greatly decreases the analytical complexity of Nginx access logs and streamlines the statistics of website access data.

Context

Many webmasters use Nginx as the server to build websites. When analyzing the website traffic data, they must perform a statistical analysis on Nginx access logs to obtain data such as the page views and the access time periods of the website. In the traditional methods such as CNZZ, a js is inserted in the frontend page and will be triggered when a user accesses the website. However, this method can only record access requests. Stream computing and offline statistics & analysis can also be used to analyze Nginx access logs, which however requires to build an environment, and is subject to imbalance between timeliness and analytical flexibility.

Log Service supports querying and analyzing real-time logs, and saves the analytical results to Dashboard, which greatly decreases the analytical complexity of Nginx access logs and streamlines the statistics of website access data. This document introduces the detailed procedure of log analysis function by analyzing the Nginx access logs.

Log format

We recommend that you use the following log_format configuration for better meeting the analytic scenarios:

```
    log_format  main  '$remote_addr - $remote_user [$time_local] "$
request" $http_host '
                '$status $request_length $body_bytes_sent "$http_refer
er" '
                '"$http_user_agent"  $request_time $upstream_r
esponse_time';
```

The meaning of each field is as follows:

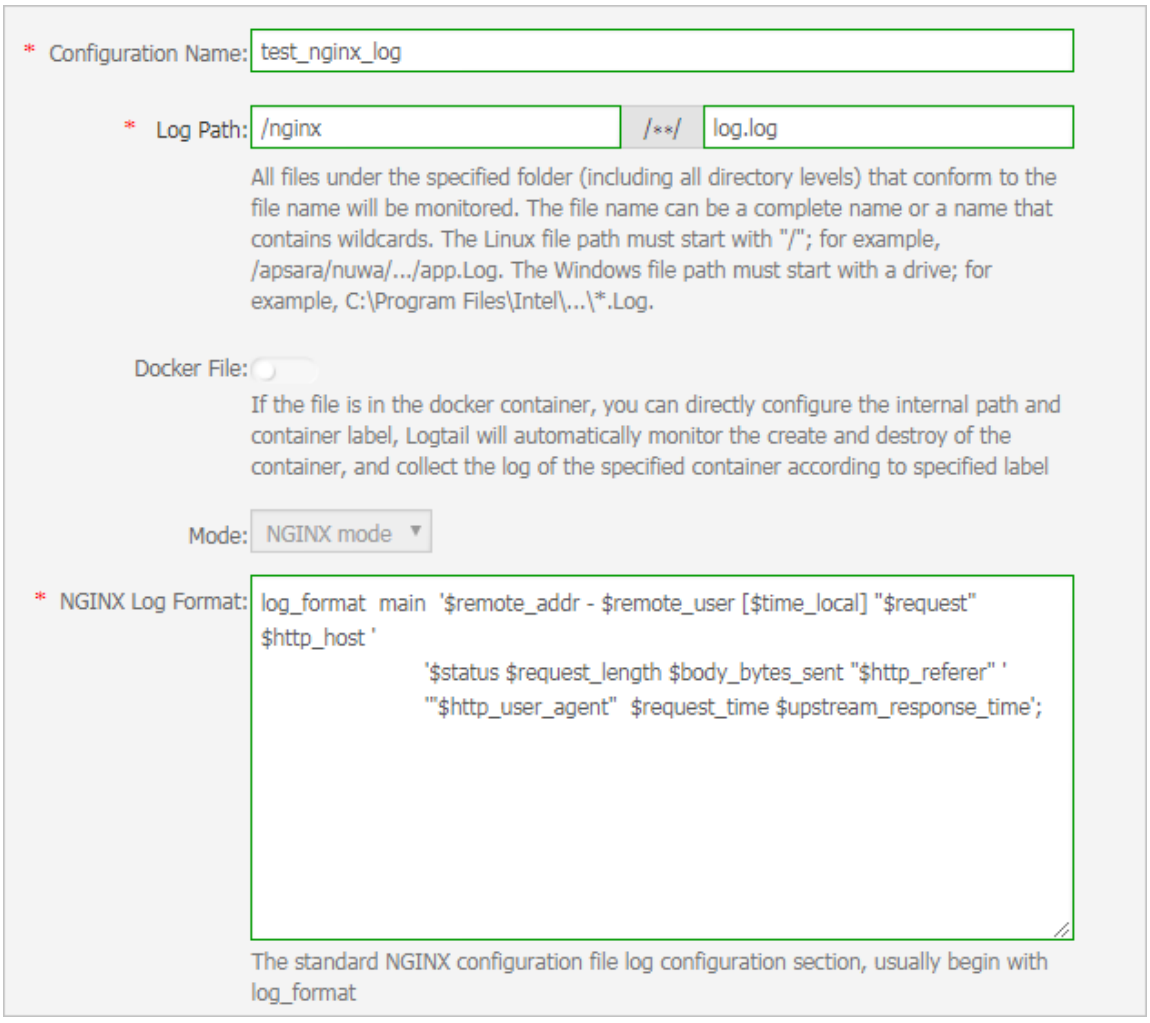
Field	Meaning
remote_addr	Client address
remote_user	The client username.
time_local	The server time.
request	The request content, including method name, address, and HTTP protocol.
http_host	The HTTP address used by the user request.
status	The returned HTTP status code.
request_length	The request size.
body_bytes_sent	The returned size.
http_referer	The referer.
http_user_agent	The client name.

Field	Meaning
request_time	The overall request latency.
upstream_response_time	The processing latency of upstream services.

Procedure

1. *Log on to the Log Service console*.
2. Creat a project.
3. On the **Logstores** page, click the data import wizard icon to access the wizard.
4. Configure the data source.
   a) Select the **NGINX Access Log** data source.
   b) Enter **Configuration name**.
   c) Enter **Log Path**.
   d) Enter **NGINX Log Format**.

   Enter your `log_format` information in the NGINX Log Format field.



e) Confirm **Nginx key**.

Log Service automatically extracts the corresponding keys.

Note:

`$request` is extracted as two keys: `request_method` and `request_uri`.

f) Optional: Configure advanced options.

Parameter	Description
Local Cache	Indicates whether to enable the **Local Cache** function. When Log Service is unavailable, logs can be cached to a local directory and then uploaded after the service is recovered. The default maximum size of logs that can be cached is 1 GB.
Topic Generation Mode	• **Null - Do not generate topic**: The default option, which indicates to set the topic as a null string and you can query logs without entering the topic. • **Machine Group Topic Attributes**: Used to clearly differentiate log data generated in different frontend servers. • **File Path Regular**: When this mode is selected, you must enter a **Custom Regular** below to extract a part of the path

Parameter	Description
	as the topic. This mode is used to distinguish the log data generated by a user or an instance.
Custom RegEx	If you choose to generate a topic in **File Path Regular** mode, enter a custom regular expression here.
Log File Encoding	• utf8: UTF-8 encoding. • gbk: GBK encoding.
Maximum Monitor Directory Depth	Indicates the maximum depth of the monitoring directory when logs are collected from the log source, that is, up to what levels logs are collected. The maximum monitoring directory depth ranges from 0 to 1000, of which 0 means only the directory at the current level is monitored.
Timeout	If a log file is not updated within the specified period of time, the system considers that the file has timed out. You can configure the following settings for **Timeout**. • Never timed out: All log files are continuously monitored and never time out. • 30 minute timeout: If a log file is not updated in 30 minutes, the system considers that the log file has timed out and no longer monitors the file.
Filter Configuration	Only logs that completely meet the filtering conditions are collected. For example,`Key:level Regex:WARNING|ERROR` indicates to only collect logs whose level is WARNING or ERROR. You can also filter logs that do not conform to a condition. For example, `Key:level Regex:^(?!. *(INFO|DEBUG))` indicates to not collect logs whose level is INFO or DEBUG. For similar examples, see *regex-exclude-word* and *regex-exclude-pattern*.

g) After completing the configurations, click **Next**.

5. Select the machine group check box and click **Apply to machine groups**.

   If you have not created a machine group, you must create one first.

6. Optional: Search, analysis, and visualization.

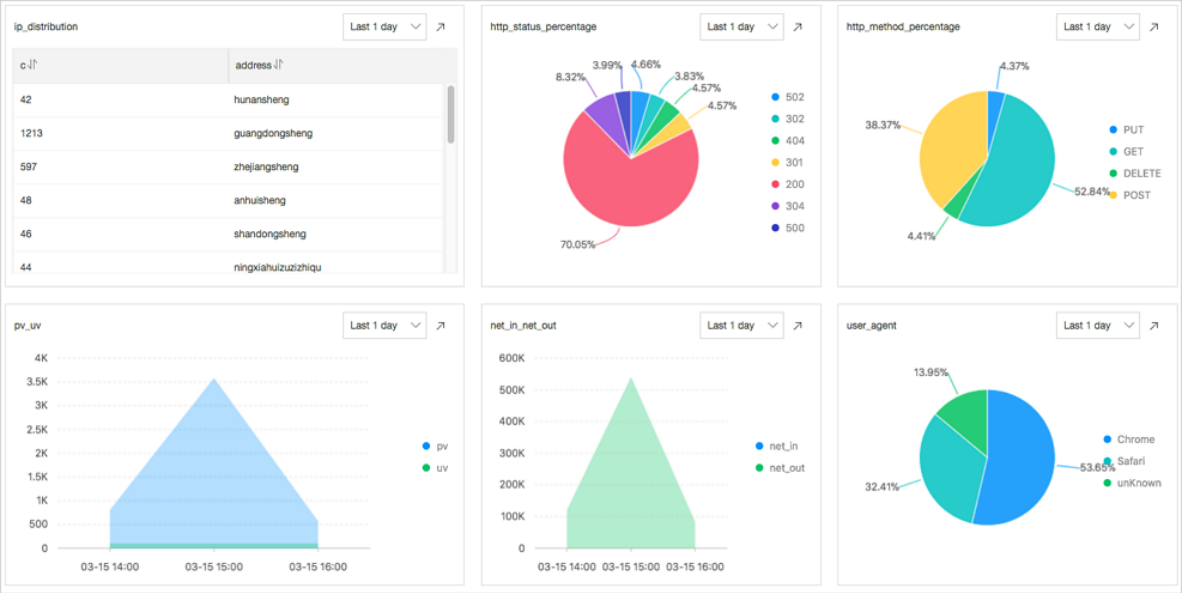
   Make sure the heartbeat statuses of the machine groups that apply the Logtail configuration are normal and you can click **Preview** on the right to obtain the collected data.

Log Service provides predefined keys for analysis and usage. You can select the actual keys (generated according to the previewed data) to map with the default keys.

Click **Next**, Log Service configures the index attributes for you and creates the `nginx-dashboard` dashboard for analysis and usage.

7. Optional: Analyze access logs.

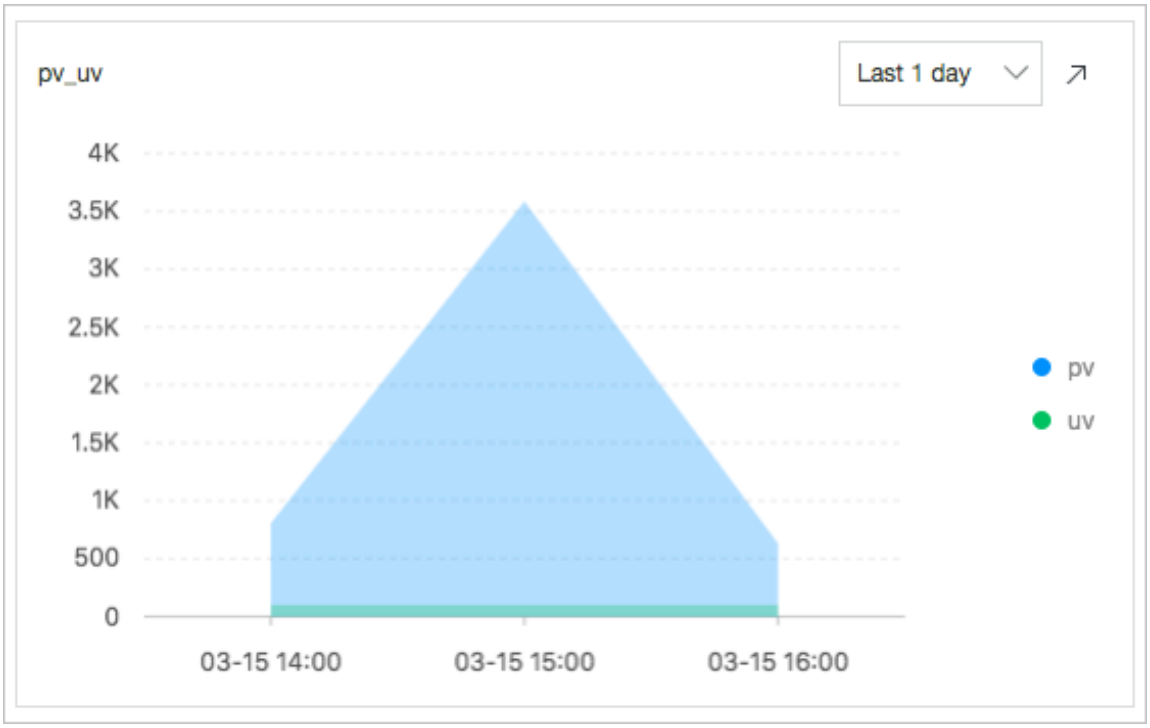
   After the index feature is enabled, you can view the analysis of each indicator on the page where dashboards are generated by default. For how to use dashboards, see *Dashboard*.



- **PV/UV statistics (pv_uv)**

  Count the numbers of PVs and UVs in the last day.

Figure 24-2: PV/UV Statistics



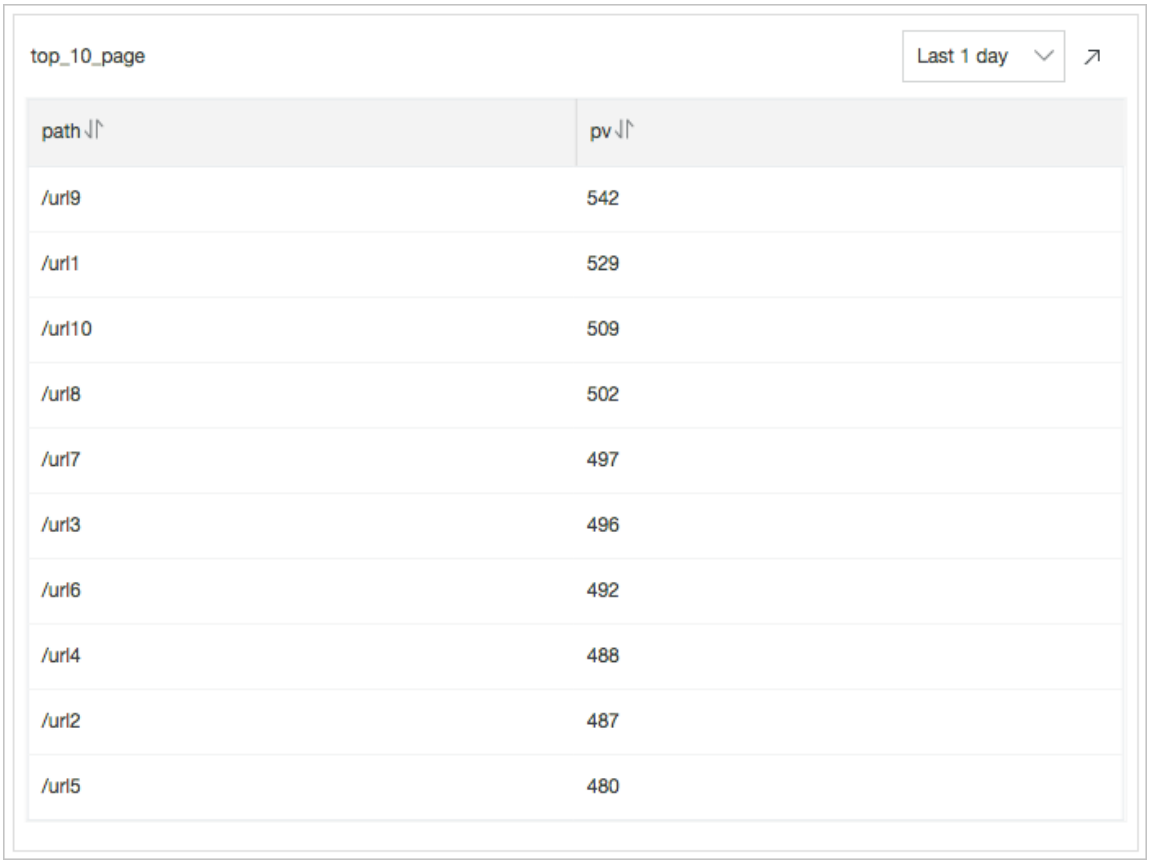
Statistical statement:

```
  * | select approx_distinct(remote_addr) as uv ,
                                count(1) as pv ,
                                date_format(date_trunc('hour',
__time__), '%m-%d %H:%i')  as time
                                group by date_format(date_trunc('
hour', __time__), '%m-%d %H:%i')
                                order by time
                                limit 1000
```

- **Count the top 10 access pages (top_page)**

  Count the top 10 pages with the most PVs in the last day.

Figure 24-3: Access Statistics



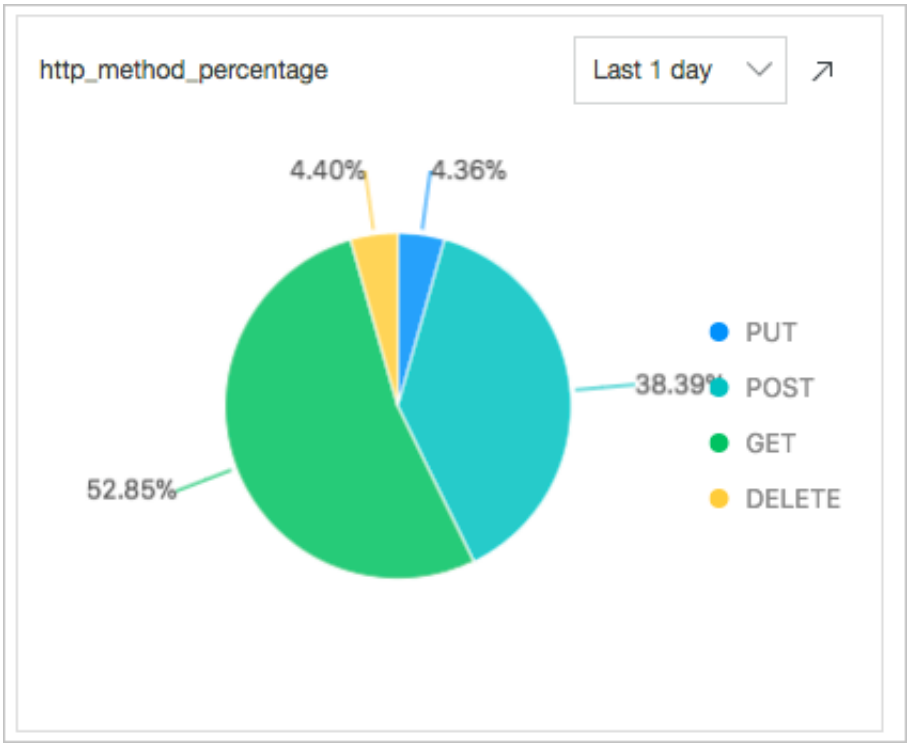
Statistical statement:

```
 * | select split_part(request_uri,'?',1) as path,
                                count(1) as pv
                                group by split_part(request_uri
,'?',1)
                                order by pv desc limit 10
```

- **Count the ratios of request methods (http_method_percentage)**

  Count the ratio of each request method used in the last day

Figure 24-4: Request method share



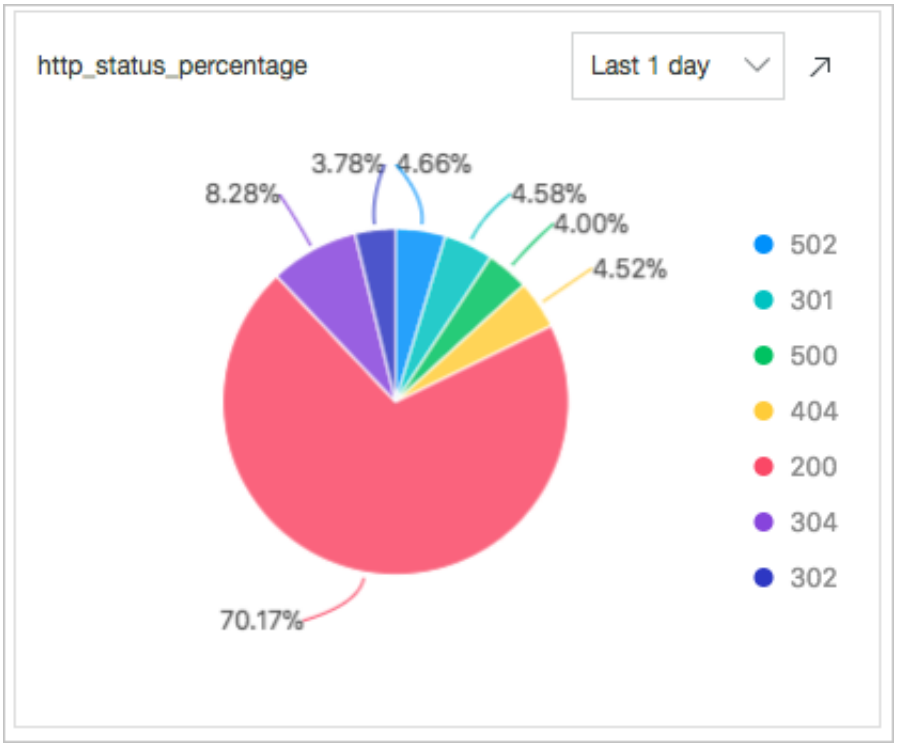
Statistical statement:

```
* | select count(1) as pv,
                                request_method
                                group by request_method
```

- **Count the ratios of request statuses (http_status_percentage)**

  Count the ratio of each request status (HTTP status code) in the last day.

Figure 24-5: Count the ratios of request statuses



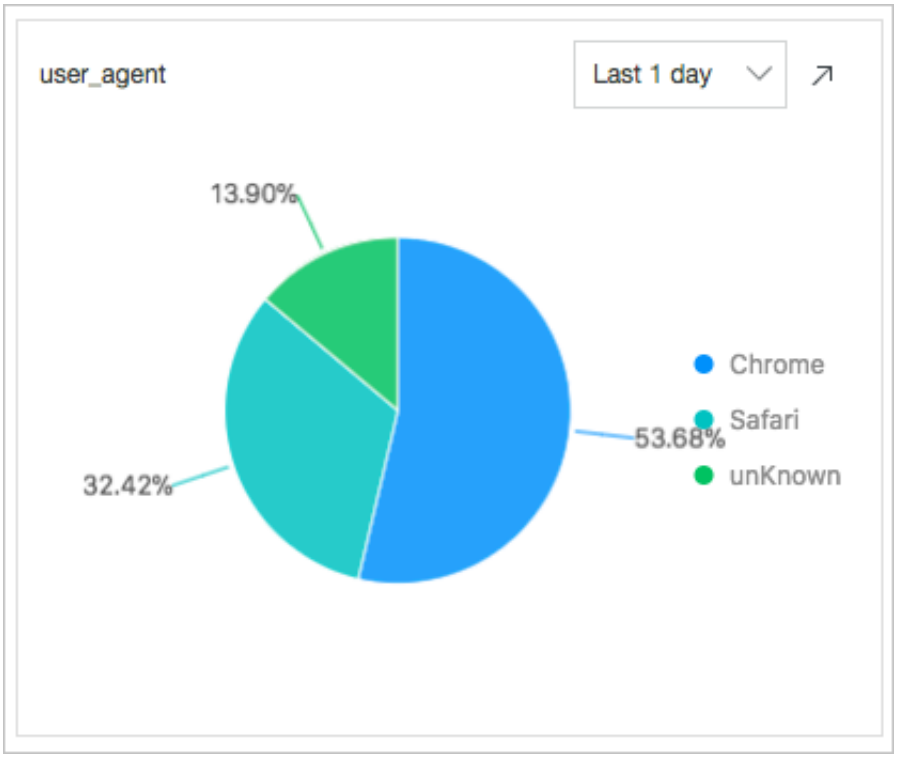
Statistical statement:

```
* | select count(1) as pv,
                                status
                                group by status
```

- **Count the ratios of request UA (user_agent)**

  Count the ratio of each browser used in the last day.

Figure 24-6: Count the ratios of request UA



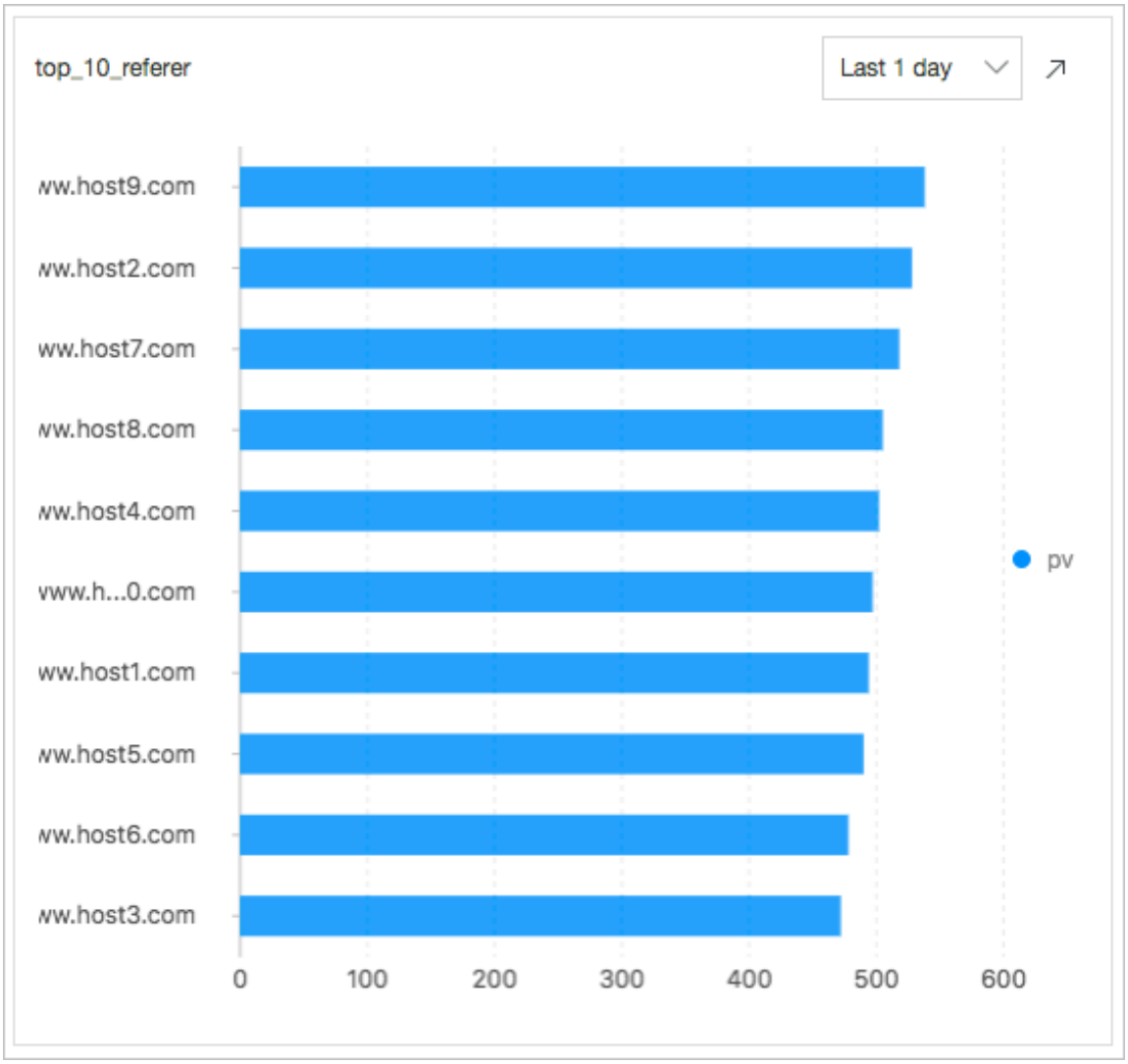
Statistical statement:

```
 * | select count(1) as pv,
                                    case when http_user_agent like '%
Chrome%' then 'Chrome'
                                    when http_user_agent like '%
Firefox%' then 'Firefox'
                                    when http_user_agent like '%Safari
%' then 'Safari'
                                    else 'unKnown' end as http_user_
agent
                                    group by  http_user_agent
                                    order by pv desc
                                    limit 10
```

- **Count the top 10 referers (top_10_referer)**

  Count the top 10 referers in the last day.

Figure 24-7: Count the top 10 referers



Statistical statement:

```
* | select count(1) as pv,
                                    http_referer
                                    group by http_referer
                                    order by pv desc limit 10
```

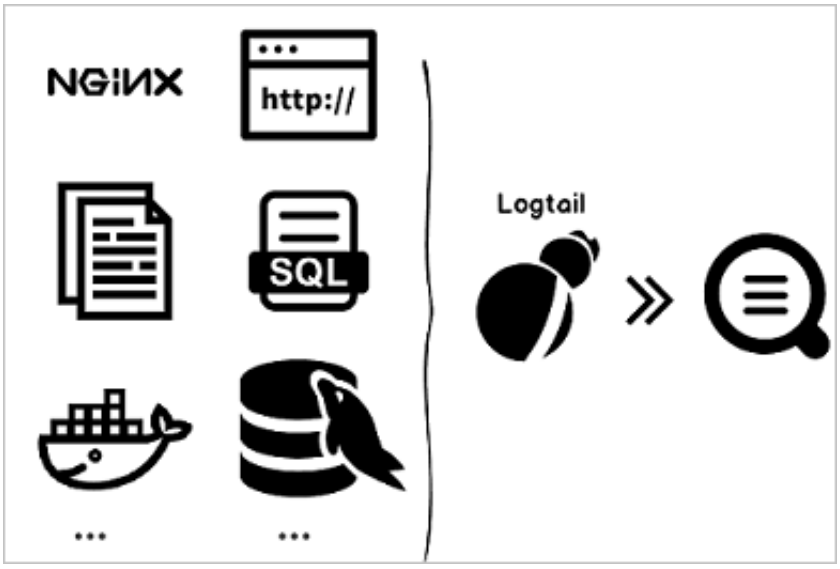
24.7 Collection by Logtail

24.7.1 Overview

24.7.1.1 Logtail overview

Log Service provides a log collection agent, Logtail access service. It allows you to collect logs from ECS instances and other servers on the console in real time. Currently, Log Service only

allows you to install Logtail on a Linux server. To collect logs from a Windows server, you need to use Logstash.



Benefits

- Non-invasive log collection based on log files: You do not need to modify any application code, and log collection does not affect the operating logic of your applications.
- Exception handling in a stable manner during the log collection process: When problems (such as the network or Log Service is abnormal, and the user data temporarily exceeds the reserved bandwidth writing limit) occur, Logtail actively retries and caches data locally to guarantee the data security.
- Central management based on Log Service: After installing Logtail, you only need to configure the devices from which the logs you want to collect and the collection method on the server, rather than logging on to the servers one by one. For Logtail installation, see *Install Logtail (for Linux)*.
- Comprehensive self-protection mechanism: To ensure that the collection agent running on the client machine does not significantly impact the performance of your services, Logtail provides a protective mechanism and strictly limits its use of CPU, memory, and network resources.

Processing capabilities and limits

See *Limits*.

Configuration process

Follow these steps to use Logtail to collect logs from servers:

1. Install Logtail. Install Logtail on the server whose logs you want to collect. For more information, see *Install Logtail (for Linux)*.
2. *Create an IP address-based machine group*. Log Service uses machine groups to manage all servers from which you want to collect logs with Logtail. Log Service allows you to define machine groups using IP addresses or user-defined identifiers. You can create a machine group as instructed when applying Logtail configurations to machine groups.
3. Create Logtail configurations and apply them to a machine group. You can use the **Data Import Wizard** to create Logtail configurations for collecting *text files* and *syslogs*, and to apply the Logtail configurations to a machine group.

After the preceding process is completed, logs of a specific type on the ECS servers are collected and sent to the selected Logstore. Historical logs are not collected. You can use the Log Service console, SDK, or API to query these logs. You can also view the Logtail collection status on each ECS server, for example, whether the collection is normal and whether any error occurs.

For all the Logtail access service operations supported by the Log Service console, see *Collect logs by Logtail*.

Docker

- Container service: See *Integrated Log Service* in **Container Service User Guide**.
- Built-in Docker of ECS and IDC: Mount the log directories from containers to the host server.
  - *Install Logtail (for Linux)*.
  - Mount the log directories from containers to the host server.
    - Method 1: Run commands. For example, if the directory on the host server is `/log/webapp` and the directory in a container is `/opt/webapp/log`, run the following command:

      ```
      docker run -d -P --name web -v /src/webapp:/opt/webapp training
      /webapp python app.py
      ```

    - Method 2: Use the orchestration template.

Note:

We recommend that you modify the Logtail startup parameters, change the checkpoint save path of Logtail, and mount the log directories to the host server. It prevents repeated collection due to checkpoint information loss when containers are released.

Terms

- **Machine group**: A machine group contains one or more machines on which logs of a specific type are collected. By applying a Logtail configuration to a machine group, Log Service collects logs from all the machines in the machine group according to the same Logtail configuration. The Log Service console allows you to manage machine groups conveniently, including operations to create, delete, add, and remove servers.
- **Logtail client**: Logtail is the agent that collects logs and runs on servers from which logs are to be collected. For more information, see *Install Logtail (for Linux)*. After installing Logtail on the server, create a Logtail configuration and apply it to a machine group.
  - In **Linux**, Logtail is installed in the `/usr/local/ilogtail` directory and starts two separate processes (a collection process and a daemon) whose names begin with ilogtail. The program running log is `/usr/local/ilogtail/ilogtail.LOG`.
- **Logtail configuration**: Logtail configuration is a collection of policies to collect logs by using Logtail. By configuring Logtail parameters such as data source and collection mode, you can customize the log collection policy for all the machines in the machine group. Logtail configurations describe how to collect a specific type of logs on servers, parse the collected logs, and send the logs to the specified Logstore of Log Service. You can add a Logtail configuration for each Logstore in the console to enable the Logstore to receive logs collected by using this Logtail configuration.

Features

The Logtail access service provides the following functions:

- **Real-time log collection:** It dynamically monitors log files, reads them in real time, and parses incremental logs. Generally, a latency of less than three seconds exists between the time when a log is generated and the time when a log is sent to Log Service.

  Note:
  The Logtail access service does not support the collection of historical data. Logs with an interval of more than five minutes between the time when a log is read and the time when a log is generated are discarded.
- **Automatic log rotation processing:** Many applications rotate log files based on the file size or generation date. During the rotation process, the original log file is renamed and a new empty log file is created. For example, the monitored `app.LOG` is rotated to generate`app.LOG.1` and `app.LOG.2`. You can specify the file (for example, `app.LOG`) to which collected logs

are written. Logtail automatically detects the log rotation process and ensures that no logs are lost during this process.

> **Note:**
> If log files are rotated multiple times within several seconds, data loss may occur.

- **Automatic handling of collection exceptions:** When data transmission fails because of exceptions such as Log Service errors, network measures, and quota exceeding the limit, Logtail actively retries based on the specific scenario. If the retry fails, Logtail writes the data to the local cache and then automatically re-sends the data later.

  > **Note:**
  > The local cache is located in the disk of your server. If the data cached locally is not received by Log Service within 24 hours, it is discarded and deleted from the local machine.

- **Flexible collection policy configuration**: You can use Logtail configuration to flexibly specify how logs are collected from a server. Specifically, you can select log directories and files, which support exact match or fuzzy match with wildcards, based on actual scenarios. You can customize the extraction method for log collection and the names of extracted fields. Log Service supports extracting logs by using regular expressions. The log data models of Log Service require that each log must have a precise timestamp. Therefore, Logtail provides custom log time formats, allowing you to extract the required timestamp information from log data of different formats.
- **Automatic synchronization of collection configuration:** Generally, after you create or update a configuration in the Log Service console, Logtail automatically accepts and brings the configuration into effect within three minutes. No collected data is lost when configuration is being updated.
- **Automatic client upgrade:** After you manually install Logtail on a server, Log Service automatically manages Logtail upgrades without manual intervention. No log data is lost during the Logtail upgrade process.
- **Status monitoring**: To prevent the Logtail client from consuming too many resources and thus affecting your services, the Logtail client monitors its consumption of CPU and memory in real time. The Logtail client is automatically restarted when its resource usage exceeds the limit to avoid affecting other operations on the machine. The Logtail client actively limits network traffic to avoid excessive bandwidth consumption.

Note:

- Log data may be lost when the Logtail client is being restarted.
- If the Logtail client exits because of an exception in its processing logic, the corresponding protection mechanism is triggered and the Logtail client is restarted to continue to collect logs. However, log data generated before the restart may be lost.

- **Transferred data signature:** To prevent data tampering during data transfer, the Logtail client proactively retrieves your AccessKey to sign all log data packets before sending them.

Note:
The Logtail client uses an HTTPS channel to obtain your AccessKey to ensure its security.

24.7.1.2 How Logtail collection works

The Logtail client collects server logs in the following steps: listen for files, read files, process logs, filter logs, aggregate logs, and send data.

After you install the Logtail client on your server and configure a Logtail Config, Logtail starts collecting logs to Log Service. Logtail collects logs by going through the following process.

1. Listen for files
2. Read files
3. Process logs
4. Filter logs
5. Aggregate logs
6. Send logs

Note:
After the Logtail collection configuration is applied to the machine group, the log files on the servers in the machine group without modification events are considered historical files. Logtail does not collect historical files in normal running mode. To collect historical logs, see *Import historical log files*.

Listen for files

After the Logtail client is installed on the server and the Logtail collection configuration is added based on the data source, the Logtail collection configuration is delivered from the server to Logtail in real time. Logtail starts listening for files based on the collection configuration.

1. Logtail successively scans log directories and files that comply with the specified file name rules based on the configured log path and maximum monitoring directory depth.

   To ensure log collection timeliness and stability, Logtail listens for the registration events of collected directories and performs periodic polling.
2. If Logtail finds that the rule-compliant log files in the specified directory are not modified after the configuration is applied, Logtail does not collect these files. If modification events are recorded for the log files, Logtail triggers the collection process and reads these files.

Read files

After determining that a log file has been updated, Logtail reads the file.

1. Logtail checks the size of a file read for the first time.
   - If the file size is less than 1 MB, Logtail reads the file from the start of the file content.
   - If the file size is larger than 1 MB, Logtail reads the last 1-MB content of the file.
2. If Logtail has read the file before, Logtail reads the file from the last checkpoint.
3. Logtail can read up to data of 512 KB at a time. Therefore, you need to limit the log size to 512 KB.

Process logs

After reading a log, Logtail divides the log into lines, parses the log, and confirms the time field of the log.

1. **Divide into lines**:

   If the Logtail collection configuration specifies a **regular expression at the beginning of the line**, the log data read by Logtail at one time is divided into lines (logs) based on the configured beginning of the line. If the beginning of the line is not set, each data block is processed as a log.
2. **Parse the log**:

   Logtail parses each log based on the collection configuration, such as those for regular expressions, delimiters, and JSON.

   Note:
   An excessively complex regular expression may lead to an abnormally high CPU usage. Therefore, we recommend that you use an efficient regular expression.
3. **Handle parsing failures**:

Logtail determines the method for handling failed parsing based on whether the *discard logs that fail to be parsed* function is enabled in the collection configuration.

- If the **function** is enabled, Logtail discards the log and reports an error, indicating that parsing has failed.
- If the **function** is disabled, Logtail uploads the original log that failed to be parsed with the Key set to **raw_log** and the Value set to the log content.

4. **Set the time field of a log**:
   - The log time is the current parsing time if the time field is not set.
   - If the time field is set:
     - The log time is extracted from the parsed log fields when the difference between the log time and the current time is less than 12 hours.
     - The log is discarded and an error is reported when the difference between the log time and the current time is greater than 12 hours.

Filter logs

After processing logs, Logtail filters them based on the *filtering settings* in the collection configuration.

- If the **filtering settings** are not configured, Logtail skips to the next step without filtering any logs.
- If the **filtering settings** are configured, Logtail traverses and verifies all the fields of each log.
  - Logs that comply with the filtering settings: Logtail collects the logs that contain all the fields configured by the filter and comply with the settings.
  - Logs that do not comply with the filtering settings: Logtail does not collect the logs that do not comply with the filtering settings.

Aggregate logs

After log filtering settings are configured, Logtail sends the logs that comply with these settings to Log Service. To reduce the number of network requests, Logtail caches the processed and filtered logs for a period and then aggregates and packages them before sending them to Log Service.

When any of the following conditions is met during caching, logs are immediately packaged and sent to Log Service.

- Log aggregation lasts more than 3s.

- The number of aggregated logs exceeds 4,096.
- The total size of aggregated logs exceeds 512 KB.

Send logs

Logtail aggregates and sends the collected logs to Log Service. You can set the `max_bytes_per_sec` and `send_request_concurrency` parameters in *startup parameter configuration* to adjust the log data sending rate and the maximum concurrency. In this case, Logtail keeps the sending rate and concurrency below the configured thresholds.

If data sending fails, Logtail retries or stops sending based on the error message.

Error message	Description	Handling method
Error 401	The error message returned when the Logtail client is not authorized to collect data.	Logtail discards the log packets.
Error 404	The error message returned when the specified project or Logstore does not exist in the Logtail collection configuration.	Logtail discards the log packets.
Error 403	The error message returned when the Shard quota exceeds the upper limit.	Logtail waits for three seconds and retries.
Error 500	The error message returned when a server exception occurred.	Logtail waits for three seconds and retries.
Network timeout	The error message returned when a network connection error occurred.	Logtail waits for three seconds and retries.

24.7.2 Installation

Before using Logtail of Log Service to collect server logs, install a Logtail agent on the server and set startup parameters as needed.

24.7.2.1 Install Logtail (for Linux)

Installing Logtail is the required step for collecting logs through Logtail. Currently, Logtail can be installed on only Linux servers.

Applicable systems:

Linux x86-64 (64-bit) servers in the following versions:

- Aliyun Linux

- Ubuntu
- Debian
- CentOS
- OpenSUSE

Procedure

1. Download the Logtail installation script.

   Run the following command to download Logtail:

   ```
   logtail.your Log Service endpoint/logtail.sh
   ```

2. Execute the installation script.

   Start the shell terminal and run the following command as an administrator to install Logtail:

   ```
   sh logtail.sh
   ```

   Note:
   Logtail installation uses the overwrite mode. If you have installed Logtail before, the installer uninstalls and deletes the `/usr/local/ilogtail` directory, and then reinstalls it.

What's next

View the Logtail version.

The following information shows that the running Logtail version is 0.9.4:

```
$ls /usr/local/ilogtail/ilogtail -lh lrwxrwxrwx 1 root root 34 Nov 3
12:00 /usr/local/ilogtail/ilogtail -> /usr/local/ilogtail/ilogtail_0.9
.4
```

Uninstall Logtail.

Download**logtail.sh** by referring to *Install Logtail*. Run the following command as an administrator in shell mode:

```
wget http://{sls data endpoint}/logtail.sh chmod 755 logtail.sh sh
logtail.sh uninstall
```

24.7.2.2 Configure startup parameters

This topic describes how to configure Logtail startup parameters. You can use this topic as parameter setting reference.

Context

The configuration of Logtail startup parameters is applicable to the following scenarios:

- If many log files are collected, excessive memory space is occupied. The metadata of each file must be maintained in memory, including the file signature, collection location, and file name.
- CPU utilization is high due to heavy log data traffic.
- A high volume of log data leads to heavy traffic sent to Log Service.
- Syslogs and TCP data streams need to be collected.

Startup configuration

- File path:

```
/usr/local/ilogtail/ilogtail_config.json
```

- File format:

  JSON

- File sample (only partial configuration items are shown)

```
{ ... "cpu_usage_limit" : 0.4, "mem_usage_limit" : 100, "max_bytes_
per_sec" : 2097152, "process_thread_count" : 1, "send_reque
st_concurrency" : 4, "streamlog_open" : false, "streamlog_pool_size_
in_mb" : 50, "streamlog_rcv_size_each_call" : 1024, "streamlog_
formats":[], "streamlog_tcp_port" : 11111, "buffer_file_num" : 25, "
buffer_file_size" : 20971520, "buffer_file_path" : "", ... }
```

Common configuration parameters

Parameter name	Parameter value	Parameter description
cpu_usage_limit	CPU usage threshold, double type, calculated per core.	For example, the value 0.4 indicates that the CPU utilization of Logtail is limited to 40% of single-core capacity. Logtail restarts automatically when the threshold is exceeded. In many cases, the single-core CPU processing capability is about 24 MB/s in easy mode and about 12 MB/s in full mode.
mem_usage_limit	In-memory usage threshold, int type, measured in MB.	For example, the value 100 indicates that the memory usage of Logtail is restricted to 100 MB. Logtail restarts automatically when the threshold is exceeded. If you need to collect more than 1000 distinct files, increase the threshold value properly.

Parameter name	Parameter value	Parameter description
max_bytes_per_sec	Traffic limit on the raw data sent by Logtail, int type , measured in bytes per second.	For example, the value 2097152 indicates that the data transfer rate of Logtail is restricted to 2 MB/s.
process_thread_count	Number of threads Logtail uses to write data to log files.	The default value is 1, which supports a write speed of 24 MB/s in easy mode and 12 MB/s in complete regex mode. Adjust the threshold only when necessary.
send_request_concurrency	By default, Logtail sends data packets asynchronously. You can set a larger asynchronous concurrency value if the write TPS is large.	By default, four asynchronous concurrencies are available. You can calculate the proper concurrency quantity based on the condition that one concurrency supports 0.5 MB/s to 1 MB/s network throughout. The actual concurrency quantity varies with the network delay.
streamlog_open	Syslog reception switch, bool type.	False indicates that syslog reception is disabled and true indicates that syslog reception is enabled.
streamlog_pool_size_in_mb	Size of syslog memory pool used to receive logs. The memory is used to cache syslog data. The unit is MB.	Logtail requests memory when it starts. Set the pool size based on the machine memory size and your needs.
streamlog_rcv_size_each_call	The cache size used each time Logtail calls the Linux socket rcv interface. The value ranges from 1024 to 8192, in bytes.	You can increase the value in the case of heavy syslog traffic.
streamlog_formats	Method of parsing received syslogs.	--
streamlog_tcp_addr	The binding address Logtail uses to receive syslogs. The default value is 0.0.0.0.	For details, see *Collect syslogs through Logtail*.
streamlog_tcp_port	The TCP port through which Logtail receives syslogs.	The default value is 11111.
buffer_file_num	When a network exception occurs or the write quota is exceeded, Logtail writes	The default version is 25.

Parameter name	Parameter value	Parameter description
	the logs that are parsed in real time to a local file (in the installation directory) and then tries to resend the logs to Log Service after recovery. This parameter indicates the maximum number of cached files.	
buffer_file_size	Maximum number of bytes of each buffered file. **`buffer_file_num * buffer_file_size`** indicates the maximum disk space available for cached files.	The default value is 20,971,520 bytes (20 MB).
buffer_file_path	Directory that stores cached files. After you modify this parameter, manually move the files named in the format of *`logtail\_buffer\_file_*`* in the old cache directory to the new directory so that Logtail can read the cached files and delete them after sending.	The default value is null, indicating that cached files are stored in the Logtail installation directory */usr/local/ilogtail*.
bind_interface	Name of the NIC bound to the local machine, for example, eth1. Only Linux version is supported.	By default, the available NICs are bound automatically. If this parameter is configured, Logtail uses only the specified NIC to upload logs.
check_point_filename	Full path of the checkpoint files. The parameter is used to customize the storage path of the checkpoint files of Logtail.	By default, the files are stored in */tmp/logtail_check_point*. We recommend that Docker users modify the checkpoint file storage path, and mount the file path to the host server to prevent repeated collection due to checkpoint information loss when containers are released. For example, set check_point_filename in Docker to `/data/logtail/check_point.dat` and add `-v /data/docker1/`

Parameter name	Parameter value	Parameter description
		`logtail:/data/logtail` to Docker startup commands. In addition, mount the *`/data/docker1/logtail`* directory on the host server to the *`/data/logtail`* directory on the Docker.

Note:

- The preceding table only lists the common startup parameters. If *`ilogtail_config.json`* has parameters not listed above, the default values are used.
- Add or modify the values of configuration parameters as required. You do not need to add unused configuration parameters to *`ilogtail_config.json`*.

Modify configuration

1. Configure *`ilogtail_config.json`* as required.

   Check that the modified configurations are JSON compatible.

2. Restart Logtail to apply the configurations.

```
/etc/init.d/ilogtaild stop /etc/init.d/ilogtaild start /etc/init.d/
ilogtaild status
```

24.7.3 Data sources

Log Service collects logs from multiple data sources, such as text logs and syslogs.

24.7.3.1 Text logs

24.7.3.1.1 Collect text logs

The Logtail client can help you easily collect logs from ECS instances through the console.

Context

After a Logstore is created, the system prompts you to go to the data access wizard. In the displayed dialog box, click **OK** to create a Logtail configuration. Alternatively, you can go to the **Logstores** page and click **Data Access Wizard** to create a Logtail configuration.

Prerequisites

You must install Logtail before you can use it to collect logs. Apsara Stack Log Service allows you to install Logtail in Linux operating systems. For the installation method, see *Install Logtail (for Linux)*.

Restrictions

- A file can only be collected by using one configuration. Use a soft link to collect multiple copies of a log file. For example, to collect files under */home/log/nginx/log* by using two configurations, use the original path for one configuration, run the `ln -s /home/log/nginx/log /home/log/nginx/link_log` command to create a soft link of this folder, and then use the soft link path for the other configuration.
- For more information about the operating systems supported by the Logtail client, see *Install Logtail (for Linux)*.

Logtail collection configuration procedure

In the Log Service console, you can configure the Logtail to collect text logs in modes such as simple mode, delimiter mode, JSON mode, and full mode. The following describes how to configure Logtail in simple mode and complete regex mode:

Procedure

1. *Log on to the Log Service console*.
2. Create a project and a Logstore. For the detailed procedures, see *Create a Project* and *Create a Logstore*.
3. In the log service console, click the project to go to the **Logstores** page.
4. Select the Logstore and click the **Data Import Wizard** icon next to it to start the data import configuration.
5. Select a data type.

   Select **Text** under **Other Sources** and then click **Next** to go to the **Configure Data Source** step.
6. Configure the data source.

   a) Specify the configuration name.

   The configuration name can only contain lowercase letters, numbers, hyphens (-), and underscores (_). It must start and end with a lowercase letter or number and must be 3 to 63 bytes in length.

   Note:

   The configuration name cannot be modified after the configuration is created.

   b) Specify the log directory and the file name.

   The directory structure supports both the complete path mode and the wildcard mode.

Note:

- The directory wildcard must be the asterisk (*) or question mark (?).
- A file can only be collected by using one configuration.

Both the complete file name and the wildcard can be used as the log file name. For file naming rules, see *Wildcard matching*.

Multi-level directory matching is set as the log search mode. This indicates that all files with compliant file names under the folder can be monitored (including all sub-directories).

- Example: `/apsara/nuwa/… /*.log` means the files whose suffix is `.log` exist in the `/apsara/nuwa` (including its recursive subdirectories).
- Example: `/var/logs/app_* … /*.log*` means the files whose file name contains `.log` exist in all of the directories that conform to the `app_*` mode (including their recursive sub-directories) under the `/var/logs` directory.

c) Specify the log collection mode.

Currently, Log Service allows you to parse logs in **NGINX Configuration**, **Simple Mode**, **Delimiter Mode**, **JSON Mode**, and **Regex Mode**. In this example, the simple mode and Regex mode are used to introduce the collection mode settings.

- Simple mode

  Currently, the simple mode refers to the single-line mode. By default, one line of data is a log, that is, two logs are separated by a line break in a log file. The system does not extract log fields (the regular expression is (. *) by default), and uses the system time of the current server as the log generation time. To configure more detailed settings, you can change the configuration to the Regex mode and modify the settings.

  In simple mode, you only need to specify the file directory and file name. Logtail collects logs line by line. It does not extract fields from the log content. In addition, the log time is set to the system time of the server when the log is crawled.

- Regex Mode

  To configure more personalized field extraction settings for log contents (such as cross-line logs and field extraction), select the **Regex Mode**.

  A. Enter **Log Sample**.

The log sample is provided so that the Log Service console can automatically extract the regular expression matching mode in the sample. Make sure to use logs in the real scenario.

B. Disable **Singleline**.

By default, the single-line mode is used, that is, two logs are separated by a line break. To collect cross-line logs (such as Java program logs), you must disable **Singleline** and then configure **Regular Expression**.

C. Configure **Regular Expression**.

This option provides two functions: automatic generation and manual input. After entering a log sample, click **Generate Automatically** to automatically generate a regular expression. If the generation fails, you can switch to the manual mode and enter the regular expression for verification.

D. Configure **Extract Field**.

To analyze and process fields separately in the log content, use the **Extract Field** function to convert the specified field to a key-value pair before sending it to Log Service. Therefore, you must specify a method for parsing the log content, which is a regular expression.

The Log Service console allows you to specify a regular expression for parsing the log content in two ways. The first way is to automatically generate a regular expression through simple interactions. You can select fields from the log sample. Then, the Log Service console automatically generates a regular expression.

In this way, you can generate the regular expression without writing it on your own. You can also manually enter a regular expression. Click **Manually Input Regular Expression** to switch to the manual input mode. After entering the regular expression, click **Validate** to validate whether the entered regular expression can parse and extract the log sample.

No matter the regular expression for parsing the log content is automatically generated or manually entered, you must name each extracted field, that is, set keys.

d) Set **Use System Time**.

Use System Time is enabled by default. If it is disabled, you must specify a certain field (value) as the time field during field extraction and name this field `time`. After setting the `time` field, you can click **Generate Automatically** in **Time Conversion Format** to generate

the method for parsing the time field. For more information about log time formats, see *Configure a time format*.

e) Optional: Set **Advanced Options**.

Set **Local Cache**, **Upload Original Log**, **Topic Generation Mode**, **Log File Encoding**, **Maximum Monitor Directory Depth**, **Timeout**, and **Filter Configuration** based on your requirements. Keep the default configurations unless otherwise required.

Parameter	Description
Local Cache	Indicates whether to enable the **Local Cache** function. When Log Service is unavailable, logs can be cached to a local directory and then uploaded after the service is recovered. The default maximum size of logs that can be cached is 1 GB.
Topic Generation Mode	• **Null - Do not generate topic**: The default option, which indicates to set the topic as a null string and you can query logs without entering the topic. • **Machine Group Topic Attributes**: Used to clearly differentiate log data generated in different frontend servers. • **File Path Regular**: When this mode is selected, you must enter a **Custom Regular** below to extract a part of the path as the topic. This mode is used to distinguish the log data generated by a user or an instance.
Custom RegEx	If you choose to generate a topic in **File Path Regular** mode, enter a custom regular expression here.
Log File Encoding	• utf8: UTF-8 encoding. • gbk: GBK encoding.
Maximum Monitor Directory Depth	Indicates the maximum depth of the monitoring directory when logs are collected from the log source, that is, up to what levels logs are collected. The maximum monitoring directory depth ranges from 0 to 1000, of which 0 means only the directory at the current level is monitored.
Timeout	If a log file is not updated within the specified period of time, the system considers that the file has timed out. You can configure the following settings for **Timeout**. • Never timed out: All log files are continuously monitored and never time out. • 30 minute timeout: If a log file is not updated in 30 minutes, the system considers that the log file has timed out and no longer monitors the file.

Parameter	Description
Filter Configuration	Only logs that completely meet the filtering conditions are collected. For example,`Key:level Regex:WARNING|ERROR` indicates to only collect logs whose level is WARNING or ERROR. You can also filter logs that do not conform to a condition. For example, `Key:level Regex:^(?!. *(INFO|DEBUG))` indicates to not collect logs whose level is INFO or DEBUG. For similar examples, see *regex-exclude-word* and *regex-exclude-pattern*.

f) After completing the settings, click **Next**.

7. Select the target machine group and click **Apply to Machine Group** to apply the configuration to the specified machine group.

   If you have not created a machine group, create one first. For more information, see *Create an IP address-based machine group*.

   Note:

   - It takes up to three minutes for the Logtail configuration to take effect after being pushed.
   - After creating Logtail configurations, you can view the Logtail configuration list, modify the Logtail configurations, or delete Logtail configurations.

24.7.3.1.2 Configure a time format

Each log in Log Service must contain a timestamp. When collecting logs from users log files, the Logtail access service must extract the timestamp string in a log record and parse it as a timestamp. Therefore, you need to specify a timestamp format for parsing.

Logtail in Linux supports all time formats provided by the strftime function. Logtail only parses and uses the timestamp strings that can be expressed in the log formats defined by the strftime function.

The timestamp strings of logs have diverse formats. To make configuration easier, the following table lists the common log time formats supported by Logtail:

Supported format	Description	Example
%a	Abbreviation of week.	Fri
%A	Full name of week.	Friday
%b	Abbreviation of month.	Jan
%B	Full name of month.	January

Supported format	Description	Example
%d	The Nth day of a month, ranging from 01 to 31 in decimal format.	07, 31
%h	Abbreviation of month, which is the same as `%b`.	Jan
%H	Hour, in 24-hour format.	22
%I	Hour, in 12-hour format.	11
%m	Month, in decimal format.	08
%M	Minute, ranging from 00 to 59 in decimal format.	59
%n	Linefeed.	Linefeed
%p	Local time in am or pm format.	AM/PM
%r	Time combination in 12-hour format, which is the same as `%I:%M:%S %p`.	11:59:59 AM
%R	Combination of hours and minutes, which is the same as `%H:%M`.	23:59
%S	Seconds, ranging from 00 to 59 in decimal format.	59
%t	Tab character.	Tab character
%y	Year (excluding century), ranging from 00 to 99 in decimal format.	04; 98
%Y	Year, in decimal format.	2004; 1998
%z	Time zone or abbreviation.	-07:00, +0800
%C	Century, ranging from 00 to 99 in decimal format.	16
%e	The Nth day of a month, ranging from 1 to 31 in decimal format. Prefix a blank to a single-digit number.	7, 31

Supported format	Description	Example
%j	The Nth day of a year, ranging from 00 to 366 in decimal format.	365
%u	Week in decimal format. It ranges from 1 to 7, and the value 1 indicates Monday.	2
%U	The Nth week of a year. The first day of a week is Sunday. It ranges from 00 to 53.	23
%V	The Nth week of a year. The first day of a week is Monday . If the first week of a month contains four or more days , this is considered the first week. Otherwise, the next week is considered the first week. It ranges from 01 to 53.	24
%w	Week in decimal format. It ranges from 0 to 6, and the value 0 indicates Sunday.	5
%W	The Nth week of a year. The first day of a week is Monday. It ranges from 00 to 53.	23
%c	Standard date and time.	If you want to specify a long or short date, use the above formats for more accurate expression.
%x	Standard date.	If you want to specify a long or short date, use the above formats for more accurate expression.
%X	Standard time.	If you want to specify a long or short date, use the above formats for more accurate expression.
%s	Unit timestamp.	1476187251

24.7.3.1.3 Generate a topic

A topic is a custom field used to mark a batch of logs. Logs in one Logstore can be grouped by log topics. You can specify a topic when writing a log or querying logs.

Log is the minimum data unit processed in Log Service. It is defined in semi-structured data mode. The specific data model consists of topic, time, content, and source. For details, see *Log Service Overview*.

A topic is a custom field used to mark a batch of logs. Logs in one Logstore can be grouped by log topics. You can specify a topic when writing a log or querying logs. For example, access logs are marked by sites, and platform users can use user IDs as the log topics and write them into logs. In this way, users can view only their own logs based on log topics. If there is no need to group logs in a Logstore, one log topic can be used for all logs. The default value of this field is a null string, which is also a valid topic.

Note:

You cannot set a topic for syslogs.

You can set or change the topic in the Log Service console.

Topic generation mode

You can set topics when collecting logs by using Logtail or when uploading data by using APIs or SDKs. At present, the following topic generation modes are supported in the Log Service console: **Null - no topic**, **Machine Group Topic Attribute**, and **File Path Regular**.

- **Null - no topic**

  When you configure Logtail to collect text files in the Log Service console, the default log topic generation mode is **Null - no topic**. That is, the topic is a null string, and logs can be directly queried without a topic.

- **Machine Group Topic Attribute**

  The **Machine Group Topic Attribute** mode is used to differentiate log data generated by different servers. If log data of different servers is stored in the same file path and the same file, you can divide machines into different machine groups when you want to differentiate the log data of different servers by topic. That is, set **Group Topic** differently for different machine groups when creating machine groups and set **Topic Generation Mode** to **Machine Group Topic Attribute**. Apply the previously created Logtail configuration to the machine groups to complete the configuration.

If **Machine Group Topic Attribute** is selected, Logtail uploads the topic attribute of the machine group to which the current machine belongs as the topic name to Log Service, when reporting data. When you perform a query by using the **LogSearch/Analytics** function, you need to specify a topic (namely the topic attribute of the target machine group) as the query condition.

- **File Path Regular**

  The **File Path Regular** mode is used to differentiate the log data generated by a user or an instance. If service logs are stored in different directories by user or instance, Log Service cannot distinguish which user or instance generates the logs when collecting log files so long as subdirectories are different and log file names are the same. In this case, you can set **Topic Generation Mode** to **File Path Regular**, enter the regular expression of the file path, and set the topic to the instance name.

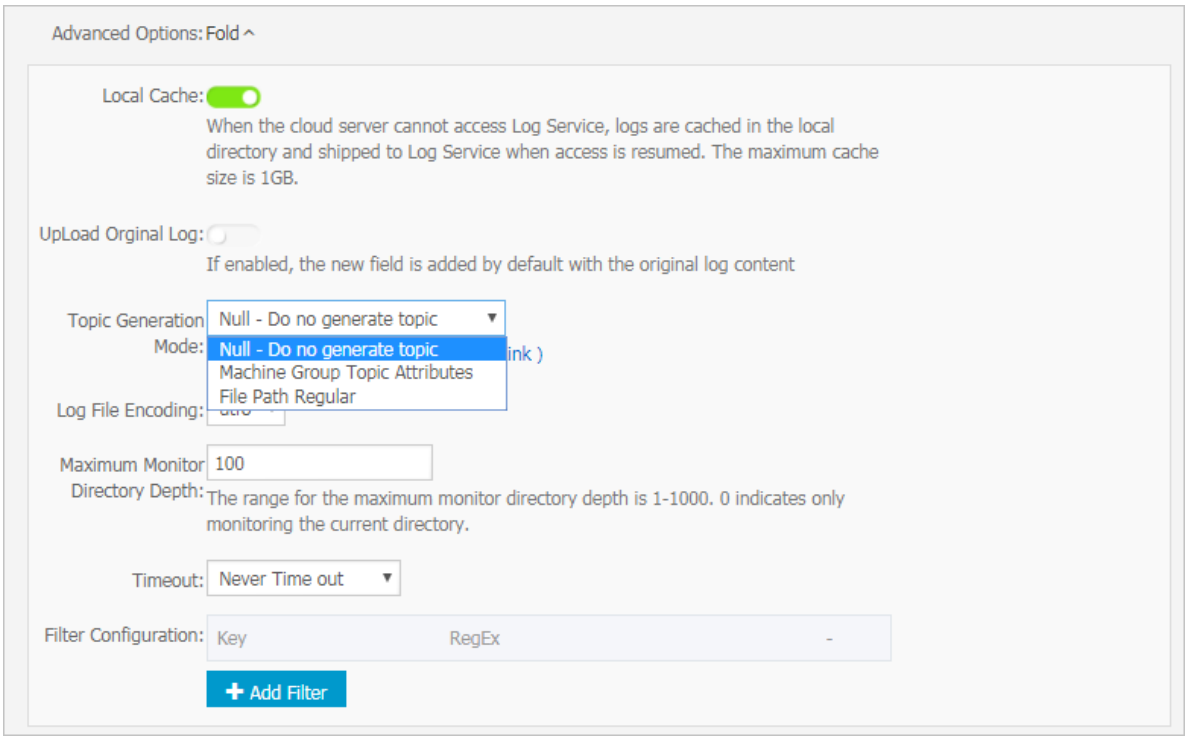
  When **File Path Regular** is selected as the topic generation mode, Logtail uploads the instance name as the topic name to Log Service when reporting data. The topic generated varies with your directory structure and configuration. You need to specify the topic name as the instance name when you performa query by using the **LogSearch/Analytics** function.

Set a log topic

Procedure

1. Configure Logtail in the Log Service console by referring to *Collect text logs*.

   If you want to set the topic generation mode to **Machine Group Topic Attribute**, set **Group Topic** in the **Create Machine Group** or **Modify Machine Group** dialog box first.
2. In Logtail configurations, expand **Advanced Options**, and set **Topic Generation Mode**.



24.7.3.1.4 Import historical log files

Logtail collects only incremental log files by default. To import historical log files, use the historical log file importing function in Logtail.

Prerequisites

- The Logtail version must be `0.16.6` or later.
- Historical files to be collected must be in the configured collection range and have never been collected by Logtail.
- The last modification time of the historical files must be earlier than the Logtail configuration time.
- The maximum latency for local event importing is one minute.
- Because local configuration loading is a special behavior, Logtail sends `LOAD_LOCAL_EVENT_ALARM` to the server to notify the user of such events.

Context

Logtail collects files based on events, which are generated during monitoring or periodical file polling. Besides, Logtail can load events from local files to drive log collection. Historical file collection is a function implemented based on local event loading.

Procedure

1. Create a collection configuration.

Create collection configurations according to *Collect text logs* and apply the configurations to the machine group. Ensure that the files are within the configured range.

2. Obtain the unique configuration ID.

   You can obtain the unique ID of the collection configuration from `/usr/local/ilogtail/user_log_config.json` as follows:

```
grep "##" /usr/local/ilogtail/user_log_config.json | awk {print $1
} "##1.0##log-config-test$multi" "##1.0##log-config-test$ecs-test"
 "##1.0##log-config-test$metric_system_test" "##1.0##log-config-test
$redis-status"
```

3. Add a local event.

   The storage path of local events is `/usr/local/ilogtail/local_event.json`. The files are of the standard JSON type, in the following format:

```
[ { "config" : "${your_config_unique_id}", "dir" : "${your_log_dir
}", "name" : "${your_log_file_name}" }, { ... } ... ]
```

   - **Configuration item**

Configurations	Description	Example
config	Unique configuration ID obtained in Step 2.	##1.0##log-config-test$ecs-test
dir	Folder of the file. **Note:** A folder must not end with `/`.	`/data/logs`
name	Log file name	`access.log.2018-08-08`

Note:

To prevent loading of invalid JSON data on Logtail, we recommend that you save local event configurations to a temporary file and copy the configurations to `/usr/local/ilogtail/local_event.json` after editing.

   - **Sample configuration**

```
$ cat /usr/local/ilogtail/local_event.json
                                    [
                                    {
                                    "config": "##1.0##log-config-test$
ecs-test",
                                    "dir": "/data/log/",
                                    "name": "access.log.2017-08-08"
                                    },
                                    {
```

```
ecs-test",                                "config": "##1.0##log-config-test$
                                          "dir": "/tmp",
                                          "name": "access.log.2017-08-09"
                                          }
                                          ]
```

- **Check whether Logtail has loaded the configuration.**

  Logtail typically loads the local configuration file to the memory and clears the content of `local_event.json` within one minute after `local_event.json` is saved locally.

  You can check whether Logtail has read events using any of the following methods:

  - If the content of `local_event.json` is cleared, Logtail has read the event information.
  - Check whether the file `/usr/local/ilogtail/ilogtail.LOG` contains the keyword `process local event`. If the content of `local_event.json` is cleared but the keyword is not found in the file, your local configuration file may be filtered out due to invalid content.
  - *Query error diagnostics* to check for `LOAD_LOCAL_EVENT_ALARM` notification.

- **Check whether the configuration is loaded but no data is collected.**

  If Logtail has loaded the configuration but no data is collected, the possible causes can be:

  - The configuration is invalid.
  - The local `config` does not exist.
  - No log files exist in the path specified in Logtail collection configuration.
  - The log file has already been collected by Logtail.

- **Determine how to re-collect data that has been collected before.**

  To collect data that has been collected before, perform the following steps:

  1. Run the `/etc/init.d/ilogtaild stop` command to stop Logtail.
  2. Search for the corresponding log file path in the `/tmp/logtail_check_point` file.
  3. Delete checkpoint(JSON object) from the log file and save the file.
  4. Add a local event as described in step 3.
  5. Run the `/etc/init.d/ilogtaild start` command to start Logtail.

24.7.3.2 Collect syslogs

Logtail supports local configuration of TCP ports to receive the syslog data transferred by syslog agents through TCP. The received data is parsed by Logtail and then forwarded to LogHub.

Prerequisites

Before collecting logs through Logtail, you must install Logtail. Apsara Stack Log Service allows you to install Logtail in Linux operating systems. For the installation method, see *Install Logtail (for Linux)*.

Step 1: Create a Logtail syslog configuration

1. *Log on to the Log Service console*.
2. Create a project and a Logstore. For the detailed procedures, see *Create a Project* and *Create a Logstore*.
3. In the Log Service console, click the target project to go to the **Logstores** page.
4. Select the Logstore and click **Data Import Wizard** next to it to start the data import configuration.
5. Select a data source type.

   Select **syslog** in **Custom Data** and click **Next**.
6. Complete the syslog configuration and click **Next**.

Parameter	Description
Configuration name	The configuration name can only contain lowercase letters, numbers, hyphens (-), and underscores (_). It must start and end with a lowercase letter or number and must be 3 to 63 bytes in length. **Note:** Once the configuration name is specified, it cannot be changed.
Tag settings	For more information about tag settings, see *Syslog collection reference*.
Advanced options	The following table lists the advanced configuration of syslog collection.

Parameter	Description
Local Cache	Indicates whether to enable the **Local Cache** function. When Log Service is unavailable, logs can be cached to a local directory and then uploaded after the service is recovered. The default maximum size of logs that can be cached is 1 GB.

Parameter	Description
Topic Generation Mode	• **Null - Do not generate topic**: The default option, which indicates to set the topic as a null string and you can query logs without entering the topic. • **Machine Group Topic Attributes**: Used to clearly differentiate log data generated in different frontend servers. • **File Path Regular**: When this mode is selected, you must enter a **Custom Regular** below to extract a part of the path as the topic. This mode is used to distinguish the log data generated by a user or an instance.
Custom RegEx	If you choose to generate a topic in **File Path Regular** mode, enter a custom regular expression here.
Log File Encoding	• utf8: UTF-8 encoding. • gbk: GBK encoding.
Maximum Monitor Directory Depth	Indicates the maximum depth of the monitoring directory when logs are collected from the log source, that is, up to what levels logs are collected. The maximum monitoring directory depth ranges from 0 to 1000, of which 0 means only the directory at the current level is monitored.
Timeout	If a log file is not updated within the specified period of time, the system considers that the file has timed out. You can configure the following settings for **Timeout**. • Never timed out: All log files are continuously monitored and never time out. • 30 minute timeout: If a log file is not updated in 30 minutes, the system considers that the log file has timed out and no longer monitors the file.
Filter Configuration	Only logs that completely meet the filtering conditions are collected. For example,`Key:level Regex:WARNING|ERROR` indicates to only collect logs whose level is WARNING or ERROR. You can also filter logs that do not conform to a condition. For example, `Key: level Regex:^(?!. *(INFO|DEBUG))` indicates to not collect logs whose level is INFO or DEBUG. For similar examples, see *regex-exclude-word* and *regex-exclude-pattern*.

7. Select the machine group and click **Apply to Machine Group** to apply the configuration to the selected machine group.

If you have not created a machine group, create one first. For more information, see *Create an IP address-based machine group*.

Step 2: Configure the Logtail protocol

Navigate to the Logtail installation directory `/usr/local/ilogtail/` on the server, locate `ilogtail_config.json`, and modify syslog-related settings as needed.

Procedure

1. Check whether Syslog is enabled.

   True indicates Syslog is enabled while False indicates Syslog is disabled.

   ```
   "streamlog_open" : true
   ```

2. Configure the size of the Syslog memory pool for storing received logs.

   Logtail requests memory of the specified size at one time when launched. Set the pool size based on the machine memory size and your needs. The unit is MB.

   ```
   "streamlog_pool_size_in_mb" : 50
   ```

3. Configure the buffer size.

   You must specify the size of the buffer that Logtail uses when calling the socket io rcv interface.

   ```
   "streamlog_rcv_size_each_call" : 1024
   ```

4. Configure the syslog format.

   ```
   "streamlog_formats":[]
   ```

5. Configure the TCP port.

   Configure the TCP port used by Logtail to receive syslogs. The port 11111 is used by default.

   ```
   "streamlog_tcp_port" : 11111
   ```

6. After the configuration is complete, restart Logtail.

   To restart Logtail, run the following commands to stop the Logtail client and then start it again.

   ```
   sudo /etc/init.d/ilogtaild stop
   sudo /etc/init.d/ilogtaild start
   ```

7. Install rsyslog.

   Visit the following official websites for more information on installing rsyslog:

   - *Ubuntu installation method*
   - *Debian installation method*

- *RHEL/CENTOS installation method*

8. Modify the settings.

   Navigate to `/etc/rsyslog.conf` and modify the settings in the file as needed. Example:

```
$WorkDirectory /var/spool/rsyslog # where to place spool files
$ActionQueueFileName fwdRule1     # unique name prefix for spool
files
$ActionQueueMaxDiskSpace 1g       # 1gb space limit (use as much as
possible)
$ActionQueueSaveOnShutdown on     # save messages to disk on
shutdown
$ActionQueueType LinkedList       # run asynchronously
$ActionResumeRetryCount -1        # infinite retries if host is down
# Defines the fields of log data
$template ALI_LOG_FMT,"0.1 sys_tag %timegenerated:::date-unixtimest
amp% %fromhost-ip% %hostname% %pri-text% %protocol-version% %app-
name% %procid% %msgid% %msg:::drop-last-lf%\n"
*. *      @@10.101.166.173:11111;ALI_LOG_FMT
```

   Note:

   In the `ALI_LOG_FMT` template, the value of the second field is `sys_tag`. This value must be the same as the one entered in step 1. This configuration indicates that all the (*.*) syslog data received`\*.\*` by this machine is formatted according to the `ALI_LOG_FMT` template, and forwarded to 10.101.166.173:11111 by using the TCP protocol. Machine 10.101.166.173 must be in the machine group selected in step 1 and configured according to step 2.

9. Start rsyslog.

```
sudo /etc/init.d/rsyslog restart
```

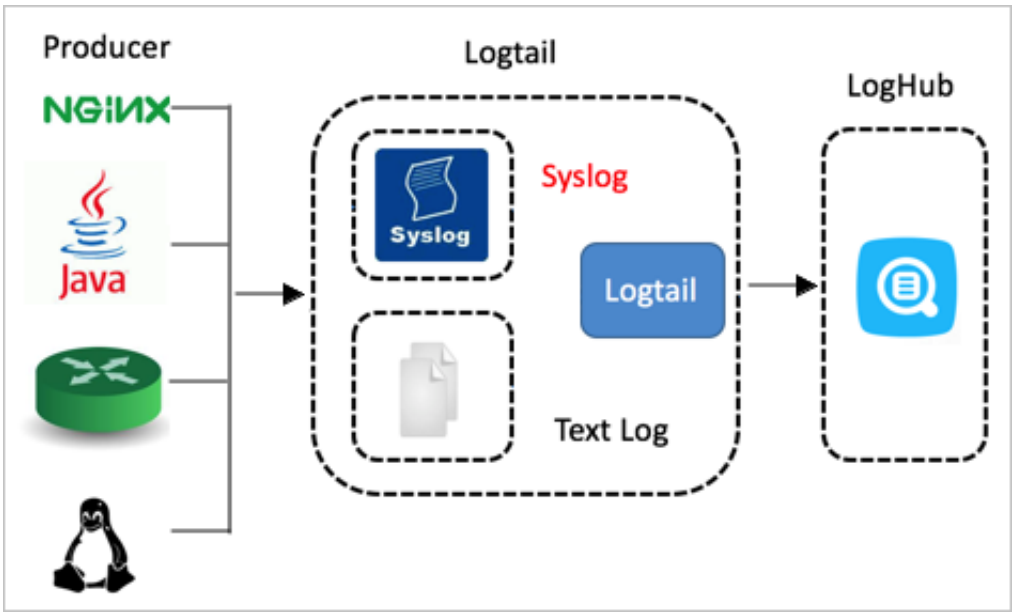
   Before starting rsyslog, check whether another syslog agent is installed on the machine, such as syslogd, sysklogd, and syslog-ng. If any of the preceding items exists, disable it.

   After completing the preceding three steps, you can now collect syslogs on the machine into Log Service.

24.7.3.3 Syslog collection reference

This document describes syslog collection reference for Logtail.

Currently, Logtail supports collection of syslogs and text files, as shown in the following figure.



Logtail collects syslogs based on TCP. For more information about how to configure Logtail to collect syslogs, see *Collect syslogs through Logtail*.

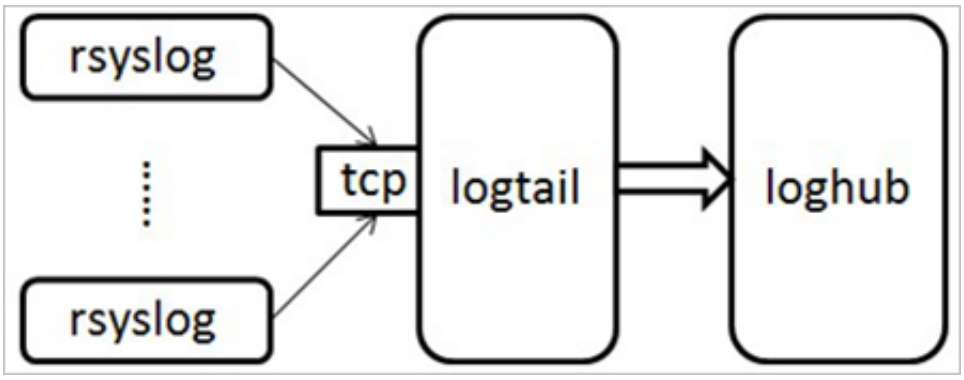
Benefits

For the syslog concept, see *Syslog*.

Compared with text files, syslogs can be directly collected to LogHub without being flushed into disks. It provides enhanced confidentiality and removes the need for parsing. A single machine delivers 80 MB/s of throughput.

How it works

Logtail supports local TCP port configurations and receives the logs forwarded by syslog agents. The following figure shows the relationship among Logtail, syslogs, and LogHub. Logtail with the TCP port enabled receives the syslogs forwarded by rsyslog or other syslog agents over TCP, parses the received logs, and forwards the logs to LogHub.



Syslog format

Logtail receives data as streams through the TCP port. If you want to parse individual logs from the data streams, ensure that the log format meets the following requirements:

- Logs are separated by linefeeds (`\n`), but do not contain linefeeds.
- Only the message body of a log can contain spaces; other fields cannot contain spaces.

The syslog format is shown as follows:

```
$version $tag $unixtimestamp $ip [$user-defined-field-1 $user-defined-field-2 $user-defined-field-n] $msg\n"
```

Meanings of the fields:

Log field	Meaning
version	Specifies the version of the log format. Logtail uses the version to parse user-defined-field.
tag	Specifies the data tag used to locate the project or Logstore. It cannot contain spaces or linefeeds.
unixtimestamp	Specifies the timestamp of the log.
ip	Specifies the IP address of the machine that corresponds to the log. If the IP address is 127.0.0.1, it is replaced with the peer address of the TCP socket when the log is sent to the server.
user-defined-field	Zero or multiple custom fields can be set. The fields cannot contain spaces or linefeeds. The braces indicate that the fields are optional.
msg	Specifies the message body of the log. The `\n` symbol appended to the message is a linefeed, but the message cannot contain linefeeds.

The following example is a log that meets the above format requirements:

```
2.1 streamlog_tag 1455776661 10.101.166.127 ERROR com.alibaba.streamlog.App.main(App.java:17) connection refused, retry
```

In addition to syslogs, Logtail can also collect other types of logs that meet the following requirements:

- The formatted logs can meet the requirements.
- Logs can be appended to the remote server over TCP.

Rules for Logtail to parse syslogs

You should add a configuration in Logtail for parsing syslogs, for example:

```
"streamlog_formats": [ {"version": "2.1", "fields": ["level", "method
"]}, {"version": "2.2", "fields": []}, {"version": "2.3", "fields": ["
pri-text", "app-name", "syslogtag"]} ]
```

Logtail identifies the corresponding user-defined-field format in streamlog_formats based on the version field. According to the sample configuration, the preceding sample log with the version field 2.1 contains two user-defined fields: level and method. The sample log is parsed in the following format:

```
{ "source": "10.101.166.127", "time": 1455776661, "level": "ERROR",
 "method": "com.alibaba.streamlog.App.main(App.java:17)", "msg": "
connection refused, retry" }
```

The version field is used to parse the user-defined-field. The tag is used to search for the project or Logstore to which the data is to be sent. The two fields are not sent to the Log Service instance as the log content. In addition, Logtail has preset some log formats in which the version field starts with "0." or "1.", for example, 0.1 or 1.1. Therefore, custom version fields cannot start with "0." or "1.".

Common Logtail syslog collection tools

- **log4j**
  - Introduce the Log4j library.

    ```
    <dependency> <groupId>org.apache.logging.log4j</groupId> <
    artifactId>log4j-api</artifactId> <version>2.5</version> </
    dependency> <dependency> <groupId>org.apache.logging.log4j</
    groupId> <artifactId>log4j-core</artifactId> <version>2.5</version
    > </dependency>
    ```

  - Introduce the Log4j configuration file `log4j_aliyun.xml`.

    ```
     <?xml version="1.0" encoding="UTF-8"?> <configuration status="OFF
    "> <appenders> <Socket name="StreamLog" protocol="TCP" host="10.
    101.166.173" port="11111"> <PatternLayout pattern="%X{version} %X
    {tag} %d{UNIX} %X{ip} %-5p %l %enc{%m}%n" /> </Socket> </appenders
    > <loggers> <root level="trace"> <appender-ref ref="StreamLog" />
     </root> </loggers> </configuration>
    ```

    10.101.166.173:11111 is the address of the machine where Logtail is located.

  - Set ThreadContext in programs.

    ```
    package com.alibaba.streamlog; import org.apache.logging.log4j.
    LogManager; import org.apache.logging.log4j.Logger; import org.
    apache.logging.log4j.ThreadContext; public class App { private
    ```

```
static Logger logger = LogManager.getLogger(App.class); public
static void main( String[] args ) throws InterruptedException {
ThreadContext.put("version", "2.1"); ThreadContext.put("tag", "
streamlog_tag"); ThreadContext.put("ip", "127.0.0.1"); while(true)
 { logger.error("hello world"); Thread.sleep(1000); } //ThreadCont
ext.clearAll(); } }
```

- **Tengine**

  Tengine can collect syslogs by ilogtail.

  Tengine uses the ngx_http_log_module for logging to the local syslog agent, which forwards the logs to rsyslog.

  For syslog configuration on Tengine, see *Configure syslog in Tengine.*

  Example:

  Send INFO-level access logs of the user type to Unix dgram(/dev/log) of the local machine and set the application tag to Nginx.

  ```
  access_log syslog:user:info:/var/log/nginx.sock:nginx
  ```

  Rsyslog configuration:

  ```
  module(load="imuxsock") # needs to be done just once input(type="
  imuxsock" Socket="/var/log/nginx.sock" CreatePath="on") $template
  ALI_LOG_FMT,"2.3 streamlog_tag %timegenerated:::date-unixtimestamp%
   %fromhost-ip% %pri-text% %app-name% %syslogtag% %msg:::drop-last-lf
  %\n" if $syslogtag == nginx then @@10.101.166.173:11111;ALI_LOG_FMT
  ```

- **Nginx**

  The collection of Nginx access logs is used as an example.

  Access log configuration:

  ```
  access_log syslog:server=unix:/var/log/nginx.sock,nohostname,tag=
  nginx;
  ```

  Rsyslog configuration:

  ```
  module(load="imuxsock") # needs to be done just once input(type="
  imuxsock" Socket="/var/log/nginx.sock" CreatePath="on") $template
  ALI_LOG_FMT,"2.3 streamlog_tag %timegenerated:::date-unixtimestamp%
   %fromhost-ip% %pri-text% %app-name% %syslogtag% %msg:::drop-last-lf
  %\n" if $syslogtag == nginx then @@10.101.166.173:11111;ALI_LOG_FMT
  ```

  For more information, visit *http://nginx.org/en/docs/syslog.html*.

- **Python syslog**

Example:

```
import logging import logging.handlers logger = logging.getLogger(
myLogger) logger.setLevel(logging.INFO) #add handler to the logger
 using unix domain socket /dev/log handler = logging.handlers.
SysLogHandler(/dev/log) #add formatter to the handler formatter =
logging.Formatter(Python: { "loggerName":"%(name)s", "asciTime":"%(
asctime)s", "pathName":"%(pathname)s", "logRecordCreationTime":"%(
created)f", "functionName":"%(funcName)s", "levelNo":"%(levelno)s",
 "lineNo":"%(lineno)d", "time":"%(msecs)d", "levelName":"%(levelname
)s", "message":"%(message)s"}) handler.formatter = formatter logger.
addHandler(handler) logger.info("Test Message")
```

24.7.4 Machine group

Log Service manages all ECS servers whose logs need to be collected using Logtail in machine groups.

After creating Logtail configurations, you can create a machine group on the **Machine Groups** page of a project in the Log Service project list. You can also go to the **Apply to Machine Group** page and click **Create Machine Group** to create a machine group.

A machine group is defined by either of the following two items:

- IP address: defines the name of the machine group and adds the internal IP addresses of servers to the group.

  You can add internal IP addresses of ECS servers to a machine group so that multiple ECS servers are directly added to the group, and centrally configure Logtail for the ECS servers.
- ID: indicates membership of the machine group, and is associated with the IDs configured on corresponding machines.

  The system consists of multiple modules, and each component of each module can be horizontally expanded separately. One module can contain multiple machines and the machine group is created for each module separately to collect logs by type. Therefore, you must set a custom ID for each module separately and configure an ID for the server of each module. For example, a common website generally consists of an HTTP request processing module, a cache module, a logic processing module, and a storage module, which can be identified as http_module, cache_module, logic_module, and store_module, respectively.

24.7.4.1 Machine groups

Log Service manages all ECS instances whose logs need to be collected by Logtail in machine groups.

A machine group is a virtual group that contains multiple servers. If you want the logs of multiple servers to be collected by Logtail clients with the same configuration, you can add the servers to a machine group and apply the Logtail configuration to the machine group.

A machine group is defined by either of the following two items:

- IP address: Add the IP addresses of all the servers to the machine group. Each server in the group can be identified by using its unique IP address.
- Custom ID: Customize an ID for the machine group and use this same custom ID for each server of the machine group.

IP address-based machine groups

You can add multiple servers to a machine group by adding their IP addresses to the machine group. Then you can configure the Logtail clients on all the servers at the same time.

- If you are using ECS instances that are not bound with hostnames and the network types of these instances remain unchanged, you can use their private IP addresses to define the machine group.
- In other cases, use the server IP address obtained automatically by the Logtail client when you define a machine group. The IP address of each server is recorded in the IP address field of the `app_info.json` server file on the server.

> **Note:**
> `app_info.json` is the file that records the internal information of the Logtail client.
> The internal information includes the server IP addresses obtained by the Logtail client automatically. Manually modifying the IP address field of this file does not change the IP addresses obtained by the Logtail client.

A Logtail client automatically obtains a server IP address by using the following methods:

- If the IP address of a server has been bound with its hostname in the `/etc/hosts` file of a server, the Logtail client automatically obtains the IP address.
- If the IP address of a server has not been bound with its hostname, the Logtail automatically obtains the IP address of the first NIC on the server.

For more information, see *Create an IP address-based machine group*.

Machine groups with user-defined identifiers

In addition to IP addresses, user-defined identifiers can also be used to dynamically define machine groups. Multiple servers in a machine group can use one user-defined identifier to implement machine group auto scaling. You only need to configure one user-defined identifier for new machines. Log Service then automatically identifies these machines and adds them to the machine group.

Typically, the system consists of multiple modules. Each module can be expanded horizontally. That is, multiple servers can be added to each module. By creating a machine group separately for each module, you can collect logs by module. Therefore, you need to create a user-defined identifier for each module, and set the machine group identifier for the servers of each module. For example, a common website generally consists of an HTTP request processing module, a cache module, a logic processing module, and a storage module, which can be identified as http_module , cache_module, logic_module, and store_module, respectively.

For more information, see *Create a machine group with a user-defined identifier*.

24.7.4.2 Create an IP address-based machine group

Log Service allows you to create IP address-based machine groups. After adding the IP addresses of servers retrieved by Logtail to an IP address-based machine group, you can use the same Logtail configuration to collect logs from the servers.

Prerequisites

- A project and a Logstore have been created.
- Ensure that there are one or more ECS instances.
- Logtail has been installed on the servers.

Procedure

1. Check the IP addresses of servers automatically retrieved by Logtail.

   The IP addresses automatically retrieved by Logtail are recorded in the ip field in the *app_info.json* file.

   On a server with Logtail installed, check the *app_info.json* file, whose path is:

   - For Linux: */usr/local/ilogtail/app_info.json*
   - For 64-bit Windows: *C:\Program Files (x86)\Alibaba\Logtail\app_info.json*
   - For 32-bit Windows: *C:\Program Files\Alibaba\Logtail\app_info.json*

2. *Log on to the Log Service console*.

3. In the left-side navigation pane, choose **LogHub - Real-time Collection** > **Logtail Machine Groups** to go to the **Machine Groups** page of the project.
4. Click **Create Machine Group** in the upper-right corner.

   Alternatively, after creating a collection configuration in the data import wizard, click **Create Machine Group** on the **Apply to Machine Group** page.
5. Create a machine group.

   a) Set **Machine Group Name**.

   The name can only contain lowercase letters, numbers, hyphens (-), and underscores (_). It must start and end with a lowercase letter or number and must be 3 to 128 bytes in length.

   b) Set **Machine Group Name**.

   > **Note:**
   > Because the machine group name cannot be modified once created, set it with caution.

   c) Select **IP Address** for **Machine Group Identifier**.

   d) Set **IP Address**.

   Enter the server IP address obtained in *1*.

   > **Note:**
   > - Make sure that the server IP address was obtained in accordance with *1*.
   > - When the machine group contains multiple servers, separate their IP addresses with linefeeds.
6. Optional: Set **Machine Group Topic**.

   For more information about the machine group topic, see *Generate a topic*.
7. Click **OK**.

What's next

After the machine group is created, you can view the machine group list, modify machine groups, view machine group status, manage machine group configurations, and delete machine groups.

24.7.4.3 Create a machine group with a user-defined identifier

In addition to IP addresses, you can also use user-defined identifiers to dynamically define machine groups.

User-defined identifiers feature considerable advantages in the following scenarios:

- In a custom network environment, such as a VPC, the IP address of one machine may conflict with that of another, resulting in Logtail management failure on the Log Service. User-defined identifiers help to avoid such issues.
- Multiple machines can share one tag for automatic scaling of machine groups. You only need to configure one user-defined identifier for new machines. Log Service then automatically identifies these machines and adds them to the machine group.

Procedure

1. Configure user-defined identifiers.

   Configure user-defined identifiers in the `/etc/ilogtail/user_defined_id` file.

   For example, set a user-defined machine identifier as follows:

   ```
   #cat /etc/ilogtail/user_defined_id
   ```

   Note:

   - You can configure multiple user-defined identifiers for a server and separate them with line breaks.
   - To use IP addresses to identify machines, delete the user_defined_id file. The updated configuration takes effect in one minute. Run the following command to delete user_defined_id:

     ```
     rm -f /etc/ilogtail/user_defined_id
     ```
   - If the directory (*`/etc/ilogtail/`*, or *`C:\LogtailData`*) or the file (*`/etc/ilogtail/user_defined_id`* or *`C:\LogtailData\user_defined_id`*) does not exist, create it manually.

   After you add, delete, or modify the user_defined_id file, the latest configuration takes effect in one minute by default.

   The configuration takes effect after Logtail restarts. Run the following command to restart Logtail immediately if necessary:

   ```
   /etc/init.d/ilogtaild stop
   /etc/init.d/ilogtaild start
   ```

2. Create a machine group.

   a) *Log on to the Log Service console*.

   b) Click the name of a project to go to the **Logstores** page.

c) On the **Machine Groups** page, click **Create Machine Group** in the upper-right corner.

d) Complete the configurations for the machine group.

- **Machine Group Name**: Enter a name of the machine group.

  The machine group name can only contain lowercase letters, numbers, hyphens (-), and underscores (_). It must start and end with a lowercase letter or number and must be 3 to 128 bytes in length.

  Note:
  Because the machine group name cannot be modified once created, set it with caution.

- **Machine Group Identifier**: Select a **user-defined identifier**.
- (Optional) **Machine Group Topic**: Enter a machine group topic. For more information, see *Generate a topic*.
- **User Defined Identifier**: Enter the user-defined identifier configured in step 1.

e) Click **OK**. When scaling up machines, you simply need to set user-defined identifiers on new servers.

3. View machine group status.

On the **Machine Groups** page, click **Machine Status** at the right of the machine group to view the list of machines that use the same user-defined identifier and their heartbeat status.

The system is usually composed of multiple modules, each of which can contain multiple machines. For example, a common website consists of a frontend HTTP request processing module, cache module, logic processing module, and storage module. Each module may be individually scaled horizontally. Thus, when new machines are added, logs must be collected in real time.

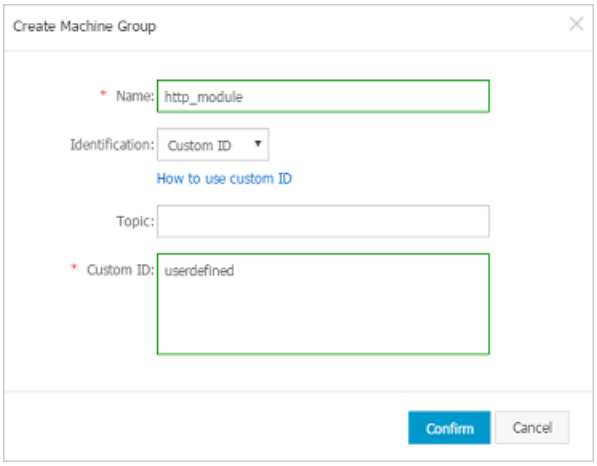
1. Create a user-defined identifier

After installing the Logtail client, enable the user-defined identifier for the server . For the modules in the preceding example, the user-defined identifier can be defined as http_module, cache_module, logic_module, and store_module.

2. Create a machine group.

Enter the corresponding user-defined identifier of the machine group in the **User-defined Identifier** field when creating the machine group. See the following

configurations of the http_module machine group. The http_module machine group is shown in the following figure:



3. Click **Check Status** for the machine group to view the list of machines that use the same user-defined identifier and their heartbeat statuses.
4. If a server with IP address 10.1.1.3 is added to the front-end module, enable userdefined-id on the new server. After the successful operation, you can view the added machine in the **Machine Group Status** dialog box.

24.7.4.4 View the machine group list

Go to the **Computer Groups** page to check the machine groups created under the current Logstore.

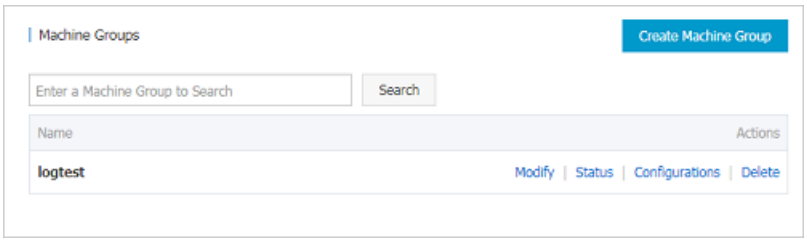
Prerequisites

1. *Create a Project*.
2. *Create a Logstore*.
3. *Create an IP address-based machine group*

Procedure

1. *Log on to the Log Service console*.
2. Click the project name to go to the **Logstores** page.
3. In the left-side navigation pane, click **Logtail Machine Group** to go to the **Machine Groups** page.

   You can view all machine groups under the project.



24.7.4.5 Modify a machine group

After creating a machine group, you can adjust the ECS instance list in the machine group whenever needed.

Prerequisites

1. *Create a Project*.
2. *Create a Logstore*.
3. *Create an IP address-based machine group*

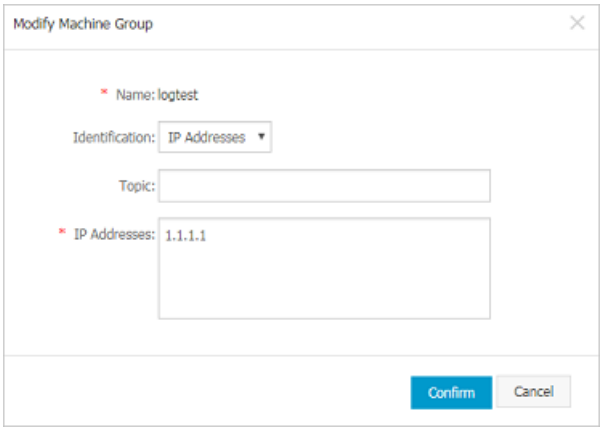
Modify a machine group

Procedure

1. *Log on to the Log Service console*.
2. Click a project name to go to the **Logstores** page.
3. In the left-side navigation pane, click **Logtail Machine Group** to go to the **Machine Groups** page.
4. Locate the machine group that you want to modify and click **Modify**.

> **Note:**
> Machine group names cannot be changed once created.

5. Modify the configuration of the machine group, and then click **OK**.



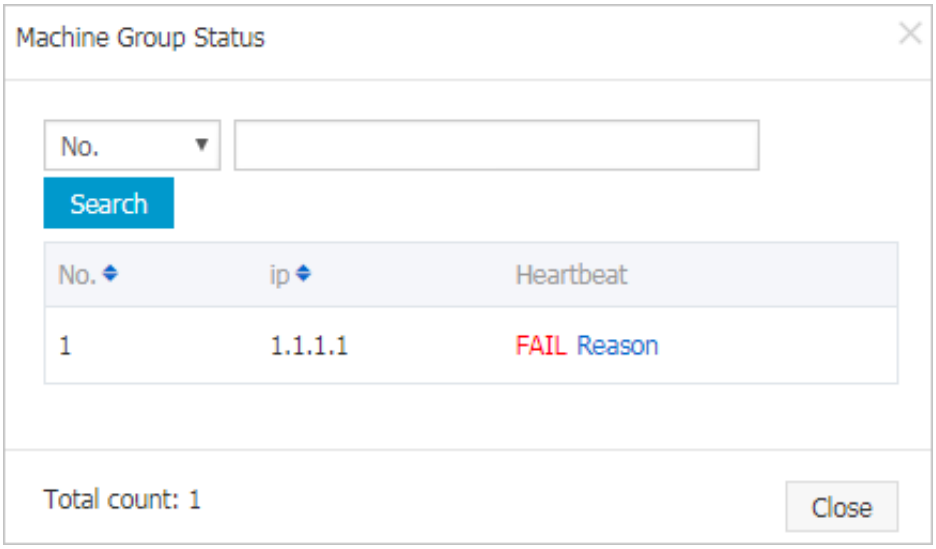
24.7.4.6 View the machine group status

To verify that the Logtail agent is successfully installed on all ECS servers in a machine group, you can view the heartbeat information of the Logtail agent.

Procedure

1. *Log on to the Log Service console*.
2. Click the name of a project to go to the **Logstores** page.
3. In the left-side navigation pane, click **Logtail Machine Group** to go to the **Machine Groups** page.
4. Select a machine group and click **Machine Status**.

   If the Logtail agent is successfully installed on all ECS servers, the heartbeat status of all ECS servers is **OK**. If the heartbeat status of an ECS server is**FAIL**, perform self-check as prompted. If the problem persists, submit a ticket for help.



24.7.4.7 Manage machine group configurations

Log Service manages all ECS instances whose logs need to be collected by Logtail in machine groups. You can enter the machine group list of a project **from the Projects page** of Log Service. Log Service allows you to create, modify, and delete machine groups, view the machine group list and status, manage configurations, and apply machine group IDs.

Prerequisites

1. *Create a Project*.
2. *Create a Logstore*.
3. *Create an IP address-based machine group*

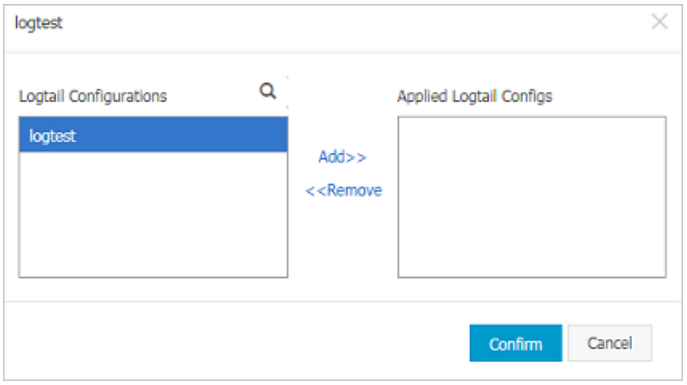
Context

Log Service manages all the servers whose logs need to be collected through machine groups. One important management item is the collection configuration of the Logtail client. For more information, see *Collect text logs* and *Collect syslogs*. You can apply or delete a Logtail configuration for or from a machine group to decide what logs are collected, how the logs are parsed, and to which Logstore the logs are sent by the Logtail on each ECS instance.

Procedure

1. *Log on to the Log Service console*.
2. Click a project name to go to the **Logstores** page.
3. In the left-side navigation pane, click **Logtail Machine Group** to go to the **Machine Groups** page.
4. Select a machine group and click **Config**.
5. Select a Logtail configuration and click **Add** or **Delete** to modify the Logtail configurations applied to machine groups.

   After a Logtail configuration is created, the configuration is issued to the Logtail client on each ECS instance in the machine group. When you remove the Logtail configuration, it is actually removed from the Logtail client.



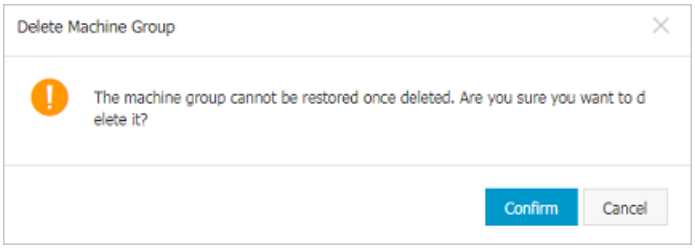
24.7.4.8 Delete a machine group

When you no longer need to collect logs from a machine group, you can delete the machine group.

Procedure

1. *Log on to the Log Service console*.
2. Create a project and a Logstore. For the detailed procedures, see *Create a Project* and *Create a Logstore*.

3. Click a project name to open the **Logstores** page. On the left-side navigation pane, click **Logtail Machine Groups**. The **Logtail Machine Groups** page is displayed.
4. Select a machine group and click **Delete**.
5. In the displayed message, click **OK**.



24.7.5 Troubleshooting

24.7.5.1 View the local log collection status

Logtail can be used to view the health status and log collection progress. This helps you to check log collection issues and customize status monitoring for log collection.

Instructions

If a Logtail client supporting status query function is installed, you can query local log collection status by entering commands on the client. For more information about how to install Logtail, see *Install Logtail (for Linux)*.

Run the `/etc/init.d/ilogtaild -h` command on the client to check whether the client supports querying local log collection status. If `logtail insight, version` is returned, it indicates that this function is supported on the Logtail client.

```
/etc/init.d/ilogtaild -h
Usage: ./ilogtaild { start | stop (graceful, flush data and save
checkpoints) | force-stop | status | -h for help}$
logtail insight, version : 0.1.0
common list :
       status all [index]
              get logtail running status
       status active [--logstore | --logfile] index [project] [
logstore]
              list all active logstore | logfile. if use --logfile,
please add project and logstore. default --logstore
       status logstore [--format=line | json] index project logstore
              get logstore status with line or json style. default --
format=line
       status logfile [--format=line | json] index project logstore
fileFullPath
              get log file status with line or json style. default --
format=line
       status history beginIndex endIndex  project logstore [
fileFullPath]
              query logstore | logfile history status.
```

```
index :   from 1 to 60. in all, it means last $(index) minutes; in
active/logstore/logfile/history, it means last $(index)*10 minutes
```

Currently, Logtail supports the following query commands, command functions, query time intervals, and time windows for result statistics:

Command	Function	Time interval to query	Time window for statistics
all	Query the running status of Logtail	Last 60 minutes	1 minute
active	Query Logstores or log files that are currently active (that is, with data collected).	Last 600 minutes	10 minutes
Logstore	Query the collection status of a Logstore.	Last 600 minutes	10 minutes
logfile	Query the collection status of a log file.	Last 600 minutes	10 minutes
history	Query the collection status of a Logstore or log file over a period of time.	Last 600 minutes	10 minutes

 **Note:**

- The `index` parameter in the command indicates the index value of the time window, which is counted from the current time. Its valid range is from 1 to 60. If the time window for statistics is one minute, windows in the last `(index, index-1]` minutes are queried. If the time window for statistics is 10 minutes, windows in the last `(10*index, 10*(index-1)]` minutes are queried.
- All query commands belong to status subcommands, so the main command is status.

all

Command format

```
/etc/init.d/ilogtaild status all [ index ]
```

 **Note:**

The all command is used to view the running status of Logtail. The index parameter is optional. If left blank, the default value is 1.

Example

```
/etc/init.d/ilogtaild status all 1
ok
/etc/init.d/ilogtaild status all 10
busy
```

Output description

Project	Description	Priority	Solution
ok	The current status is normal.	N/A	No action is required.
busy	The current collection speed is high, and Logtail is running properly.	N/A	No action is required.
many_log_files	The large number of log files are being collected.	Low	Check whether the configuration contains files that do not need to be collected.
process_block	Current log parsing is blocked.	Low	Check whether the generation of logs is too fast. If this output persists, *Configure startup parameters* as needed to change the CPU usage threshold or the concurrent sending threshold on the network.
send_block	Current sending is blocked.	Relatively high	Check whether the generation of logs is too fast and whether the network status is normal. If this output persists, *Configure startup parameters* as needed to change the CPU usage threshold or the concurrent

Project	Description	Priority	Solution
			sending threshold on the network.
send_error	Failed to upload log data.	High	To troubleshoot the issue, see *Configure startup parameters*.

active command

Command format

```
/etc/init.d/ilogtaild status active [--logstore] index
 /etc/init.d/ilogtaild status active --logfile index project-name
logstore-name
```

Note:

- The `active [--logstore] index` command is used to query Logstores that are currently active. The `--logstore` parameter can be omitted without changing the meaning of the command.
- The `active --logfile index project-name logstore-name` command is used to query all active log files in a Logstore for a project.
- The active command is used to query active log files level by level. We recommend that you first locate the currently active Logstore and then query active log files in this Logstore.

Example

```
/etc/init.d/ilogtaild status active 1
sls-zc-test : release-test
sls-zc-test : release-test-ant-rpc-3
sls-zc-test : release-test-same-regex-3

/etc/init.d/ilogtaild status active --logfile 1 sls-zc-test release-
test
/disk2/test/normal/access.log
```

Output description

- If you run the `active --logstore index` command, all currently active Logstores are displayed in the format of `project-name : logstore-name`. If you run the `active --logfile index project-name logstore-name` command, the complete paths of active log files are displayed.
- A Logstore or log file with no log collection activity in the current query window does not appear in the output.

logstore command

Command format

```
/etc/init.d/ilogtaild status logstore [--format={line|json}] index
project-name logstore-name
```

Note:

- The logstore command is used to output the collection statuses of the specified project and Logstore in LINE or JSON format.
- If the`--format=` parameter is not configured, `--format=line` is selected by default. The echo information is output in LINE format. *Note*: The `--format` parameter must be placed after `logstore`.
- If this Logstore does not exist or has no log collection activity in the current query window, you get an empty output in LINE format or a `null` value in JSON format.

Example

```
/etc/init.d/ilogtaild status logstore 1 sls-zc-test release-test-same
time_begin_readable : 17-08-29 10:56:11
time_end_readable : 17-08-29 11:06:11
time_begin : 1503975371
time_end : 1503975971
project : sls-zc-test
logstore : release-test-same
status : ok
config : ##1.0##sls-zc-test$same
read_bytes : 65033430
parse_success_lines : 230615
parse_fail_lines : 0
last_read_time : 1503975970
read_count : 687
avg_delay_bytes : 0
max_unsend_time : 0
min_unsend_time : 0
max_send_success_time : 1503975968
send_queue_size : 0
send_network_error_count : 0
send_network_quota_count : 0
send_network_discard_count : 0
send_success_count : 302
send_block_flag : false
sender_valid_flag : true
/etc/init.d/ilogtaild status logstore --format=json 1 sls-zc-test
release-test-same
{
   "avg_delay_bytes" : 0,
   "config" : "##1.0##sls-zc-test$same",
   "last_read_time" : 1503975970,
   "logstore" : "release-test-same",
   "max_send_success_time" : 1503975968,
   "max_unsend_time" : 0,
```

```
    "min_unsend_time" : 0,
    "parse_fail_lines" : 0,
    "parse_success_lines" : 230615,
    "project" : "sls-zc-test",
    "read_bytes" : 65033430,
    "read_count" : 687,
    "send_block_flag" : false,
    "send_network_discard_count" : 0,
    "send_network_error_count" : 0,
    "send_network_quota_count" : 0,
    "send_queue_size" : 0,
    "send_success_count" : 302,
    "sender_valid_flag" : true,
    "status" : "ok",
    "time_begin" : 1503975371,
    "time_begin_readable" : "17-08-29 10:56:11",
    "time_end" : 1503975971,
    "time_end_readable" : "17-08-29 11:06:11"
}
```


Output description

Keyword	Description	Unit
status	The overall status of this Logstore. For specific statuses , descriptions, and change methods, see the following table.	N/A
time_begin_readable	The start time of reading.	N/A
time_end_readable	The end time of reading.	N/A
time_begin	The start time of statistics collection.	UNIX timestamp (seconds)
time_end	The end time of statistics collection.	UNIX timestamp (seconds)
project	The project name.	N/A
Logstore	The Logstore name.	N/A
config	The collection configuration name, which is globally unique and consisted of `##1.0##`, project, `$`, and config.	N/A
read_bytes	The number of logs read in the window.	Bytes
parse_success_lines	The number of successfully parsed log lines in the window.	Line

Keyword	Description	Unit
Parse_fail_lines	The number of log lines that failed to be parsed in the window.	Line
last_read_time	The last reading time in the window.	UNIX timestamp (seconds)
read_count	The number of times of reading logs in the window.	Number of times
avg_delay_bytes	The average of the differences between the current offset and the file size each time logs are read in the window.	Bytes
max_unsend_time	The maximum time that unsent data packets are in the send queue when the window ends. The value is 0 when the queue is empty.	UNIX timestamp (seconds)
min_unsend_time	The minimum time that unsent data packets are in the send queue when the window ends. The value is 0 when the queue is empty.	UNIX timestamp (seconds)
max_send_success_time	The maximum time that data is successfully sent in the window.	UNIX timestamp (seconds)
send_queue_size	The number of unsent data packets in the current send queue when the window ends.	-
send_network_error_count	The number of data packets that failed to be sent in the window because of network errors.	-
send_network_quota_count	The number of data packets that failed to be sent in the window because the quota is exceeded.	-
send_network_discard_count	The number of discarded data packets in the window because	-

Keyword	Description	Unit
	of data exceptions or insufficient permissions.	
send_success_count	The number of successfully sent data packets in the window.	-
send_block_flag	Whether the send queue is blocked when the window ends.	N/A
sender_valid_flag	Whether the send flag of this Logstore is valid when the window ends. TRUE indicates the flag is valid, and FALSE indicates it is disabled due to network errors or quota errors.	N/A

Logstore status

Status	Definition	Processing method
ok	The status is normal.	No action is required.
process_block	Log parsing is blocked.	Check whether the generation of logs is too fast. If this output persists, *Configure startup parameters* as needed to change the CPU usage threshold or the concurrent sending threshold on the network.
parse_fail	Log parsing failed.	Check whether the log format is consistent with the log collection configuration.
send_block	Current sending is blocked.	Check whether the generation of logs is too fast and whether the network status is normal. If this output persists, *Configure startup parameters* as needed to change the CPU usage threshold or the concurrent sending threshold on the network.

Status	Definition	Processing method
sender_invalid	An exception occurred during log data sending.	Check the network status. If the network is normal, see *Troubleshoot log collection errors* in Query diagnosis errors to troubleshoot the issue.

logfile command

Command format

```
/etc/init.d/ilogtaild status logfile [--format={line|json}] index
project-name logstore-name fileFullPath
```

Note:

- The logfile command is used to output the collection status of a specific log file in LINE or JSON format.
- If the `--format=` parameter is not configured, `--format=line` is selected by default. The echo information is output in LINE format.
- If this log file is unavailable or has no log collection activity in the current query window, you get an empty output in LINE format or a `null` value in JSON format.
- The `--format` parameter must be placed behind `logfile`.
- The `filefullpath`must be a full path name.

Example

```
/etc/init.d/ilogtaild status logfile 1 sls-zc-test release-test-same /
disk2/test/normal/access.log
time_begin_readable : 17-08-29 11:16:11
time_end_readable : 17-08-29 11:26:11
time_begin : 1503976571
time_end : 1503977171
project : sls-zc-test
logstore : release-test-same
status : ok
config : ##1.0##sls-zc-test$same
file_path : /disk2/test/normal/access.log
file_dev : 64800
file_inode : 22544456
file_size_bytes : 17154060
file_offset_bytes : 17154060
read_bytes : 65033430
parse_success_lines : 230615
parse_fail_lines : 0
last_read_time : 1503977170
read_count : 667
```

```
avg_delay_bytes : 0
/etc/init.d/ilogtaild status logfile --format=json 1 sls-zc-test
release-test-same /disk2/test/normal/access.log
{
   "avg_delay_bytes" : 0,
   "config" : "##1.0##sls-zc-test$same",
   "file_dev" : 64800,
   "file_inode" : 22544456,
   "file_path" : "/disk2/test/normal/access.log",
   "file_size_bytes" : 17154060,
   "last_read_time" : 1503977170,
   "logstore" : "release-test-same",
   "parse_fail_lines" : 0,
   "parse_success_lines" : 230615,
   "project" : "sls-zc-test",
   "read_bytes" : 65033430,
   "read_count" : 667,
   "read_offset_bytes" : 17154060,
   "status" : "ok",
   "time_begin" : 1503976571,
   "time_begin_readable" : "17-08-29 11:16:11",
   "time_end" : 1503977171,
   "time_end_readable" : "17-08-29 11:26:11"
}
```

Output description

Keyword	Description	Unit
status	The collection status of this log file in the current query window. See the status of logstore command.	N/A
time_begin_readable	The start time of reading.	N/A
time_end_readable	The end time of reading.	N/A
time_begin	The start time of statistics collection.	UNIX timestamp (seconds)
time_end	The end time of statistics collection.	UNIX timestamp (seconds)
project	The project name.	N/A
logstore	The Logstore name.	N/A
file_path	The path of the log file.	N/A
file_dev	The device ID of the log file.	N/A
file_inode	The inode of the log file.	N/A
file_size_bytes	The size of the last scanned file in the window.	Bytes
read_offset_bytes	The parsing offset of this file.	Bytes

Keyword	Description	Unit
config	The collection configuration name, which is globally unique and consisted of `##1.0##`, project, `$`, and config.	N/A
read_bytes	The number of logs read in the window.	Bytes
parse_success_lines	The number of successfully parsed log lines in the window.	Line
Parse_fail_lines	The number of log lines that failed to be parsed in the window.	Line
last_read_time	The last reading time in the window.	UNIX timestamp (seconds)
read_count	The number of times of reading logs in the window.	Number of times
avg_delay_bytes	The average of the differences between the current offset and the file size each time logs are read in the window.	Bytes

history command

Command format

```
/etc/init.d/ilogtaild status history beginIndex endIndex project-name
logstore-name [fileFullPath]
```

Note:

- The history command is used to query the collection status of a Logstore or log file over a period of time.
- `beginIndex` and `endIndex` represent the start and end values for the code query window index respectively. `beginIndex <= endIndex`.
- If `fileFullPath` is not entered, the collection information of the Logstore is queried. If this parameter is entered, the collection information of the log files is queried.

Example

```
/etc/init.d/ilogtaild status history 1 3 sls-zc-test release-test-same
 /disk2/test/normal/access.log
```

```
        begin_time          status      read  parse_success
parse_fail      last_read_time  read_count  avg_delay    device
inode  file_size  read_offset
 17-08-29 11:26:11              ok   62.12MB         231000
0   17-08-29 11:36:11         671         0B     64800  22544459    18
.22MB      18.22MB
 17-08-29 11:16:11              ok   62.02MB         230615
0   17-08-29 11:26:10         667         0B     64800  22544456    16
.36MB      16.36MB
 17-08-29 11:06:11              ok   62.12MB         231000
0   17-08-29 11:16:11         687         0B     64800  22544452    14
.46MB      14.46MB
$/etc/init.d/ilogtaild status history 2 5 sls-zc-test release-test-
same
        begin_time          status      read  parse_success
parse_fail      last_read_time  read_count  avg_delay  send_queue
network_error  quota_error  discard_error  send_success  send_block
send_valid          max_unsend          min_unsend    max_send_success
 17-08-29 11:16:11              ok   62.02MB         230615
0   17-08-29 11:26:10         667         0B           0
0            0              0           300       false        true
70-01-01 08:00:00   70-01-01 08:00:00   17-08-29 11:26:08
 17-08-29 11:06:11              ok   62.12MB         231000
0   17-08-29 11:16:11         687         0B           0
0            0              0           303       false        true
70-01-01 08:00:00   70-01-01 08:00:00   17-08-29 11:16:10
 17-08-29 10:56:11              ok   62.02MB         230615
0   17-08-29 11:06:10         687         0B           0
0            0              0           302       false        true
70-01-01 08:00:00   70-01-01 08:00:00   17-08-29 11:06:08
 17-08-29 10:46:11              ok   62.12MB         231000
0   17-08-29 10:56:11         692         0B           0
0            0              0           302       false        true
70-01-01 08:00:00   70-01-01 08:00:00   17-08-29 10:56:10
```

Output description

- This command is used to view historical collection information of a Logstore or log file in the form of list, one line for each window.
- For the description of each output field, see the `logstore` and `logfile` commands.

Returned values

Normal returned value

0 is returned when all command inputs are valid (**including failure to run a query on a log store or log file**), for example:

```
/etc/init.d/ilogtaild status logfile --format=json 1 error-project
error-logstore /no/this/file
null
echo $?
0
/etc/init.d/ilogtaild status all
ok
echo $?
```

```
0
```

Abnormal returned value

A non-zero return value indicates an exception. See the following for details:

Returned values	Type	Output	Troubleshooting
10	Invalid command or missing parameters	`invalid param, use -h for help.`	Enter `-h` to view help.
1	The query goes beyond the 1–60 time window	`invalid query interval`	Enter `-h` to view help.
1	Failed to query the specified time window	`query fail, error : $(error)`. For more information, see *error interpretation*.	This issue may occur when the startup time of Logtail is less than the query time span. For other cases, open a ticket.
1	No matching query window time	`no match time interval, please check logtail status`	Check whether Logtail is running. For other cases, open a ticket.
1	No data in the query window	`invalid profile, maybe logtail restart`	Check whether Logtail is running. For other cases, open a ticket.

Example

```
/etc/init.d/ilogtaild status nothiscmd
invalid param, use -h for help.
echo $?
10
/etc/init.d/ilogtaild status/all 99
invalid query interval
echo $?
1
```

Usage scenarios

You can use Logtail health check to understand the overall status of Logtail, and perform collection progress query to obtain related metrics during collection. With the obtained information , you can monitor log collection in a customized manner.

Monitor the running status of Logtail

Monitor the running status of Logtail by using the `all` command.

How it works: The current status of Logtail is queried every minute. If Logtail is in the `process_block`, `send_block`, or `send_error` state for 5 minutes, an alarm is triggered.

You can adjust the alert duration and the range of statuses to be monitored based on the importance of log collection in specific scenarios.

Monitor the log collection progress

Monitor the collection progress of a Logstore by using the `logstore` command.

How it works: The `logstore` command is called every 10 minutes to get status information about this Logstore. If `avg_delay_bytes` is over 1 MB (1024 × 1024) or `status` is not `ok`, an alarm is triggered.

The `avg_delay_bytes` alarm threshold can be adjusted based on the log collection traffic.

Determine whether collection of a log file is complete

Determine whether collection of a log file is complete by using the `logfile` command.

How it works: After writing to the log file stops, the `logfile` command is called every 10 minutes to obtain the status information of this file. If this file shows the same value for `read_offset_bytes` and `file_size_bytes`, it means that collection of this log file is complete.

Troubleshoot log collection issues

If log collection is delayed on a server, use the `history` command to find related collection information on this server.

1. If `send_block_flag` is TRUE, it indicates that the log collection delays because of the network.
   - If `send_network_quota_count` is greater than 0, you must *Split a shard* of the Logstore.
   - If `send_network_error_count` is greater than 0, you must check the network connectivity.
   - If no related network error occurs, you must adjust the *concurrent transmission limit and traffic limit* of Logtail.
2. Sending-related parameters are normal, but the `avg_delay_bytes` value is higher than a normal one.

- The average log parsing speed can be calculated by using `read_bytes` to determine whether log generation traffic is normal.
- The *configuration parameters* of Logtail can be adjusted as needed.

3. `parse_fail_lines` is greater than 0.

   Check whether the parsing configurations for log collection can be mapped to all logs.

24.7.5.2 Query error diagnostics

Errors may occur during log collection by using Logtail, such as regular expression parsing failures, incorrect file paths, and traffic exceeding the shard service capability. Log Service provides the diagnosis function for diagnosing Logtail log collection errors on your own.

Procedure

1. Go to the error diagnostics page.

   Log on to the Log Service console. Select the specified project to go to the **Logstores** page. Then, click **Diagnosis** in the **Log Collection Mode** column.

2. View log collection errors.

   You can view the list of Logtail log collection errors of a specified Logstore in the Log Collection Error dialog box.

3. Query log collection errors of a specified machine

   To query all log collection errors occurred to a specific machine, enter the IP address of the machine in the search box on the query page. Logtail reports errors every five minutes.

   After an error is rectified, check whether the error is reported again based on the error time statistics after the service recovers. Historical errors are displayed before expiration. You can ignore these errors and check only the new errors reported after error rectification.

Error type	Description	Processing method
LOGFILE_PERMINSSION_ALARM	Logtail has no permission to read the specified file.	Check the Logtail startup account on the server. We recommend that you start Logtail as the root user.
SPLIT_LOG_FAIL_ALARM	The regular expression cannot match with the beginning of a row and cannot separate logs into rows.	Check the correctness of the row start regular expression. If the log contains only one row, you can set the row start regular expression to `.*`.

Error type	Description	Processing method
MULTI_CONFIG_MATCH_ALARM	Only one file can be collected by a Logtail configuration at one time.	Check whether a file is collected by multiple Logtail configurations. If yes, delete the redundant configurations.
REGEX_MATCH_ALARM	The log content does not match the regular expression in regular expression mode.	Copy the log sample from the error content for re-matching and generate a new regular expression for parsing.
PARSE_LOG_FAIL_ALARM	Logtail fails to parse logs because the log format does not conform to the definition in the parsing modes such as JSON and delimiter.	Click the error to view relevant details.
CATEGORY_CONFIG_ALARM	The Logtail collection configuration is invalid.	A common error is that the file path fails to be extracted as a topic by a regular expression. For other errors, submit a ticket.
LOGTAIL_CRASH_ALARM	Logtail has crashed because it exceeds the upper limit of machine resource usage.	For more information, see *Configure startup parameters* to modify the upper limits of CPU utilization and memory usage. Open a ticket if you have any questions.
REGISTER_INOTIFY_FAIL_ALARM	Logtail fails to register log listening in Linux possibly because Logtail does not have permissions to access the folder or the folder has been deleted.	Check whether Logtail has the folder access permission and whether the folder exists.
DISCARD_DATA_ALARM	This error is caused by the insufficient CPU resources configured for Logtail or the network bandwidth throttling.	Modify the upper limit of the CPU usage or limits on concurrent incoming traffic to Log Service by following the instructions provided in *Configure startup parameters*. You can also open a ticket for additional support.

Error type	Description	Processing method
SEND_DATA_FAIL_ALARM	1. The main account does not create any AccessKey. 2. The Logtail client cannot connect to Log Service, or the network link quality is bad. 3. The writing quota of Log Service is insufficient.	1. Refer to *View the key pair* to view AccessKey. 2. Check the local configuration file /usr/local/ilogtail/ ilogtail_config.json. Run the curl <end point> command and check whether any output is returned. 3. Add shards to the Logstore to support writing larger data volumes.
PARSE_TIME_FAIL_ALARM	Logtail fails to parse the time field based on the time parsing expression.	Configure the time parsing expression correctly based on the log time.
REGISTER_INOTIFY_FAIL_ALARM	Logtail fails to register inotify watcher for the log directory.	Check whether the log monitoring directory exists. If the directory exists, check the directory permission setting.
SEND_QUOTA_EXCEED_ALARM	The traffic of writing logs exceeds the limit.	*View the key pair* in the console.
READ_LOG_DELAY_ALARM	Log collection lags behind log generation. In normal cases, this is because the CPU resources for configuring Logtail are insufficient or the incoming traffic to Log Service is restricted.	Modify the upper limit of the CPU usage or limits on concurrent incoming traffic to Log Service by following the instructions. You can also open a ticket for additional support.
DROP_LOG_ALARM	Log collection lags behind log generation, and unprocessed log rotations outnumbers 20. In normal cases, this is because the CPU resources for configuring Logtail are insufficient or the incoming traffic to Log Service is restricted.	Modify the upper limit of the CPU usage or limits on concurrent incoming traffic to Log Service by following the instructions provided in *Configure startup parameters*. You can also open a ticket for additional support.
LOGDIR_PERMINSSION_ALARM	Logtail has no permission to read the log monitoring directory.	Check whether the log monitoring directory exists. If the directory exists, check the directory permission setting.

Error type	Description	Processing method
ENCODING_CONVERT_ALARM	Logtail fails to convert the encoding.	Check whether the log encoding format configuration is consistent with the log encoding format.
OUTDATED_LOG_ALAR	Logs expire with a time lag of more than 12 hours. Possible causes: Log parsing lags behind by more than 12 hours, the user-defined time field is incorrectly configured, or the time output of the logging program is abnormal.	Check whether READ_LOG_DELAY_ALARM exists. If yes, handle the error according to the instructions of READ_LOG_DELAY_ALARM. If not, check the time field. If the time field is correctly configured, check whether the time output of the logging program is normal. If you have doubts, submit a ticket.
STAT_LIMIT_ALARM	The number of files in the Logtail Config directory exceeds the upper limit.	Check whether the log collection configuration directory contains many files and subdirectories, and properly configure the monitored root directory and the maximum monitoring depth of the directory.
DROP_DATA_ALARM	Flushing logs into the local disk times out when exiting the process and the logs unflushed are discarded.	Generally, this error is caused by the severe collection obstruction. You can modify the upper limit of CPU usage or limits on concurrent incoming traffic to Log Service by following the instructions provided in *Configure startup parameters*. You can also open a ticket for additional support.
INPUT_COLLECT_ALARM	An exception occurred when collecting the input sources.	Fix the error as instructed by the error message.
HTTP_LOAD_ADDRESS_ALARM	The entered HTTP address is invalid.	Check the validity of the address.

Error type	Description	Processing method
HTTP_COLLECT_ALARM	An exception occurred when collecting HTTP.	Troubleshoot the error as instructed by the error message. Generally, this error is caused by timeout.
FILTER_INIT_ALARM	An exception occurred when initiating the filter.	Generally, this error is caused by the invalid regular expression of the filter. Handle the error as instructed.
INPUT_CANAL_ALARM	An exception occurred when running MySQL binlog.	Troubleshoot the error as instructed by the error message. The canal service may be restarted when the configuration is updated. Therefore, you can ignore the service restart error.
CANAL_INVALID_ALARM	An exception in the internal status of MySQL binlog.	This error is typically due to meta information inconsistency caused by table schema information change during running. Check whether the table schema is being modified when the error is reported. If not, submit a ticket.
MYSQL_INIT_ALARM	An exception occurred when initiating MySQL.	Fix the error as instructed by the error message.
MYSQL_CHECKPOING_ALARM	An exception in the MySQL checkpoint format.	Check whether the checkpoint configuration is modified. If not, submit a ticket.
MYSQL_TIMEOUT_ALARM	MySQL query times out.	Check whether MySQL server and network connection are normal.
MYSQL_PARSE_ALARM	Logtail fails to parse the MySQL query results.	Check whether the checkpoint format configured on MySQL is consistent with the format of corresponding fields.

> **Note:**
> To check all the complete log rows discarded due to a parsing failure, log on to the machine and check the `/usr/local/ilogtail/ilogtail.LOG` file.

24.7.5.3 Troubleshoot log collection errors

If an error occurred when you use Logtail to collect logs, perform the following steps for troubleshooting:

Procedure

1. Check whether the primary account is configured with an AccessKey.
   a) *Log on to the Log Service console*.
   b) On the homepage of the console, click the profile picture of the currently logged-on user in the upper-right corner. Then, click **Personal Information**.
   c) On the **Personal Information** page, click **AccessKey**. In the displayed dialog box, click **OK**. Check whether the current account contains an AccessKey.
2. Check whether the Logtail heartbeat of the machine group is normal.

   Log on to the Log Service console. On the **Machine Groups** page, check the status of the machine group. If the heartbeat status is OK, perform the next step; if the heartbeat status is FAIL, proceed with the troubleshooting.

   The causes of Logtail heartbeat failure include:

   - **Logtail is not installed**

     Check the client status in Linux:

     ```
     sudo /etc/init.d/ilogtaild status
     ```

     If the Logtail client is not installed, see *Install Logtail (for Linux)* to install Logtail on the server for log collection.
   - **Incorrect parameters are selected during installation**

     As Log Service operates by region, you must specify the correct endpoint when installing the Logtail client. Check the configuration of your installed clients in the following paths:

     — For Linux: `/usr/local/ilogtail/ilogtail_config.json`

     Perform the following checks:

- Check whether the network ingress connected to the client is in the same region as your Log Service project. For the list of network ingresses, see *Endpoints*.
- Check whether the correct domain is selected based on the network environment of your machine. If an internal domain is selected for a VPC environment, the client connection will fail. You can telnet to the domain configured in *ilogtail_config.json*, such as `telnet logtail.cn-hangzhou-intranet.log.aliyuncs.com 80`

- **An incorrect IP address or user ID is configured on the server.**

  Generally, Logtail obtains the IP address of a machine in one of the following ways:

  - If host name binding is configured in the */etc/hosts* file, confirm the bound IP address. Run the `hostname` command to view the host name.
  - If no host name is bound, Logtail obtains the IP address of the first NIC of the local computer.

  View the IP addresses on the server in the following paths:

  - For Linux: */usr/local/ilogtail/app_info.json*

  If the IP addresses in the machine group on the server are inconsistent with those obtained by the Logtail client, make the following changes as needed:

  - If incorrect IP addresses of the machine group are entered on the server, modify and save the correct IP addresses of the machine group. Then check the IP addresses again one minute later.
  - If the network configuration (for example, */etc/hosts*) of the machines is modified, restart Logtail to obtain new IP addresses.

  Run the following command to restart Logtail if necessary:

  - For Linux: `sudo /etc/init.d/ilogtaild stop; sudo /etc/init.d/ilogtaild start`

3. Check whether the collection configuration is created and applied to the machine group

   After you confirm that the Logtail client status is normal, check the following configurations:

   a) Check whether the Logtail configuration is created

      Check that the log monitoring directory and log file name match those on the machines. The directory structure supports both the complete path mode and the wildcard mode.

   b) Check whether Logtail configuration is applied to the machine group

Check the Logtail machine group by choosing **Config** and check that the target configuration is applied to the machine group.

4. Check for collection errors

   If Logtail is properly configured, check whether new data is generated in real time in the log file. As Logtail collects incremental data only, it does not read inventory files if the files are not updated. If the log file is updated but the updates cannot be queried in the Log Service, troubleshoot the problem following the steps below:

   - **Collection error diagnostic**

     View *Query error diagnostics*, and fix these errors based on the error types reported by Logtail.

   - **View Logtail logs**

     Client logs include key information and all warning and error logs. To query complete and real-time errors, view client logs in the following paths:

     — Linux: `/usr/local/ilogtail/ilogtail.LOG`

   - **Usage exceeds the limit**

     To collect large volumes of logs, files, or data, you can modify the Logtail startup parameters for higher log collection throughput. For more information about how to make adjustments, see *Configure startup parameters*.

   If the problem persists, submit a ticket to Log Service engineers and attach key information collected during troubleshooting.

24.7.6 Limits

This document describes restrictions on Logtail, including restrictions on file collection, resources, and error handling.

Table 24-3: File collection restriction

Category	Limits
File encoding	UTF8/GBK encoding of log files is supported , and UTF8 encoding is recommended for better processing performance. If log files are encoded in other encoding formats, errors such as garbled characters and data losses occur.
Log file size	Not limited.

Category	Limits
Log file rotation	Supported. Files named as `.log*` or `.log` are supported.
Log collection behavior upon log parsing block	When log parsing is blocked, Logtail retains the open state of the log file descriptor (FD). If log file rotation occurs multiple times during the block, Logtail attempts to keep the log parsing sequence of each rotation. If more than 20 unparsed logs are rotated, Logtail does not process subsequent log files. For more information, see *Related technical documents*.
Soft link	Monitored directories can be soft links.
Single log size	The maximum size of a single log is 512 KB. If multiple-line logs are divided by the regular expression at the beginning of the line, the maximum size of each log is still 512 KB. If the size of a log exceeds 512 KB, the log is forcibly split into multiple parts for collection. For example, if the size of a log is 1025 KB, the first 512 KB is processed, then the 512 KB in the middle is processed, and lastly the 1 KB in the end is processed.
Regular expression	Regular expressions can be Perl-compatible regular expressions.
Multiple collection configurations for the same file	Not supported. You are advised to collect log files in a Logstore and configure multiple subscriptions. If this function is required, configure soft links for log files to bypass the restriction.
File opening behavior	Logtail retains the open state of a file to be collected. Logtail closes the file if the file does not have any modification within five minutes (in case that rotation does not occur).
First log collection behavior	Logtail collects only incremental log files. If modifications are found in a file for the first time and the file size exceeds 1 MB, Logtail collects the logs from the last 1 MB. Otherwise, Logtail collects the logs from the beginning. If a log file is not modified after the configuration is issued , Logtail does not collect this file.

Category	Limits
Non-standard text log	If a log contains \0 lines, the log is truncated at the position of the first \0 line.

Table 24-4: Checkpoint management

Item	Capabilities and limits
Checkpoint timeout interval	If a file has not been modified for more than 30 days, the checkpoint is deleted.
Checkpoint save policy	Regular save every 15 minutes. Files are automatically saved when the program exits.
Checkpoint save path	The default save path is `/tmp/logtail_checkpoint`. To modify this path, see *Configure startup parameters*.

Table 24-5: Limits on configuration

Item	Capabilities and limits
Configuration update	Updated configuration takes effect with a delay of about 30s.
Dynamic configuration loading	Supported. The configuration update does not affect other collections.
Number of configured tasks	Theoretically unlimited. We recommend that the number of collection configurations on a server is no more than 100.
Multi-tenant isolation	Collection configurations for different tenants are isolated.

Table 24-6: Limits on resources and performance

Item	Capabilities and limits
Log processing throughput	The default traffic of raw logs is limited to 2 MB/s. (Data is uploaded after it is encoded and compressed, with a general compression ratio of 5 to 10 times.) Logs may be lost if the traffic limit is exceeded. To adjust the parameter, see *Configure startup parameters*.
Maximum performance	Single-core capability: The maximum processing capability is 100 MB/s for logs in simple mode, 20 MB/s by default for logs in regular mode (depending on the complexity of regular expressions), 40 MB/s for logs in delimiter mode, and 30 MB/s for logs in JSON mode. After multiple

Item	Capabilities and limits
	processing threads are enabled, the performance can be improved by 1.5 to 3 times.
Number of monitored directories	Logtail actively restricts the number of monitored directories to avoid excessive consumption of user resources. When the monitoring upper limit is reached, Logtail stops monitoring more directories and log files. Logtail monitors a maximum of 3,000 directories, including subdirectories.
Default resource restriction	By default, Logtail occupies up to 40% of CPU and 256 MB of memory. If logs are generated at a high speed, you can modify the parameters by referring to *Configure startup parameters*.
Resource out-of-limit processing policy	If the resources occupied by Logtail in 3 minutes exceed the upper limit, Logtail is forced to restart, which may cause loss or duplication of data.

Table 24-7: Limits on error handling

Item	Capabilities and limits
Network error handling	If the network connection is abnormal, Logtail actively retries and automatically adjusts the retry interval.
Handling of resource quota exceeding	If the data transmission rate exceeds the maximum quota of Logstore, Logtail blocks log collection and automatically retries. *Related technical documents*.
Maximum retry period for timeout	If data transmission fails for more than 6 successive hours, Logtail discards the data.
Status self-check	Logtail automatically restarts in the case of an exception, for example, abnormal exit of a program or resource limit exceeding.

Table 24-8: Other limits

Item	Capabilities and limits
Log collection delay	Except for block status, the delay in log collection by Logtail does not exceed one second after logs are flushed to a disk.
Log upload policy	Logtail automatically aggregates logs in the same file before uploading the logs. Log uploading is triggered if the number of logs exceeds 2,000, the total size of the log file exceeds 2 MB, or the log collection duration exceeds 3s.

24.8 Other collection methods

24.8.1 Use LogStash to collect logs

Log Service supports collection of server logs through LogStash and upload of data to Log Service through a plug-in.

At present, Log Service supports collection of logs through APIs, SDKs and LogStash. As an open source log management tool, LogStash can quickly collect and define distributed and diversified logs, and transmit them to a specified location, for example, a server or file. You can install LogStash and plug-ins on ECS instances and IDC machines or virtual machines of other cloud vendors and perform a simple configuration to easily migrate server logs to the cloud.

After you install LogStash and related plug-ins on the machines, and configure file directories and Log Service projects or Logstores, LogStash automatically traces changes of log files, collects log files in real time, parses the log files, and sends them to Log Service.

Log Service supports entry of data through LogStash. It provides the following features:

- Collect logs of various types on machines and support data sources such as files, TCP, and Syslog.
- Access the account security system and support data signature transmission and access permission control through secret key pairs.
- Support batch transmission of logs to reduce TPS costs arising out of entry into Log Service.
- Compress log data and enter the data to Log Service to reduce the occupied network egress bandwidth.

24.8.1.1 Logstash overview

Log Service supports collection of server logs through Logstash and upload of data to Log Service through a plug-in.

At present, Log Service supports collection of logs through APIs, SDKs and Logstash. As an open source log management tool, Logstash can quickly collect and define distributed and diversified logs, and transmit them to a specified location, for example, a server or file. You can install Logstash and plug-ins on ECS instances and IDC machines or virtual machines of other cloud vendors and perform a simple configuration to easily migrate server logs to the cloud.

After you install Logstash and related plug-ins on the machines, and configure file directories and Log Service projects or Logstores, Logstash automatically traces changes of log files, collects log files in real time, parses the log files, and sends them to Log Service.

Log Service supports data writing through Logstash. It provides the following features:

- Collect logs of various types on machines and support data sources such as files, TCP, and Syslog.
- Access the account security system and support data signature transmission and access permission control through private key pairs.
- Support batch transmission of logs to reduce TPS costs arising out of writing to Log Service.
- Compress log data and write the data to Log Service to reduce the occupied network egress bandwidth.

24.8.1.2 Quick installation

You can select a default mode to quickly install LogStash on your server.

Context

Log Service provides an installation package based on LogStash 2.2.2, which integrates JRE 1.8, Log Service write plug-in, and NSSM 2.24. The deployment procedure using this package is simpler than *Custom Installation*. If you have complex requirements, you can select custom installation.

Procedure

1. Download the *installation package* and decompress it on drive C.
2. Check that the program start path of LogStash is `C:\logstash-2.2.2-win\bin\logstash.bat`.

24.8.1.3 Custom installation

You can install LogStash by custom and configure installation items as required during installation.

Context

When you have other requirements for installation and configuration of LogStash, you can select custom installation to modify the default installation and configuration.

Procedure

1. Install Java.
   a. Download the installation package.

      Go to *Java website*, download the JDK, and double-click the JDK to install it.
   b. Set environment variables.

      Add or modify environment variables in advanced system settings.

- **PATH:** *C:\Program Files\Java\jdk1.8.0_73\bin*
- **CLASSPATH:** *C:\Program Files\Java\jdk1.8.0_73\lib;C:\Program Files\Java\jdk1.8.0_73\lib\tools.jar*
- **JAVA_HOME:** *C:\Program Files\Java\jdk1.8.0_73*

c. Perform verification.

Run `PowerShell` or `cmd.exe` for verification.

```
PS C:\Users\Administrator> java -version
java version "1.8.0_73"
Java(TM) SE Runtime Environment (build 1.8.0_73-b02)
Java HotSpot(TM) 64-Bit Server VM (build 25.73-b02, mixed mode)
PS C:\Users\Administrator> javac -version
javac 1.8.0_73
```

2. Install LogStash.

a. Download the installation package.

Download LogStash from the official website: *On the LogStash homepage*, select LogStash 2.2 or a later version.

b. Install LogStash.

Decompress *logstash-2.2.2.zip* to the *C:\logstash-2.2.2* directory.

Check that the program start path of LogStash is *C:\logstash-2.2.2\bin\logstash.bat*.

3. Install the plug-in used by LogStash to write logs to Log Service.

Install the plug-in online or offline based on the network environment of the machine.

- Online installation.

  The plug-in is hosted on RubyGems. For more information, click *https://rubygems.org/gems/logstash-output-logservice*.

  Run `PowerShell` or `cmd.exe` to go to the LogStash installation directory. Run the following command to install LogStash:

  ```
  PS C:\logstash-2.2.2> .\bin\plugin install logstash-output-logservice
  ```

- Offline installation.

  Download LogStash from the official website: Go to the *logstash-output-logservice page* and click **Download** in the lower right corner.

If you cannot access the Internet on the machine on which logs are collected, copy the downloaded gem package to the *C:\logstash-2.2.2* directory on the machine. Run *PowerShell* or *cmd.exe* to go to the LogStash installation directory. Run the following command to install LogStash:

```
PS C:\logstash-2.2.2> .\bin\plugin install C:\logstash-2.2.2\
logstash-output-logservice-0.2.0.gem
```

- Perform verification.

```
PS C:\logstash-2.2.2> .\bin\plugin list
```

LogStash can be found in the list of plug-ins installed on the local computer. logstash-output -logservice.

4. Install NSSM.

Download NSSM from the official website: Go to *NSSM official website* and download NSSM.

After you download the installation package to the local computer, decompress it to *C:\logstash-2.2.2\nssm-2.24*.

24.8.1.4 Set LogStash to a Windows service

For convenient automatic log collection, you can set LogStash to a Windows service to maintain running of LogStash in the background and automatic start upon startup of the computer.

Context

Start *logstash.bat* under PowerShell. The LogStash process works in the foreground. It is generally used for configuration test and collection commissioning. You are advised to set LogStash to a Windows service to maintain running of LogStash in the background and automatic start upon startup of the computer.

In addition, you can start, stop, modify, and delete a service through the command line. For more information about how to use NSSM, see *NSSM official documents*.

Add the LogStash service

This operation is generally performed when LogStash is deployed the first time. Skip this operation if LogStash is added.

Run the following commands to add the LogStash service:

- 32-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe install logstash "C:\
logstash-2.2.2-win\bin\logstash.bat" "agent -f C:\logstash-2.2.2-win
\conf"
```

- 64-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe install logstash "C:\
logstash-2.2.2-win\bin\logstash.bat" "agent -f C:\logstash-2.2.2-win
\conf"
```

Start the LogStash service

If the configuration file is updated in the *conf* directory of LogStash, stop the LogStash service first. Then start the LogStash service.

Run the following commands to start the LogStash service:

- 32-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe start logstash
```

- 64-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe start logstash
```

Stop the LogStash service

Run the following commands to stop the LogStash service:

- 32-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe stop logstash
```

- 64-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe stop logstash
```

Modify the LogStash service

Run the following commands to modify the LogStash service:

- 32-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe edit logstash
```

- 64-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe edit logstash
```

Delete the LogStash service

Run the following commands to delete the LogStash service:

- 32-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe remove logstash
```

- 64-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe remove logstash
```

24.8.1.5 Create a LogStash collection configuration

Log Service enables you to collect server log data through Logstash, and create Logstash collection configurations for you data sources. After completing the configurations, Logstash pushes logs to Log Service in real time.

Related plug-ins

- **logstash-input-file**

  The plug-in is used to collect log files in tail mode. For more information, see *logstash-input-file*.

  Note:
  path indicates the file path, which must use Unix separators, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.

- **logstash-output-logservice**

  You can use the plug-in to collect logs to Log Service.

Parameter	Description
endpoint	Entry of Log Service, for example, *http://regionid.example.com*.
project	Name of a Log Service project.
logstore	Logstore name.
topic	Log topic name. By default, the log topic is set to null.

Parameter	Description
source	Log source. If this parameter is set to null, the IP address of the local computer is automatically used. Otherwise, the set value prevails.
access_key_id	Account secret key ID.
access_key_secret	Account secret key.
max_send_retry	Maximum number of retries when packets cannot be transmitted to Log Service due to an exception. If retry fails, packets are discarded. The retry interval is 200 ms.

Procedure

1. Create a collection configuration.

   After you add a configuration file to the `C:\logstash-2.2.2-win\conf\` directory, restart LogStash to make the configuration file take effect.

   You can create a configuration file in the format `*.conf` for each type of logs. You are advised to save the configuration file in the `C:\logstash-2.2.2-win\conf\` directory for convenient management.

   Note:
   The configuration file must be UTF-8 encoded without BOM. You can modify the file encoding format by using Notepad++.

   - IIS log

     See *Use LogStash to collect IIS logs*.
   - CSV log

     The system time when logs are collected is used as the log upload time. For details, see *Use LogStash to collect CSV logs*.
   - Default log time

     Take the format of CSV logs as an example. The time in log content is used as the log upload time. For details, see *Use LogStash to collect CSV logs*.
   - General log

     By default, the system time when logs are collected is used as the uploaded log time. Log fields are not parsed. Single-line logs and multiline logs are supported. For details, see *Use LogStash to collect other logs*.

2. Verify configuration syntax.

   a. Run `PowerShell` or `cmd.exe` to go to the LogStash installation directory. Run the following command to verify the configuration:

   ```
   PS C:\logstash-2.2.2-win\bin> .\logstash.bat agent --configtest --
   config C:\logstash-2.2.2-win\conf\iis_log.conf
   ```

   b. Modify the collection configuration file. Add the temporary configuration item rubydebug in the output phase to output collected results to the console. During configuration, set the type field.

   ```
   output { if [type] == "***" { stdout { codec => rubydebug }
   logservice { ... } } }
   ```

   c. Start PowerShell or cmd.exe, change the current directory to the installation directory of Logstash, and run `PowerShell` or `cmd.exe`. Run the following command:

   ```
   PS C:\logstash-2.2.2-win\bin> .\logstash.bat agent -f C:\logstash-
   2.2.2-win\conf
   ```

   After verification is complete, end the *`logstash.bat`* process and delete the temporary configuration item rubydebug.

What's next

Start *`logstash.bat`* under PowerShell. The Logstash process works in the foreground. It is generally used for configuration test and collection commissioning. You are advised to set LogStash to a Windows service so as to maintain running of LogStash in the background and automatic start upon startup of the computer. For more information about how to set LogStash to a Windows service, see *Set LogStash to a Windows service*.

24.8.1.6 Advanced functions

Log Service supports collection of server logs through LogStash and upload of data to Log Service through a plug-in.

LogStash provides *https://www.elastic.co/guide/en/logstash/current/index.html* to meet personalized requirements, for example:

- *grok*: Parse log content into multiple fields through the structure of regular expression.
- *json line*, *json*: Support structured parsing of JSON logs.
- *date*: Support parsing and conversion of fields related to date and time.
- *multiline*: Define more complex multi-line logs.
- *kv*: Support structured parsing of Key-Value logs.

24.8.1.7 LogStash error handling

After LogStash is configured to collect logs, if an error occurs during log collection, select a handling method based on the error type.

If the following collection errors occur when LogStash collects logs, handle the errors according to corresponding suggestions:

- Garbled characters are found in Log Service.

  By default, LogStash supports UTF-8 encoding. Check whether input files are correctly encoded.

- Errors displayed on the console

  When the following error is displayed on the console: `io/console not supported; tty will not be manipulated`, if product functions are not affected, ignore the error.

  For other errors, you are advised to refer to Google or LogStash forum for help.

24.8.2 SDK collection

24.8.2.1 Producer Library

LogHub Producer Library is a write LogHub class library for Java applications with high concurrency. Both Producer Library and Consumer Library package LogHub read and write requests to lower the threshold for data collection and consumption.

Features

- Provides an asynchronous sending interface, thus ensuring thread security.
- Adds configurations for multiple projects.
- Configures the number of network IO threads used for sending.
- Configures the number and size of logs merged into a package.
- Controllable memory usage. In other words, when the memory usage reaches the threshold you configured, the send interface of Producer will be blocked until idle memory is available.

Benefits

- Logs collected by agents not flushed into a disk: Data is directly sent to the server through the network after the data is generated.
- High-concurrency write operations on the agent: For example, there are more than one hundred write operations within one second.

- Agent computing logically separated from I/O: Logging does not affect the computing time used .

In the above scenarios, Producer Library helps you reduce program development costs, aggregate multiple write requests, and asynchronously send them to the LogHub server. During the process, you can configure parameters for batch aggregation, the server exception processing logic, and so on.

Comparison of the above access methods:

Access method	Advantage/Disadvantage	Target scenario
SDK direct transmission	Logs are not flushed into a disk, but are directly sent to the server. Switching between the network I/O and program I/O needs to be properly processed.	Logs are not flushed into a disk.
Producer Library	Logs are not flushed into a disk, but are combined and asynchronously sent to the server at a large throughput.	Logs are not flushed into a disk. The QPS is high on the agent.

Procedure

- *Java Producer*
- *Log4J1.XAppender (based on Java Producer)*
- *Log4J2.XAppender (based on Java Producer)*
- *LogBack Appender (based on Java Producer)*
- *C Producer*
- *C Producer Lite*

24.8.2.2 Log4j Appender

Alibaba Cloud Log Log4j Appender enables you to set the log output destination to Log Service.

Loghub Log4j Appender

Apache Log4j is an open source project which allows you to set the log output destination to the console, file, GUI component, socket server, NT event recorder, and Unix Syslog daemon. You can set the output format and level of each log to control log generation at a smaller granularity. Configurations can be performed using a configuration file without the need to modify the code of applications.

For the download address and usage, see *Github*.

24.8.2.3 C Producer Library

LogHub not only supports the Java version of Producer Library, but also supports the C version of Producer Library and Producer Lite Library, providing you with a concise, high-performance, and low-resource-consumption one-stop log collection solution across platforms.

For the project address on GitHub and more details, see

- *C Producer Library (recommended for the server)*
- *C Producer Lite Library (recommended for IoT and smart devices)*

24.8.3 Common log formats

Common log formats are listed to provide reference for you to configure Logtail to collect these logs.

24.8.3.1 Overview

Log Service lists common log formats to provide you with a reference for configuring log collection through Logtail.

24.8.3.2 Apache logs

The Apache log format and directory are specified in the `/etc/apache2/httpd.conf` configuration file.

Log format

The "combined" and "common" log formats are defined in the Apache log configuration file.

- Combined format:

```
LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}
i\"" combined
```

- Common format:

```
LogFormat "%h %l %u %t \"%r\" %>s %b"
```

The following statement writes logs into the specified file in the "combined" format.

```
CustomLog "/var/log/apache2/access_log" combined
```

Field description

Field format	Meaning
%a	remote_ip

Field format	Meaning
%A	local_ip
%B	size
%b	size
%D	time_taken_ms
%h	remote_host
%H	protocol
%I	ident
%m	method
%p	port
%P	pid
“%q”	url_query
“%r”	request
%s	status
%>s	status
%t	time
%T	time_taken
%u	remote_user
%U	url_stem
%v	server_name
%V	canonical_name
%I	bytes_received
%O	bytes_sent
“%{User-Agent}i”	user_agent
“%{Referer}i”	referer

Sample log

```
192.168.1.2 - - [02/Feb/2016:17:44:13 +0800] "GET /favicon.ico HTTP/1.
1" 404 209 "http://localhost/x1.html" "Mozilla/5.0 (Macintosh; Intel
```

```
Mac OS X 10_11_3) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/48.0.
2564.97 Safari/537.36"
```

Configure Logtail to collect Apache logs.

1. Create a project and a Logstore. For detailed instructions, see *Create a Project* and *Create a Logstore*.
2. On the **Logstores** page, click the **Data Import Wizard** icon to access the wizard.
3. Select a data type.

   Select **Text File** and click **Next**.
4. Configure the data source.

   a. Enter the configuration name and log path, and set the log collection mode to **Full Mode**.

   b. Enter a log sample and enable **Extract Field**.

   c. Select fields to generate a regular expression, and manually adjust it.

      Log Service can automatically parse the selected sample log. That is, when you select fields, it can automatically generate a regular expression. There may be possible minor changes in the log data format, so you need to click **Manually Input Regular Expression** to adjust the automatically generated regular expression. This makes it suitable for all the log formats that may be encountered during collection.

      After modifying the regular expression, click **Validate**. Extracted results are displayed if the regular expression is correct.

   d. Enter keys corresponding to the log extraction results.

      Choose a descriptive field name for each extraction result. For example, choose "time" for the time field. Enable **Use System Time** and click **Next**.

   After configuring Logtail, apply the configuration to the machine group to collect Apache logs.

24.8.3.3 Nginx logs

The Nginx log format and directory are specified in the */etc/nginx/nginx.conf* configuration file.

Nginx log format

The configuration file defines the print format of Nginx logs, namely, the main format:

```
log_format main $remote_addr - $remote_user [$time_local] "$request"
 $request_time $request_length  $status $body_bytes_sent "$http_refer
er"  "$http_user_agent";
```

The statement uses the "main" log format and the written file name.

```
access_log /var/logs/nginx/access.log main
```

Field description

Field name	Meaning
remoteaddr	IP address of the agent.
remote_user	User name of the agent.
request	Request URL and HTTP protocol.
status	Request status.
bodybytessent	Number of bytes (not including the size of the response header) sent to the agent. The number of bytes indicated by this variable is the same as that indicated by bytes_sent in modlogconfig of the Apache module.
connection	Serial number of a connection.
connection_requests	Number of requests obtained by one connection
msec	Time when the log is written, which is measured in seconds and accurate to milliseconds
pipe	Whether requests are sent pipelined over HTTP. If yes, the pipe value is p; otherwise, the value is a period (.).
httpreferer	Source page of the access request.
“http_user_agent”	Browser information of the agent, which should be enclosed by double quotation marks.

Field name	Meaning
requestlength	Request length, which includes the request line , request header, and request body.
request_time	Time when the request is processed, which is measured in seconds and accurate to milliseconds. The time starts when the first byte is read from the agent until logs are written after the last character is sent to the agent.
[$time_local]	Local time when the general log format is applied, which must be enclosed by braces.

Sample log

```
192.168.1.2 - - [10/Jul/2015:15:51:09 +0800] "GET /ubuntu.iso HTTP/1.0
" 0.000 129 404 168 "-" "Wget/1.11.4 Red Hat modified"
```

Configure Logtail to collect Nginx logs

1. Create a project and a Logstore. For detailed instructions, see *Create a Project* and *Create a Logstore*.
2. On the **Logstores** page, click **Data Import Wizard** to access the wizard.
3. Select a data type.

   Select **Text File** and click **Next**.
4. Select a data type.

   Select **NGINX Access Log** and click **Next**.
5. Configure the data source.

   a. Enter **Configuration Name** and **Logs Directory Path** .

   b. Enter the Nginx log format.

      Enter the log configuration part of the standard Nginx configuration file. It typically starts with `log_format`. Log Service automatically reads the Nginx key.

   c. Configure **Advanced Options** as required and click **Next**.

      For the description of advanced options, see *Advanced options*.

After configuring Logtail, apply the configuration to the machine group to collect Nginx logs.

24.8.3.4 Python log

Python logging module provides a general logging system for use by third-party modules or applications. The logging module provides different log levels and records logs by different methods including file, HTTP GET/POST, SMTP, and Socket. You can customize a log recording method as required. The logging module has the same mechanism as Log4j except that they have different implementation details. The logging module provides the logger, handler, filter, and formatter features.

Python log format

The formatter specifies the log output format. The formatter constructor requires two parameters for construction: message format string and message date string. The parameters are optional.

Python log format:

```
import logging import logging.handlers LOG_FILE = tst.log handler =
logging.handlers.RotatingFileHandler(LOG_FILE, maxBytes = 1024*1024,
backupCount = 5) # Instantiate handler fmt = %(asctime)s - %(filename
)s:%(lineno)s - %(name)s - %(message)s formatter = logging.Formatter
(fmt) # Instantiate formatter handler.setFormatter(formatter) # Add
formatter logger = logging.getLogger(tst) for handler # Obtain logger
logger.addHandler(handler) with name tst # Add handler logger.setLevel
(logging.DEBUG) logger.info(first info message) logger.debug(first
debug message) for logger
```

Field description

The formatter is configured in the `%(key)s` format, that is, substitution of dictionary keywords. The following keywords are provided:

Format	Meaning
%(name)s	Name of the logger that generates logs.
%(levelno)s	Numerical log levels, including debug, info, warning, error, and critical.
%(levelname)s	Text log levels, including ‘debug’, ‘info’, ‘ warning’, ‘eerror’, and ‘critical’.
%(pathname)s	Complete path (if available) to the source file that contains the statement for log output.
%(filename)s	File name.
%(module)s	Name of the module that contains the statement for log output.
%(funcName)s	Name of the function that calls the log output function.

Format	Meaning
%(lineno)d	Line of the code (if available) that contains the statement for calling the log output function.
%(created)f	Log creation time, in the Unix time format, expressed by the number of seconds passed since 1970-1-1 00:00:00 UTC.
%(relativeCreated)d	Difference (in milliseconds) between the log creation time and the loading time of the logging module.
%(asctime)s	Log creation time. The default format is like 2003-07-08 16:49:45,896. The number following the comma (,) indicates the number of milliseconds.
%(msecs)d	Log creation time, in milliseconds.
%(thread)d	Thread ID (if available).
%(threadName)s	Thread name (if available).
%(process)d	Process ID (if available).
%(message)s	Log information,

Sample log

Output sample log:

```
2015-03-04 23:21:59,682 - log_test.py:16 - tst - first info message
2015-03-04 23:21:59,682 - log_test.py:17 - tst - first debug message
```

Configure Logtail to collect Python logs

For more information about how to configure Logtail to collect Python logs, see *Apache logs*. Set configuration options based on your network deployment and the actual situation.

The automatically generated regular expression is based on the sample log but does not cover every log type. Therefore, you need to tune the regular expression after it is generated.

Common Python logs and their regular expressions:

- Sample log:

```
2016-02-19 11:03:13,410 - test.py:19 - tst - first debug message
```

Regular expression:

```
(\d+-\d+-\d+\s\S+)\s+-\s+([^:]+):(\d+)\s+-\s+(\w+)\s+-\s+(.*)
```

- Log format:

```
%(asctime)s - %(filename)s:%(lineno)s - %(levelno)s %(levelname)
s %(pathname)s %(module)s %(funcName)s %(created)f %(thread)d %(
threadName)s %(process)d %(name)s - %(message)s
```

Sample log:

```
2016-02-19 11:06:52,514 - test.py:19 - 10 DEBUG test.py test <module
> 1455851212.514271 139865996687072 MainThread 20193 tst - first
debug message
```

Regular expression:

```
(\d+-\d+-\d+\s\S+)\s-\s([^:]+):(\d+)\s+-\s+(\d+)\s+(\w+)\s+(\S+)\s
+(\w+)\s+(\S+)\s+(\S+)\s+(\d+)\s+(\w+)\s+(\d+)\s+(\w+)\s+-\s+(.*)
```

24.8.3.5 Log4j log

Log4j logs include Log4j 1 logs and Log4j 2 logs. This document describes how to configure a regular expression for collecting Log4j 1 logs based on the default configurations.

Access method

Log Service supports the following methods to collect Log4j logs:

- LogHub Log4j Appender
- Logtail

Collect Log4j logs by using Loghub Log4j Appender

For more information, see *Log4j Appender* .

Collect Log4j logs by using Logtail

Log4j logs include Log4j 1 logs and Log4j 2 logs. This document describes how to configure a regular expression for collecting Log4j 1 logs based on the default configurations. To collect Log4j 2 logs, you need to modify the default configurations and print the complete date.

```
<Configuration status="WARN"> <Appenders> <Console name="Console"
target="SYSTEM_OUT"> <PatternLayout pattern="%d{yyyy-MM-dd HH:mm:ss:
SSS zzz} [%t] %-5level %logger{36} - %msg%n"/> </Console> </Appenders
> <Loggers> <Logger name="com.foo.Bar" level="trace"> <AppenderRef ref
```

```
="Console"/> </Logger> <Root level="error"> <AppenderRef ref="Console
"/> </Root> </Loggers> </Configuration>
```

For more information about how to configure Logtail to collect Log4j logs, see *Apache logs*. Set configuration options based on your network deployment and the actual situation.

The automatically generated regular expression is based on the sample log but does not cover every log type. Therefore, you need to tune the regular expression after it is generated.

The following shows the sample log in the default format of Log4j that is printed to a file:

```
2013-12-25 19:57:06,954 [10.207.37.161] WARN impl.PermanentTairDaoImpl
 - Fail to Read Permanent Tair,key:e:470217319319741_1,result:com
.example.tair.Result@172e3ebc[rc=code=-1, msg=connection error or
timeout,value=,flag=0]
```

Match the starting line of a multiline log (the beginning of a line is expressed by IP address information):

```
\d+-\d+-\d+\s.*
```

Regular expression used to extract log information:

```
(\d+-\d+-\d+\s\d+:\d+:\d+,\d+)\s\[([^\]]*)\]\s(\S+)\s+(\S+)\s-\s(.*)
```

Time conversion format:

```
%Y-%m-%d %H:%M:%S
```

Sample log extraction result:

Key	Value
time	2013-12-25 19:57:06,954
ip	10.207.37.161
level	WARN
class	impl.PermanentTairDaoImpl
message	Fail to Read Permanent Tair,key:e:470217319319741_1,result:com.example.tair.Result@172e3ebc[rc=code=-1, msg=connection error or timeout,value=,flag=0]

24.8.3.6 Node.js log

By default, Node.js logs are printed to the console, which makes data collection and troubleshooting inconvenient. The Log4js package is provided to print logs to files and customize log formats, which makes data collection and sorting easy.

```
var log4js = require(log4js); log4js.configure({ appenders: [ { type
: file, //Output filename: logs/access.log, maxLogSize: 1024, backups
:3, category: normal } ] }); var logger = log4js.getLogger(normal);
logger.setLevel(INFO); logger.info("this is a info msg"); logger.error
("this is a err msg");
```

Log format

The Log4js log that is output to a text file is shown below:

```
[2016-02-24 17:42:38.946] [INFO] normal - this is a info msg [2016-02-
24 17:42:38.951] [ERROR] normal - this is a err msg
```

Log4js has six output levels: trace, debug, info, warn, error, and fatal, in ascending order of severity.

Use Logtail to collect Node.js logs

For more information about how to configure Logtail to collect Node.js logs, see *Apache logs*. Set configuration options based on your network deployment and the actual situation.

The automatically generated regular expression is based on the sample log but does not cover every log type. Therefore, you need to tune the regular expression after it is generated. You can refer to the following Node.js sample logs to write a correct and complete regular expression for logs.

Common Node.js logs and their regular expressions:

- Node.js sample log 1
  - Sample log:

    ```
    [2016-02-24 17:42:38.946] [INFO] normal - this is a info msg
    ```

  - Regular expression:

    ```
    \[([^]]+)]\s\[([^\]]+)]\s(\w+)\s-(.*)
    ```

  - Extracted fields:

    `time`, `level`, `loggerName`, and `message`

- Node.js sample log 2:

- Sample log:

```
[2016-01-31 12:02:25.844] [INFO] access - 42.120.73.203 - - "GET
 /user/projects/ali_sls_log?ignoreError=true HTTP/1.1" 304 - "http
:// aliyun.com/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_3)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/48.0.2564.97 Safari/
537.36"
```

- Regular expression:

```
\[([^]]+)]\s\[(\w+)]\s(\w+)\s-\s(\S+)\s-\s-\s"([^"]+)"\s(\d+)
[^"]+("[^"]+)"\s"([^"]+).*
```

- Extracted fields:

  `time`, `level`, `loggerName`, `ip`、`request`, `status`, `referer`, and `user_agent`

24.8.3.7 WordPress log

Default WordPress log format

Raw sample log:

```
172.64.0.2 - - [07/Jan/2016:21:06:39 +0800] "GET /wp-admin/js/password
-strength-meter.min.js?ver=4.4 HTTP/1.0" 200 776 "http://wordpress
.c4a1a0aecdb1943169555231dcc4adfb7.cn-hangzhou.alicontainer.com/wp-
admin/install.php" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537
.36"
```

Match the starting line of a multiline log (the beginning of a line is expressed by IP information):

```
\d+\.\d+\.\d+\.\d+\s-\s.*
```

Regular expression used to extract log information:

```
(\S+) - - \[([^\]]*)] "(\S+) ([^"]+)" (\S+) (\S+) "([^"]+)" "([^"]+)"
```

Time conversion format:

```
%d/%b/%Y:%H:%M:%S
```

Sample log extraction result:

Key	Value
ip	10.10.10.1
time	07/Jan/2016:21:06:39 +0800
method	GET

Key	Value
url	/wp-admin/js/password-strength-meter.min.js?ver=4.4 HTTP/1.0
status	200
length	776
ref	http://wordpress.c4a1a0aecdb1943169555231dcc4adfb7.cn-hangzhou.alicontainer.com/wp-admin/install.php
user-agent	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537.36

24.8.3.8 Delimiter log

This section describes the default format of delimiter logs to provide you with a reference for configuring delimiter log collection through Logtail.

Log introduction

Delimiter logs use line breaks as the boundary. Each line is a log. The fields of each log are connected by fixed delimiters, including tabs, spaces, vertical lines (|), commas (,), semicolons (;), and other single characters. Fields containing delimiters are enclosed in double quotation marks (").

Common delimiter logs include CSV and TSV formatted logs.

Log formats

A delimiter log is divided into several fields by delimiters, and supports two modes: **single character** and **multiple characters**.

- **Single character mode**

  Single character mode divides logs by matching single characters, such as tabs (`\t`), spaces, vertical lines (`|`), commas (`,`), and semicolons (`;`).

  > **Note:**
  > The double quotation mark (`"`) cannot be a delimiter, but is used as the quote of the default single character delimiter.

  Single character delimiters are often contained in log fields. To prevent log fields from being divided incorrectly, a double quotation mark (`"`) is used as the quote to isolate the log field. If

double quotation marks (") are found in the log content but not used as the quote, they must be escaped and processed as `""`. You can either use a double quotation mark (`"`) in field border as the quote, or use double quotation marks (`""`) as field data. For other situations, use other modes, such as simple mode and full mode, to parse fields because the situations do not meet the format definition of delimiter logs.

- **Double quotation mark (") used as the quote**

  When the double quotation mark (`"`) is used as the quote, fields containing delimiters must be enclosed in a pair of quotes. The quote must be located adjacent to the delimiter. Modify the format if any spaces, tabs, and other characters exist between them.

  For example, comma (`,`) is the delimiter and double quotation mark (") is the quote. The log format is `1997,Ford,E350,"ac, abs, moon",3000.00`. This log can be parsed into five fields, `1997`, `Ford`, `E350`, `ac, abs, moon`, and `3000.00`. `ac, abs, moon`, which is enclosed in quotes, is considered as a complete field.

- **Double quotation marks (") used as a part of log field**

  When double quotation marks (”) are used as a part of the log field instead of the quote, they must be escaped and processed as `""`. The marks are restored when the fields are being parsed, that is, restoring `""` to `"`.

  For example, when commas are used as delimiters and double quotation marks and commas are a part of a field, enclose the field with a pair of quotes and escape the double quotation marks into `""`. The log format after the processing is as follows: `1999,Chevy,"Venture ""Extended Edition, Very Large""",,5000.00`. This log can be parsed into five fields: `1999`, `Chevy`, `Venture "Extended Edition, Very Large"`, a blank field, and `5000.00`.

- **Multiple character mode**

  In **multiple character mode**, a delimiter can contain two or three characters, such as `||`, `&&&`, `^_^`. In this mode, logs are parsed completely by matching delimiters and you do not need to use the quote to enclose the logs.

  Note:

  Make sure that the full match of the delimiter does not appear in the log field. Otherwise, the field will be divided incorrectly.

For example, if the delimiter is set to `&&`, the log `1997&&Ford&&E350&&ac&abs&moon&&3000.00` can be parsed into five fields: `1997`, `Ford`, `E350`, `ac&abs&moon`, and `3000.00`.

Sample log

- **Log with single-character delimiters**

```
05/May/2016:13:30:28,10.10.10.1,"POST /PutData? Category=YunOsAccou
ntOpLog&AccessKeyId=<yourAccessKeyId>&Date=Fri%2C%2028%20Jun%202013
%2006%3A53%3A30%20GMT&Topic=raw&Signature=<yourSignature> HTTP/1.1",
200,18204,aliyun-sdk-java
05/May/2016:13:31:23,10.10.10.2,"POST /PutData? Category=YunOsAccou
ntOpLog&AccessKeyId=<yourAccessKeyId>&Date=Fri%2C%2028%20Jun%202013
%2006%3A53%3A30%20GMT&Topic=raw&Signature=<yourSignature> HTTP/1.1",
401,23472,aliyun-sdk-java
```

- **Log with multi-character delimiters**

```
05/May/2016:13:30:28&&10.200.98.220&&POST /PutData? Category=
YunOsAccountOpLog&AccessKeyId=<yourAccessKeyId>&Date=Fri%2C%2028%
20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=<yourSignature
> HTTP/1.1&&200&&18204&&aliyun-sdk-java
05/May/2016:13:31:23&&10.200.98.221&&POST /PutData? Category=
YunOsAccountOpLog&AccessKeyId=<yourAccessKeyId>&Date=Fri%2C%2028%
20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=<yourSignature
> HTTP/1.1&&401&&23472&&aliyun-sdk-java
```

Configure Logtail to collect delimiter logs

For the detailed procedure of collecting Python logs through Logtail, see *Apache logs*. Select the corresponding configuration based on your network deployment and the actual situation.

1. Create a project and a Logstore. For the detailed procedures, see *Create a Project* and *Create a Logstore*.
2. Click **Data Import Wizard** on the **Logstores** page.
3. Select a data type.

   Select the **Text File**and click **Next**.
4. Configure a data source.

   a. Enter Configuration Name and Log Path. Then, select **Delimiter Mode** as the log collection mode.

   b. Enter the log sample and select the delimiter.

   Select a proper separator based on the log format. Otherwise, parsing may fail.

   c. Specify the keys in the log extraction result.

After you enter a sample log and select a separator, Log Service extracts fields of the log based on the delimiter and defines the fields as values. You need to specify keys for the values.

The preceding sample log uses commas (`,`) as delimiters and contains six fields. The key values are: time, ip, url, status, latency, and user-agent.

d. Specifies the log time.

You can select to use the system time or a log field (such as the time field, 05/May/2016:13:30:29) as the log time. For how to configure the date format, see *Configure a time format*.

e. After the configuration is applied to the machine group, preview logs in the console to check whether logs are successfully collected.

24.8.3.9 JSON logs

This topic describes the default format of JSON logs to provide you with a reference for configuring Logtail to collect JSON logs.

A JSON log can be written in two types of structures:

- Object: a collection of name-value pairs
- Array: an ordered list of values

Logtail supports JSON logs of the object type. Logtail automatically extracts the keys and values from the first layer of an object as the names and values of fields respectively. The field value can be an object, array, or basic types such as string or number.

Logtail does not support automatically parsing non-object data such as JSON arrays. You need to use regular expressions to extract the fields or use the simple mode to collect logs by line.

Sample log

```
{"url": "POST /PutData? Category=YunOsAccountOpLog&AccessKeyId=<
yourAccessKeyId>&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT
&Topic=raw&Signature=<yourSignature> HTTP/1.1", "ip": "10.200.98.220
", "user-agent": "aliyun-sdk-java", "request": {"status": "200", "
latency": "18204"}, "time": "05/May/2016:13:30:28"}
{"url": "POST /PutData? Category=YunOsAccountOpLog&AccessKeyId=<
yourAccessKeyId>&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT
&Topic=raw&Signature=<yourSignature> HTTP/1.1", "ip": "10.200.98.210
```

```
", "user-agent": "aliyun-sdk-java", "request": {"status": "200", "
latency": "10204"}, "time": "05/May/2016:13:30:29"}
```

Configure Logtail to collect JSON logs

For the complete process of collecting JSON logs by using Logtail, see *Apache logs*. Select an appropriate configuration based on your network deployment and the actual situation.

1. Create a project and a Logstore. For the detailed procedures, see *Create a Project* and *Create a Logstore*.
2. On the **Logstores** page, click the **Data Import Wizard** icon.
3. Select a data type.

   Select **Text File** and click **Next**.
4. Configure a data source.

   a. Enter the configuration name and Log Path, and set log collection mode to **JSON mode**.

   b. Determine whether to use the system time as the log time based on your requirements. You can choose to enable or disable **Use system time**.

   - Enable **Use System Time**

     Enabling this function means to use the time when Log Service collects the log as the log time, instead of extracting the time fields in the log.
   - Disable **Use System Time**

     Disabling this function means to extract the time fields from the log as the log time.

     If you choose to disable the **Use System Time** function, you must define the key of the extracted time field, and the time conversion format. For example, the `time` field (05/May/2016:13:30:29) in an object can be extracted to indicate the log generation time. For details about how to set the date format, see *Configure a time format*.
5. After the configuration is applied to the machine group, preview logs in the console to check whether logs are successfully collected.

24.8.3.10 ThinkPHP log

ThinkPHP is a Web application development framework based on the PHP language.

ThinkPHP log format

The log print format in ThinkPHP is as follows:

Sample log

```
[ 2016-05-11T21:03:05+08:00 ] 10.10.10.1 /index.php
INFO: [ app_init ] --START--
INFO: Run Behavior\BuildLiteBehavior [ RunTime:0.000014s ]
INFO: [ app_init ] --END-- [ RunTime:0.000091s ]
INFO: [ app_begin ] --START--
INFO: Run Behavior\ReadHtmlCacheBehavior [ RunTime:0.000038s ]
INFO: [ app_begin ] --END-- [ RunTime:0.000076s ]
INFO: [ view_parse ] --START--
INFO: Run Behavior\ParseTemplateBehavior [ RunTime:0.000068s ]
INFO: [ view_parse ] --END-- [ RunTime:0.000104s ]
INFO: [ view_filter ] --START--
INFO: Run Behavior\WriteHtmlCacheBehavior [ RunTime:0.000032s ]
INFO: [ view_filter ] --END-- [ RunTime:0.000062s ]
INFO: [ app_end ] --START--
INFO: Run Behavior\ShowPageTraceBehavior [ RunTime:0.000032s ]
INFO: [ app_end ] --END-- [ RunTime:0.000070s ]
ERR: D model class not found for method instantiation
```

The log that is printed using this method is as follows:

Configure Logtail to collect ThinkPHP logs

For how to configure Logtail to collect Python logs, see `<xref href="LogService_user_guide_0047.dita"format="dita"/>`. Select configurations based on your network deployment and the actual situation.

The automatically generated regular expression is based on the sample log but does not cover every log type. Therefore, you need to tune the regular expression after it is generated.

ThinkPHP logs are multiline logs in varying modes, and the following fields can be extracted from these logs: time, IP address of the visitor, accessed URL, and printed message. Because the message mode is not fixed, the message is packaged into a field that contains multiple lines of information.

Parameters for configuring Logtail to collect ThinkPHP logs:

Regular expression at the beginning of the line

```
\[\s\d+-\d+-\w+:\d+:\d+\+\d+:\d+\s.*
```

Regular expression:

```
\[\s(\d+-\d+-\w+:\d+:\d+)[^:]+:\d+\s]\s+(\S+)\s(\S+)\s+(.*)
```

Time expression:

```
%Y-%m-%dT%H:%M:%S
```

24.8.3.11 Use LogStash to collect IIS logs

Before using LogStash to collect IIS logs, modify the configuration file to parse IIS log fields.

Sample log

View IIS log configurations, select the W3C format (default field setting), and save the format to put it into effect.

```
2016-02-25 01:27:04 112.74.74.124 GET /goods/list/0/1.html - 80 - 66.
249.65.102 Mozilla/5.0+(compatible;+Googlebot/2.1;++http://www.google.
com/bot.html) 404 0 2 703
```

Collection configuration

```
input { file { type => "iis_log_1" path => ["C:/inetpub/logs/LogFiles
/W3SVC1/*.log"] start_position => "beginning" } } filter { if [type
] == "iis_log_1" { #ignore log comments if [message] =~ "^#" { drop
 {} } grok { # check that fields match your IIS log settings match =>
 ["message", "%{TIMESTAMP_ISO8601:log_timestamp} %{IPORHOST:site} %{
WORD:method} %{URIPATH:page} %{NOTSPACE:querystring} %{NUMBER:port
} %{NOTSPACE:username} %{IPORHOST:clienthost} %{NOTSPACE:useragent}
 %{NUMBER:response} %{NUMBER:subresponse} %{NUMBER:scstatus} %{NUMBER
:time_taken}"] } date { match => [ "log_timestamp", "YYYY-MM-dd HH
:mm:ss" ] timezone => "Etc/UTC" } useragent { source=> "useragent
" prefix=> "browser" } mutate { remove_field => [ "log_timestamp
"] } } } output { if [type] == "iis_log_1" { logservice { codec => "
json" endpoint => "***" project => "***" logstore => "***" topic =>
 "" source => "" access_key_id => "***" access_key_secret => "***"
max_send_retry => 10 } } }
```

Note:

- The configuration file must be UTF-8 encoded without BOM. You can modify the file encoding format by using Notepad++.
- *path* indicates the file path, which must use Unix separators, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.

- The *type* field must be modified and saved in the file. If multiple Logstash configuration files exist on one computer, ensure that the setting of type is unique in the configuration files. Otherwise, data may not be properly processed.

Related plug-ins: *file* and *grok*.

Restart LogStash to make the configuration take effect.

Create a configuration file in the *conf* directory. Restart LogStash to make the setting take effect. For more information about how to restart LogStash, see *Set LogStash to a Windows service*.

24.8.3.12 Use LogStash to collect CSV logs

Before using LogStash to collect CSV logs, modify the configuration file to parse CSV log fields.

Context

The system time when CSV logs are collected or the time in log content can be used as the log upload time. Based on different definitions of log time, configure LogStash to collect CSV logs in two modes.

Use the system time as the log upload time.

- **Sample log**

```
10.116.14.201,-,2/25/2016,11:53:17,W3SVC7,2132,200,0,GET,project/
shenzhen-test/logstore/logstash/detail,C:\test\csv\test_csv.log
```

- **Collection configuration**

```
input { file { type => "csv_log_1" path => ["C:/test/csv/*.log"]
start_position => "beginning" } } filter { if [type] == "csv_log_1
" { csv { separator => "," columns => ["ip", "a", "date", "time",
 "b", "latency", "status", "size", "method", "url", "file"] } } }
output { if [type] == "csv_log_1" { logservice { codec => "json"
endpoint => "***" project => "***" logstore => "***" topic => ""
source => "" access_key_id => "***" access_key_secret => "***"
max_send_retry => 10 } } }
```

Note:

- The configuration file must be UTF-8 encoded without BOM. You can download Notepad++ to modify the file encoding format.
- *path* indicates the file path, which must use Unix separators, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.

- The *type* field must be modified and saved in the file. If multiple LogStash configuration files exist on one computer, ensure that the setting of *type* is unique in the configuration files. Otherwise, data may not be properly processed.

Related plug-ins: *file* and *csv*.

- **Restart LogStash to make the modification take effect.**

Create a configuration file in the *conf* directory. See *Set LogStash to a Windows service*.

Restart LogStash to make the setting take effect.

Use the time in log content as the log upload time.

- **Sample log**

```
10.116.14.201,-,Feb 25 2016 14:03:44,W3SVC7,1332,200,0,GET,project/
shenzhen-test/logstore/logstash/detail,C:\test\csv\test_csv_withtime
.log
```

- **Collection configuration**

```
input { file { type => "csv_log_2" path => ["C:/test/csv_withtime
/*.log"] start_position => "beginning" } } filter { if [type] == "
csv_log_2" { csv { separator => "," columns => ["ip", "a", "datetime
", "b", "latency", "status", "size", "method", "url", "file"] } date
 { match => [ "datetime" , "MMM dd YYYY HH:mm:ss" ] } } } output {
if [type] == "csv_log_2" { logservice { codec => "json" endpoint =>
 "***" project => "***" logstore => "***" topic => "" source => ""
access_key_id => "***" access_key_secret => "***" max_send_retry =>
10 } } }
```

Note:

- The configuration file must be UTF-8 encoded without BOM. You can download Notepad++ to modify the file encoding format.
- *path* indicates the file path, which must use Unix separators, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.
- The *type* field must be modified and saved in the file. If multiple LogStash configuration files exist on one computer, ensure that the setting of *type* is unique in the configuration files. Otherwise, data may not be properly processed.

Related plug-ins: *file* and *csv*.

- **Restart LogStash to make the configuration take effect.**

Create a configuration file in the *conf* directory. Restart LogStash make the configuration take effect. For more information about how to restart LogStash, see *Set LogStash to a Windows service*.

24.8.3.13 Use LogStash to collect other logs

Before using LogStash to collect logs, modify the configuration file to parse log fields.

Use the system time as the log upload time.

- **Sample log**

```
2016-02-25 15:37:01 [main] INFO com.aliyun.sls.test_log4j - single
line log 2016-02-25 15:37:11 [main] ERROR com.aliyun.sls.test_log4j
 - catch exception ! java.lang.ArithmeticException: / by zero at com
.aliyun.sls.test_log4j.divide(test_log4j.java:23) ~[bin/:?] at com.
aliyun.sls.test_log4j.main(test_log4j.java:13) [bin/:?] 2016-02-25
15:38:02 [main] INFO com.aliyun.sls.test_log4j - normal log
```

- **Collection configuration**

```
input { file { type => "common_log_1" path => ["C:/test/multiline
/*.log"] start_position => "beginning" codec => multiline {
pattern => "^\d{4}-\d{2}-\d{2} \d{2}:\d{2}:\d{2}" negate => true
auto_flush_interval => 3 what => previous } } } output { if [type
] == "common_log_1" { logservice { codec => "json" endpoint =>
 "***" project => "***" logstore => "***" topic => "" source => ""
access_key_id => "***" access_key_secret => "***" max_send_retry =>
10 } } }
```

Note:

- The configuration file must be UTF-8 encoded without BOM. You can download Notepad++ to modify the file encoding format.
- *path* indicates the file path, which must use Unix separators, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.
- The *type* field must be modified and saved in the file. If multiple LogStash configuration files exist on one computer, ensure that the setting of *type* is unique in the configuration files. Otherwise, data may not be properly processed.

Related plug-ins: *file* and *multiline* (For a single-line log file, remove the codec => multiline line.)

- **Restart LogStash to make the configuration take effect.**

Create a configuration file in the *conf* directory. Restart LogStash make the configuration take effect. For more information about how to restart LogStash, *Set LogStash to a Windows service*.

24.9 Query and analysis

24.9.1 Indexing and query

Log Service provides real-time LogSearch and Analytics capabilities for a large number of logs. When indexing is disabled, raw data is sequentially consumed based on shards, which is similar to ordered consumption of messages in Kafka. When indexing is enabled, you can collect statistics on and query log data in addition to ordered consumption.

Benefits

- Real-time: Logs can be analyzed immediately after they are written.
- Fast:
  - Query: Billions of data entries can be processed and queried within one second (with five conditions).
  - Analysis: Hundreds of millions of data entires can be aggregated and analyzed within one second (with aggregation by five dimensions and the GroupBy condition).
- Flexible: Query and analysis conditions can be changed as required to obtain results in real time.
- Ecological enrichment: In addition to the reporting, dashboard, and fast analysis features provided by the console, LogSearch and Analytics also seamlessly integrate with Grafana, DataV, and Jaeger and support RESTful APIs and JDBC.

Indexing

The indexing feature of Log Service can sort the values of log data in one or more columns, allowing you to quickly access the log data collected by Log Service. Before using LogSearch and Analytics, you must collect log data and *Configure an index* on the log data.

Log Service indexing is divided into **full-text indexing** and **key-value indexing**.

- **Full-text indexing**: In this mode, indexing is enabled for the full content of a log. The values of all the keys in the log are queried by default, and the log can be queried if any of the keys matches the keyword.
- **Key-value indexing**: In this mode, you can set different indexes for different keys. After setting key-value indexes for a log, you can query specific keys to narrow down the query scope.

To use **key-value indexing**, you must specify the field data type. Currently, Log Service supports the *Text type*, *Value type*, and *JSON type* types. For more information about these index data types, see *Overview*.

Terms

When LogSearch/Analytics (indexing) is disabled, raw data can be consumed by shard sequentially, similar to Kafka. After LogSearch/Analytics is enabled, statistics and query of log data are also supported.

Data types

You can configure the type of each key in a log (full text index is a special key, whose value is the log). Currently, Log Service supports the following data types.

Category	Type	Description	Example
Basic	*Text type*	The text type and supports the combination of keyword and fuzzy match as well as Chinese word segmentation.	`uri:"login*" method:"post"`
Basic	*Value type*	The numeric type that supports interval queries.	`status>200, status in [200, 500]`
Basic	*Value type*, *JSON type*	The numeric type that supports floating-point numbers.	`price>28.95, t in [20.0, 37]`
Combination	*JSON type*	Indicates that the index is a JSON field that supports nested queries. The field type is Text by default. You can set indexes of the Text, Long, and Double types for the b elements at layer a in the a.b path format. The fields adopt the configured types.	`level0.key>29.95 level0.key2:"action"`
Combination	*Text type*	Indicates that the full content of the log is queried as text.	`error and "login fail"`

Syntax of LogSearch/Analytics

Query: It consists of Search and Analytics, which are separated using `|`.

`$Search |$Analytics`

- Search: It is a search criteria, which can be generated by using keywords, fuzzy match conditions, values, ranges, and combinations. If it is blank or an asterisk (*), all data is used.
- Analytics: It calculates and collects statistics on search results or full data.

Note:

Both Search and Analytics are optional. If Search is empty, all the data in the specified period is not filtered and the results are counted directly. If Analytics is empty, the query results are returned and no statistics are collected.

Restrictions

If the data volume of logs to be searched for is very large (for example, the time span is long and there are more than 10 billion log entries), the data cannot be searched completely by one query request. In this case, Log Service returns the existing data and notifies you that the query result is incomplete.

At the same time, the server caches the query results generated in the last 15 minutes. When the query result is partially cached, the server continues to scan log data that has not been cached. To reduce your workload of merging multiple query results, Log Service merges the hit query results in the cache and the new hit results of the current query and returns them to you.

Therefore, Log Service enables you to get the final result by calling the interface repeatedly with the same parameters.

24.9.2 Real-time analysis

Log Service supports aggregate functions. This service combines the query function with SQL computing to calculate the query result.

Syntax example:

```
status>200 |select avg(latency),max(latency) ,count(1) as c GROUP BY
method  ORDER BY c DESC  LIMIT 20
```

Basic syntax:

```
[search query] | [sql query]
```

The SEARCH condition and computing condition are separated by a vertical bar (|). This syntax indicates to filter the logs you need from the logs by using the search query and perform SQL query calculation for these logs. The search query syntax is specific to Log Service. For more information, see *Query syntax*.

Prerequisites

To use the analysis function, you must click **Enable** for the SQL-related fields in **Search and Analysis Config**.

- If you do not enable analysis function, computing function of up to 10,000 rows of data per shard is provided, and the delay is relatively high.
- With the Enable Analytics switch turned on, Log Service provides the quick analysis in seconds .
- Only works for new data when function is enabled.
- No additional charges are incurred after the Enable Analytics is turned on.

Supported SQL syntax

Log Service supports the following SQL syntax. For details, click the specific links.

- Aggregate functions in the SELECT statement:
  - *General aggregate functions*
  - *Map functions*
  - *Estimating functions*
  - *Mathematical statistical functions*
  - *Mathematical functions*
  - *String functions*
  - *URL functions*
  - *Date and time functions*
  - *Regular expression functions*
  - *JSON functions*
  - *Type conversion functions*
  - *Arrays*
  - *Binary string functions*
  - *Bit operation*
  - *Comparison functions and operators*
  - *Lambda function*
  - *Logical function*
  - *Geospatial functions*
- *GROUP BY syntax*
- *Window functions*
- *HAVING syntax*
- *ORDER BY syntax*

- *LIMIT syntax*
- *CASE WHEN syntax*

Syntax

The SQL syntax structure is as follows:

- The FROM clause and WHERE clause are not required in each SQL statement. The default FROM clause specifies the current Logstore from which to query data and the default WHERE clause defines the condition as search query.
- The supported clauses include SELECT, GROUP BY, ORDER BY [ASC,DESC], LIMIT, and HAVING.
- Only the first 10 results are returned by default. To return more results, add limit n to the statement, for example: `* | select count(1) as c, ip group by ip order by c desc limit 100`.

Built-in fields

Log Service has built-in fields for statistics. These built-in fields are automatically added when you configure any valid column.

Field name	Type	Definition
`__time__`	bigint	The generation time of a log.
`__source__`	varchar	The source IP of the log. Note : This field is source when you query. The underlines (__) is added before and after source only in SQL.
`__topic__`	varchar	The topic of the log.

Restrictions

1. The highest concurrency of each project is 5.
2. A single column varchar has the maximum length of 512 and will be truncated if the length exceeds 512.
3. 100 rows of data are returned by default, and paging is not supported. If you want more data to be returned, use *LIMIT syntax*.

Example

Count the hourly PV, UV, and maximum delay corresponding to a user request, with the highest delay of 10:

```
*|select date_trunc('hour',from_unixtime(__time__)) as time,
     count(1) as pv,
     approx_distinct(userid)  as uv,
     max_by(url,latency) as top_latency_url,
     max(latency,10) as top_10_latency
     group by 1
     order by time
```

24.9.3 Disable an index

You can disable an index when you do not use the query and analysis function of Log Service.

Procedure

1. *Log on to the Log Service console*.
2. Click the name of a project.
3. Select a Logstore and click **Search** in the **LogSearch** column.
4. Click **Index Attributes** > **Disable** in the upper-right corner.

24.9.4 Index types

24.9.4.1 Overview

Log Service allows you to set indexes for the full text or some fields of the collected logs. If you set an index for the full text of a log, the value used to query this log is the content of the entire log. If you set indexes for some fields of a log, you can set the data type of each key used in queries.

Data type

The following table describes the supported index types.

Category	Type	Description	Example
Basic	*Text type*	The text type that supports the combination of keyword and fuzzy match as well as Chinese word segmentation.	`uri:"login*" method:"post"`
Basic	*Value type*	The numeric type that supports interval queries.	`status>200, status in [200 , 500]`
Basic	*Value type*, *JSON type*	The numeric type that supports floating-point numbers.	`price>28.95, t in [20.0, 37 ]`

Category	Type	Description	Example
Combination	*JSON type*	Indicates that the index is a JSON field that supports nested queries. The field type is Text by default. You can set indexes of the Text, Long, and Double types for the b elements at layer a in the a.b path format. The fields adopt the configured types.	`level0.key>29.95 level0.key2:"action"`
Combination	*Text type*	Indicates that the full content of the log is queried as text.	`error and "login fail"`

Example

The following log includes time and other four keys.

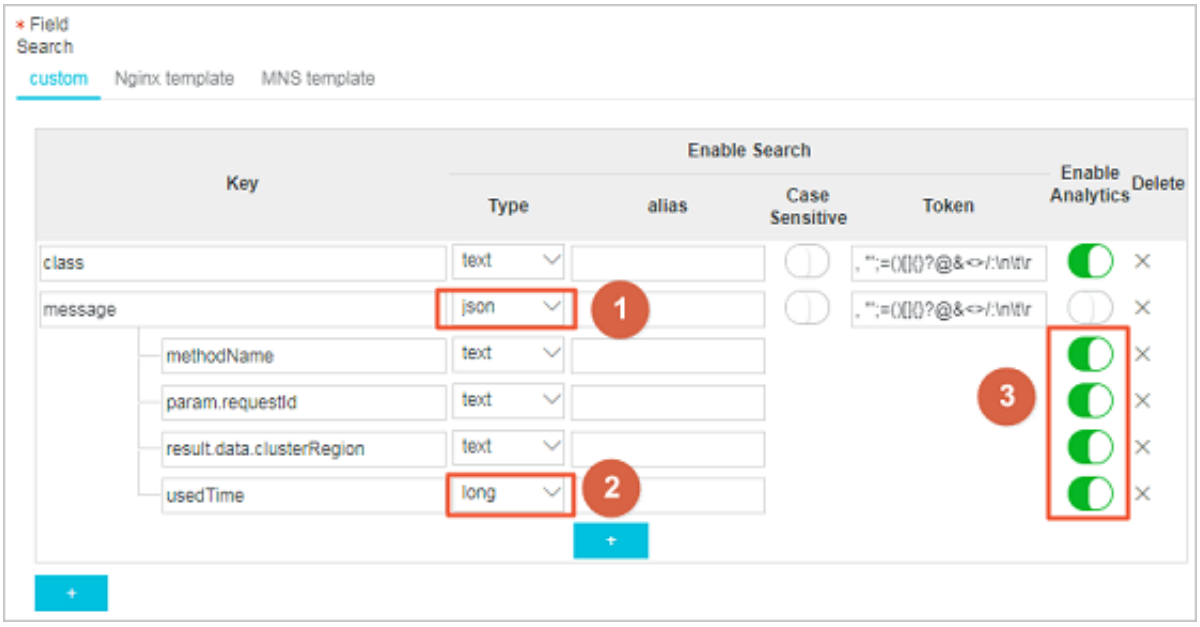
No.	Key	Type
0	time	-
1	class	text
2	status	long
3	latency	double
4	message	JSON

```
0. time:2018-01-01 12:00:00
  1. class:central-log
  2. status:200
  3. latency:68.75
  4. message:
  {
      "methodName": "getProjectInfo",
      "success": true,
      "remoteAddress": "1.1.1.1:11111",
      "usedTime": 48,
      "param": {
              "projectName": "ali-log-test-project",
              "requestId": "d3f0c96a-51b0-4166-a850-f4175dde7323"
      },
      "result": {
          "message": "successful",
          "code": "200",
          "data": {
              "clusterRegion": "ap-southeast-1",
              "ProjectName": "ali-log-test-project",
              "CreateTime": "2017-06-08 20:22:41"
          },
          "success": true
      }
```

```
  }
```

You can set indexes for a log as follows:

Figure 24-8: Index setting



In the preceding figure:

- (1) indicates query of all the data of the String and Boolean types in JSON fields.
- (2) indicates query of data of the Long type.
- (3) indicates SQL analysis of configured fields.

Example:

1. **Query of data of the String and Boolean types**

   - No configurations in the JSON field are needed.
   - JSON maps and arrays are automatically expanded. You can query fields that are multi-level nested with each level separated by a period (.).

   ```
   class : cental*
   message.traceInfo.requestId : 92.137_1518139699935_5599
   message.param.projectName : ali-log-test-project
   message.success : true
   ```

2. **Query of data of the Double and Long types**

   The fields in a JSON field must be configured separately and must not be contained in an array .

   ```
   latency>40
   ```

```
message.usedTime > 40
```

3. **Combination query**

```
class : cental* and message.usedTime > 40 not message.param.
projectName:ali-log-test-project
```

24.9.4.2 Text type

Similar to search engines, text data is queried based on term matching. Therefore, you must configure word segmentation, case sensitivity, and enable Chinese word segmentation.

Configuration instructions

Case sensitivity

Case sensitivity for raw log query. For example, if the raw log is "internalError":

- If the parameter is set to **False** (case insensitive), the sample log can be located with the keyword either "INTERNALERROR" or "internalerror".
- If the parameter is set to **True** (case sensitive), the sample log can be located only with the keyword "internalError".

Word segmentation

You can separate the contents of a raw log into several keywords by using a word segmentation.

For example, when we query the following log content:

```
/url/pic/abc.gif
```

- If no word segmentation is set, the string is considered as an individual word`/url/pic/abc.gif`. You can only query this log by using the complete string or fuzzy match such as`/url/pic/*`.
- If `/`is set as the word segmentation, the raw log is separated into three words: `url`, `pic`, and `abc.gif`. You can find the log by query or fuzzy query with any of the keywords `url`, `abc.gif`, and `pi*`, or with `/url/pic/abc.gif` (segmented into `url, pic, and abc.gif` in query).
- If the word segmentation is set to `/.`, the raw log is separated into four words: `url`, `pic`, `abc`, and `gif`.

Note:

You can extend the query range by setting appropriate word segmentations.

Contains Chinese characters

If the log contains Chinese characters, enable Chinese word segmentation. For example, for the following log content:

```
buyer:用户小李飞刀lee
```

With the default word segmentation ":", the raw log content is segmented into two words: `buyer` and `用户小李飞刀lee`. If you search for `用户`, `Lee` will not be returned. If you enable the option of **Chinese character included**, the Log Service analyzer analyzes the meaning of the Chinese words and segments the log content into five words: `buyer`, `用户`, `小李`, `飞刀`, and `lee`. You can locate the log with either the keyword `飞刀` or `小李飞刀` (resolved into: `小李 and 飞刀`).

Note:
The function of Chinese word segmentation somehow compromises the write speed. Set the option with caution based on your need.

Full text index

By default, full text query (index) considers all the fields and keys of a log, except the time field, as text data, and does not need to specify keys. For example, the following log is composed of four fields (time/status/level/message):

```
[20180102 12:00:00] 200,error,some thing is error in this field
```

- time:2018-01-02 12:00:00
- level:"error"
- status:200
- message:"some thing is error in this field"

After enabling full text index, the following text data is assembled in the "key:value + space" mode. For example:

```
status:200 level:error message:"some thing is error in this field"
```

Note:

- Prefix is not required for full text query. Enter error as the keyword, both level field and message field meet the query condition.

- You must set a word segmentation for the full text query. If a space is set as the word segmentation, status:200 is considered as a phrase. If ":" is set as the word segmentation, status and 200 are considered as two independent phrases.
- Numbers are processed as texts. For example, you can use the keyword 200 to query this log. The time field is not processed as a text.
- You can query this log if you enter a key such as "status".

24.9.4.3 Value type

When configuring indexes, you can configure a field as the value type and query the key by using a value range.

Configuration instructions

Supported types: `long` (long integer) or `double` (decimal). After configuring a field as the value type, you can only query the key by using a value range.

Example

To query the longkey whose key range is (1000 2000], use the following methods.

- Use values to query the longkey:

```
longKey > 1000 and longKey <= 2000
```

- Use an interval to query the longkey:

```
longKey in (1000 2000]
```

For more syntax, see *Query syntax*.

24.9.4.4 JSON type

JSON is a combined data type consisting of Text, Boolean, Number, Array, and Map.

Configuration instructions

Text-type data

JSON fields of the Text and Boolean types are automatically identified.

For example, the following JSON keys can be searched for using `jsonkey.key1:"text_value"` and `jsonkey.key2:true`.

```
jsonkey: {
   key1:text_value,
   key2:true,
   key3:3.14
```

```
}
```

Number-type data

You can search for data of the Double and Long types in non-JSON arrays by setting a type and specifying a path.

For example, the statement used to search for the jsonkey.key3 field of the Double type is as follows:

```
 jsonkey.key3 > 3
```

Non-fully valid JSON

Non-fully valid JSON data is parsed until the invalid content appears.

for example:

```
"json_string":
{
     "key_1" :  "value_1",
     "key_map" :
      {
             "key_2" : "value_2",
             "key_3" : "valu
```

The data following key_3 is truncated and lost. The log with missing data is correctly parsed until the `json_string.key_map.key_2` field.

Note

- JSON object and JSON array are not supported.
- Fields cannot appear in JSON arrays.
- Fields of the Boolean type can be converted to the text type.

Query syntax

The parent path prefix in JSON is required to search for a specified key. The query syntax for the text and numerical types is similar to other types. For more information, see *Query syntax*.

24.9.5 Query syntax and functions

24.9.5.1 Query syntax

To help you query logs more effectively, Log Service provides a set of query syntaxes used to express query conditions. You can specify query conditions through the GetLogs and

GetHistograms APIs on Log Service or on the query page of the Log Service console. This section details the query condition syntax.

Index type

Log Service supports creating an index for the Logstore in two modes:

- Full-text indexing: The entire line of log is queried as a whole, without differentiating the key and value (Key, Value).
- Key value indexing: Query is performed when Key is specified, for example, *`FILE:app`*, *`Type: action`*. All the contained strings under this key will be hit.

Syntax keyword

LogSearch query conditions support the following keywords:

Name	Meaning
and	Binary operator. The format is `query1 and query2`, indicating the intersection of the query results of `query1` and `query2`. If there is no syntax keyword between words, the relation between words is `and` by default.
or	Binary operator. The format is . `query1 or query2` , indicating the intersection of the query results of `query1` and `query2`.
not	Binary operator. The format is `query1 not query2`, indicating a result that meets `query1` and does not meet `query2`, that is, `query1-query2`. If only `not query1` exists, it indicates that logs that do not contain the query results of `query1` are selected.
(,)	The left and right brackets are used to merge one or multiple sub -queries into one query to increase the priority of query in the brackets.
:	Used to query the key-value pair. `term1:term2` forms a key-value pair. If the key or value contains reserved characters such as spaces and colons (:)`:`, quotation marks `""` are required to enclose the entire key or value.
"	Convert a keyword into a common query character. Any term in the left and right quotation marks will be queried and will not be used as a syntax keyword. Or all the terms in the left and right quotation marks are regarded as a whole in the key-value query.

Name	Meaning
\	Escape character. Used to escape quotation marks. The quotation marks after escaping indicate the symbols themselves and are not considered as escape characters, for example, `"\""`.
|	Pipeline operator, indicating more computing based on the previous computing, for example, query1 | timeslice 1h | count.
timeslice	Time slice operator indicates the length of time during which the data is regarded as a whole for computing, and the use methods are timeslice 1h, timeslice 1m, and timeslice 1s, which respectively indicate 1 hour, 1 minute, and 1 second as a whole. For example, query1 | timeslice 1h | count indicates querying the query condition, and the total number of times with 1 hour as the time slice is returned.
count	Count operator, indicating the number of logs.
*	Fuzzy query keyword, used to replace zero or more characters. For example, if `que*` is used in a query, all the hit words starting with `que` will be returned. NOTE: Up to 100 results meeting the keyword are returned for the query.
?	Fuzzy query keyword, used to replace one character. For example, if `qu?ry` is used in a query, all the hit words starting with `qu`, ending with `ry`, and with a character in the middle are returned.
`__topic__`	Query the data under a certain topic. Under the new syntax, the data of zero or more topics can be queried in the query, for example, `__topic__:mytopicname`.
`__tag__`	Query a tag value under a tag key, for example, `__tag__:tagkey:tagvalue`.
source	Query data of an IP address, for example, `source:127.0.0.1`.
>	Query the logs with the value of a field greater than a specific number, for example, `latency > 100`.
>=	Query the logs with the value of a field greater than or equal to a specific number, for example, `latency >= 100`.
<	Query the logs with the value of a field smaller than a specific number, for example, `latency < 100`.
<=	Query the logs with the value of a field smaller than or equal to a specific number, for example, `latency <= 100`.

Name	Meaning
=	Query the logs with the value of a field equal to a specific number, for example, `latency = 100`.
in	Query the logs with a field falling in a specific range. Brackets ([]) are used to indicate closed intervals and parentheses (()) are used to indicate open intervals, with two numbers enclosed and separated by spaces. For example, `latency in [100 200]` or `latency in (100 200]`.

Note:

- Syntax keywords are case-insensitive.
- Priorities of syntax keywords are sorted in the descending order as `: > " > ( ) > and not > or`.
- Log Service also reserves the right to use the following keywords. If you need to use these keywords, enclose the keywords with double quotation marks: `sort asc desc group by avg sum min max limit`.
- If the full text index and key value index have different word segmentation characters when they are configured, data cannot be queried using the full text query method.
- To perform a numeric query, set the data type of the queried column to double or long. If no data type is set or the syntax used for the numeric range query is incorrect, Log Service translates the query condition into a full text index, which may lead to an unexpected result.
- If you change the data type of a column from text to numeric, only the = query is supported for the data prior to this change.

Query example

1. Logs that contains a and b at the same time: `a and b` or `a b`
2. Logs that contain a or b: `a or b`
3. Logs that contain a but no b: `a not b`
4. Those in all the logs that contain no a: `not a`
5. Query the logs that contain a and b, but no c: `a and b not c`
6. Logs that contain a or b and must contain c: `(a or b ) and c`
7. Logs that contain a or b, but no c: `(a or b ) not c`
8. Logs that contain a and b and may contain c: `a and b or c`
9. Logs with the FILE field containing apsara: `FILE:apsara`

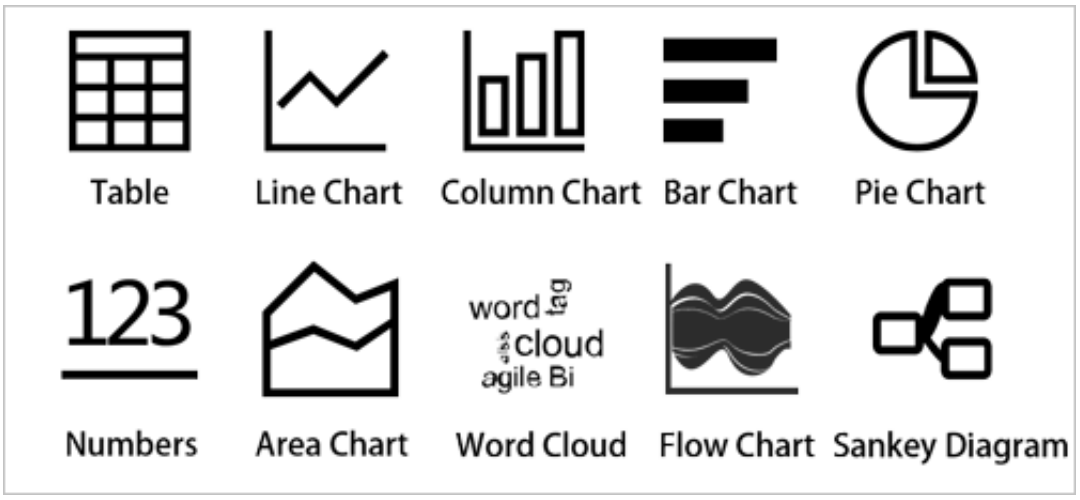
10. Logs whose FILE field contains apsara and shennong: `FILE:"apsara shennong"`, `FILE: apsara FILE: shennong`, or `FILE:apsara and FILE:shennong`
11. Logs containing and: `and`
12. Logs with the FILE field containing apsara or shennong: `FILE:apsara or FILE:shennong`
13. Logs with the file info field containing apsara: `"file info":apsara`
14. Logs that contain quotation marks: `\"`
15. Logs starting with shen: `shen*`
16. Query all the logs starting with shen under the FILE field: `FILE:shen*`
17. Query the logs starting with shen, ending with ong and with a character in the middle: `shen? ong`
18. Query all the logs starting with shen and aps: `shen* and aps*`
19. Query the distribution of logs starting with shen, with the time slice of 20 minutes: `shen* | timeslice 20m | count`
20. Query all the data under topic1 and topic2: `__topic__:topic1 or __topic__ : topic2`
21. Query all the data of tagvalue2 under tagkey1: `__tag__ : tagkey1 : tagvalue2`
22. A query for all the data with a latency greater than or equal to 100 and less than 200 can be written in either of the following ways: `latency >=100 and latency < 200` or `latency in [100 200)`.
23. A query for all the requests with a latency greater than 100 must be written in the following way: `latency > 100`.
24. Query logs that do not contain crawler and logs with http_referer not containing opx: <codeph> latency > 100</codeph>.
25. Query logs with the cdnIP field being null: <codeph>latency > 100</codeph>.
26. Query logs without the cdnIP field: `not cdnIP:*`.
27. Query logs with the cdnIP field: `cdnIP:*`.

Specified or cross-topic query

Each Logstore can be divided into one or more subspaces according to the topic. During query, the query range can be limited for the specified topic to increase the speed. Therefore, the user with the level-2 classification requirement for the LogStore is recommended to use topic to divide the LogStore.

When one or more topics are specified to perform query, query is implemented in the topic meeting the conditions only. However, if no topic is entered, the data under all the topics is queried by default.

For example, topics are used to classify logs under different domain names:



Topic query syntax:

- The data under all topics can be queried. The data of all topics is queried if no topic is specified in the query syntax and parameters.
- Topic can be queried in the query. The query syntax is `__topic__:topicName`. The old mode is still supported at the same time. The topic is specified in the URL parameter.
- Multiple topics can be queried, for example, `__topic__:topic1 or __topic__:topic2` indicates the union of data under topic1 and topic2.

Fuzzy search

Log Service support fuzzy search. Specify a word within 64 characters, and add fuzzy search keywords such as `*` and `?` in the middle or in the end of the word. 100 eligible words will be searched out, in the meantime, all the logs eligible and contain the 100 words will be returned

Note:

- Prefix must be specified when query logs, that is, the word can not begin with `*` and `?`.
- Precise the specified word, you will get a more accurate result.
- Fuzzy search cannot be used to search for words that exceeds 64 characters. It is recommended that you specified a word under 64 characters.

24.9.5.2 Context query

When you expand a log file, each log records an event, and they do not exist independently. Several consecutive logs can be used to review the occurrence process of the whole event sequence.

Log context query specifies the log source (machine + files) and a log in it, and searches a number of records (the text above) before this log and a number of logs (the text below) after this log in the original file. Particularly, it is an effective way for clarifying the problem cause and effect under the DevOps scenario.

The Log Service console provides a dedicated page for query. You can use a browser to view the context information in the original file of the specified log, in a way similar to turning the pages of the original log file. By viewing the context information of a specified log, you can quickly identify related fault information during service troubleshooting, and locate problems with ease.

Scenarios

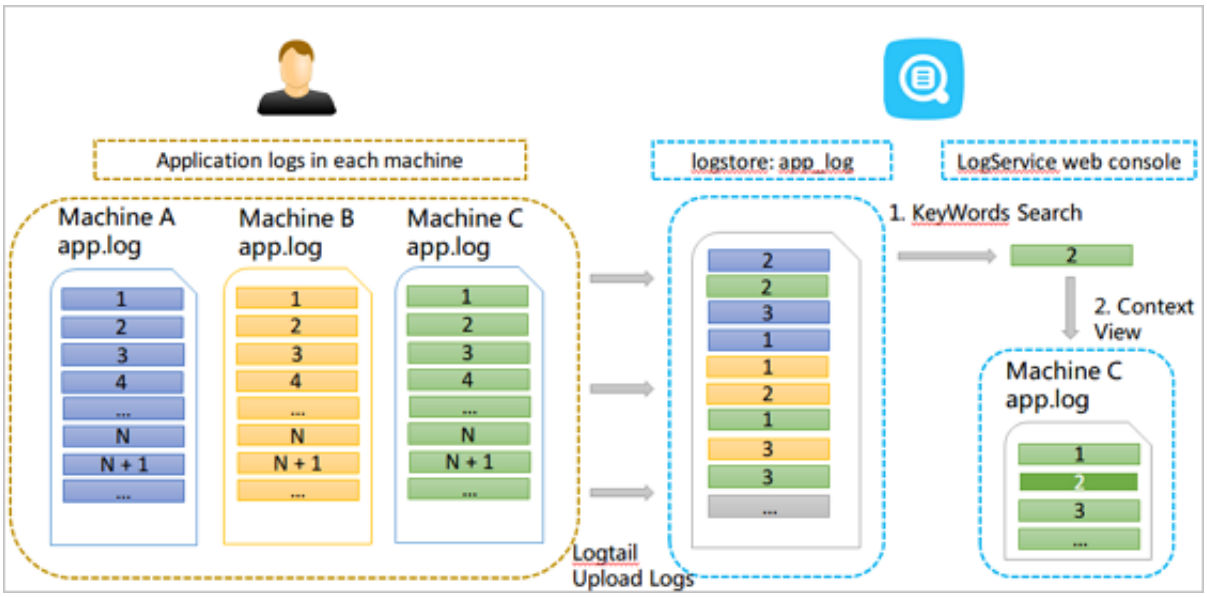
For example, the O2O take-out website will record the transaction track of an order in the program log on the server:

User Logon > **Browse Items** > **Click Items** > **Add to Shopping Cart** > **Place an Order** > **Pay for the Order** > **Deduct Payment** > **Generate Order**

If you fail to place the order, the O&M personnel need to identify the cause quickly. For a conventional context query, the administrator adds the machine logon permission to related members, and then the investigator logs on to each machine where applications are deployed in sequence and uses the order ID as the keyword to search application log files and identify the cause of a failed order.

Log Service allows you to troubleshoot in the following approach:

1. Install the log collection client Logtail on the server, and add the machine group and log collection configuration on the console. Then, Logtail starts to upload the incremental logs.
2. Access the console log query page of Log Service, specify the time range, and find the order failure log according to the order ID.
3. Page up with the found error log as benchmark till other related log information is found (for example: credit card deduction fails).



Benefits

- There is no intrusion into the application, and no need to change the log file format.
- The specified log context information of any machine and file can be viewed in the Log Service console, avoiding the trouble of logging on to each machine to view the log file.
- In combination with the time of event occurrence, if the context query is performed after the suspicious log is located quickly in the specified time range of the Log Service console, you can always get twice the result with half the effort.
- You do not need to worry about data loss caused by insufficient server storage or log file rotation, and can view historical data in the Log Service console at any time.

Prerequisites

- *Use Logtail to collect logs* To upload data to the Logstore, only machine group creation and collection configuration are required. You can also use producer-related SDKs for uploading, such as Producer Library,

  Log4J, LogBack, and C-Producer Library.
- Enable indexing.

Note:
Currently, context query does not support syslogs.

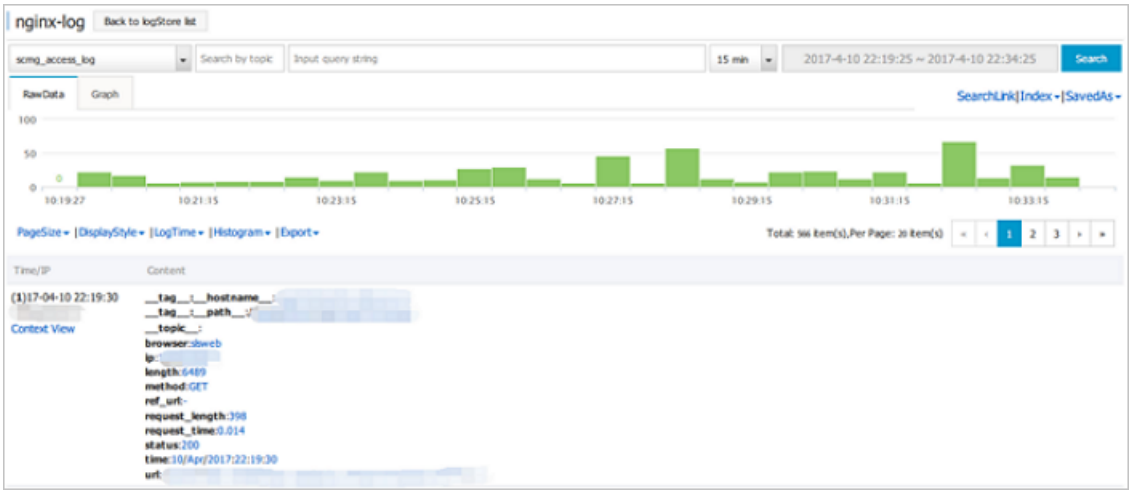
Procedure

1. *Log on to the Log Service console*.

2. Click the name of a project.
3. On the **Logstores** page, select a Logstore, and click **Search** in the **LogSearch** column to go to the search page.

   If there is a **Context View** link to the left of a log returned on the query results page, the log supports context query.
4. Enter your search and analytic syntax, select a time range, and click **Search**.

   If there is a **Context View** link to the left of a log returned on the query results page, the log supports context query.



5. Select a log and click **Context View**. On the page that appears on the right, view the context log of the target log.
6. Scroll with the mouse on the page to view the context information of the selected log. To view the historic or current information, click **Earlier** or **Later**.

24.9.5.3 Other functions

In addition to the statement-based query capability, the query and analysis function of Log Service provides the following extended functions for the query optimization:

Raw logs

After the index is enabled, enter the keywords in the search box and select the search time range. Then, click **Search** to view the histogram of the log quantity, the raw logs, and the statistical graph.

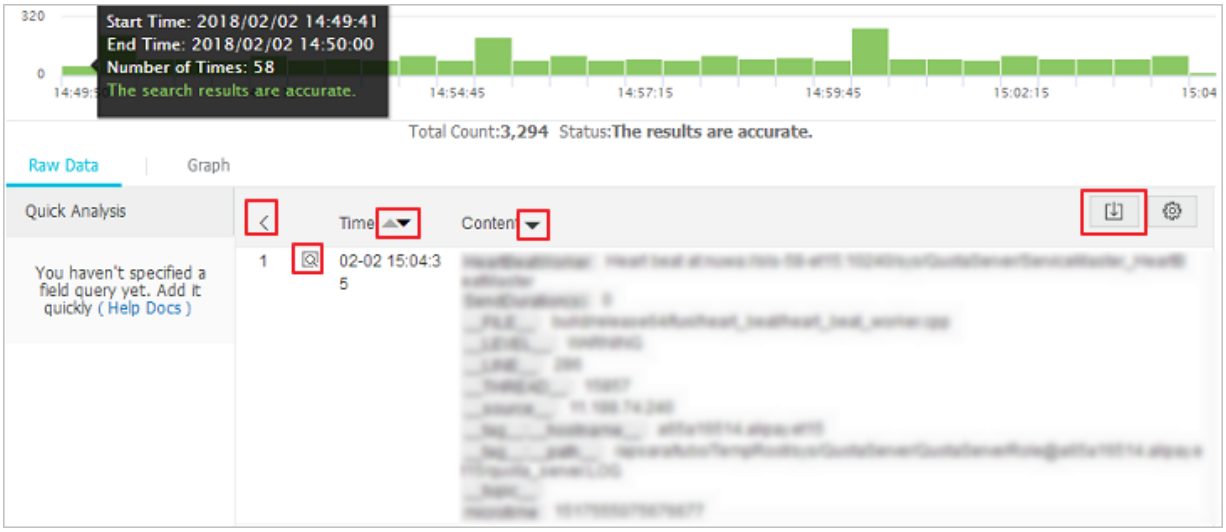
The histogram of the log quantity displays the time-based distribution of log search hit counts . With the histogram, you can view the log quantity changes over a certain period of time. By

clicking the rectangular area to narrow down the time range, you can view the information about the log hits within the specified time range to refine the display of the log search results.

On the Raw Log tab, you can view the content of hit logs in time order.

- Click the triangle symbol beside the **Time** column to switch between **chronological** order and **reverse chronological** order.
- Click the triangle symbol beside the **Content**, you can switch between **Display with Line Breaks** or **Display in One Line**.
- Click the value **Keyword** in the log content to view all the log content which contains the keyword.
- On the Raw Log tab, click **Download** in the upper-right corner to download search results in CSV format. Click **Set** to add columns named after fields to the raw log results so that you can view the target fields of each raw log in the new columns.
- Click **Context** to view the 15 logs preceding the log and the 15 logs following the log. For more information, see *Context query*.

> **Note:**
> Currently, the context query feature is only applicable to the data uploaded by Logtail.



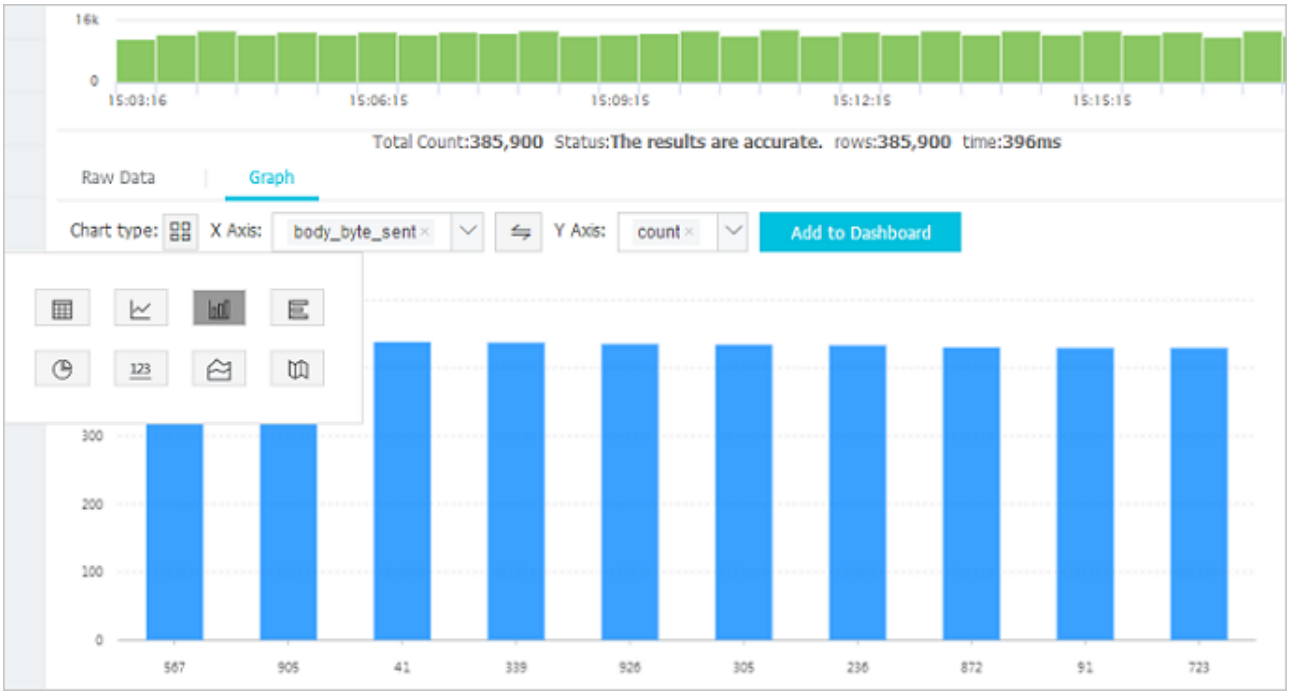
Statistical chart

After you enable indexing and enter a search analysis statement, you can view log statistics on the **Aggregation** tab.

- Data can be displayed in the following ways: tables, line charts, column charts, bar charts, pie charts, numeric values, area charts, and maps.

You can select an appropriate statistical graph type based on the actual statistical analysis needs.

- You can adjust the display content of axes X and Y to obtain the display results that meet your needs.
- Add the analysis results to **Dashboard**. For more information, see *Dashboard*.



Context query

The Log Service console provides a dedicated page for query. You can use a browser to view the context information in the original file of the specified log, in a way similar to turning the pages of the original log file. By viewing the context information of a specified log, you can quickly identify related fault information during service troubleshooting, and locate problems with ease. For more information, see *Context query*.

Quick analysis

The quick analysis function of Log Service supports a quick interactive query. This service allows you to quickly analyze the distribution of a field over a period of time and reduce the cost of indexing key data. For more information, see *Quick analysis*.

Saved Search

On the query page, click **Saved Search** in the upper right corner to save your current query action as a quick query. Next time you can initiate the query action on the **Saved Search** tab on the left without entering the query statement manually.

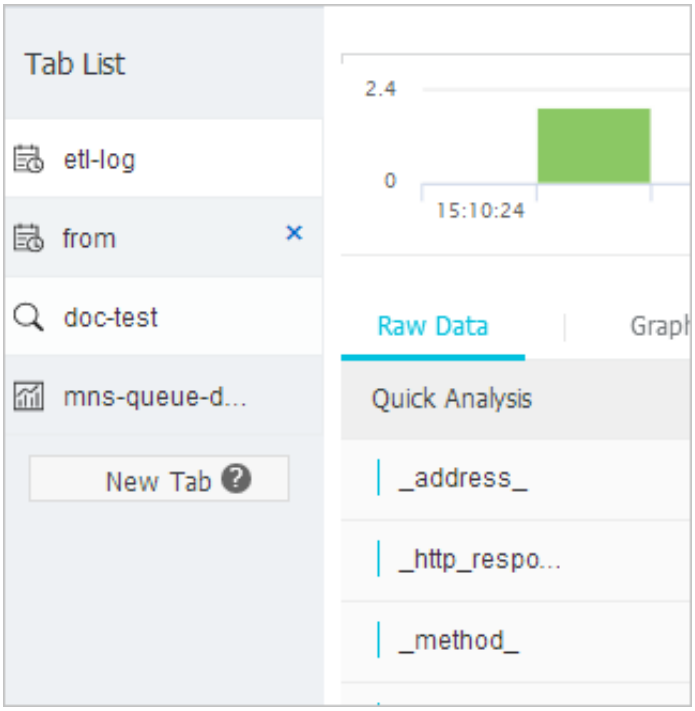
The quick query condition can be used by alarm policies. If the quick query is added to the **Tab**, it can be accessed directly on the tab.

Tag

Log Service provides a tag list on the left side of the query page. You can add the following three data pages to the tag list:

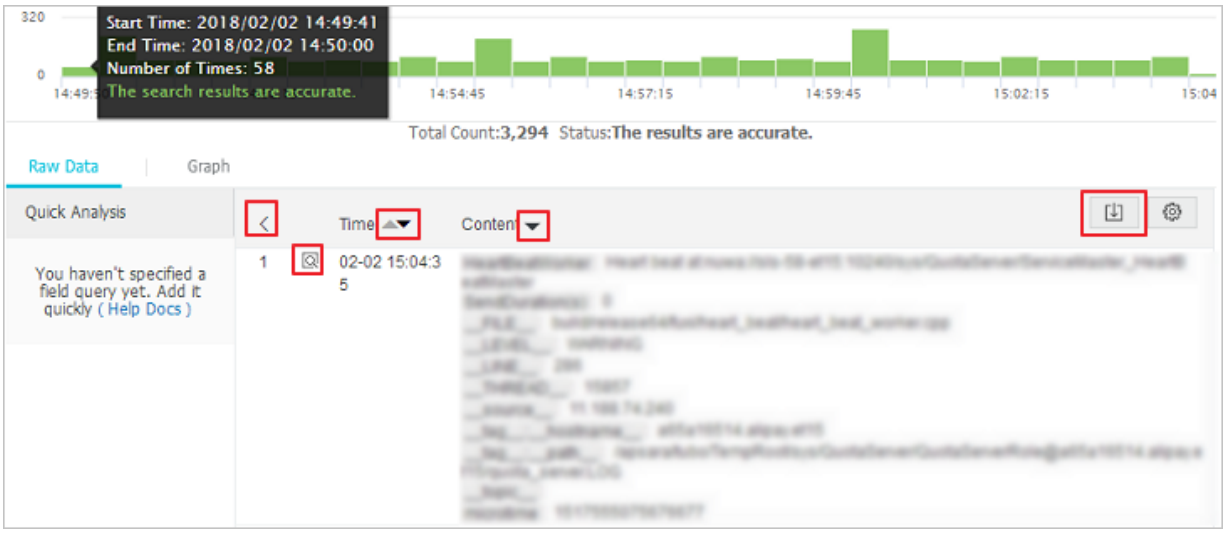
- Logstore
- Saved Search
- Dashboard

You can access Logstores, saved quick queries, and dashboards in the tag list with ease. Click **Add Tab** in the label list. In the menu that appears on the right, select the Logstore, quick query, or dashboard to be added as a tag. To delete a tagl, click the X symbol next to it in the tag list.



Dashboard

Log Service provides the dashboard feature to visually display search analysis statements. For more information, see *Dashboard*.



Save as alarm

Log Service allows you to configure alerting based on your **LogSearch Results**. You can configure the alarm rules so that specific alarm content can be sent to you by using in-site notifications or DingTalk messages.

Configuration process:

1. Configure Savedsearch.
2. Configure the alarm rules.
3. Configure notification type.
4. View alarm records.

For more information, see *Set an alarm*.

24.9.5.4 Quick analysis

The quick analysis function of Log Service supports a quick interactive query. This service allows you to quickly analyze the distribution of a field over a period of time and reduce the cost of indexing key data.

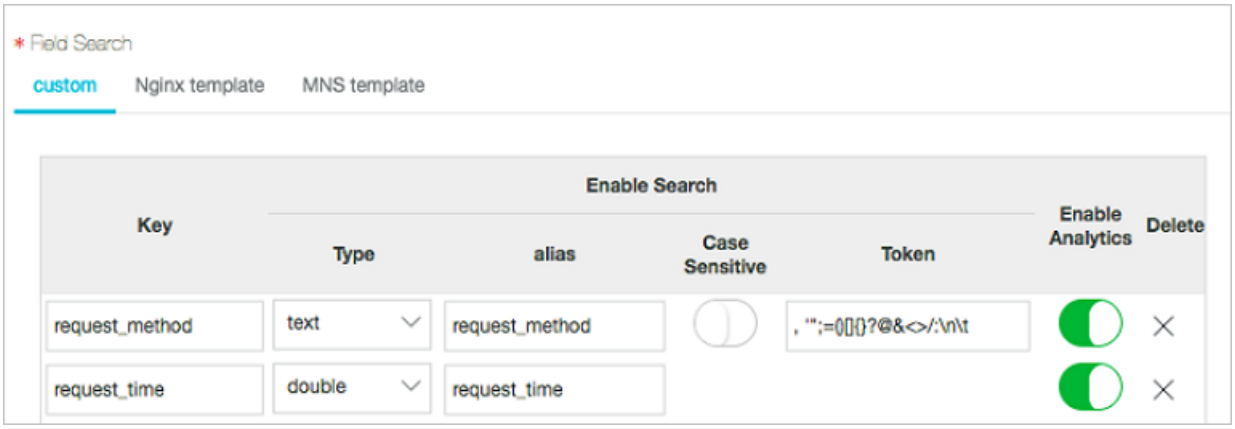
Features

- Supports grouping statistics for the first 10 of the first 10,000 pieces of data of `Text` fields.
- Supports Quick generation of the `approx_distinct` query statement for `Text` fields.
- Supports histogram statistics for the approximate distribution of `long` or `double` fields.
- Supports quick search for the maximum value, minimum value, average, or sum for `long` or `double` fields.
- Supports generating query statements based on quick analysis and query.

The user must specify the field query properties before using the quick analysis.

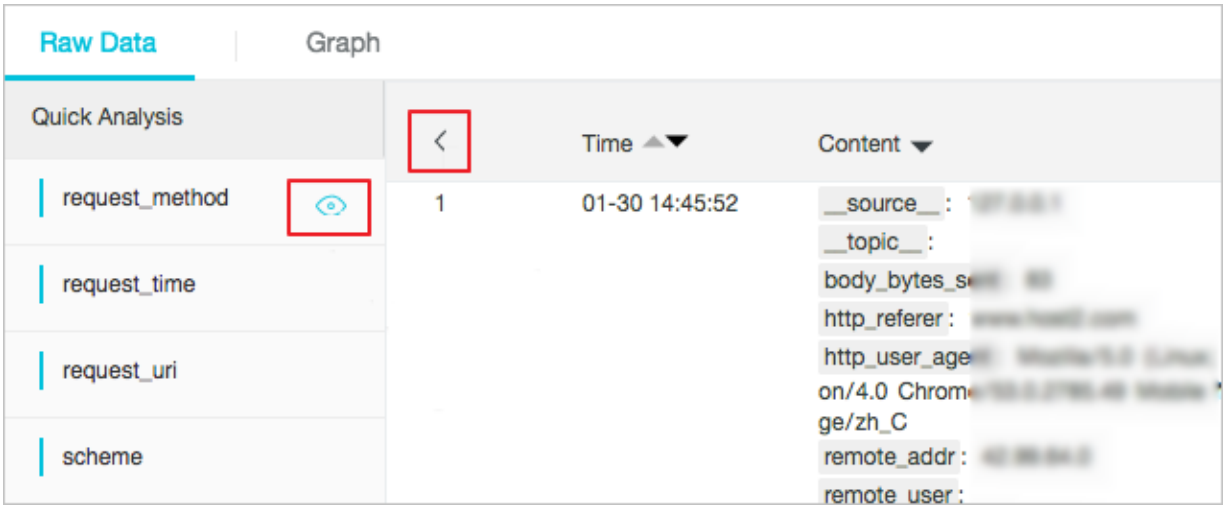
1. For specified field query, you must enable the index to activate the query and analysis function.
2. Set the `key` in the log as the field name and set the type, alias, and separator.

   If the access log contains `request_method` and `request_time` , you can configure the settings as follows:



User guide

After field query setting, go to the query page, click the **Raw Logs** tab, and view the fields in the left-side **Quick Analysis** column. Click the button above the sequence number to hide the page. Click the **eye** button to start quick analysis based on the **Current Temporal Interval** and **Current $Search** condition.



Text type

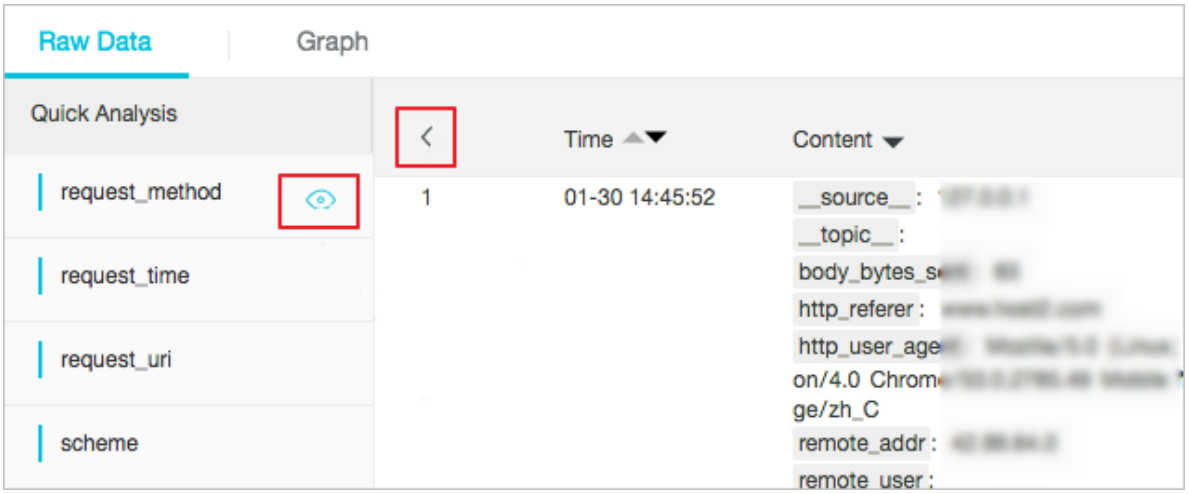
- **Group statistics for the Text type**

Click the **eye** at the right of the filed to quickly group the first 1,000 pieces of data of this **Text** field and return the ratio of the first 10 pieces.

The query statement is as follows:

```
$Search | select count(1) as pv , "${keyName}" from ( select "${
keyName}" from log limit 10000) group by "${keyName}" order by pv
desc limit 10
```

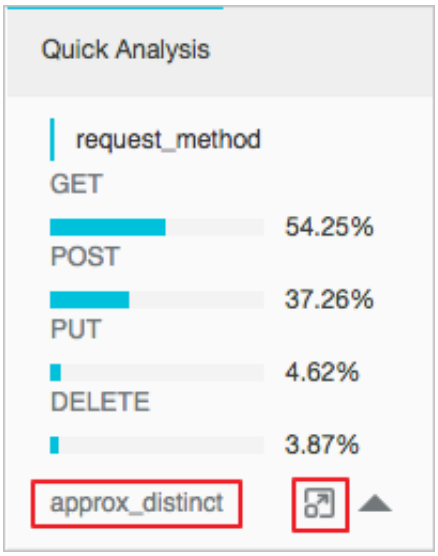
`request_method` returns the following results based on the grouping statistics, where `GET` requests are in the majority.



- **Check the number of unique entries of the field**

  Under the target fields in **Quick Analysis**, click **approx_distinct** to check the number of unique entries for `${keyName}`.

  `request_method`returns the following results based on the grouping statistics, where `GET` requests are in the majority.



- **Extend the query statement of grouping statistics to the search box**

  Click the button at the right of **approx_distinct** to extend the query statement of grouping statistics to the search box for further operations.

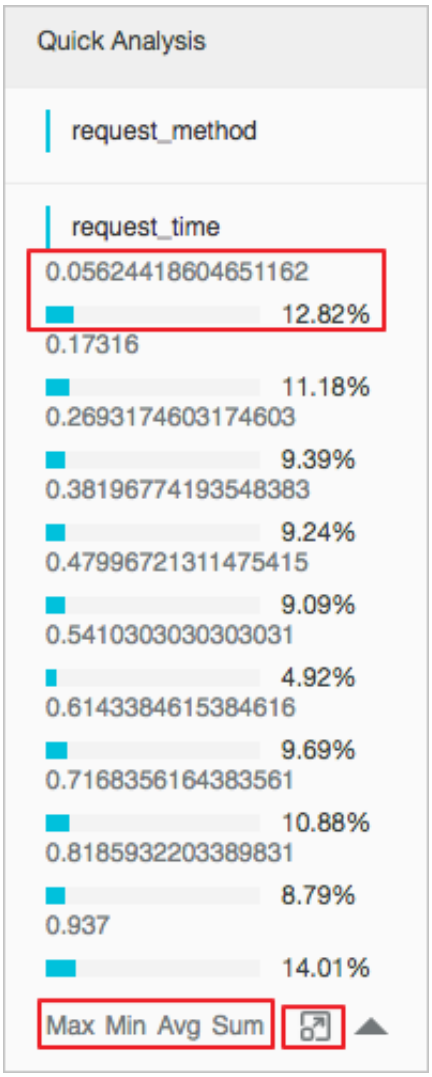
long/double

- **Histogram statistics for the approximate distribution**

  Grouping statistics is of little significance for the `long/double` fields, which have multiple type values. Therefore, histogram statistics for the approximate distribution is adopted by using 10 buckets.

  ```
  $Search | select numeric_histogram(10, ${keyName})
  ```

  `request_time` returns the following result based on the histogram statistics for the approximate distribution. You can see that the request time is mostly distributed around 0.056.



- **`MaxMinAvgSum`Statement Quick Analysis**

  Click `Max`, `Min`, `Avg`, and `Sum` under the target field to quickly search for the maximum value, minimum value, average, and sum of all `Max`.

- **Extend the query statement of grouping statistics to the search box**

  Click the button at the right of `Sum` to extend the query statement of the histogram statistics for the approximate distribution to the search box for further operations.

24.9.5.5 Saved search

Saved search is a one-click query and analysis function provided by Log Service.

Prerequisites

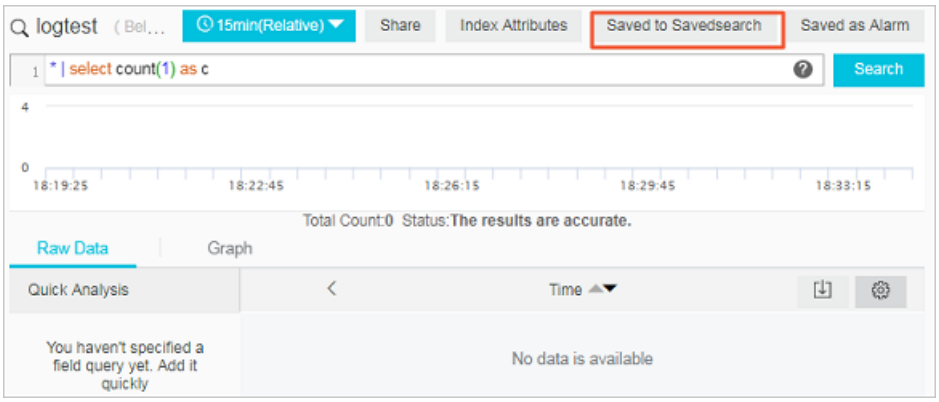
You have enabled and configured **Index**.

Context

If you need to frequently view the results of a query and analysis statement, save the statement as a saved search. In later result searches, you only need to click the name of the saved search on the left side of the search page. You can also use this saved search condition in alarm rules. Log Service executes the statement of this saved search periodically and sends an alarm notification when the search result meets the pre-set condition of the statement.

Procedure

1. *Log on to the Log Service console*.
2. Click a project name.
3. Click **Search** in the **LogSearch** column on the **Logstores** page.
4. Enter your query analysis statement, set the time range, and click **Search & Analysis**.
5. Click **Save Search** in the upper-right corner of the page.



6. Configure saved search attributes.
   a) Select an **Operation Type**.
      - **Save Search**: Saves the current query statement as a new saved search.
      - **Modify Saved Search**: Modify an existing saved search.
   b) Set **Saved Search Name**.
      - The name can only contain lowercase letters, numbers, hyphens (-), and underscores (_).
      - The name must start and end with a lowercase letter or a number.
      - This name must be a string of 3 to 63 characters.
   c) Confirm **Logstores**, **Topic**, and **Query**.

      If **Logstore** and **Topic** do not meet your requirements, return to the search page to access the target Logstore and enter your query statement, and then click **Save Search** again.
7. Click **OK**.

24.9.6 Analysis syntax and functions

This section lists the common syntax and functions supported by the search analysis feature of Log Service so as to provide reference for you.

24.9.6.1 General aggregate functions

The search analysis feature of Log Service supports general aggregate functions.

The search analysis feature of Log Service supports log analysis using general aggregate functions. The detailed statement and meaning are as follows:

Statement	Meaning	Example
`arbitrary(x)`	Returns a value in column x randomly.	`latency > 100 | select arbitrary(method)`
`avg(x)`	Calculates the arithmetic mean of all the values in column x.	`latency > 100 | select avg(latency)`
`checksum(x)`	Calculates the checksum of all the values in a column and returns the base64-encoded value.	`latency > 100 | select checksum(method)`
`count(*)`	Calculates the number of rows in a column.	-
`count(x)`	Calculates the number of non-null values in a column.	`latency > 100 | count(method)`
`count_if(x)`	Calculates the number of x = true.	`latency > 100 | count(url like '%abc')`
`geometric_mean(x)`	Calculates the geometric mean of all the values in a column.	`latency > 100 | select geometric_mean(latency)`
`max_by(x,y)`	Returns the value of column x when column y has the maximum value.	The method for the maximum latency: `latency>100 | select max_by(method, latency)`
`max_by(x,y,n)`	Returns the values of column x corresponding to n rows with maximum values in column y.	The method for the top 3 rows with maximum latency: `latency > 100 | select max_by(method,latency,3)`
`min_by(x,y)`	Returns the value of column x when column y has the minimum value.	The method for the minimum latency:`* | select min_by(x,y)`

Statement	Meaning	Example
`min_by(x,y,n)`	Returns the values of column x corresponding to n rows with minimum values in column y.	Search the methods(x) for the minimum 3(n) latency(y) values: `* | select max_by(method,latency,3)`
`max(x)`	Returns the maximum value.	`latency > 100| select max(inflow)`
`min(x)`	Returns the minimum value.	`latency > 100| select min(inflow)`
`sum(x)`	Returns the sum of all the values in column x.	`latency > 10 | select sum(inflow)`
`bitwise_and_agg(x)`	Do the AND calculation to all the values in a column.	-
`bitwise_or_agg(x)`	Do the OR calculation to all the values in a column.	-

24.9.6.2 Map functions

The search analysis feature of Log Service supports map functions.

The search analysis feature of Log Service supports log analysis using map functions. The detailed statement and meaning are as follows:

Statements	Meaning	Example
Subscript operator []	Obtains the results of a key in a map.	-
histogram(x)	Calculates the count grouped by the value of column x. Performs GROUP BY according to each value of column x and calculates the count. The syntax is equivalent to `select count group by x`.	`latency > 10 | histogram(status)` is equivalent to `latency > 10 | select count(1) group by status`.
map_agg(Key,Value)	Returns a map of key, value, and shows the random latency of each method.	`latency > 100 | select map_agg(method,latency)`
multimap_agg(Key,Value)	Returns a multi-value map of key, value, and returns all the latency for each method.	`latency > 100 | select multimap_agg(method, latency)`

cardinality(x) → bigint	Obtains the size of the map.	-
element_at(map<K, V>, key) → V	Obtains the value corresponding to the key.	-
map() → map<unknown, unknown>	Returns an empty map.	-
map(array<K>, array<V>) → map<K,V>	Converts two arrays into 1-to-1 maps.	SELECT map(ARRAY[1,3], ARRAY[2,4]); – {1 -> 2, 3 -> 4}
map_from_entries(array<row<K, V>>) → map<K,V>	Converts a multidimensional array into a map.	SELECT map_from_entries(ARRAY[(1, 'x'), (2, 'y')]); – {1 -> 'x', 2 -> 'y'}
map_entries(map<K, V>) → array<row<K,V>>	Converts an element in a map into an array.	SELECT map_entries(MAP(ARRAY[1, 2], ARRAY['x', 'y'])); – [ROW(1, 'x'), ROW(2, 'y')]
map_concat(map1<K, V>, map2<K, V>, …, mapN<K, V>) → map<K,V>	The Union of multiple maps is required, if a key exists in multiple maps, take the first one.	-
map_filter(map<K, V>, function) → map<K,V>	Refer to the lambda map_filter function.	-
transform_keys(map<K1, V>, function) → MAP<K2,V>	Refer to the lambda transform_keys function.	-
transform_values(map<K, V1>, function) → MAP<K,V2>	Refer to the lambda transform_values function.	-
map_keys(x<K, V>) → array<K>	Obtains all the keys in the map and returns an array.	-
map_values(x<K, V>) → array<V>	Obtains all values in the map and returns an array.	-
map_zip_with(map<K, V1>, map<K, V2>, function<K, V1, V2, V3>) → map<K,V3>	Refer to power functions in Lambda.	-

24.9.6.3 Estimating functions

The query and analysis function of Log Service supports estimating functions.

The query and analysis function of Log Service supports analyzing logs by using estimating functions. The specific statements and meanings are as follows.

Statements	Meaning	Example
`approx_distinct(x)`	Estimates the number of unique values in column x.	-
`approx_percentile(x, percentage)`	Sorts the column x and returns the value approximately at the given percentage position.	Returns the value at the half position: `approx_percentile(x,0.5)`
`approx_percentile(x, percentages)`	It is similar to the preceding statement, but you can specify multiple percentages to return the values at each specified percentage position.	`approx_percentile(x, array(0.1,0.2))`
`numeric_histogram( buckets, Value)`	Makes statistics on the value column in different buckets. Divides the value column into buckets number of buckets and returns the key and count of each bucket, which is equivalent to `select count group by`.	For post requests, divides the delay into 10 barrels, and returns the size of each bucket: method: `method:POST | select numeric_histogram(10, latency)`

24.9.6.4 Mathematical statistical functions

The query and analysis function of Log Service supports analyzing logs by using mathematical statistical functions. The specific statements and meanings are as follows.

Statements	Meaning	Example
`corr(y, x)`	Returns the correlation coefficient of two columns. The result is from 0 to 1.	`latency>100| select corr(latency,request_size)`
`covar_pop(y, x)`	Calculates the population covariance.	`latency>100| select covar_pop(request_size, latency)`
`covar_samp(y, x)`	Calculates the sample covariance.	`latency>100| select covar_samp(request_size,latency)`

Statements	Meaning	Example
`regr_intercept(y, x)`	Returns the linear regression intercept of input values. y is the dependent value, and x is the independent value.	`latency>100| select regr_intercept(request_size,latency)`
`regr_slope(y,x)`	Returns the linear regression slope of input values. y is the dependent value, and x is the independent value.	`latency>100| select regr_slope(request_size,latency)`
`stddev(x)` or `stddev_samp(x)`	Returns the sample standard deviation of column x.	`latency>100| select stddev(latency)`
`stddev_pop(x)`	Returns the population standard deviation of column x.	`latency>100| select stddev_pop(latency)`
`variance(x)` or `var_samp(x)`	Calculates the sample variance of column x.	`latency>100| select variance(latency)`
`var_pop(x)`	Calculates the population variance of column x.	`latency>100| select variance(latency)`

24.9.6.5 Mathematical functions

The query and analysis function of Log Service supports mathematical functions.

The query and analysis function of Log Service supports analyzing logs by using mathematic al functions. By combining query statements with mathematical functions, you can perform mathematical calculation to the log query results.

Mathematical operators

Mathematical operators such as the plus sign (+), minus sign (-), asterisk (*), slash (/), and percent sign (%) are supported. They can be used in the SELECT clause.

Example:

```
*|select avg(latency)/100 , sum(latency)/count(1)
```

Description of mathematical functions

Log Service supports the following mathematical functions:

Function Name	Meaning
`abs(x)`	Returns the absolute value of column x.
`cbrt(x)`	Returns the cube root of column x.

Function Name	Meaning
`ceiling ( x )`	Returns the number rounded up to the nearest integer of column x.
`cosine_similarity(x,y)`	Returns the cosine similarity between the sparse vectors x and y.
`degrees`	Converts radians to degrees.
`e()`	Returns the constant Euler's number.
`exp(x)`	Returns Euler's number raised to the power of x.
`floor(x)`	Returns x rounded down to the nearest integer.
`from_base(string,radix)`	Returns the value of string interpreted as a base-radix number.
`ln(x)`	Returns the natural logarithm of x.
`log2(x)`	Returns the base-2 logarithm of x.
`log10(x)`	Returns the base-10 logarithm of x
`log(x,b)`	Returns the base-b logarithm of x.
`pi()`	Returns π.
`pow(x,b)`	Returns x raised to the power of b.
`radians(x)`	Converts angle x in degrees to radians.
`rand()`	Returns a pseudo-random value in the range 0 .0 <= x < 1.0.
`random(0,n)`	Returns a pseudo-random number between 0 and n (exclusive).
`round(x)`	Returns x rounded to the nearest integer.
`round(x, y)`	Returns x rounded to y decimal places. For example, round(1.012345,2) = 1.01.
`sqrt(x)`	Returns the square root of x.
`to_base(x, radix)`	Returns the base-radix representation of x.
`truncate(x)`	Returns x rounded to integer by dropping digits after decimal point.
`acos(x)`	Returns the arc cosine of x.
`asin(x)`	Returns the arc sine of x.
`atan(x)`	Returns the arc tangent of x.

Function Name	Meaning
`atan2(y,x)`	Returns the arc tangent of y/x.
`cos(x)`	Returns the cosine of x.
`sin(x)`	Returns the sine of x.
`cosh(x)`	Returns the hyperbolic cosine of x.
`tan(x)`	Returns the tangent of x.
`tanh(x)`	Returns the hyperbolic tangent of x.
`infinity()`	Returns the constant representing positive infinity.
`is_infinity(x)`	Determine if x is finite.
`is_finity(x)`	Determine if x is infinite.
`is_nan(x)`	Determine if x is not-a-number.

24.9.6.6 String functions

The search analysis feature of Log Service supports string functions.

The search analysis feature of Log Service supports log analysis by using string functions. The detailed statement and meaning are as follows:

Function Name	Meaning
`length(x)`	Field length.
`levenshtein_distance(string1, string2)`	Returns the minimum Levenshtein distance between two strings.
`lower(string)`	Converts a string to lowercase characters.
`ltrim(string)`	Removes the leading whitespace.
`replace(string, search)`	Removes search from the string.
`replace(string, search,rep)`	Replaces search with rep in string.
`reverse(string)`	Reverse a string.
`rtrim(string)`	Removes trailing whitespace from the string.
`split(string,delimeter,limit)`	Splits the given string into substrings given a delimiter. Generates results in an array with subscripts starting with 1.
`split_part(string,delimeter,offset)`	Splits the string into substrings and returns the substrings in an array. The offset-th string will

Function Name	Meaning
	be returned. Generates results in an array with subscripts starting with 1.
`strpos(string, substring)`	Returns the starting position of the substring within the string. The position starts with 1. If not found, 0 is returned.
`substr(string, start)`	Returns substrings of the string with subscripts starting with 1.
`substr(string, start, length)`	Returns substrings of the string with subscripts starting with 1.
`trim(string)`	Removes leading and trailing whitespace from the string.
`upper(string)`	Converts the string to uppercase.
`concat(string,string......)`	Joins two or more strings into one.
`hamming_distance (string1,string2)`	Returns the Hamming distance between two strings.

Note:

The strings are enclosed by single quotation marks. A column name is enclosed by double quotation marks. For example: In a = abc, a = string abc; in a = abc, column a = column abc.

24.9.6.7 Date and time functions

Log Service supports time functions and date functions. You can use the date and time functions introduced in this document in the analysis syntax.

Date and time

1. unixtime: The number of seconds since January 1, 1970 in the type of int. For example, `1512374067` indicates the time `Mon Dec 4 15:54:27 CST 2017`. In Log Service, The built-in time `__time__` in each log of Log Service is of this type.
2. timestamp type: Indicates the time in the format of string. For example, `2017-11-01 13:30:00`.

Date Functions

Log Service supports the following common date functions:

Function Name	Meaning	Example
`current_date`	Returns the current date.	`latency>100| select current_date`
`current_time`	hour:minute; second, millisecond time zone	`latency>100| select current_time`
`current_timestamp`	Returns the result combined by current_date and current_time .	`latency>100| select current_timestamp`
`current_timezone()`	Returns the time zone.	`latency>100| select current_timezone()`
`from_iso8601_timestamp(string)`	Converts an ISO8601 to a timestamp with time zone.	`latency>100| select from_iso8601_timestamp(iso8601)`
`from_iso8601_date(string)`	Converts an ISO8601 to a date .	`latency>100| select from_iso8601_date(iso8601)`
`from_unixtime(unixtime)`	Converts a Unix time to a timestamp.	`latency>100| select from_unixtime(1494985275)`
`from_unixtime(unixtime, string)`	Converts a Unix time to a timestamp using the string as the time zone.	`latency>100| select from_unixtime (1494985275,Asia/Shanghai)`
`localtime`	Returns the current time.	`latency>100| select localtime`
`localtimestamp`	Returns the current timestamp.	`latency>100| select localtimestamp`
`now()`	Equivalent to `current_timestamp`.	-
`to_unixtime(timestamp)`	Converts a timestamp to a Unix time.	`*| select to_unixtime(2017-05-17 09:45:00.848 Asia/Shanghai)`

Time Functions

MySQL time formats

Log Service supports the following MySQL time formats: %a, %b, and %y.

Function Name	Meaning	Example
`date_format(timestamp, format)`	Formats timestamp into a string using format.	`latency>100| select date_format (date_parse (2017-05-17 09:45:00,%Y-%m-%d %H:%i:%S), %Y-%m-%d) group by method`
`date_parse(string, format)`	Parses the string into a timestamp using format.	`latency>100|select date_parse(2017-05-17 09:45:00,%Y-%m-%d %H:%i:%S) group by method`

Table 24-9: Format Description

Format	Description
%a	Days of the week in abbreviated form (Sun .. Sat).
%b	Months in abbreviated form (Jan .. Dec).
%c	Month, numerical type (1 .. 12) [4].
%D	The day of the month with the suffix (0th, 1st, 2nd, 3rd, …).
%d	The day of the month (01 .. 31) [4].
%e	The day of the month (1 .. 31) [4].
%H	The hour (00 .. 23).
%h	The hour (01 .. 12).
%I	The hour in 12-hour format (01 .. 12).
%i	The minute (00 .. 59).
%j	The day of the year (001 .. 366).
%k	The hour (0 .. 23).
%l	The hour (1 .. 12).
%M	The month in English (January .. Dece mber).
%m	The month in number (01 .. 12) [4].
%p	AM or PM.
%r	The time in 12-hour format. The format is hh: mm:ss followed by AM or PM.

Format	Description
%S	The second (00 .. 59).
%s	The second (00 .. 59).
%T	The time in 24-hour format (hh:mm:ss).
%U	The week of the year (00 .. 53).
%u	The week of the year (00 .. 53).
%V	The week of the year (01 .. 53).
%v	The week of the year (01 .. 53), where Monday is the first day of the week; used with %x.
%W	The name of a day in a week (Sunday .. Saturday).
%w	The day of the week (0 .. 6). Sunday is the day 0.
%Y	The year.
%y	The year. Double digit.
%%	%escape character

Time period alignment functions

Log Service supports time period alignment functions, which can be aligned according to seconds , minutes, hours, days, months, and years. Time period alignment functions are usually used when statistics are made according to time.

Function syntax:

```
date_trunc(unit, x)
```

Parameters:

The optional values for Unit are as follows (x is `2001-08-22 03:04:05.000`):

Unit	Converted result
second	2001-08-22 03:04:05.000
minute	2001-08-22 03:04:00.000
hour	2001-08-22 03:00:00.000
day	2001-08-22 00:00:00.000
week	2001-08-20 00:00:00.000

Unit	Converted result
month	2001-08-01 00:00:00.000
quarter	2001-07-01 00:00:00.000
year	2001-01-01 00:00:00.000

x can be a timestamp type or Unix time.

date_trunc is only applicable to statistics at a fixed time interval. For statistics based on flexible time dimensions, for example, every 5 minutes, perform GROUP BY according to the mathematic al modulus method.

```
* | SELECT count(1) as pv, __time__ - __time__% 300 as minute5groupby
minute5 limit 100
```

In the preceding formula, `%300` indicates to make the modulus and alignment every five minutes.

Date function example

The following is a comprehensive example using time formats:

```
*|select date_trunc(minute , __time__) as t, truncate (avg(latency
) ) , current_date group by t order by t desc limit 60
```

24.9.6.8 URL functions

The search analysis feature of Log Service supports URL functions.

URL functions support extracting fields from standard URL paths. A standard URL is as follows:

```
[protocol:][//host[:port]][path][?query][#fragment]
```

Common URL functions

Function Name	Meaning	Example
`url_extract_fragment(url)`	Extracts the fragment from a URL and the result is of varchar type.	`*| select url_extract_fragment(url)`
`url_extract_host(url)`	Extracts the host from a URL and the result is of varchar type.	`*|select url_extract_host(url)`
`url_extract_parameter(url, name)`	Extracts the value of the name parameter in the query from a URL and the result is of varchar type.	`*|select url_extract_parameter(url)`

Function Name	Meaning	Example
`url_extract_path(url)`	Extracts the path from a URL and the result is of varchar type.	`*|select url_extract_path(url)`
`url_extract_port(url)`	Extracts the port from a URL and the result is of bigint type.	`*|select url_extract_port(url)`
`url_extract_protocol(url)`	Extracts the protocol from a URL and the result is of varchar type.	`*|select url_extract_protocol(url)`
`url_extract_query(url)`	Extracts the query from a URL and the result is of varchar type.	`*|select url_extract_query(url)`
`url_encode(value)`	Encodes a URL.	`*|select url_encode(url)`
`url_decode(value)`	Decodes a URL.	`*|select url_decode(url)`

24.9.6.9 Regular expression functions

The query and analysis function of Log Service supports regular expression functions.

A regular expression function parses a string and returns the needed substrings.

The common regular expression functions and the meanings are as follows:

Function name	Meaning	Example
`regexp_extract_all(string, pattern)`	Returns all the substrings that match the regular expression in the string as a string array.	The result of `|*SELECT regexp_extract_all(5a 67b 890m, \d+)` is `[5,67,890]`, and the result of `*|SELECT regexp_extract_all(5a 67a 890m, (\d+)a)` is `[5a,67a]`.
`regexp_extract_all(string, pattern, group)`	Returns the part of the string that hits the regular () part of the group, returns the result as an array of strings.	The result of `*| ` SELECT regexp_extract_all('5a 67a 890m', '(\d+)a',1)` is `['5','67']`.
`regexp_extract(string, pattern)`	Returns the first substring that hits the regular expression in the string.	The result of `*|SELECT regexp_extract(5a 67b 890m, \d+)` is 5.

Function name	Meaning	Example
`regexp_extract(string, pattern,group)`	Returns the first substring in the (group)th () that hits the regular expression in the string .	The result of `*|SELECT regexp_extract(5a 67b 890m, (\d+)([a-z]+),2)` is `b`.
`regexp_like(string, pattern)`	Determines if the string matches the regular expression and returns a bool result. The regular expression is allowed to match part of the string.	The result of `*|SELECT regexp_like(5a 67b 890m, \d+m)` is true.
`regexp_replace(string, pattern, replacement)`	Replaces the part that matches the regular expression in the string with replacement.	The result of `*|SELECT regexp_replace(5a 67b 890m, \d+,a)` is `aa ab am`.
`regexp_replace(string, pattern)`	Removes the part that matches the regular expression in the string, which is equivalent to `regexp_replace(string,patterm,)`.	The result of `*|SELECT regexp_replace(5a 67b 890m, \d+)` is `a b m`.
`regexp_split(string, pattern)`	Splits the string to an array by using the regular expression.	The result of `*|SELECT regexp_split(5a 67b 890m, \d+)` is `[a,b,m]`.

24.9.6.10 JSON functions

The query and analysis function of Log Service supports JSON functions.

JSON functions can parse a string as the JSON type and extract the fields in JSON. JSON mainly has the following two structures: map and array. If a string fails to be parsed as the JSON type, the returned value is null.

Log Service supports the following common JSON functions:

Function Name	Meaning	Example
`json_parse(string)`	Converts the string into the JSON type.	`SELECT json_parse([1, 2, 3])`The result is an array of the JSON type.
`json_format(json)`	Converts the JSON type into a string.	`SELECT json_format(json_parse([1, 2, 3]))`The result is a string.

Function Name	Meaning	Example
`json_array_contains(json, value)`	Judges whether a value of the JSON type or a string (with the content of a JSON array) contains a specific value.	`SELECT json_array_contains(json_parse([1, 2, 3]), 2) or SELECT json_array_contains([1, 2, 3], 2)`
`json_array_get(json_array, index)`	Like `json_array_contains`, but obtains the element of a subscript of a JSON array.	`SELECT json_array_get(["a", "b", "c"], 0) returns a`
`json_array_length(json)`	returns the size of JSON array.	`SELECT json_array_length([1, 2, 3]) returns result 3.`
`json_extract(json, json_path)`	indicates to extract values from a JSON object. The JSON path syntax is similar to `$.store.book[0].title`. A JSON object is returned.	`SELECT json_extract(json, $.store.book);`
`json_extract_scalar(json, json_path)`	is similar to `json_extract`, but returns a string.	-
`json_size(json, json_path)`	Returns the size of JSON object or array.	`SELECT json_size([1, 2, 3]) returns result 3.`

24.9.6.11 Type conversion functions

The search analysis feature of Log Service supports type conversion functions.

Log Service supports data types such as Long, Double, and Textin the configurations. Supported types for query include Bigint, Double, Varchar, Timestamp, and Int.

The type conversion function forcibly converts a column to the specified data type:

```
try_cast(value AS type) → type
```

24.9.6.12 GROUP BY syntax

The search analysis feature of Log Service supports the GROUP BY syntax.

GROUP BY supports multiple columns and indicating the corresponding KEY by using the SELECT column alias.

Example:

```
method:PostLogstoreLogs |select avg(latency),projectName,date_trunc(
hour,__time__) as hour group by projectName,hour
```

The alias hour indicates the third SELECT column `date_trunc(hour,__time__)`. This is very helpful for complex queries.

GROUP BY supports GROUPING SETS, CUBE, and ROLLUP.

Example:

```
method:PostLogstoreLogs |select avg(latency) group by cube(projectNam
e,logstore) method:PostLogstoreLogs |select avg(latency) group by
GROUPING SETS ( ( projectName,logstore), (projectName,method)) method
:PostLogstoreLogs |select avg(latency) group by rollup(projectName,
logstore)
```

Example

Perform GROUP BY according to time

Each log has a built-in time column `__time__`. When the statistical function of any column is activated, the statistics will be automatically made for the time column.

Use the `date_trunc` function to align the time column to hour, minute, day, month, and year. `date_trunc` accepts an aligned unit and a Unix time or timestamp type column, such as `__time__`.

- PV statistics per hour and per minute

  ```
  * | SELECT count(1) as pv , date_trunc(hour,__time__) as hour
  group by hour order by hour limit 100 * | SELECT count(1) as pv
   , date_trunc(minute,__time__) as minute group by minute order by
  minute limit 100
  ```

  Note:

  limit 100 indicates that up to 100 rows can be retrieved. If the LIMIT statement is not added, up to 10 rows of data can be retrieved by default.

- date_trunc is only applicable to statistics at a fixed time interval. For statistics based on flexible time dimensions, for example, every 5 minutes, perform GROUP BY in mod.

  ```
  * | SELECT count(1) as pv, __time__ - __time__% 300 as minute5 group
   by minute5 limit 100
  ```

  In the preceding formula, `%300` indicates making the modulus and alignment every five minutes.

Retrieve non-agg columns in GROUP BY

In standard SQL, if the GROUP BY syntax is used during the SELECT operation, the system only selects the original content of the SELECT GROUP BY column, or performs aggregate computing on any columns. Retrieving content from non-GROUP BY columns is not allowed.

For example, the following syntax is invalid. Because b is a non-GROUP BY column, the system cannot determine which row of b to output during GROUP BY based on a.

```
*|select a, b , count(c) gropu by a
```

Instead, you can use the `arbitrary` function to output b:

```
*|select a, arbitrary(b), count(c) gropu by a
```

24.9.6.13 Window functions

The query and analysis function of Log Service supports window functions.

Window functions are used for cross-row calculation. SQL aggregate functions calculate the results of only one row or aggregate all rows into one row for calculation. Window functions support cross-row calculation and fill the calculation results in each row.

Syntax of window functions:

```
SELECT key1, key2, value, rank() OVER (PARTITION BY key2 ORDER BY
value DESC) AS rnk FROM orders ORDER BY key1,rnk
```

Core part:

```
rank() OVER (PARTITION BY KEY1 ORDER BY KEY2 DESC)
```

rank() is an aggregate function.You can use any function in analysis syntax or the function listed in this document. PARTITION BY indicates the buckets based on which values are calculated.

Special aggregate functions used in windows

Function Name	Meaning
rank()	Sorts data based on a specific column in a window and returns the serial numbers in the window.
row_number()	Returns the row numbers in the window.
first_value(x)	Returns the first value in the window. It is typically used to obtain the maximum value after values are sorted in the window.

Function Name	Meaning
last_value(x)	Opposite to first_value.
nth_value(x, offset)	Number of a specific offset in the window.
lead(x,offset,defaut_value)	Value of the No. offset row after a certain row in xth column in the window. If that row does not exist, use the default_value.
lag(x,offset,defaut_value)	Value of the No. offset row before a certain row in xth column in the window. If that row does not exist, use the default_value.

Example

- Rank the salaries of employees in their respective departments

```
* | select department, persionId, sallary , rank() over(PARTITION
 BY department order by sallary desc) as sallary_rank order by
department,sallary_rank
```

Response results:

department	persionId	sallary	sallary_rank
dev	john	9000	1
dev	Smith	8000	2
dev	Snow	7000	3
dev	Achilles	6000	4
Marketing	Blan Stark	9000	1
Marketing	Rob Stark	8000	2
Marketing	Sansa Stark	7000	3

- Calculate the salaries of employees as percentages in their respective departments

```
* | select department, persionId, sallary *1.0 / sum(sallary) over(
PARTITION BY department ) as sallary_percentage
```

Response results:

department	persionId	sallary	sallary_percentage
dev	john	9000	0.3
dev	Smith	8000	0.26
dev	Snow	7000	0.23

department	persionId	sallary	sallary_percentage
dev	Achilles	6000	0.2
Marketing	Blan Stark	9000	0.375
Marketing	Rob Stark	8000	0.333
Marketing	Sansa Stark	7000	0.29

- Calculate the daily UV increase over the previous day

```
* | select day ,uv, uv *1.0 /(lag(uv,1,0) over() ) as diff_perce
ntage from ( select approx_distinct(ip) as uv, date_trunc(day,
__time__) as day from log group by day order by day asc )
```

Response results:

day	uv	diff_percentage
2017-12-01 00:00:00	100	null
2017-12-02 00:00:00	125	1.25
2017-12-03 00:00:00	150	1.2
2017-12-04 00:00:00	175	1.16
2017-12-05 00:00:00	200	1.14
2017-12-06 00:00:00	225	1.125
2017-12-07 00:00:00	250	1.11

24.9.6.14 HAVING syntax

The search analysis feature of Log Service supports the HAVING syntax of standard SQL, which is used with GROUP BY to filter GROUP BY results.

Format:

```
method :PostLogstoreLogs |select avg(latency),projectName group by
projectName HAVING avg(latency) > 100
```

Differences between HAVING and WHERE

HAVING filters the results of aggregate computing after GROUP BY, and WHERE filters raw data between aggregate computing operations.

Example

Calculate the average rainfall of each province where temperature is above 10°C, and display only the provinces with average rainfall above 100 mL in the final results:

```
* | select avg(rain) ,province where teporature > 10groupby province
having avg(rain) > 100
```

24.9.6.15 ORDER BY syntax

The query and analysis function of Log Service supports the ORDER BY syntax.

ORDER BY is used to sort results. Currently, you can only sort results by one column.

Syntax format:

```
orderby Column name [desc|asc]
```

Example:

```
method :PostLogstoreLogs |select avg(latency) as avg_latency,
projectName group by projectName HAVING avg(latency) > 5700000 order
by avg_latency desc
```

24.9.6.16 LIMIT syntax

The query and analysis function of Log Service supports the LIMIT syntax.

LIMIT is followed by a number to restrict the maximum number of lines in the results. If no LIMIT statement is added, only 10 lines are returned by default.

Note:
LIMIT OFFSET and LINES syntaxes are not supported.

Example:

```
*| select avg(latency) as avg_latency , methodgroupbymethodorderbyavg_
latencydesclimit100
```

24.9.6.17 CASE WHEN syntax

The query and analysis function of Log Service supports the CASE WHEN syntax.

The CASE WHEN syntax is supported for classification of continuous data. For example, you can extract information from http_user_agent and classify the information into Android and iOS types:

```
SELECT CASE WHEN http_user_agent like %android% then android WHEN
http_user_agent like %ios% then ios ELSE unknown END as http_user_
agent, count(1) as pv group by http_user_agent
```

Example

- Proportion of requests with status code 200 in all requests:

  ```
  * | SELECT sum( CASE WHEN status =200 then 1 ELSE 0 end ) *1.0 /
  count(1) as status_200_percentage
  ```

- Distribution of latencies:

  ```
  * | SELECT ` CASE WHEN latency < 10 then s10 WHEN latency < 100
  then s100 WHEN latency < 1000 then s1000 WHEN latency < 10000 then
   s10000 else s_large end as latency_slot, count(1) as pv group by
  latency_slot
  ```

IF syntax

The IF syntax is logically equivalent to the CASE WHEN syntax.

```
CASE WHEN condition THEN true_value [ ELSE false_value ] END
```

- if(condition, true_value)

  — If the condition is true, the true_value column is returned; otherwise, null is returned.

- if(condition, true_value, false_value)

  — If the condition is true, the true_value column is returned; otherwise, the false_value column is returned.

COALESCE syntax

The coalesce function returns the first non-null value of multiple columns.

```
coalesce(value1, value2[,...])
```

NULLIF syntax

If value1 equals value2, null is returned; otherwise, value1 is returned.

```
nullif(value1, value2)
```

TRY syntax

The TRY syntax captures some underlying exceptions, for example, returning null for an incorrect division by zero.

```
try(expression)
```

24.9.6.18 Nested subquery

The query and analysis function of Log Service supports nested subquery.

You can use a nested SQL query in complicated query scenarios where a single SQL layer does not meet the requirement.

The difference between nested subquery and non-nested query is that you need to specify the from condition in the SQL statement. You need to specify the `from log` keyword in the query to read raw data from logs.

Example:

```
* | select sum(pv) from ( select count(1) as pv from log group by
method )
```

24.9.6.19 Arrays

The query and analysis function of Log Service supports arrays.

Statement	Meaning	Example
Subscript operator []	Obtains a certain element from the array.	-
Connection operator ||	Connects two arrays into one.	`SELECT ARRAY [1] || ARRAY [2]; – [1, 2]` `SELECT ARRAY [1] || 2; – [1, 2]`

Statement	Meaning	Example
		`SELECT 2 || ARRAY [1]; — [2, 1]`
array_distinct	Obtain the distinct elements in the array by means of array deduplication.	-
array_intersect(x, y)	Obtains the intersection of array x and array y.	-
array_union(x, y) → array	Obtains the union of array x and array y.	-
array_except(x, y) → array	Returns an array of elements in x but not in y , without duplicates.	-
array_join(x, delimiter, null_replacement) → varchar	Concatenates the elements of the given array using the delimiter and an optional string to replace nulls.	-
array_max(x) → x	Obtains the maximum value in array x.	
array_min(x) → x	Obtains the minimum value in array x.	-
array_position(x, element) → bigint	Returns the position of the first occurrence of the element in array x (or 0 if not found).	-
array_remove(x, element) → array	Removes all elements that equal element from array x.	-
array_sort(x) → array	Sorts and returns the array x . The elements of x must be orderable. Null elements will be placed at the end of the returned array.	-
cardinality(x) → bigint	Returns the cardinality (size) of the array x.	-
concat(array1, array2, …, arrayN) → array	Concatenates arrays.	-
contains(x, element) → boolean	Returns TRUE if the array x contains the element.	-

Statement	Meaning	Example
filter(array, function) → array	This is a Lambda function. See filter() in *Lambda function*.	-
flatten(x) → array	Flattens an array(array(T)) to an array(T) by concatenating the contained arrays.	-
reduce(array, initialState, inputFunction, outputFunction) → x	See *Lambda function* reduce.	-
reverse(x) → array	Returns an array which has the reversed order of array x.	-
sequence(start, stop) → array	Generate a sequence of integers from start to stop, incrementing by 1 if start is less than or equal to stop, otherwise -1.	-
sequence(start, stop, step) → array	Generate a sequence of dates from start to stop, incrementing by step.	-
sequence(start, stop, step) → array	Generate a sequence of timestamps from start to stop , incrementing by step. The type of step can be either INTERVAL DAY TO SECOND or INTERVAL YEAR TO MONTH.	-
shuffle(x) → array	Generate a random permutation of the given array x.	-
slice(x, start, length) → array	Subsets array x starting from index start (or starting from the end if start is negative) with a length of length.	-
transform(array, function) → array	See *Lambda function* transform().	-
zip(array1, array2[, ...]) → array	Merges the given arrays, element-wise, into a single array of rows. The M-th element of the N-th argument will be the N-th field of the	`SELECT zip(ARRAY[1, 2], ARRAY['1b', null, '3b']); – [ROW(1, '1b`

Statement	Meaning	Example
	M-th output element. If the arguments have an uneven length, missing values are filled with NULL.	`'), ROW(2, null), ROW( null, '3b')]`
zip_with(array1, array2, function) → array	See *Lambda function* zip_with.	-

24.9.6.20 Binary string functions

The query and analysis function of Log Service supports binary string functions.

The binary string type varbinary is different from the string type varchar.

Statement	Description
Connection function ||	The result of `a || b` is `ab`.
length(binary) → bigint	Returns the length of binary in bytes.
concat(binary1, …, binaryN) → varbinary	Returns the concatenation of binary1, binary2, ..., binaryN. This function provides the same functionality as the SQL-standard concatenation operator (||).
to_base64(binary) → varchar	Encodes the binary string into a base64 string representation.
from_base64(string) → varbinary	Decodes binary data from the base64 encoded string.
to_base64url(binary) → varchar	Encodes binary data into a base64 string representation using the URL safe alphabet.
from_base64url(string) → varbinary	Decodes binary data from the base64 encoded string using the URL safe alphabet.
to_hex(binary) → varchar	Encodes the binary string into a hex string representation.
from_hex(string) → varbinary	Decodes binary data from the hex encoded string.
to_big_endian_64(bigint) → varbinary	Encodes bigint in a 64-bit 2's complement big endian format.
from_big_endian_64(binary) → bigint	Decodes bigint value from a 64-bit 2's complement big endian binary.
md5(binary) → varbinary	Computes the md5 hash of the binary string.

Statement	Description
sha1(binary) → varbinary	Computes the sha1 hash of the binary string.
sha256(binary) → varbinary	Computes the sha256 hash of the binary string.
sha512(binary) → varbinary	Computes the sha512 hash of the binary string.
xxhash64(binary) → varbinary	Computes the xxhash64 hash of the binary string.

24.9.6.21 Bit operation

The query and analysis function of Log Service supports bit operation functions.

Statement	Description	Example

bit_count(x, bits) → bigint	Collects the number of 1 in the binary expression of x.	`SELECT bit_count(9, 64); – 2`
		`SELECT bit_count(9, 8); – 2`
		`SELECT bit_count(-7, 64); – 62`
		`SELECT bit_count(-7, 8); – 6`

bitwise_and(x, y) → bigint	Performs the AND operation on x and y in binary.	-
bitwise_not(x) → bigint	Calculates the opposite values of all bits of x in binary.	-
bitwise_or(x, y) → bigint	Performs the OR operation on x and y in binary.	-
bitwise_xor(x, y) → bigint	Performs the XOR operation on x and y in binary.	-

24.9.6.22 Comparison functions and operators

The search analysis feature of Log Service supports comparison functions and operators.

Comparison functions and operators

A comparison operation compares the values of two parameters, which can be used for any comparable types, such as int, bigint, double, and text.

Comparison operators

A comparison operator is used to compare two parameter values. During the comparison, if the logic is true, TRUE is returned. Otherwise, FALSE is returned.

Operators	Meaning
<	Less than
>	Greater than
<=	Smaller than or equal to
>=	Greater than or equal to
=	Equal to
<>	Not equal to
!=	Not equal to

Range operator

The BETWEEN operator in WHERE clause is used to select data within a given range of values.

- If the logic is true, TRUE is returned. Otherwise, FALSE is returned.

  Example: `SELECT 3 BETWEEN 2 AND 6;`The logic is true, and TRUE is returned.

  The previous example is equivalent to `SELECT 3 >= 2 AND 3 <= 6;`.
- You can combine the BETWEEN operator with the NOT operator to find rows whose column values are not in a range of values.

  Example: `SELECT 3 NOT BETWEEN 2 AND 6;`The logic is false, and FALSE is returned.

  The previous example is equivalent to `SELECT 3 < 2 OR 3 > 6;`.
- If the value of any parameter is NULL, NULL is returned.

IS NULL and IS NOT NULL

These operators are used to determine whether a parameter value is NULL.

IS DISTINCT FROM and IS NOT DISTINCT FROM

These operators are like the comparison operators, but they can determine whether a NULL value exists.

Example:

```
SELECT NULL IS DISTINCT FROM NULL; -- false SELECT NULL IS NOT
DISTINCT FROM NULL; -- true
```

As described in the following table, DISTINCT can be used to compare parameter values under multiple conditions.

a	b	a = b	a <> b	a DISTINCT b	a NOT DISTINCT b
1	1	TRUE	FALSE	FALSE	TRUE
1	2	FALSE	TRUE	TRUE	FALSE
1	NULL	NULL	NULL	TRUE	FALSE
NULL	NULL	NULL	NULL	FALSE	TRUE

GREATEST and LEAST

These operators are used to obtain the maximum and minimum values across many columns.

Example:

```
select greatest(1,2,3) ; -- 3 is returned.
```

Comparison conditions: ALL, ANY, and SOME

Comparison conditions are used to determine whether a parameter meets the specified conditions .

- ALL is used to determine whether a parameter meets all the conditions. If the logic is true, TRUE is returned. Otherwise, FALSE is returned.
- ANY is used to determine whether a parameter meets any of the conditions. If the logic is true, TRUE is returned. Otherwise, FALSE is returned.
- Same as ANY, SOME is used to determine whether a parameter meets any of the conditions.
- ALL, ANY, and SOME must follow the comparison operators.

As described in the following table, ALL and ANY support comparison and determination under multiple conditions.

Expression	Meaning
A = ALL (…)	When A is equal to all values, TRUE is returned.
A <> ALL (…)	When A is not equal to all values, TRUE is returned.
A < ALL (…)	When A is smaller than all values, TRUE is returned.
A = ANY (…)	When A is equal to any value, TRUE is returned. It is equivalent to A IN (...).

Expression	Meaning
A <> ANY (…)	When A is not equal to any value, TRUE is returned.
A < ANY (…)	When A is smaller than the greatest value, TRUE is returned.

Example:

```
SELECT hello = ANY (VALUES hello, world); -- true SELECT 21 < ALL (
VALUES 19, 20, 21); -- false SELECT 42 >= SOME (SELECT 41 UNION ALL
SELECT 42 UNION ALL SELECT 43); -- true
```

24.9.6.23 Lambda function

The query and analysis function of Log Service supports Lambda functions.

Lambda expressions

Lambda expressions are written with

```
->
```

.

Example:

```
x -> x + 1
(x, y) -> x + y
x -> regexp_like(x, 'a+')
x -> x[1] / x[2]
x -> IF(x > 0, x, -x)
x -> COALESCE(x, 0)
x -> CAST(x AS JSON)
x -> x + TRY(1 / 0)
```

Most MySQL expressions can be used in Lambda.

filter(array<T>, function<T, boolean>) → ARRAY<T>

Filters data from an array and obtains only elements for which the function returns TRUE.

Example:

```
SELECT filter(ARRAY [], x -> true); -- []
SELECT filter(ARRAY [5, -6, NULL, 7], x -> x > 0); -- [5, 7]
SELECT filter(ARRAY [5, NULL, 7, NULL], x -> x IS NOT NULL); -- [5, 7]
```

map_filter(map<K, V>, function<K, V, boolean>) → MAP<K,V>

Filters data from a map and obtains only element pairs for which the function returns TRUE.

Example:

```
SELECT map_filter(MAP(ARRAY[], ARRAY[]), (k, v) -> true); -- {}
SELECT map_filter(MAP(ARRAY[10, 20, 30], ARRAY['a', NULL, 'c']), (k, v
) -> v IS NOT NULL); -- {10 -> a, 30 -> c}
SELECT map_filter(MAP(ARRAY['k1', 'k2', 'k3'], ARRAY[20, 3, 15]), (k,
v) -> v > 10); -- {k1 -> 20, k3 -> 15}
```

reduce(array<T>, initialState S, inputFunction<S, T, S>, outputFunction<S, R>) → R

The reduce function retrieves each element in the array from the initial state, calculates inputFunct ion(s,t) based on the state S, and generates a new state. It finally applies outputFunction to output the final state S to result R.

1. Initial state S
2. Retrieves each element T.
3. Calculates inputFunction(S,T) to generate a new state S.
4. Repeats Steps 2 and 3 to the last element and generate a new state.
5. Uses the final state S to obtain the final output result R.

Example:

```
SELECT reduce(ARRAY [], 0, (s, x) -> s + x, s -> s); -- 0
SELECT reduce(ARRAY [5, 20, 50], 0, (s, x) -> s + x, s -> s); -- 75
SELECT reduce(ARRAY [5, 20, NULL, 50], 0, (s, x) -> s + x, s -> s);
 -- NULL
SELECT reduce(ARRAY [5, 20, NULL, 50], 0, (s, x) -> s + COALESCE(x, 0
), s -> s); -- 75
SELECT reduce(ARRAY [5, 20, NULL, 50], 0, (s, x) -> IF(x IS NULL, s, s
 + x), s -> s); -- 75
SELECT reduce(ARRAY [2147483647, 1], CAST (0 AS BIGINT), (s, x) -> s
 + x, s -> s); -- 2147483648
SELECT reduce(ARRAY [5, 6, 10, 20], -- calculates arithmetic average:
10.25
              CAST(ROW(0.0, 0) AS ROW(sum DOUBLE, count INTEGER)),
              (s, x) -> CAST(ROW(x + s.sum, s.count + 1) AS ROW(sum
DOUBLE, count INTEGER)),
              s -> IF(s.count = 0, NULL, s.sum / s.count));
```

transform(array<T>, function<T, U>) → ARRAY<U>

Calls function for each element in the array to generate the new result U.

Example:

```
SELECT transform(ARRAY [], x -> x + 1); -- []
SELECT transform(ARRAY [5, 6], x -> x + 1); -- [6, 7] --Increments
each element by 1.
SELECT transform(ARRAY [5, NULL, 6], x -> COALESCE(x, 0) + 1); -- [6,
1, 7]
SELECT transform(ARRAY ['x', 'abc', 'z'], x -> x || '0'); -- ['x0', '
abc0', 'z0']
```

```
SELECT transform(ARRAY [ARRAY [1, NULL, 2], ARRAY[3, NULL]], a ->
filter(a, x -> x IS NOT NULL)); -- [[1, 2], [3]]
```

transform_keys(map<K1, V>, function<K1, V, K2>) → MAP<K2,V>

Applies the function for each key of the map to generate a new key.

Example:

```
SELECT transform_keys(MAP(ARRAY[], ARRAY[]), (k, v) -> k + 1); -- {}
SELECT transform_keys(MAP(ARRAY [1, 2, 3], ARRAY ['a', 'b', 'c']), (k
, v) -> k + 1); -- {2 -> a, 3 -> b, 4 -> c}  --Increments each key by
1.
SELECT transform_keys(MAP(ARRAY ['a', 'b', 'c'], ARRAY [1, 2, 3]), (k
, v) -> v * v); -- {1 -> 1, 4 -> 2, 9 -> 3}
SELECT transform_keys(MAP(ARRAY ['a', 'b'], ARRAY [1, 2]), (k, v) -> k
 || CAST(v as VARCHAR)); -- {a1 -> 1, b2 -> 2}
SELECT transform_keys(MAP(ARRAY [1, 2], ARRAY [1.0, 1.4]), -- {one ->
1.0, two -> 1.4}
                      (k, v) -> MAP(ARRAY[1, 2], ARRAY['one', 'two'])[
k]);
```

transform_values(map<K, V1>, function<K, V1, V2>) → MAP<K, V2>

Applies the function for all values in the map, converts V1 to V2, and generates a new map `<K, V2>`.

```
SELECT transform_values(MAP(ARRAY[], ARRAY[]), (k, v) -> v + 1); -- {}
SELECT transform_values(MAP(ARRAY [1, 2, 3], ARRAY [10, 20, 30]), (k,
v) -> v + 1); -- {1 -> 11, 2 -> 22, 3 -> 33}
SELECT transform_values(MAP(ARRAY [1, 2, 3], ARRAY ['a', 'b', 'c']), (
k, v) -> k * k); -- {1 -> 1, 2 -> 4, 3 -> 9}
SELECT transform_values(MAP(ARRAY ['a', 'b'], ARRAY [1, 2]), (k, v) -
> k || CAST(v as VARCHAR)); -- {a -> a1, b -> b2}
SELECT transform_values(MAP(ARRAY [1, 2], ARRAY [1.0, 1.4]), -- {1 ->
one_1.0, 2 -> two_1.4}
                        (k, v) -> MAP(ARRAY[1, 2], ARRAY['one', 'two
'])[k] || '_' || CAST(v AS VARCHAR));
```

zip_with(array<T>, array<U>, function<T, U, R>) → array<R>

Merges two arrays, and specifies the elements of the new array using the function. Element T in the first array and element U in the second array are used to generate the new result R.

Example:

```
SELECT zip_with(ARRAY[1, 3, 5], ARRAY['a', 'b', 'c'], (x, y) -> (y, x
)); --Transposes the positions of the elements of the first and second
 arrays to generate a new array. Result: [ROW('a', 1), ROW('b', 3),
ROW('c', 5)]
SELECT zip_with(ARRAY[1, 2], ARRAY[3, 4], (x, y) -> x + y); -- Result:
 [4, 6]
```

```
SELECT zip_with(ARRAY['a', 'b', 'c'], ARRAY['d', 'e', 'f'], (x, y) -
> concat(x, y)); --Concatenates the elements of the first and second
arrays to generate a new string. Result: ['ad', 'be', 'cf']
```

map_zip_with(map<K, V1>, map<K, V2>, function<K, V1, V2, V3>) → map<K, V3>

Merges two maps, uses values V1 and V2 to generate V3 based on each key, and generates a new map `as follows:K,V3>`.

```
SELECT map_zip_with(MAP(ARRAY[1, 2, 3], ARRAY['a', 'b', 'c']),
                    MAP(ARRAY[1, 2, 3], ARRAY['d', 'e', 'f']),
                    (k, v1, v2) -> concat(v1, v2)); --Merges values
which have the same map keys.-- {1 -> ad, 2 -> be, 3 -> cf}
SELECT map_zip_with(MAP(ARRAY['k1', 'k2'], ARRAY[1, 2]),
                    MAP(ARRAY['k2', 'k3'], ARRAY[4, 9]),
                    (k, v1, v2) -> (v1, v2)); --Generates an array
using the two values.-- {k1 -> ROW(1, null), k2 -> ROW(2, 4), k3 ->
ROW(null, 9)}
SELECT map_zip_with(MAP(ARRAY['a', 'b', 'c'], ARRAY[1, 8, 27]),
                    MAP(ARRAY['a', 'b', 'c'], ARRAY[1, 2, 3]),
                    (k, v1, v2) -> k || CAST(v1/v2 AS VARCHAR)); --
Concatenates the key values and division results of the two values --
 {a -> a1, b -> b4, c -> c9}
```

24.9.6.24 Logical function

The query and analysis function of Log Service supports logical functions.

Logical operators

Table 24-10: Logical operators

Operator	Description	Example
AND	Returns TRUE only when both the left and right operands are TRUE.	a AND b
OR	Returns TRUE if either the left or right operand is TRUE.	a OR b
NOT	Returns TRUE only when the right operand is FALSE.	NOT a

NULL involved in logical operation

The following table lists the true values when the values of a and b are TRUE, FALSE, and NULL, respectively.

Table 24-11: True value table 1

a	b	a AND b	a OR b
TRUE	TRUE	TRUE	TRUE
TRUE	FALSE	FALSE	TRUE
TRUE	NULL	NULL	TRUE
FALSE	TRUE	FALSE	TRUE
FALSE	FALSE	FALSE	FALSE
FALSE	NULL	FALSE	NULL
NULL	TRUE	NULL	TRUE
NULL	FALSE	FALSE	NULL
NULL	NULL	NULL	NULL

Table 24-12: True value table 2

a	NOT a
TRUE	FALSE
FALSE	TRUE
NULL	NULL

24.9.6.25 Column alias

The query and analysis function of Log Service supports setting column alias.

Context

In the SQL standard, a column name must consist of letters, digits, and underlines and start with a letter.

If you have configured a column name not conforming to the SQL standard (such as User-Agent) during log collection configuration, you need to name an alias for the column on the statistic properties configuration page for query. The alias is used for SQL statistical analysis only. The original name is used in underlying storage. Therefore, you must use the original name when you perform a search.

In addition, you can give the column an alias to replace the original name for query when the column name is long.

Table 24-13: Alias example

Original column name	Alias
User-Agent	ua
User.Agent	ua
123	col
abceefghijklmnopqrstuvw	a

24.9.6.26 Geospatial functions

The query and analysis function of Log Service supports Geospatial functions.

Concept of geometry

Geospatial functions support geometries in the Well-Known Text (WKT) format.

Table 24-14: Geometry format

Geometries	WKT format
Point	`POINT (0 0)`
LineString	`LINESTRING (0 0, 1 1, 1 2)`
Polygon	`POLYGON ((0 0, 4 0, 4 4, 0 4, 0 0), (1 1, 2 1, 2 2, 1 2, 1 1))`
MultiPoint	`MULTIPOINT (0 0, 1 2)`
MultiLineString	`MULTILINESTRING ((0 0, 1 1, 1 2), (2 3, 3 2, 5 4))`
MultiPolygon	`MULTIPOLYGON (((0 0, 4 0, 4 4, 0 4, 0 0), (1 1, 2 1, 2 2, 1 2, 1 1)), ((-1 -1, -1 -2, -2 -2, -2 -1, -1 -1)))`
GeometryCollection	`GEOMETRYCOLLECTION (POINT(2 3), LINESTRING (2 3, 3 4))`

Constructors

Table 24-15: Constructor Description

Function	Description
ST_Point(double, double) → Point	Returns a geometry type point object with the given coordinate values.
ST_LineFromText(varchar) → LineString	Returns a geometry type linestring object from WKT representation.
ST_Polygon(varchar) → Polygon	Returns a geometry type polygon object from WKT representation.
ST_GeometryFromText(varchar) → Geometry	Returns a geometry type object from WKT representation.
ST_AsText(Geometry) → varchar	Returns the WKT representation of the geometry.

Operators

Function	Description
ST_Boundary(Geometry) → Geometry	Returns the closure of the combinatorial boundary of this geometry.
ST_Buffer(Geometry, distance) → Geometry	Returns the geometry that represents all points whose distance from the specified geometry is less than or equal to the specified distance.
ST_Difference(Geometry, Geometry) → Geometry	Returns the geometry value that represents the point set difference of the given geometries.
ST_Envelope(Geometry) → Geometry	Returns the bounding rectangular polygon of a geometry.
ST_ExteriorRing(Geometry) → Geometry	Returns a line string representing the exterior ring of the input polygon.
ST_Intersection(Geometry, Geometry) → Geometry	Returns the geometry value that represents the point set intersection of two geometries.
ST_SymDifference(Geometry, Geometry) → Geometry	Returns the geometry value that represents the point set symmetric difference of two geometries. Returns the non-intersecting parts of two geometrics.

Relationship tests

Function	Description
ST_Contains(Geometry, Geometry) → boolean	Returns TRUE if and only if no points of the second geometry lie in the exterior of the first geometry, and at least one point of the interior of the first geometry lies in the interior of the second geometry. Returns FALSE if the second geometry is on the boundaries of the first geometry.
ST_Crosses(Geometry, Geometry) → boolean	Returns TRUE if the supplied geometries have some, but not all, interior points in common.
ST_Disjoint(Geometry, Geometry) → boolean	Returns TRUE if the give geometries do not spatially intersect – if they do not share any space together.
ST_Equals(Geometry, Geometry) → boolean	Returns TRUE if the given geometries represent the same geometry.
ST_Intersects(Geometry, Geometry) → boolean	Returns TRUE if the given geometries spatially intersect in two dimensions (share any portion of space).
ST_Overlaps(Geometry, Geometry) → boolean	Returns TRUE if the given geometries share space, are of the same dimension, but are not completely contained by each other.
ST_Relate(Geometry, Geometry) → boolean	Returns TRUE if first geometry is spatially related to second geometry.
ST_Touches(Geometry, Geometry) → boolean	Returns TRUE if the given geometries have at least one point in common, but their interiors do not intersect.
ST_Within(Geometry, Geometry) → boolean	Returns TRUE if first geometry is completely inside second geometry. Returns FALSE if boundary intersection exists.

Accessors

Function	Description
ST_Area(Geometry) → double	Returns the 2D Euclidean area of a geometry.
ST_Centroid(Geometry) → Geometry	Returns the point value that is the mathematic al centroid of a geometry.

Function	Description
ST_CoordDim(Geometry) → bigint	Return the coordinate dimension of the geometry.
ST_Dimension(Geometry) → bigint	Returns the inherent dimension of this geometry object, which must be less than or equal to the coordinate dimension.
ST_Distance(Geometry, Geometry) → double	Returns the two-dimensional cartesian minimum distance (based on spatial ref) between two geometries in projected units.
ST_IsClosed(Geometry) → boolean	Returns TRUE if the linestring's start and end points are coincident.
ST_IsEmpty(Geometry) → boolean	Returns TRUE if this Geometry is an empty geometrycollection, polygon, point etc.
ST_IsRing(Geometry) → boolean	Returns TRUE if and only if the line is closed and simple.
ST_Length(Geometry) → double	Returns the length of a linestring or multi-linestring using Euclidean measurement on a two-dimensional plane (based on spatial ref) in projected units. Returns the length of a linestring or multi-linestring. The length is a prediction specific to a two-dimensional plane based on spatial reference using Euclidean measurement.
ST_XMax(Geometry) → double	Returns X maxima of a bounding box of a geometry.
ST_YMax(Geometry) → double	Returns Y maxima of a bounding box of a geometry.
T_XMin(Geometry) → double	Returns X minima of a bounding box of a geometry.
ST_YMin(Geometry) → double	Returns Y minima of a bounding box of a geometry.
ST_StartPoint(Geometry) → point	Returns the first point of a LineString geometry as a Point.
ST_EndPoint(Geometry) → point	Returns the last point of a LineString geometry as a Point.
ST_X(Point) → double	Return the X coordinate of the point.
ST_Y(Point) → double	Return the Y coordinate of the point.

Function	Description
ST_NumPoints(Geometry) → bigint	Returns the number of points in a geometry.
ST_NumInteriorRing(Geometry) → bigint	Returns the cardinality of the collection of interior rings of a polygon.

24.9.6.27 JOIN syntax

JOIN associates the fields of multiple tables. In Log Service, JOIN is applicable to a single Logstore, between Logstore and RDS, and between Logstores. This document describes how to use JOIN across Logstores.

Procedure

1. *Download* the latest version of Python SDK.
2. Use the GetProjectLogs API for query.

SDK example

```
#!/usr/bin/env python
#encoding: utf-8
import time,sys,os
from aliyun.log.logexception import LogException
from aliyun.log.logitem import LogItem
from aliyun.log.logclient import LogClient
from aliyun.log.getlogsrequest import GetLogsRequest
from aliyun.log.getprojectlogsrequest import GetProjectLogsRequest
from aliyun.log.putlogsrequest import PutLogsRequest
from aliyun.log.listtopicsrequest import ListTopicsRequest
from aliyun.log.listlogstoresrequest import ListLogstoresRequest
from aliyun.log.gethistogramsrequest import GetHistogramsRequest
from aliyun.log.index_config import *
from aliyun.log.logtail_config_detail import *
from aliyun.log.machine_group_detail import *
from aliyun.log.acl_config import *
if __name__=='__main__':
    token = None
    endpoint = "http://***************.log.aliyuncs.com"
    accessKeyId  = '***************'
    accessKey='***************'
    client = LogClient(endpoint, accessKeyId, accessKey,token)
    logstore = "meta"
    # In the query statements, specify two Logstores, their respective
 time ranges, and the key for Logstore association.
    req = GetProjectLogsRequest(project,"select count(1) from
sls_operation_log  s join  meta m on s.__date__ >'2018-04-10 00:00:00
' and s.__date__ <'2018-04-11 00:00:00' and m.__date__ >'2018-04-23 00
:00:00' and m.__date__ <'2018-04-24 00:00:00' and s.projectid = cast(m
.ikey as varchar)");
    res = client.get_project_logs(req)
    res.log_print();
```

```
    exit(0)
```

24.9.7 Advanced analytics

24.9.7.1 Excellent analysis cases

This section introduce the excellent analysis cases.

Case list

1. *Trigger an alarm when the error rate in the recent 5 minutes exceeds 40%.*
2. *Trigger an alarm when traffic decreases sharply*
3. *Calculate the average latency of each bucket set by data range.*
4. *Return percentage data included in GROUP BY results.*
5. *Count the number of entries that satisfy the condition.*

Trigger an alarm when the error rate in the recent 5 minutes exceeds 40%.

Count the percentage of Error 500 every minute; trigger an alarm when the percentage exceeds 40% for the past 5 minutes.

```
status:500 | select __topic__, max_by(error_count,window_time)/1.0/sum
(error_count) as error_ratio, sum(error_count) as total_error from (
select __topic__, count(*) as error_count , __time__ - __time__ % 300
 as window_time from log group by __topic__, window_time ) group by
```

```
__topic__ having max_by(error_count,window_time)/1.0/sum(error_count)
 > 0.4 and sum(error_count) > 500 order by total_error desc limit 100
```

Trigger an alarm when traffic decreases sharply

Count traffic every minute; trigger an alarm when the recent traffic decreases sharply. Data from the past one minute does not cover a full minute; therefore, divide the statistical value by (max(**time**) - min(**time**)) for normalization to calculate the average traffic per minute.

```
* | SELECT SUM(inflow) / (max(__time__) - min(__time__)) as inflow_per
_minute, date_trunc(minute,__time__) as minute group by minute
```

Calculate the average latency of each bucket set by data range.

```
* | select avg(latency) as latency , case when originSize < 5000 then
 s1 when originSize < 20000 then s2 when originSize < 500000 then s3
when originSize < 100000000 then s4 else s5 end as os group by os
```

Return percentage data included in GROUP BY results.

List the count results of different departments and related percentage data. The query combines subquery and window functions. sum(c) over() indicates the sum of values in all rows.

```
* | select department, c*1.0/ sum(c) over () from(select count(1) as c
, department from log groupby department)
```

Count the number of entries that satisfy the condition.

To perform counting by URL feature in the URL, you can use the CASE WHEN syntax, or the simpler count_if syntax.

```
* | select count_if(uri like %login) as login_num, count_if(uri like %
register) as register_num, date_format(date_trunc(minute, __time__), %
m-%d %H:%i) as time group by time order by time limit 100
```

24.9.7.2 Optimize a query

The analysis efficiency varies from one query to another. Common methods of query optimization are as follows:

1. *Avoid running Group By on string columns if possible*
2. *List fields with relatively large dictionary values on top when running Group By on multiple columns*
3. *Use estimating functions*
4. *Retrieve required columns in SQL and do not read all columns if possible*
5. *Place non-GROUP BY columns in an aggregate function if possible.*

Avoid running Group By on string columns if possible

Running Group By on strings may result in a large amount of hash calculations, which account for more than 50% of total calculations.

For example, consider the following two queries:

```
* | select count(1) as pv , date_trunc(hour,__time__) as time group by
 time * | select count(1) as pv , from_unixtime(__time__-__time__%3600
) as time group by __time__-__time__%3600
```

Both Query 1 and Query 2 calculate the log count value every hour. However, Query 1 converts time into a string, for example, `2017-12-12 00:00:00`, and then runs Group By on this string. Query 2 runs Group By on the on-the-hour time value and then converts the result into a string. Query 1 is less efficient than Query 2 because the former needs to hash strings.

List fields with relatively large dictionary values on top when running Group By on multiple columns

For example, there are 13 provinces with 100 million users.

```
Quick: * | select province,uid,count(1)groupby province,uid Slow: * |
select province,uid,count(1)groupby uid,province
```

Use estimating functions

Estimating functions provide much stronger performance than accurate calculation. In estimation, accuracy is compromised to an acceptable extent for fast calculation.

```
Quick: * |select approx_distinct(ip) Slow: * | select count(distinct(
ip))
```

Retrieve required columns in SQL and do not read all columns if possible

Use the query syntax to retrieve all columns. To speed up calculation, retrieve only the required columns in SQL if possible.

```
Quick: * |select a,b c Slow:* |select*
```

Place non-GROUP BY columns in an aggregate function if possible.

For example, a user ID exactly corresponds to a user name. Therefore, use only userid in running GROUP BY.

```
Quick: * | select userid, arbitrary(username), count(1)groupby userid
Slow: * | select userid, username, count(1)groupby userid,username
```

24.9.8 Log analysis through JDBC

In addition to RESTful APIs, you can also use JDBC and standard SQL-92 for log querying and analysis.

Connection parameters

Connection parameters	Example	Description
host	regionid.example.com	Access point
port	10005	Use 10005 as the default port.
user	bq2sjzesjmo86kq	AccessKey ID
password	4fdO1fTDDuZP	AccessKey

Connection parameters	Example	Description
database	sample-project	The project under your account.
table	sample-logstore	The Logstore under the project.

The following is an example of connection through a MySQL command:

```
mysql -hcn-shanghai-intranet.log.aliyuncs.com -ubq2sjzesjmo86kq -
p4fdO1fTDDuZP -P10005
use sample-project; //Use a project.
```

Prerequisites

You must use the AccessKey of the main account or a sub-account to access the JDBC interface. The sub-account must belong to the project owner and have the project-level read permission.

Syntax

Precautions

The WHERE condition must contain `__date__`or `__time__` to limit the time range of query. The type of `__date__` is timestamp, and the type of `__time__` is bigint.

Example:

- `__date__ > '2017-08-07 00:00:00' and __date__ < '2017-08-08 00:00:00'`
- `__time__ > 1502691923 and __time__ < 1502692923`

At least one of the preceding conditions must be met.

Filtering syntax

The filtering syntax in the WHERE condition is as follows:

Meaning	Example	Description
String search	`key = "value"`	Results after word segmentation are queried.
String fuzzy search	`key = "valu*"`	Results of fuzzy match after word segmentation are queried.
Value comparison	`num_field > 1`	Comparison operators including `>`, `>=`, `=`, `<` and `<=`.
Logic operations	`and or not`	For example, `a = "x" and b ="y"` or a = "x" and not b `="y"`.

Meaning	Example	Description
Full-text search	`__line__ ="abc"`	Full-text index search requires the special key (`__line__`).

Computation syntax

For the supported operators, see *Query syntax and functions*.

SQL-92 syntax

The SQL-92 syntax is a combination of filtering and computation syntax.

The following query is used as an example:

```
status>200 |select avg(latency),max(latency) ,count(1) as c GROUP BY
method  ORDER BY c DESC  LIMIT 20
```

The filter part and time condition in the query can be combined into a new query condition based on standard SQL-92 syntax.

```
select avg(latency),max(latency) ,count(1) as c from sample-logstore
 where status>200 and __time__>=1500975424 and __time__ < 1501035044
GROUP BY  method  ORDER BY c DESC  LIMIT 20
```

Access Log Service through JDBC

Program call

Developers can use the MySQL syntax to connect to Log Service in any program that supports MySQL connector. For example, JDBC or Python MySQLdb can be used.

Example:

```
import com.mysql.jdbc.*;
import java.sql.*;
import java.sql.Connection;
import java.sql.ResultSetMetaData;
import java.sql.Statement;
public class testjdbc {
    public static void main(String args[]){
        Connection conn = null;
        Statement stmt = null;
        try {
            //STEP 2: Register JDBC driver
            Class.forName("com.mysql.jdbc.Driver");
            //STEP 3: Open a connection
            System.out.println("Connecting to a selected database
...");
            conn = DriverManager.getConnection("jdbc:mysql://cn-
shanghai-intranet.log.aliyuncs.com:10005/sample-project","accessid","
accesskey");
            System.out.println("Connected database successfully...") ;
            //STEP 4: Execute a query
```

```
            System.out.println("Creating statement...");
            stmt = conn.createStatement();
            String sql = "SELECT method,min(latency,10)  as c,max
(latency,10) from sample-logstore where  __time__>=1500975424 and
__time__ < 1501035044 and latency > 0  and latency < 6142629 and  not
 (method='Postlogstorelogs' or method='GetLogtailConfig') group by
method " ;
            String sql-example2 = "select count(1) ,max(latency),
avg(latency), histogram(method),histogram(source),histogram(status),
histogram(clientip),histogram(__source__) from  test10 where __date__
  >'2017-07-20 00:00:00'  and  __date__ <'2017-08-02 00:00:00' and
__line__='abc#def' and latency < 100000 and (method = 'getlogstorelogS
' or method='Get**' and method <> 'GetCursorOrData' )";
            String sql-example3 = "select count(1) from  sample-
logstore where     __date__  >       '2017-08-07 00:00:00' and
__date__ <     '2017-08-08 00:00:00' limit 100";
            ResultSet rs = stmt.executeQuery(sql);
            //STEP 5: Extract data from result set
            while(rs.next()){
                //Retrieve by column name
                ResultSetMetaData data = rs.getMetaData();
                System.out.println(data.getColumnCount());
                for(int i = 0;i < data.getColumnCount();++i) {
                    String name = data.getColumnName(i+1);
                    System.out.print(name+":");
                    System.out.print(rs.getObject(name));
                }
                System.out.println();
            }
            rs.close();
        } catch (ClassNotFoundException e) {
            e.printStackTrace();
        } catch (SQLException e) {
            e.printStackTrace();
        } catch (Exception e) {
            e.printStackTrace();
        } finally {
            if (stmt ! = null) {
                try {
                    stmt.close();
                } catch (SQLException e) {
                    e.printStackTrace();
                }
            }
            if (conn ! = null) {
                try {
                    conn.close();
                } catch (SQLException e) {
                    e.printStackTrace();
                }
            }
        }
    }
}
```

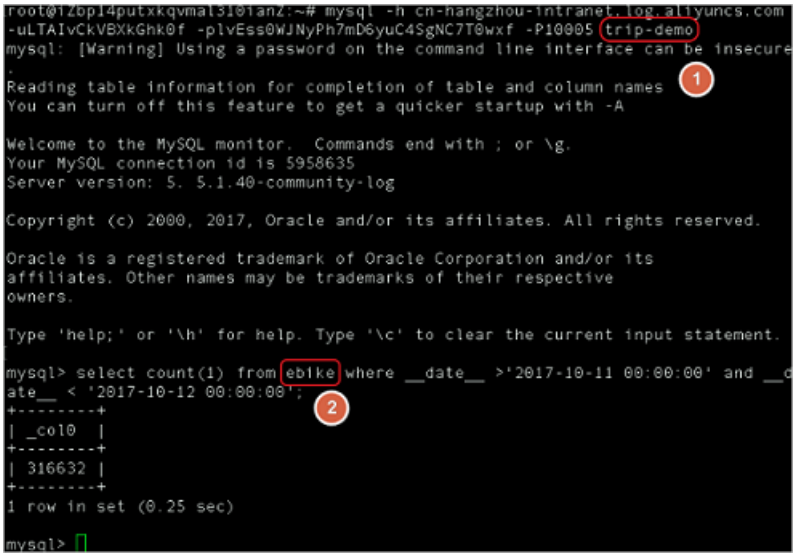
Tool calling

As shown in *Figure 24-9: Connection example*, in the classic network intranet or VPC environment, you can use the MySQL client to connect to Log Service.

Note:

1. Enter your project name at (1).
2. Enter your Logstore name at (2).

Figure 24-9: Connection example



24.10 Alarms

24.10.1 Overview

Log Service allows you to configure the alarming function based on Saved Search to monitor service status in real time.

Function implementation

Log Service implements the alarming function based on Saved Search. Log on to the Log Service console and go to the **query page**. Then, set an alarm rule and specify the attribute, check condition, and alarm action for the rule on the page. *Set an alarm*. Then, Log Service periodically queries the collected log data and sends an alarm notification when the query results meet the predefined condition, realizing real-time monitoring of the service status.

- **Alarm rule attribute**: Set the time and interval of data checks so that Log Service can check as scheduled.
- **Check condition**: Set the comparison field, comparison operator, and check threshold. The comparison operator and the check threshold comprise the check condition. An alarm

notification is sent when the query results based on the comparison field meet the check condition.

- **Alarm action**: Set the notification method and alarm content. We recommend that you configure **notification center** as the notification method. If you use this method, alarms are sent to the contacts specified in the notification center through website notices, text messages, and emails.

Saved Search in alarms

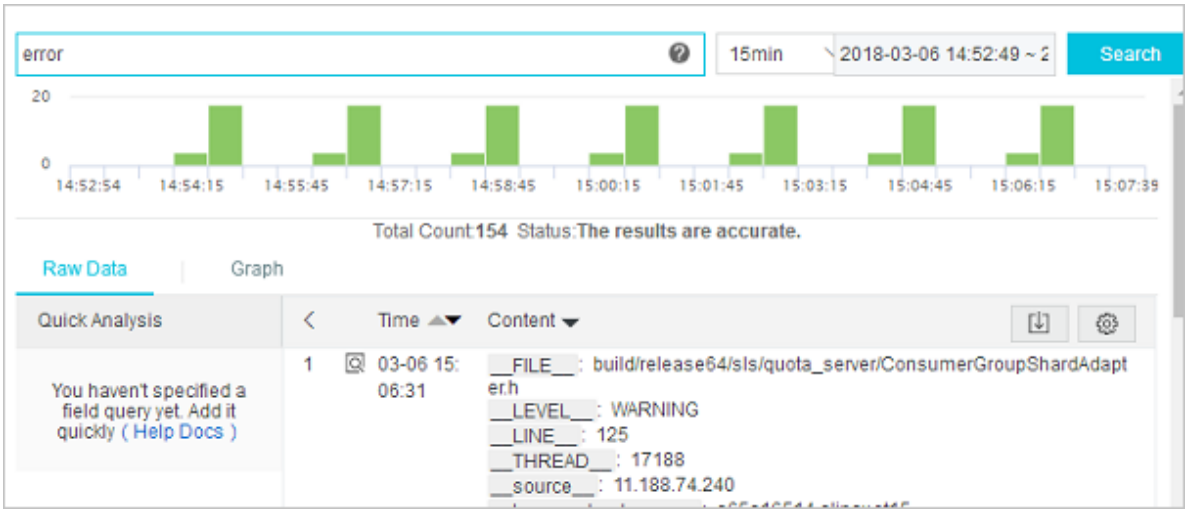
Saved Search supports query statements and query and analysis statements.

- Query statement: returns the log data that matches the query condition.
- Query and analysis statement: collects statistics on the logs that match the query condition and returns statistical results.
- **Query statement**

  For example, you want to query the data that contains errors in the last 15 minutes, where the condition is error and a total of 154 records are found. The content of each record is a combination of key and value. You can set an alarm condition for the value under a specific key .

  Note:
  If the number of query results exceeds 10 in a single query, only the first 10 results are retrieved by the alarm rules, and an alarm is triggered when any of the 10 results meets the condition.

Figure 24-10: Query statement



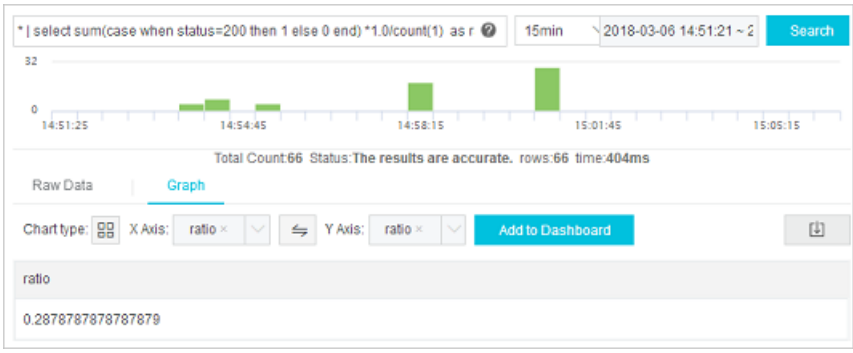
- **Query and analysis statement**

  For example, the following statement is used to query the proportion of logs with the status code 200 to all the logs. For more information about query syntax, see *Query syntax*.

  ```
  * | select sum(case when status=200 then 1 else 0 end) *1.0/count(1)
    as ratio
  ```

  Therefore, you can set the alarm condition as `ratio < 0.9`, indicating that an alarm is triggered when the proportion of logs with the status code 200 to all the logs is less than 90%.

  Figure 24-11: Query and analysis statement



24.10.2 Set an alarm

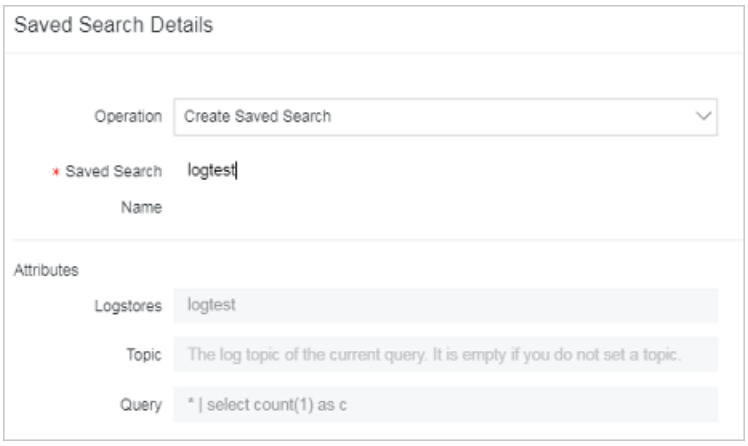
You can save Saved Search as an alarm on the query page so that Log Service performs a scheduled check and sends an alarm when the alarm condition is met.

Prerequisites

- Log data has been collected.
- *Configure an index*.

Procedure

1. *Log on to the Log Service console*.
2. On the **Logstores** page, click **Search** in the **LogSearch** column.
3. Enter a query statement in the search box.
4. Set a time range to be queried and click **Search**.
5. Click **Save As Alarm** in the upper-right corner of the page.



6. Configure alarm rules and click **OK**.

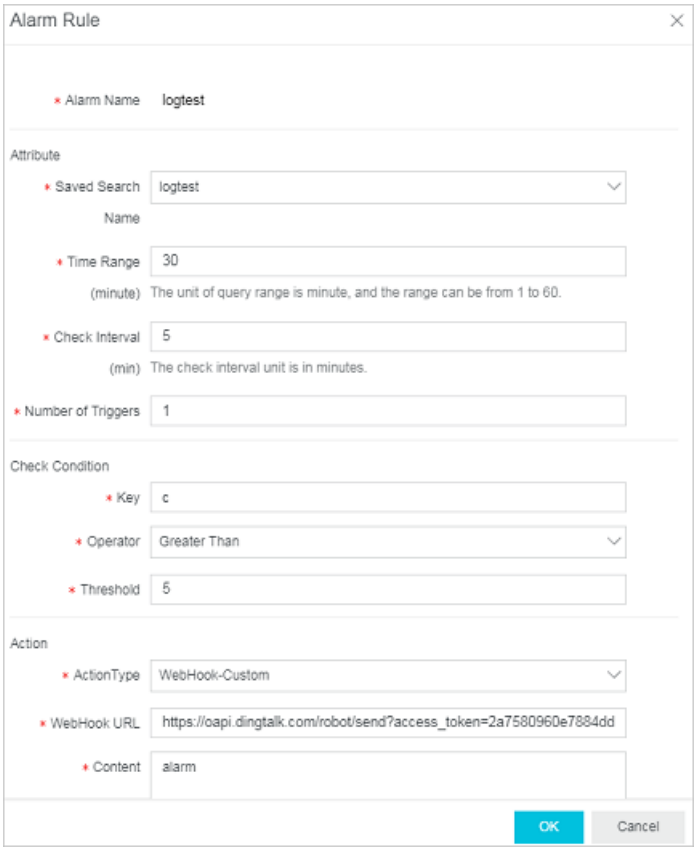
Rule	Description
Alarm rule name	The name of the alarm rule. This name must be a string of 3 to 63 characters.
Alarm rule attributes	
Saved Search name	The name of Saved Search used by the alarm. Saved Search can be set to: • **Current query**, indicating that the current query statement is used as Saved Search; • Another existing Saved Search.
Data query time (minutes)	The time range of the data read by the server during each alarm check. For example, if it is set to 1, the server queries the data generated within the last 1 minute of the check time. The data query time is in minutes and ranges from 1 to 60. **Note:** Currently, when performing an alarm check, the server only processes the first 10 data records generated during the time range for sampling purpose.
Check interval (minutes)	The interval at which the server performs an alarm check. The check interval is in minutes and ranges from 1 to 1,440.
Trigger times	The number of consecutive alarm check triggers. An alarm notification is sent when the specified trigger count is reached. Value range: 1 to 10,000. For example, if the **check interval (minutes)** is set to 1 and the **trigger times** is set to 2, the server performs a check every one minute and sends an alarm if the results of two

Rule	Description
	consecutive checks meet the alarm condition. The minimum interval between alarm notifications is two minutes.
Check condition	
Key name	The key name for alarming in the log content.
Comparison operator	The comparison operator in the check condition, which can be of the numeric or character type.
Check threshold	The comparative value in the check condition, which is combined with the **comparison operator** to determine whether the Saved Search results meet the alarm condition.
Alarm action	
Notification type	The method for sending alarm notifications. When the configured alarm rule is triggered, Log Service sends an alarm based on the predefined notification method. Alarms are only sent through the **WebHook-custom** method. That is, notifications are sent to the custom WebHook link through the Post method.
WebHook address	The URL of WebHook.
Notification content	The content of an alarm notification. The maximum length of the content is 500 characters.

Table 24-16: Comparison operators

Operation	Description	Example
>	Checks whether the column value is greater than the specified value.	$count > 0
<	Checks whether the column value is less than a value.	$count<200
>=	Checks whether the column value is greater than or equal to the specified value.	$count>=0
<=	Checks whether the column value is smaller than or equal to the specified value.	$count<=0
like	A matched substring.	$project like "admin"

Operation	Description	Example
regex	A string that matches with the regular expression.	$project regex match "^/S+$"



Result

You can view specific alarm results after creating alarm rules.

1. On the **Logstores** page, choose **Search/Analytics** > **Alarm** from the left-side navigation pane.
2. Click **View** for an alarm rule to view the specific alarm records under the rule.

Alarm status:

- **Success**: indicates that the rule is successfully executed and the standard to trigger the alarm is displayed in **Trigger Details**.
- **Failure**: indicates that the rule failed during the query, alarm rule matching, or notification phase. In this case, you can view **Trigger Details** for more information.
  - Query failed: indicates that the query syntax is incorrect.
  - Query call failed: indicates that the query failed to be called. We recommend that you check your network connectivity.

- Failed to call the rule: indicates that the rule failed to be called We recommend that you check the format consistency between the rule parameters and returned data.

24.10.3 Notification methods

Currently, alarm notifications are sent by using WebHook custom methods.

An alarm notification contains the following items:

Item	Description
Uid	The ID of an Alibaba Cloud account, namely, the AliUid.
Project	The name of the Log Service project for which an alarm is generated.
Trigger	The name of the alarm.
Condition	The **check condition** configured for the alarm.
Message	The content of the alarm notification.
Context	The query results. If the alarm notification is sent through the notification center, the maximum context length is 100 bytes. If the alarm notification is sent through the WebHook DingTalk Chatbot or a WebHook custom method, the maximum context length is 1,000 bytes.

WebHook-custom method

The alarm notification method can be set to WebHook-custom. When an alarm is triggered, the alarm notification is sent to the custom WebHook address by using the Post method.

Procedure

1. Log on to the Log Service console and *Set an alarm*.
2. Set **Notification Type** to **WebHook-Custom**.
3. Enter a custom WebHook address in **WebHook Address**, and then enter the **notification content**.

   When an alarm is triggered, the alarm content is sent to the custom WebHook address by using the Post method.

   Example:

```
{"uid":
            "13415134513","project":"ali-cn","trigger":"oplog_alert
","condition":"3413 >
            3000", "message":"PV count down 30%",
```

```
        "context":"c:3413"}
```

24.11 Log consumption

24.11.1 Preview logs

Log preview is a regular type of log consumption. The Log Service console provides a dedicated preview page to help you preview a portion of the logs in the LogStore directly in your browser.

Procedure

1. *Log on to the Log Service console*.
2. Click the name of a project to go to the **Logstores** page.
3. On the **Logstores** page, select a Logstore and click **Preview** in the **Log Consumption** column.
4. On the **logstore** page that appears, select a shard, specify the time range, and click **Preview**.

   The **logstore** page displays the log data of the first 10 packets in the specified time range.

24.11.2 Consumption by consumer groups

24.11.2.1 Consumption by consumer groups

Log Service provides the log consumption method of ConsumerGroup.

The consumer library is an advanced method to consume logs in Log Service. It provides the consumer group feature to abstract and manage consumers. Different from SDKs for data reading , the consumer group allows you to focus on the service logic without paying attention to Log Service implementation details and the load balancing and failover between consumers.

Basic concepts

Before using the consumer library, get to know the consumer group and consumer.

- Consumer group

  A consumer group consists of multiple consumers. The consumers in the same consumer group consume data in the same Logstore. The data consumed by consumers is not repeated.
- Consumer

Consumers form a consumer group and consume data. Consumers in the same consumer group must have different names.

In Log Service, a Logstore has several shards. The consumer library allocates shards to the consumers in a consumer group based on the following principles:

- Each shard is allocated to only one consumer.
- A consumer may have multiple shards.

After a consumer joins a consumer group, the shard subordination in this group is adjusted for load balancing, but the preceding principles remain unchanged. The allocation process is transparent.

The consumer library also saves checkpoints to enable consumption to resume from the breakpoint after programs recover. This avoids repeated data consumption.

Instructions for use

Maven dependency

```
<dependency> <groupId>com.google.protobuf</groupId> <artifactId>
protobuf-java</artifactId> <version>2.5.0</version> </dependency> <
dependency> <groupId>com.aliyun.openservices</groupId> <artifactId
>aliyun-log</artifactId> <version>0.6.11</version> </dependency> <
dependency> <groupId>com.aliyun.openservices</groupId> <artifactId>
loghub-client-lib</artifactId> <version>0.6.15</version> </dependency>
```

main .java file

```
public class Main { // Domain name of Log Service. Set it according to
 actual information. private static String sEndpoint = "cn-hangzhou.
log.aliyuncs.com"; // Name of a Log Service project. Set it according
 to actual information. private static String sProject = "ali-cn
-hangzhou-sls-admin"; // Name of a Logstore. Set it according to
actual information. private static String sLogstore = "sls_operat
ion_log"; // Name of a consumer group. Set it according to actual
information. private static String sConsumerGroup = "consumerGroupX
"; // AccessKey for data consumption. Set it according to actual
information. private static String sAccessKeyId = ""; private static
String sAccessKey = ""; public static void main(String []args) throws
 LogHubClientWorkerException, InterruptedException { // The consumer
 names (the second parameter) in the same consumer group must be
different. The names of consumer groups may be the same. Different
 consumer names are used to start multiple processes on several
machines to consume data in the same Logstore in the load balancing
manner. The consumer group names can be differentiated by the machine
 IP addresses. maxFetchLogGroupSize (the ninth parameter) indicates
the number of log groups retrieved from the server at a time. You can
keep the default value, or adjust it in the range (0,1000] as needed.
LogHubConfig config = new LogHubConfig(sConsumerGroup, "consumer_1",
sEndpoint, sProject, sLogstore, sAccessKeyId, sAccessKey, LogHubConf
ig.ConsumePosition.BEGIN_CURSOR); ClientWorker worker = new ClientWork
er(new SampleLogHubProcessorFactory(), config); Thread thread = new
```

```
Thread(worker); //The ClientWorker instance runs automatically after
 the thread is executed and extends the Runnable interface. thread
.start(); Thread.sleep(60 * 60 * 1000); //The shutdown function of
the ClientWorker instance is called to exit the consumption instance
. The associated thread is stopped automatically. worker.shutdown
(); //Multiple asynchronous tasks are generated when the ClientWorker
 instance is running. You are advised to wait 30s until all tasks are
exited after shutdown. Thread.sleep(30 * 1000); } }
```

SampleLogHubProcessor.java file

```
public class SampleLogHubProcessor implements ILogHubProcessor {
private int mShardId; // Records the last persistent checkpoint
time. private long mLastCheckTime = 0; public void initialize(int
shardId) { mShardId = shardId; } // Master logic of data consumptio
n. All the exceptions must be captured and cannot be thrown. public
 String process(List<LogGroupData> logGroups, ILogHubCheckPointTra
cker checkPointTracker) { // Print the retrieved data simply. for
(LogGroupData logGroup: logGroups){ FastLogGroup flg = logGroup.
GetFastLogGroup(); System.out.println(String.format("\tcategory\t:
\t%s\n\tsource\t:\t%s\n\ttopic\t:\t%s\n\tmachineUUID\t:\t%s", flg.
getCategory(), flg.getSource(), flg.getTopic(), flg.getMachineUUID
())); System.out.println("Tags"); for (int tagIdx = 0; tagIdx < flg
.getLogTagsCount(); ++tagIdx) { FastLogTag logtag = flg.getLogTags
(tagIdx); System.out.println(String.format("\t%s\t:\t%s", logtag
.getKey(), logtag.getValue())); } for (int lIdx = 0; lIdx < flg.
getLogsCount(); ++lIdx) { FastLog log = flg.getLogs(lIdx); System.
out.println("--------\nLog: " + lIdx + ", time: " + log.getTime() +
 ", GetContentCount: " + log.getContentsCount()); for (int cIdx = 0
; cIdx < log.getContentsCount(); ++cIdx) { FastLogContent content =
log.getContents(cIdx); System.out.println(content.getKey() + "\t:\t"
 + content.getValue()); } } } long curTime = System.currentTimeMillis
(); // Writes checkpoints to the server every 30s. If a ClientWork
er instance crashes during the 30s period, // the new ClientWorker
instance consumes data starting from the last checkpoint. Duplicate
data may exist. if (curTime - mLastCheckTime > 30 * 1000) { try { //
When the parameter is set to TRUE, checkpoints are updated to the
server immediately; when the parameter is set to FALSE, checkpoints
 are cached locally. The default update interval is 60s. checkPoint
Tracker.saveCheckPoint(true); } catch (LogHubCheckPointException e
) { e.printStackTrace(); } mLastCheckTime = curTime; } return null
; } // The ClientWorker instance calls this function upon exit,
during which you can perform cleanup. public void shutdown(ILogHubChe
ckPointTracker checkPointTracker) { //Save the consumption breakpoint
 to Log Service. try { checkPointTracker.saveCheckPoint(true); }
catch (LogHubCheckPointException e) { e.printStackTrace(); } } }
class SampleLogHubProcessorFactory implements ILogHubProcessorFact
ory { public ILogHubProcessor generatorProcessor() { // Generate a
consumption instance. return new SampleLogHubProcessor(); } }
```

Run the preceding code to print all the data in a Logstore. If you allow multiple consumers to consume the same Logstore, you can modify the program based on the comment to add different consumer names by using the same consumer group name, and start another consumption process.

Constraints and exception diagnostic

Up to 10 consumer groups can be created for each Logstore. The error `ConsumerGroupQuotaExceed` is returned when this limit is exceeded.

We recommend that you configure Log4j for the consumer program to throw error messages in consumer groups for troubleshooting. If you save the log4j.properties file to the resources directory and execute the program, the following error message is displayed:

```
[WARN ] 2018-03-14 12:01:52,747 method:com.aliyun.openservices.loghub
.client.LogHubConsumer.sampleLogError(LogHubConsumer.java:159) com.
aliyun.openservices.log.exception.LogException: Invalid loggroup count
, (0,1000]
```

You can refer to a simple example of log4j.properties configuration:

```
log4j.rootLogger = info,stdout log4j.appender.stdout = org.apache.
log4j.ConsoleAppender log4j.appender.stdout.Target = System.out log4j.
appender.stdout.layout = org.apache.log4j.PatternLayout log4j.appender
.stdout.layout.ConversionPattern = [%-5p] %d{yyyy-MM-dd HH:mm:ss,SSS}
method:%l%n%m%n
```

24.11.2.2 View consumer group status

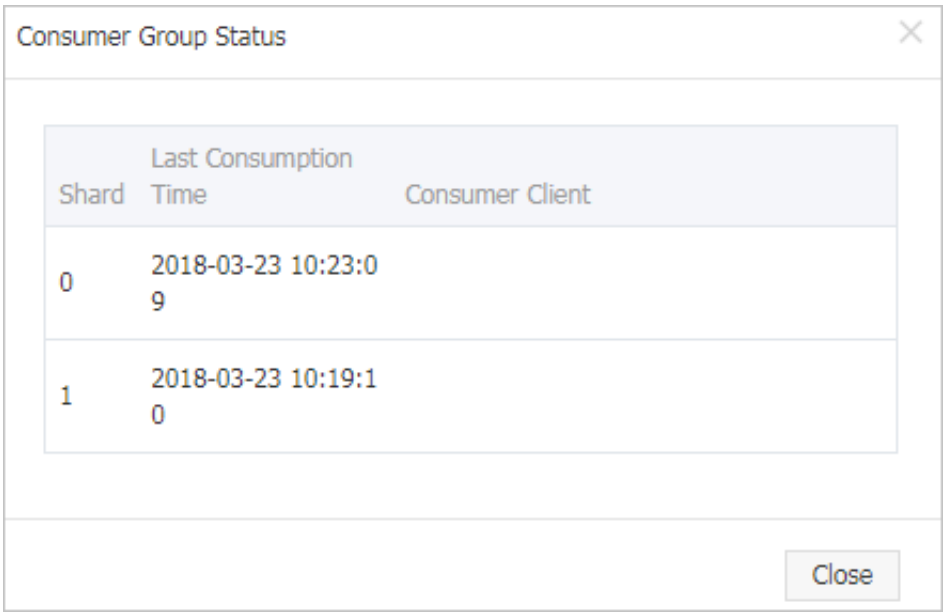
Log Service provides the log consumption method of ConsumerGroup.

Consumption by consumer groups is an advanced real-time data consumption mode. It provides automatic Logstore consumption load balancing for multiple consumer instances. Both Spark Streaming and Storm use ConsumerGroup as their basic mode.

Procedure

1. *Log on to the Log Service console*.
2. Click the name of a project.
3. In the left-side navigation pane, click **LogHub - Consume** > **Consumer Group**.
4. On the **Consumer Group** page, select a Logstore to check whether the collaborative consumption function is enabled.

5. Select a consumer group, and click **Consumption Status** to view the progress of data consumption for each shard.



As shown in the figure above, the page displays four shards of the Logstore, corresponding to four consumers. The most recent data consumption time is shown for each consumer in the second column. With data consumption time, you can determine if the current data processing can keep up with data production. If processing seriously lags behind (i.e. data consumption is slower than data production), you should consider increasing the number of consumers.

Here, we will use the Java SDK as an example to show how to get consumption status using API.

```
package test;
import java.util.ArrayList;
import com.aliyun.openservices.log.Client;
import com.aliyun.openservices.log.common.Consts.CursorMode;
import com.aliyun.openservices.log.common.ConsumerGroup;
import com.aliyun.openservices.log.common.ConsumerGr
oupShardCheckPoint;
import com.aliyun.openservices.log.exception.LogException;
public class ConsumerGroupTest {
    static String endpoint = "";
    static String project = "";
    static String logstore = "";
    static String accesskeyId = "";
    static String accesskey = "";
    public static void main(String[] args) throws LogExcepti
on {
        Client client = new Client(endpoint, accesskeyId,
accesskey);
        //Retrieve all consumer groups in this Logstore. If
no consumer group exists, the consumerGroups length is 0.
        ArrayList<ConsumerGroup>  consumerGroups;
        try{
            consumerGroups = client.ListConsumerGroup(
project, logstore).GetConsumerGroups();
        }
```

```
        catch(LogException e){
            if(e.GetErrorCode() == "LogStoreNotExist")
                System.out.println("this logstore does not
have any consumer group");
            else{
                //internal server error branch
            }
            return;
        }
        for(ConsumerGroup c: consumerGroups){
            //Print consumer group properties, including
names, heartbeat timeout, and whether or not the consumption
 is in order.
            System.out.println("Name: " + c.getConsume
rGroupName());
            System.out.println("Heartbeat timeout: " + c.
getTimeout());
            System.out.println("Consumption in order: " + c.
isInOrder());
            for(ConsumerGroupShardCheckPoint cp: client.
GetCheckPoint(project, logstore, c.getConsumerGroupName()).
GetCheckPoints()){
                System.out.println("shard: " + cp.getShard
());
                //Please format, this time returns the exact
 time to milliseconds, the length of the integer
                System.out.println("Last data consumption
time: " + cp.getUpdateTime());
                System.out.println("Consumer name: " + cp.
getConsumer());
                String consumerPrg = "";
                if(cp.getCheckPoint().isEmpty())
                    consumerPrg = "Consumption not started";
                else{
                    //UNIX timestamp. Measured in seconds.
Format the value upon output.
                    try{
                        int prg = client.GetPrevCursorTime
(project, logstore, cp.getShard(), cp.getCheckPoint()).
GetCursorTime();
                        consumerPrg = "" + prg;
                    }
                    catch(LogException e){
                        if(e.GetErrorCode() == "InvalidCur
sor")
                            consumerPrg = "Invalid. The
previous consumption time has exceeded the data lifecycle in
 the Logstore.";
                        else{
                            //internal server error
                            throw e;
                        }
                    }
                }
                System.out.println("Consumption progress: "
 + consumerPrg);
                String endCursor = client.GetCursor(project
, logstore, cp.getShard(), CursorMode.END).GetCursor();
                int endPrg = 0;
                try{
```

```
                    endPrg = client.GetPrevCursorTime(
project, logstore, cp.getShard(), endCursor).GetCursorTime
();
                }
                catch(LogException e){
                    //do nothing
                }
                //UNIX timestamp. Measured in seconds.
Format the value upon output.
                System.out.println("The arrival time of the
last piece of data: " + endPrg);
            }
        }
    }
}
```

24.11.3 Use Flink to consume logs

Log Service provides the Flink log connector for connecting to Flink.

The Flink log connector provided by Log Service is used to connect to Flink. It consists of the consumer and producer.

The consumer reads data from Log Service. It supports the exactly-once semantics and shard-based load balancing.

The producer writes data into Log Service. When using the connector, you must add the Maven dependency to the project:

```
<dependency>
            <groupId>org.apache.flink</groupId>
            <artifactId>flink-streaming-java_2.11</artifactId>
            <version>1.3.2</version>
            </dependency>
            <dependency>
            <groupId>com.aliyun.openservices</groupId>
            <artifactId>flink-log-connector</artifactId>
            <version>0.1.7</version>
            </dependency>
            <dependency>
            <groupId>com.google.protobuf</groupId>
            <artifactId>protobuf-java</artifactId>
            <version>2.5.0</version>
            </dependency>
            <dependency>
            <groupId>com.aliyun.openservices</groupId>
            <artifactId>aliyun-log</artifactId>
            <version>0.6.10</version>
            </dependency>
            <dependency>
            <groupId>com.aliyun.openservices</groupId>
            <artifactId>log-loghub-producer</artifactId>
            <version>0.1.8</version>
```

```
            </dependency>
```

Prerequisites

1. *Log on to the Log Service console*.
2. An AccessKey pair is available, and a project and a Logstore have been created. For detailed procedures, see *Procedure*.

Log consumer

In the connector, the Flink log consumer provides the capability of subscribing to a specific Logstore in Log Service to achieve the exactly-once semantics. During this process, you do not need to concern about the change of the number of shards in the Logstore.

Each sub-task in Flink consumes some shards in the Logstore. If shards in the Logstore are split or merged, shards consumed by the sub-task change accordingly.

Associated APIs

The Flink log consumer uses the following Log Service APIs:

- GetCursorOrData

  You can call this operation to pull data from shards. Note that frequent calls to this API may cause data to exceed the shard quota of Log Service. You can use ConfigConstants.LOG_FETCH_DATA_INTERVAL_MILLIS and ConfigConstants.LOG_MAX_NUMBER_PER_FETCH to control the call interval and the number of logs pulled during each call. For more information about the shard quota, see *Split a shard*.

  ```
  configProps.put(ConfigConstants.LOG_FETCH_DATA_INTERVAL_MILLIS, "100
  ");
                          configProps.put(ConfigConstants.LOG_MAX_NU
  MBER_PER_FETCH, "100");
  ```

- ListShards

  You can call this operation to obtain the list of all shards and shard status in a Logstore. If your shards are frequently split and merged, you can adjust the call interval to find shard changes in time.

  ```
  // Call ListShards every 30s
                          configProps.put(ConfigConstants.LOG_SHARDS
  _DISCOVERY_INTERVAL_MILLIS, "30000");
  ```

- CreateConsumerGroup

You can call this operation only when consumption progress monitoring is enabled. You can call this operation to create a consumer group to synchronize the checkpoint.

- ConsumerGroupUpdateCheckPoint

  You can call this operation to synchronize snapshots of Flink to a consumer group of Log Service.

Procedure

1. Configure the startup parameter.

```
Properties configProps = new Properties();
                    // Set the name of the domain used to access Log
Service.
                    configProps.put(ConfigConstants.LOG_ENDPOINT, "cn-
hangzhou.log.aliyuncs.com");
                    // Set the AccessKey
                    configProps.put(ConfigConstants.LOG_ACCESSSKEYID,
 "");
                    configProps.put(ConfigConstants.LOG_ACCESSKEY,
 "");
                    // Set the Log Service project
                    configProps.put(ConfigConstants.LOG_PROJECT, "ali-
cn-hangzhou-sls-admin");
                    // Set the Log Service Logstore
                    configProps.put(ConfigConstants.LOG_LOGSTORE, "
sls_consumergroup_log");
                    // Set the start position to consume Log Service
                    configProps.put(ConfigConstants.LOG_CONSUM
ER_BEGIN_POSITION, Consts.LOG_END_CURSOR);
                    // Set the message deserialization method for Log
Service
                    RawLogGroupListDeserializer deserializer = new
RawLogGroupListDeserializer();
                    final StreamExecutionEnvironment env = StreamExec
utionEnvironment.getExecutionEnvironment();
                    DataStream<RawLogGroupList> logTestStream = env.
addSource(
                    new FlinkLogConsumer<RawLogGroupList>(deserializer
, configProps));
```

The preceding is a simple consumption example. As java.util.Properties is used as the configuration tool, configurations of all consumers can be located in ConfigConstants.

Note:
The number of sub-tasks in the Flink stream is independent from that of shards in the Log Service Logstore. If the number of shards is greater than that of sub-tasks, each sub-task consumes multiple shards exactly once. If the number of shards is smaller than that of sub-tasks, some sub-tasks are idle until new shards are generated.

2. Set the consumption start position.

You can set the start position for consuming a shard on the Flink log consumer. By setting ConfigConstants.LOG_CONSUMER_BEGIN_POSITION, you can set whether to consume a shard from its header or tail or at a specific time point. The connector also supports consumption restoration from a specific consumer group. The specific values are as follows:

- Consts.LOG_BEGIN_CURSOR: indicates that the shard is consumed from its header, that is, from the earliest data of the shard.
- Consts.LOG_END_CURSOR: indicates that the shard is consumed from its tail, that is, from the latest data of the shard.
- Consts.LOG_FROM_CHECKPOINT: indicates that the shard is consumed from the saved checkpoint in a specific consumer group. The consumer group is specified by ConfigConstants .LOG_CONSUMERGROUP.
- UnixTimestamp: a string of an integer value, which is expressed in seconds from 1970-01-01 00:00:00 UTC. It indicates that the shard is consumed from this time point.

Examples of the preceding four values are as follows:

```
configProps.put(ConfigConstants.LOG_CONSUMER_BEGIN_POSITION, Consts.
LOG_BEGIN_CURSOR);
                configProps.put(ConfigConstants.LOG_CONSUMER_BEGIN_P
OSITION, Consts.LOG_END_CURSOR);
                configProps.put(ConfigConstants.LOG_CONSUMER_BEGIN_P
OSITION, "1512439000");
                configProps.put(ConfigConstants.LOG_CONSUMER_BEGIN_P
OSITION, Consts.LOG_FROM_CHECKPOINT);
```

Note:

If you have configured consumption restoration from the StateBackend of Flink when you start the Flink task, the connector ignores the preceding configurations and uses the checkpoint saved in StateBackend.

3. (Optional) Configure consumption progress monitoring.

The Flink log consumer supports consumption progress monitoring. The consumption progress indicates the real-time consumption position of each shard, which is expressed using the timestamp.

```
configProps.put(ConfigConstants.LOG_CONSUMERGROUP, "your consumer
group name");
```

Note:

The preceding code is optional. If it is set, the consumer creates a consumer group first. If the consumer group already exists, no further operation is required. Snapshots in the consumer are automatically synchronized to the consumer group of Log Service. You can view the consumption progress of the consumer in the Log Service console.

4. Configure support for disaster tolerance and exactly-once semantics.

If the checkpoint function of Flink is enabled, the Flink log consumer periodically stores the consumption progress of each shard. When a job fails, Flink restores the log consumer and starts consumption from the latest checkpoint that is stored.

The period of writing checkpoint defines the maximum amount of data to be rolled back (that is, re-consumed) when a failure occurs. The code is as follows:

```
final StreamExecutionEnvironment env = StreamExecutionEnvironment.
getExecutionEnvironment();
                // Enable the exactly-once semantics on Flink
                env.getCheckpointConfig().setCheckpointingMode(
CheckpointingMode.EXACTLY_ONCE);
                // Store the checkpoint every 5s
                env.enableCheckpointing(5000);
```

For details about the Flink checkpoints, see the Flink documentation *Checkpoints*.

Log Producer

The Flink log producer writes data into Alibaba Cloud Log Service.

Note:

The producer supports only the Flink at-least-once semantics. It means that when a job failure occurs, data written into Log Service may be duplicated but never lost.

Procedure

1. Initialize the producer.

a. Initialize properties for the producer.

The initialization process for the producer is similar to that for the consumer. The producer contains the following parameters. Set these parameters to the default values in general conditions or to custom values as required.

```
// The number of I/O threads used for sending data. The default
value is 8.
                        ConfigConstants.LOG_SENDER_IO_THREAD_COUNT
                        // The time when the log data is cached.
The default value is 3000.
                        ConfigConstants.LOG_PACKAGE_TIMEOUT_MILLIS
                        // The number of logs in the cached
package. The default value is 4096.
                        ConfigConstants.LOG_LOGS_COUNT_PER_PACKAGE
                        // The size of the cached package. The
default value is 3 Mb.
                        ConfigConstants.LOG_LOGS_BYTES_PER_PACKAGE
                        // The total memory size that the job can
use. The default value is 100 Mb.
                        ConfigConstants.LOG_MEM_POOL_BYTES
```

The preceding parameters are optional. You can retain the default values.

b. Reload LogSerializationSchema to define the method for serializing data to RawLogGroup.

RawLogGroup is a collection of logs. For the meaning of each field, see the **Data model** section in the *Log Service Developer Guide*.

To use the shardHashKey function of Log Service, specify the shard to which data is written . You can use LogPartitioner to generate the HashKey of data.

Example:

```
FlinkLogProducer<String> logProducer = new FlinkLogProducer<String
>(new SimpleLogSerializer(), configProps);
                        logProducer.setCustomPartitioner(new
LogPartitioner<String>() {
                        // Generate a 32-bit hash value
                        public String getHashKey(String element) {
                        try {
                        MessageDigest md = MessageDigest.
getInstance("MD5");
                        md.update(element.getBytes());
                        String hash = new BigInteger(1, md.digest
()).toString(16);
                        while(hash.length() < 32) hash = "0" +
hash;
                        return hash;
                        } catch (NoSuchAlgorithmException e) {
                        }
                        return  "00000000000000000000000000000000
00000000000000000000000000000000";
                        }
```

```
                        });
```

Note:

LogPartitioner is optional. If this parameter is not set, data is randomly written into a shard.

2. In the following example, a simulated string is written to Log Service:

```
// Serialize data to the data format of Log Service
                    class SimpleLogSerializer implements LogSeriali
zationSchema<String> {
                    public RawLogGroup serialize(String element) {
                    RawLogGroup rlg = new RawLogGroup();
                    RawLog rl = new RawLog();
                    rl.setTime((int)(System.currentTimeMillis() /
1000));
                    rl.addContent("message" element);
                    rlg.addLog(rl);
                    return rlg;
                    }
                    }
                    public class ProducerSample {
                    public static String sEndpoint = "cn-hangzhou.
log.aliyuncs.com";
                    public static String sAccessKeyId = "";
                    public static String sAccessKey = "";
                    public static String sProject = "ali-cn-hangzhou
-sls-admin";
                    public static String sLogstore = "test-flink-
producer";
                    private static final Logger LOG = LoggerFactory.
getLogger(ConsumerSample.class);
                    public static void main(String[] args) throws
Exception {
                    final ParameterTool params = ParameterTool.
fromArgs(args);
                    final StreamExecutionEnvironment env =
StreamExecutionEnvironment.getExecutionEnvironment();
                    env.getConfig().setGlobalJobParameters(params);
                    env.setParallelism(3);
                    DataStream<String> simpleStringStream = env.
addSource(new EventsGenerator());
                    Properties configProps = new Properties();
                    // Set the domain to access Log Service
                    configProps.put(ConfigConstants.LOG_ENDPOINT,
sEndpoint);
                    // Set the AccessKey to access Log Service
                    configProps.put(ConfigConstants.LOG_ACCESSSKEYID
, sAccessKeyId);
                    configProps.put(ConfigConstants.LOG_ACCESSKEY,
sAccessKey);
                    // Set the Log Service project into which logs
are written
                    configProps.put(ConfigConstants.LOG_PROJECT,
sProject);
                    // Set the Log Service LogStore into which logs
are written
                    configProps.put(ConfigConstants.LOG_LOGSTORE,
sLogstore);
                    FlinkLogProducer<String> logProducer = new
FlinkLogProducer<String>(new SimpleLogSerializer(), configProps);
```

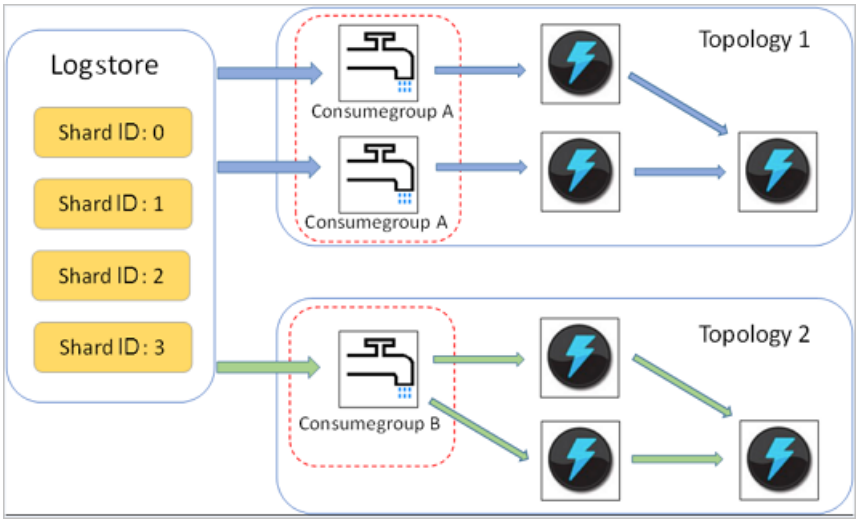
```
                    simpleStringStream.addSink(logProducer);
                    env.execute("flink log producer");
                    }
                    // Simulate log generation
                    public static class EventsGenerator implements
SourceFunction<String> {
                    private boolean running = true;
                    @Override
                    public void run(SourceContext<String> ctx)
throws Exception {
                    long seq = 0;
                    while (running) {
                    Thread.sleep(10);
                    ctx.collect((seq++) + "-" + RandomStringUtils.
randomAlphabetic(12));
                    }
                    }
                    @Override
                    public void cancel() {
                    running = false;
                    }
                    }
                    }
```

24.11.4 Storm consumption

LogHub of Log Service provides an efficient and reliable log channel for collecting log data through Logtail and SDKs. You can access real-time systems such as Spark Streaming and Storm to consume the data written to LogHub.

The LogHub Storm spout feature is provided to read data from LogHub in real time, reducing Storm users cost for LogHub consumption.

Basic architecture and process



- Enclosed in the red dotted boxes in the preceding figure are LogHub Storm spouts. Each Storm topology has a group of spouts that read all data from a Logstore. The spouts in different topologies are independent of each other.
- Each topology is identified by a unique LogHub consumer group name. The LogHub client library is used for load balancing and automatic failover among the spouts in the same topology .
- Spouts read data from LogHub in real time, send the data to the bolt nodes of the topology, and save consumption endpoints as checkpoints to the LogHub server periodically.

Restrictions

- To prevent misuse, each Logstore supports up to five consumer groups. You can use the DeleteConsumerGroup API of the Java SDK to delete unused consumer groups.
- We recommend that the number of spouts be equal to the number of shards. Otherwise, a single spout may be unable to process a large amount of data.
- If a shard contains a large amount of data which exceeds the processing capability of a single spout, you can use the shard split API to lower down the per-shard data volume.
- The dependency on the Storm ACK mechanism is mandatory in LogHub spouts, to confirm that spouts send messages correctly to bolts. Therefore, bolts must call ACK for such confirmation.

Example

- **Spout (used for topology creation)**

```
 public static void main( String[] args ) { String mode = "Local
"; // Use the local test mode. String conumser_group_name = ""; //
Each topology must be assigned a unique consumer group name,
which contains 3 to 63 characters including letters a-z, numbers
 0-9, underlines (_), and hyphens (-). The consumer group must
 be specified and must start and end with lowercase letters or
numbers. String project = ""; // Project of Log Service. String
logstore = ""; // Logstore of Log Service. String endpoint = ""; //
Domain name of Log Service. String access_id = ""; // AccessKey
 of the user. String access_key = ""; // Configurations required
 for creating a LogHub Storm spout. LogHubSpoutConfig config =
new LogHubSpoutConfig(conumser_group_name, endpoint, project,
logstore, access_id, access_key, LogHubCursorPosition.END_CURSOR
); TopologyBuilder builder = new TopologyBuilder(); // Create a
 LogHub Storm spout. LogHubSpout spout = new LogHubSpout(config
); // In the actual situation, the number of spouts may be equal
to the number of Logstore shards. builder.setSpout("spout", spout
, 1); builder.setBolt("exclaim", new SampleBolt()).shuffleGrouping
("spout"); Config conf = new Config(); conf.setDebug(false); conf
.setMaxSpoutPending(1); // The serialization method LogGroupDa
taSerializSerializer of LogGroupData must be configured explicitly
 when Kryo is used for data serialization and deserialization.
LogGroupDataSerializSerializer Config.registerSerialization(conf
, LogGroupData.class, LogGroupDataSerializSerializer.class); if (
```

```
mode.equals("Local")) { logger.info("Local mode..."); LocalClust
er cluster = new LocalCluster(); cluster.submitTopology("test-
jstorm-spout", conf, builder.createTopology()); try { Thread.sleep
(6000 * 1000); //waiting for several minutes } catch (Interrupte
dException e) { // TODO Auto-generated catch block e.printStackTrace
(); } cluster.killTopology("test-jstorm-spout"); cluster.shutdown
(); } else if (mode.equals("Remote")) { logger.info("Remote mode
..."); conf.setNumWorkers(2); try { StormSubmitter.submitTopology
("stt-jstorm-spout-4", conf, builder.createTopology()); } catch (
AlreadyAliveException e) { // TODO Auto-generated catch block e.
printStackTrace(); } catch (InvalidTopologyException e) { // TODO
Auto-generated catch block e.printStackTrace(); } } else { logger.
error("invalid mode: " + mode); } } }
```

- **Sample code of the bolts that consume data. (Only the content of each log is printed.)**

```
public class SampleBolt extends BaseRichBolt { private static
final long serialVersionUID = 4752656887774402264L; private static
 final Logger logger = Logger.getLogger(BaseBasicBolt.class);
private OutputCollector mCollector; @Override public void prepare(@
SuppressWarnings("rawtypes") Map stormConf, TopologyContext context
, OutputCollector collector) { mCollector = collector; } @Override
 public void execute(Tuple tuple) { String shardId = (String) tuple
 .getValueByField(LogHubSpout.FIELD_SHARD_ID); @SuppressWarnings("
unchecked") List<LogGroupData> logGroupDatas = (ArrayList<LogGroupDa
ta>) tuple.getValueByField(LogHubSpout.FIELD_LOGGROUPS); for (
LogGroupData groupData : logGroupDatas) { // Each log group consists
 of one or more logs. LogGroup logGroup = groupData.GetLogGroup();
for (Log log : logGroup.getLogsList()) { StringBuilder sb = new
StringBuilder(); // Each log has a time field and multiple key-value
 pairs. int log_time = log.getTime(); sb.append("LogTime:").append(
log_time); for (Content content : log.getContentsList()) { sb.append
("\t").append(content.getKey()).append(":") .append(content.getValue
()); } logger.info(sb.toString()); } } // The dependency on the
Storm ACK mechanism is mandatory in LogHub spouts, to confirm that
 spouts send messages correctly to bolts. //Therefore, bolts must
call ACK for such confirmation. mCollector.ack(tuple); } @Override
```

```
public void declareOutputFields(OutputFieldsDeclarer declarer) { //
do nothing } }
```

Maven

Use the following code for versions earlier than Storm 1.0 (for example, 0.9.6):

```
<dependency> <groupId>com.aliyun.openservices</groupId> <artifactId>
loghub-storm-spout</artifactId> <version>0.6.5</version> </dependency>
```

Use the following code for Storm 1.0 and later versions:

```
<dependency> <groupId>com.aliyun.openservices</groupId> <artifactId
>loghub-storm-1.0-spout</artifactId> <version>0.1.2</version> </
dependency>
```

24.11.5 Spark Streaming consumption

Log Service supports consuming logs in real time by using Spark Streaming.

E-MapReduce has implemented a set of universal APIs for Spark Streaming to calculate real-time LogHub data consumption. To download SDKs, go to *GitHub*.

24.11.6 Consumption by StreamCompute

StreamCompute can be used to consume data directly in LogHub after data sources of the LogHub type are created.

StreamCompute can be used to consume data directly in LogHub after data sources of the LogHub type are created.

```
CREATE STREAM TABLE source_test_galaxy ( $schema ) WITH ( type='loghub
',
endpoint=$endpoint, accessId=$loghub_access_id, accessKey=$loghub_acc
ess_key, projectName=$project, logstore=$logstore );
```

Table 24-17: Parameter list

Parameter	Description
`$schema`	The keys in logs that are mapped to the columns in the StreamCompute table, for example, `name STRING, age STRING, id STRING`.
`$endpoint`	Your endpoint.
`$loghub_access_id`	AccessID of the account (or subaccount) with the read permission.

Parameter	Description
`$loghub_access_key`	Accesskey of the account (or subaccount) with the read permission.
`$project`	Project where data is located.
`$logstore`	Logstore where data is located.

Example:

```
CREATE STREAM TABLE source_test_galaxy ( name STRING, age STRING, id
 STRING ) WITH ( type='loghub', endpoint='http://cn-hangzhou-intranet
.log.aliyuncs.com', accessId='mock_access_id', accessKey='mock_acces
s_key', projectName='ali-cloud-streamtest', logstore='stream-test' );
```

25 Domain Name System (DNS)

25.1 What is Apsara Stack DNS?

Apsara Stack DNS provides basic domain name translation and scheduling services for VPC environments. You can perform the following operations through Apsara Stack DNS in your VPC:

- Access other ECS servers deployed in VPCs.
- Access cloud service instances provided by Apsara Stack.
- Access custom enterprise business systems.
- Access Internet services and business.
- Establish network connections between Apsara Stack DNS and user-created DNS through a leased line.

You can perform the following operations through Apsara Stack DNS:

- Manage domain names
- Manage DNS records
- Manage default forwarding

25.2 Log on to the DNS console

The following example uses Google Chrome to demonstrate how to log on to the Domain Name System (DNS) console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.

- The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
- You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click **LOGIN** to go to the **Dashboard** page.
5. On the top menu bar, choose **Console** > **Domain Name System**.

25.3 Global internal domains

25.3.1 Overview

Apsara Stack DNS provides data management for internal domain names. You can register, search, and delete internal domain names and add remarks. You can also add, delete, and modify DNS records.

Supported DNS record types include A, AAAA, CNAME, NS, MX, TXT, SRV, and PTR.

Internal domain name management can translate internal domain names for servers deployed in a VPC. The DNS endpoints are deployed based on anycast, which ensures the continuity of services if errors occur.

25.3.2 View internal domain names

You can view internal domain names in the Apsara Stack console.

Procedure

1. *Log on to the DNS console*.
2. Click **Global Internal Domains** to view all internal domains.

25.3.3 Add a domain name

You can add a domain name in the Apsara Stack console.

Procedure

1. *Log on to the DNS console*.

2. On the **Global Internal Domains** page, click **Create Domain Name** in the upper-right corner.
3. In the dialog box that appears, enter a name in the `Domain Name` filed and click **OK**.

25.3.4 Add remarks for a domain name

You can add remarks for a domain name in the Apsara Stack console.

Context

This operation allows you to add remarks for the use of a domain name, such as the hostname and internal information system.

Procedure

1. *Log on to the DNS console*.
2. On the **Global Internal Domains** page, select the domain name that you want to add remarks to, and click **Description** in the Actions column.
3. In the dialog box that appears, enter your remarks in the Description field and click **OK**.

25.3.5 Delete a domain name

You can delete a domain name in the Apsara Stack console.

Procedure

1. *Log on to the DNS console*.
2. On the **Global Internal Domains** page, select the domain name that you want to delete, and click **Delete** in the Actions column.
3. In the dialog box that appears, click **OK** to delete the domain name.

25.3.6 Delete multiple domain names

You can delete multiple domain names at the same time in the Apsara Stack console.

Procedure

1. *Log on to the DNS console*.
2. On the **Global Internal Domains** page, select the domain names that you want to delete, and click **Delete Domain Names** in the upper-right corner.
3. In the dialog box that appears, click **OK** to delete the domain names.

25.3.7 Manage DNS records

You can manage DNS records in the Apsara Stack console.

Procedure

1. *Log on to the DNS console*.
2. On the **Global Internal Domains** page, select the domain name that you manage, and click **Manage Resource Records** in the Actions column.
3. On the **Manage Resource Records** page, click **Add Resource Record Set** in the upper-right corner.
4. In the **Add Resource Record Set** dialog box that appears, enter a hostname in the `Name` field, select a record `Type` and a `Resolution Policy`, specify a number in the `TTL` field, enter data entries in the **Data** field, and click **OK**.
5. After you have added a DNS record, you can perform the following operations:
   - **Add remarks for a DNS record**

     Select a DNS record that you want to add remarks to and click **Description** in the Actions column. Enter the remarks in the Description field and click **OK**.
   - **Delete a DNS record**

     Select a DNS record that you want to delete, click **Delete** in the upper-right corner or in the Actions column, and click **OK** to delete the DNS record.
   - **Modify a DNS record**

     Select a DNS record that you want to modify and click **Change** in the Actions column. In the dialog box that appears, enter the required information and click **OK**.
   - **Delete multiple DNS records**

     Select the DNS records that you want to delete, click **Delete** in the upper-right corner, and click **OK** in the dialog box that appears.

25.4 Global forwarding domains

25.4.1 Overview

All operations of this feature require administrator privileges.

All operations of this feature require administrator privileges.

Apsara Stack DNS can forward requests to other DNS servers for resolution.

Domain name forwarding includes two forwarding modes: forward all requests (without recursion) and forward all requests (with recursion).

- When the forward all requests (without recursion) mode is set, the system only uses forwarders for resolution. If the domain name cannot be resolved or the request is timed out, then a massage is returned to the client indicating that the request fails.
- When the forward all requests (with recursion) mode is set, The system uses forwarders for resolution first. If the domain name cannot be resolved, then the request is sent to the local DNS server for resolution.

25.4.2 View a forwarding domain

You can search for and view a forwarding domain in the Apsara Stack console. This operation requires administrator privileges.

Procedure

1. *Log on to the DNS console*.
2. Click **Global Forward Domains**.
3. Enter a domain name in the **Domain Name** field and click **Search** to search for and view the domain name.

25.4.3 Add a domain name

You can add a domain name in the Apsara Stack console. This operation requires administrator privileges.

Procedure

1. *Log on to the DNS console*.
2. On the **Global Forward Domains** page, click **Create Domain Name**.
3. In the dialog box that appears, enter a name in the `Domain Name` field, select a `Forwarding Mode`, enter one or multiple IP addresses in the `Forwarder IP Addresses` field, and click **OK**.

25.4.4 Add remarks for a domain name

You can add remarks for a domain name in the Apsara Stack console. This operation requires administrator privileges.

Context

This operation allows you to add remarks for the use of a domain name, such as the hostname and internal system information.

Procedure

1. *Log on to the DNS console*.
2. Click **Global Forward Domains**.
3. Select the domain name that you want to add remarks to and click **Description**.
4. Enter the remarks in the Description field and click **OK**.

25.4.5 Change forwarding settings

You can change forwarding settings for a domain name in the Apsara Stack console. This operation requires administrator privileges.

Procedure

1. *Log on to the DNS console*.
2. Click **Global Forward Domains**.
3. Select the domain name that you want to modify and click **Change**.
4. In the dialog box that appears, select a `Forwarding Mode`, enter one or multiple IP addresses in the `Forwarder IP Addresses` field, and click **OK**.

25.4.6 Delete a domain name

You can delete a domain name in the Apsara Stack console. This operation requires administrator privileges.

Procedure

1. *Log on to the DNS console*.
2. Click **Global Forward Domains**.
3. Select the domain name that you want to delete and click **Delete** in the Actions column.
4. In the dialog box that appears, click **OK** to delete the domain name.

25.4.7 Delete multiple domain names

You can delete multiple domain names at the same time in the Apsara Stack console. This operation requires administrator privileges.

Procedure

1. *Log on to the DNS console*.

2. Click **Global Forward Domains**.
3. Select the domain names that you want to delete and click **Delete Domain Names** in the upper-right corner.
4. In the dialog box that appears, click **OK** to delete the domain names.

25.5 Global default forwarding

25.5.1 Enable default forwarding

You can enable default forwarding in the Apsara Stack console. This operation requires administrator privileges.

Procedure

1. *Log on to the DNS console*.
2. Click **Global Default Forwarding**.
3. Click **Enable** in the Actions column.
4. In the dialog box that appears, select a `Forwarding Mode`, enter one or multiple IP addresses in the `Forwarder IP Addresses` filed, and click **OK**.

   After you have enabled default forwarding, **`Enable Default Forwarding`** is set to **`ON`** and the forwarding mode and forwarders are displayed on the page.

25.5.2 Change default forwarding settings

You can change default forwarding settings in the Apsara Stack console. This operation requires administrator privileges.

Procedure

1. *Log on to the DNS console*.
2. Click **Global Default Forwarding**.
3. Click **Change** in the Actions column.
4. In the dialog box that appears, select a `Forwarding Mode`, enter one or multiple IP addresses in the `Forwarder IP Addresses` field, and click **OK**.

25.5.3 Disable default forwarding

You can disable default forwarding in the Apsara Stack console. This operation requires administrator privileges.

Procedure

1. *Log on to the DNS console*.
2. Click **Global Default Forwarding**.
3. Click **Disable** in the Actions column.
4. In the dialog box that appears, click **OK** to disable default forwarding.

26 API Gateway

26.1 Product overview

API gateway is a complete API hosting service. It helps you use APIs to provide capabilities, services, and data to your partners. You can also publish APIs to the API marketplace for other developers to purchase and use.

- API Gateway provides a range of mechanisms to enhance security and reduce risks arising from APIs. These mechanisms include attack prevention, replay prevention, request encryption , identity authentication, permission management, and request throttling.
- API Gateway provides a full range of API lifecycle management functions, including creating, testing, publishing, and unpublishing APIs. It also generates SDKs and API documentation to improve API management and iteration efficiency.
- API Gateway provides convenient O&M functions to reduce API O&M costs, including monitoring, alarms, and log analysis.

API Gateway maximizes capability multiplexing. It allows enterprises to share capabilities and focus more on their core businesses, which benefits all parties involved.

26.2 Quick start for consumers

26.2.1 Overview

You can use API Gateway to call the API services enabled by other Alibaba Cloud users or third-party service providers. API Gateway provides a series of management and support services.

Call an API based on the following conditions:

- **API:** The API that you call is clearly defined by API parameters.
- **Application:** The application that you use to call the API has a key pair that uniquely identifies you.
- **Authorization relationship between the API and application:** An application can be used to call an API only when the application has been granted the permission to call that API. This permission is granted through authorization.

26.2.2 Step 1: Obtain the API document

API providers need to authorize your applications to use their APIs in the API Gateway console. Provide your application IDs (AppId) to the API providers so that the providers can authorize your

applications. For more information about applications, see *Create an application*. Assume that you have created an application and the API provider has authorized your application to use their APIs.

Procedure

1. Log on to the API Gateway console.
2. Click the **Applications** tab to go to the **Applications** tab page.

   The application that you created is displayed in the application list.
3. Click the application name to go to the application details page.

   The application details page consists of the Basic Information, **AppKey**, and **Callable APIs** areas.

   On the application details page:

   - The **AppKey** area shows the **AppKey** and **AppSecret** of the application. Your API request must contain the AppKey and AppSecret. API Gateway verifies your identity based on this key pair.
   - The **Callable APIs** area shows the APIs that the application has been authorized to use. If the API provider has authorized the application to use their APIs, the corresponding APIs are displayed in the callable API list. In the callable API list, click the management icon in the Actions column corresponding to an API and choose **View Details** from the shortcut menu to view the API details.

26.2.3 Step 2: Create an application

Applications are the identities that you use to call APIs. You can own multiple applications that are authorized to use different APIs based on your service requirements. Instead of user accounts, applications are authorized to use APIs. In the API Gateway console, you can create, change, or delete applications, view application details including callable APIs, and manage keys for applications.

Each application has a key pair made up of an **AppKey** and **AppSecret**. This key pair works similar to the way an account and password works. When calling an API, you must include the **AppKey** as a parameter in the request. The **AppSecret** is used to calculate the signature string. API Gateway verifies your identity based on the key pair. Before using an application to call an API, ensure that the application has been authorized to use the API. Both authorization and verification are performed on the application.

You can log on to the API Gateway console to create applications on the **Applications** tab page.

Procedure

1. Log on to the API Gateway console.
2. Click the **Applications** tab.
3. Click **Create Application**.
4. Set parameters and click **Create**.

   The application name must be globally unique. It can contain English letters, numbers, and underscores (_). It must start with a letter and be 4 to 26 characters in length.

   After an application is created, the system automatically assigns an **Appkey** and an **AppSecret** to it. You need to use the **AppSecret** to calculate the signature string. When calling an API, you must include the string in the request. API Gateway verifies your identity based on the signature.

   On the **Applications** tab page, click the application name to go to the application details page. The **AppKey** and **AppSecret** information is displayed in the lower part of the tab page. If the key pair is lost, you can reset it.

26.2.4 Step 3: Obtain authorization

Authorization is the process of authorizing an application to call an API. Your application must be authorized to call an API.

Provide your application ID (AppID) to the API provider so that the provider can authorize your application. After authorization is complete, log on to the API Gateway console.

You can view the API in the **Callable APIs** list on the application details page.

Only the API provider has the permission to authorize applications to call APIs.

26.2.5 Step 4: Call the API

You can use a self-compiled HTTP or HTTPS request to call the API.

Part 1: Request

Request address

```
http://e710888d3ccb4638a723ff8d03837095-cn-qingdao.aliapi.com/demo/
post
```

Request method

```
POST
```

Request body

```
FormParam1=FormParamValue1&FormParam2=FormParamValue2
//HTTP Request Body
```

Request header

```
Host: e710888d3ccb4638a723ff8d03837095-cn-qingdao.aliapi.com
Date: Mon, 22 Aug 2016 11:21:04 GMT
User-Agent: Apache-HttpClient/4.1.2 (java 1.6)
Content-Type: application/x-www-form-urlencoded; charset=UTF-8
//The request body type. Set the type based on the actual request
content.
Accept: application/json
//The response body type. Some APIs return data based on the specified
 response body type. We recommend that you set this Header parameter
manually. If you do not set this parameter, some HTTP clients set it
to */* by default. This can cause a signature error.
X-Ca-Request-Mode: debug
//Whether to enable the Debug mode. This parameter is case-insensitive
. The Debug mode is disabled by default. It is usually enabled during
API debugging.
X-Ca-Version: 1
//The API version number. Only version 1 is supported. This parameter
is optional. Default value: 1.
X-Ca-Signature-Headers: X-Ca-Request-Mode,X-Ca-Version,X-Ca-Stage,X-Ca
-Key,X-Ca-Timestamp
//Custom Header parameters that are used for signature calculation
. The server reads Header parameters based on this setting during
signature calculation. Content-Type, Accept, Content-MD5, and Date
are part of the basic signature structure, and are not included in
custom Header parameters. For more information, see request signature
instructions.
X-Ca-Stage: RELEASE
//The stage of the requested API. Values: TEST, PRE, and RELEASE. This
 parameter is case-insensitive. An API provider can decide to which
stage the API is to be published. When you call an API that has not
been published, an error message is returned, indicating an invalid
URL or that the API is not found.
X-Ca-Key: 60022326
//The request AppKey. AppKeys are generated in the API Gateway console
. An application can call an API only when the application has been
authorized to use the API.
```

```
X-Ca-Timestamp: 1471864864235
//The request timestamp. It is the current time, in milliseconds. A
timestamp is valid for 15 minutes.
X-Ca-Nonce: b931bc77-645a-4299-b24b-f3669be577ac
//The unique identifier of the request. The combination of AppKey,
API, and Nonce must be unique for requests within 15 minutes. This
parameter must be used together with the timestamp to prevent replay.
X-Ca-Signature: FJleSrCYPGCU7dMlLTG+UD3Bc5Elh3TV3CWHtSKh1Ys=
//The request signature.
CustomHeader: CustomHeaderValue
//Custom Header parameters. The preceding example serves only as a
reference. You can set several custom Header parameters based on the
API definition.
```

Part 2: Response

Status code

```
400
//The response status code. If the value is greater than or equal to
 200 and smaller than 300, the request is successful. If the value
is greater than or equal to 400 and smaller than 500, a client error
 has occurred. If the value is greater than 500, a server error has
occurred.
```

Response header

```
X-Ca-Request-Id: 7AD052CB-EE8B-4DFD-BBAF-EFB340E0A5AF
//The unique ID of the request. After receiving a request, API Gateway
 generates a request ID and returns the request ID to the client
through the response header. We recommend that the request ID be
recorded by both the client and the backend service for troublesho
oting and tracking.
X-Ca-Error-Message: Invalid Url
//The error message returned by API Gateway. When a request fails, API
 Gateway returns the error message to the client through the response
header.
X-Ca-Debug-Info: {"ServiceLatency":0,"TotalLatency":2}
//The Debug message that is returned when the Debug mode is enabled
. The message may be changed in the future and is used only for
reference during debugging.
```

To call an API, you must attach a signature to the HTTP or HTTPS request. For more information about the signature encryption algorithm, see *Request signature instructions*.

26.3 Quick start for providers

26.3.1 Overview

This document guides you through creating and publishing an API.

This document guides you through performing the following operations:

1. Create an API group

2. Bind domain names and certificates
3. Create an API
4. Publish an API
5. Authorize applications to use APIs

26.3.2 Create a group

You can create API groups in the API Gateway console. You can create up to 50 groups.

Procedure

1. Log on to the API Gateway console.
2. Click the **Groups** tab.
3. On the Groups tab page that appears, click **Create Group**.
4. In the Create Group dialog box that appears, set Department, Project, and other required parameters. Click **Create**.

   Group names must be globally unique. A group name must be 4 to 50 characters in length. It can contain English letters, numbers, and underscores (_) and must start with an English letter.

26.3.3 Create an API

Procedure

1. Log on to the API Gateway console.
2. Click the **API** tab.
3. Click **Create API**.
4. Set the basic information of the API and click **Next**.

Parameter	Description
Groups	The group to which the API belongs. APIs are managed by groups. Before you create an API, you must create a group. Select a group from the **Groups** drop-down list.
API Name	The API name.
Authentication Mode	The authentication mode of API requests. Values: **Alibaba Cloud Applications** and **None**. • **Alibaba Cloud Applications**: This authentication mode requires the requester

Parameter	Description
	to pass the application authentication to call this API. • **None**: This authentication mode allows any user who knows the request definition of the API to initiate a request. API Gateway directly forwards the requests to your backend service without verifying the identity of the requesters.
Description	The description of the API.

5. Define API requests. Define the settings of the requests that users can send to call the API, including the related protocols, request paths, HTTP methods, and parameters.

Parameter	Description
Network Protocol	The protocol that can be used to call the API. Both HTTP and HTTPS are supported.
Custom Domain Name	The independent domain name that has been bound to the group to which the API belongs.
Second-Level Domain Name	The second-level domain name of the group to which the API belongs.
URL Suffix	The API request path. It corresponds to the service host. The request path can be different from the backend service path. You can provide any valid and semantically -correct path for users. You can configure dynamic parameters in the request path. API Gateway can map these parameters to Query , Header, or other locations before sending the requests to the backend service.
HTTP Method	The HTTP method supported by the API. Values: PUT, GET, POST, PATCH, DELETE , and HEAD.
Request Parameters	The request parameters of the API. These parameters need to be set by the users. You can define the request parameters in Header, Query, Body, or Path (Parameter Path). If you have defined dynamic parameters in Path, specify how to set these dynamic parameters when your define request parameters. The

Parameter	Description
	following parameter types are supported: String, Number, and Boolean.
Parameter verification rules	Click the management icon in the Actions column corresponding to a request parameter, and choose **Configure Advanced Settings** from the shortcut menu. In the Configure Advanced Settings dialog box that appears, you can configure parameter verification rules, such as Maximum Length and Enumeration. API Gateway pre-verifies requests based on the verification rules. The requests with invalid parameters are not sent to your backend service. This greatly reduces the work load of the backend service.

6. Configure the backend service and click **Next**.

   This section defines mappings between frontend and backend parameters, and specifies the API backend service configurations. The backend service configurations include the backend service address, backend service path, backend response timeout period, parameter mappings, constant parameters, and system parameters. After receiving user requests, API Gateway converts the format of the requests to the format required by the backend service based on the backend service configuration. Then, API Gateway forwards the requests to the backend service.

    **Note:**
   You can enter the following parameters: dynamic parameters in Path, Header parameters, Query parameters, Body parameters (non-binary), constant parameters, and system parameters. Each parameter name must be globally unique. For example, you are not allowed to enter a parameter named **name** in both Header and Query.

   a) Configure basic backend service information.

Parameter	Description
Backend Service URL	The backend service host. It can be a domain name or an address in the format of `http(s)://host:port`.
URL Suffix	The actual request path of your API service on the backend server. If you

Parameter	Description
	have configured dynamic parameters in the backend path, you must specify the corresponding request parameters and their locations by declaring the mapping.
Timeout	The maximum amount of time that API Gateway waits for a response from the backend service. API Gateway sends a request to the backend service and waits for a response. The maximum timeout period is 30 seconds. If API Gateway does not receive a response from the backend service within the timeout period, it stops waiting and returns an error.
Mock	You can mock expected responses to return to API callers during the project development process. For more information, see Mock an API.

b) Configure backend service parameters.

API Gateway can set up mappings between frontend and backend parameters, including names and locations. API Gateway can map a request parameter at any location (Path, Header, Query, or Body) to a backend service parameter at a different location. In this way, you can package your backend services into standard APIs. This part declares the frontend-to-backend API mapping.

Note:
The frontend and backend parameters must be globally unique.

c) Configure constant parameters.

To configure API Gateway to add the **`apigateway`** tag to each request it forwards to your backend service, you can configure the tag as a constant parameter. Constant parameters are invisible to users. After a constant parameter is configured, API Gateway automatically adds this parameter to the specified location of the received requests before sending the requests to your backend service.

d) Configure system parameters.

API Gateway does not send its system parameters to you by default. To obtain these parameters, configure their locations and names in the API. The following table lists the system parameters.

Parameter	Description
CaClientIp	The IP address of the client that sends the request
CaDomain	The domain name that is used to send the request
CaRequestHandleTime	The request time (UTC)
CaAppId	The ID of the application that sends the request
CaRequestId	The request ID
CaApiName	The API name
CaHttpSchema	The protocol that is used to call the API, which is HTTP or HTTPS
CaProxy	The proxy (AliCloudApiGateway)

7. Define the response and click **Create**.

   You can set Response Content Type, Success Response Example, and Error Response Example, and define error codes. API Gateway does not parse responses, but forwards them to the API requester.

26.3.4 Publish an API

After creating an API, you need to debug, test, and publish it.

- When you use a second-level or independent domain name to access an API published to an environment, you must specify the environment to be requested in the request header.
- If you publish an API to the test or release environment where the API already has a version running, the newly published version replaces the running version to take effect in real time. However, all historical versions and definitions are recorded so you can roll the API back to an earlier version.
- You can unpublish an API in the test or release environment. After the API is unpublished , its binding or authorization relationships with the policy, keys, or applications persist and will automatically take effect when the API is published again. Remove these relationships separately if you no longer need them.

Step 1: Test the API

You can create an application and authorize the application to use this API. Then, use this application to simulate real user requests.

You can write code based on real request scenarios or use the SDK samples provided by API Gateway to call the API.

You can publish an API to the test or release environment. If no independent domain name is bound to the group to which the API belongs, you can test or call the API by using the second-level domain name. Specify an environment in your request. If no environment is specified in your request, the API published to the release environment is called by default.

Step 2: Publish the API

After testing the API, you can publish it.

You can use API Gateway to manage versions of APIs in the test or release environment. You can publish or unpublish an API and change its version. The version change takes effect in real time.

1. Log on to the API Gateway console.
2. Click the **API** tab.
3. Locate the API that you want to publish, click the management icon in the Actions column, and choose **Publish** from the shortcut menu.
4. In the dialog box that appears, select an environment, enter a description, and click **OK**.

26.3.5 Authorize applications to call APIs

You need to authorize an application before the application can call your API. After you publish an API to the release environment, you must authorize the applications of users before they can use the API. You can perform an authorization or deauthorization operation to establish or remove an authorization relationship between an API and an application. API Gateway verifies the authorization relationship.

Note:

- You can authorize one or more applications to use one or more APIs.
- If an API has been published to both the test and release environments but only the test environment is selected, applications are authorized to use only the API in the test environment.
- You can locate an application based on its ID or name provided by the user.

- To revoke the authorization from an application under an API, go to the Authorization tab page of the API. Select the application from the application list. Then, click the management icon in the Actions column and choose Deauthorize from the shortcut menu, or click Deauthorize in the upper-right corner.

Applications indicate requesters' identities. When you or your customers test or call an API, you must create an application as the requester's identity and then authorize the application to use the API.

Note:
Authorizations are environment-specific. The same application must be authorized to use the same API in both the test and release environments to avoid errors caused by inconsistency between the authorized environment and requested environment.

Procedure

1. Log on to the API Gateway console.
2. Click the **API** tab.
3. Locate the API that you want to authorize applications to use, click the management icon in the Actions column, and choose **Authorize** from the shortcut menu.
4. In the dialog box that appears, set Environment.
5. Select applications from the left-side list, click **right arrow** to add the selected applications to the right-side list, and click **OK**.

   You can enter an application ID or name to search for applications.

What's next

You can view the authorization information of APIs or revoke the authorization from an application under an API.

In the API list, click the management icon in the **Actions** column corresponding to an API, and choose View Details from the shortcut menu. Click the **Authorization** tab. On the Authorization tab page, you can view the applications that have been authorized to use this API.

You can select one or more application IDs and click **Deauthorize** in the upper-right corner to revoke the authorizations from the selected applications under the API.

26.4 Call an API

26.4.1 Manage applications

26.4.1.1 Create an application

Applications are the identities that you use to call APIs. You can own multiple applications that are authorized to use different APIs based on your service requirements. Instead of user accounts, applications are authorized to use APIs. In the API Gateway console, you can create, change, or delete applications, view application details including callable APIs, and manage keys for applications.

Each application has a key pair made up of an **AppKey** and **AppSecret**. This key pair works similar to the way an account and password works. When calling an API, you must include the **AppKey** as a parameter in the request. The **AppSecret** is used to calculate the signature string. API Gateway verifies your identity based on the key pair. Before using an application to call an API, ensure that the application has been authorized to use the API. Both authorization and verification are performed on the application.

You can log on to the API Gateway console to create applications on the **Applications** tab page.

Procedure

1. Log on to the API Gateway console.
2. Click the **Applications** tab.
3. Click **Create Application**.
4. Set parameters and click **Create**.

   The application name must be globally unique. It can contain English letters, numbers, and underscores (_). It must start with a letter and be 4 to 26 characters in length.

   After an application is created, the system automatically assigns an **Appkey** and an **AppSecret** to it. You need to use the **AppSecret** to calculate the signature string. When calling an API, you must include the string in the request. API Gateway verifies your identity based on the signature.

   On the **Applications** tab page, click the application name to go to the application details page. The **AppKey** and **AppSecret** information is displayed in the lower part of the tab page. If the key pair is lost, you can reset it.

26.4.1.2 View application details

You can view the details of applications.

Procedure

1. Log on to the API Gateway console.
2. Click the **Applications** tab to go to the **Applications** tab page.
3. Click the name of the application that you want to view.

   View the basic information, AppKey, and callable APIs of the application. The callable APIs are the APIs that this application has been authorized to use.

26.4.1.3 Change an application

You can change existing applications.

Procedure

1. Log on to the API Gateway console.
2. Click the **Applications** tab to go to the **Applications** tab page.
3. Locate the application that you want to change, click the management icon in the Actions column, and choose **Change** from the shortcut menu.
4. Change the application information and click **OK**.

26.4.1.4 Delete an application

You can delete existing applications.

Procedure

1. Log on to the API Gateway console.
2. Click the **Applications** tab to go to the **Applications** tab page.
3. Locate the application that you want to delete, click the management icon in the Actions column, and choose **Delete** from the shortcut menu.
4. In the message that appears, click **OK**.

26.4.2 View existing APIs

You can view existing APIs in the API Gateway console.

Procedure

1. Log on to the API Gateway console.
2. Click the **API** tab.

26.4.3 Authorization

Authorization is the process of authorizing an application to call an API. Your application must be authorized to call an API.

Provide your application ID (AppID) to the API provider so that the provider can authorize your application. After authorization is complete, log on to the API Gateway console.

You can view the API in the **Callable APIs** list on the application details page.

Only the API provider has the permission to authorize applications to call APIs.

26.4.4 Encrypt the signature

When you call an API, you need to construct the signature string and place the calculated string in the request header. API Gateway performs symmetric signature calculation to verify the identity of the requester.

- The calculated signature string is attached to the request header.
- You need organize the request parameters into StringToSign based on *request signature instructions*. Then, use the algorithm provided in the SDK sample to calculate the signature. The result is the preceding calculated signature string.
- Both HTTP and HTTPS requests must carry signatures.

For more information about the organization method of StringToSign, see *Request signature instructions*. You only need to change the AppKey and AppSecret in the SDK sample to your own AppKey and AppSecret. Organize StringToSign based on request signature instructions. Then, you can use the signature string to initiate a request.

26.4.5 Request signature instructions

Domain name

- Each API belongs to an API group, and each API group has a unique domain name. The domain names are independent domain names that are bound to API groups by the service provider. API Gateway uses domain names to locate API groups.
- Domain names are in the format of *`www.[independent domain name].com/[Path]?[HTTPMethod]`*.
- API Gateway locates a API group by domain name, and locates the unique API by the combination of Path and HTTPMethod.
- After you purchase an API, you can obtain the API documentation from the **Purchased APIs** list in the API Gateway console. If you have not purchased an API, you need to require

the API provider to authorize your application to call the API. Then, you can obtain the API documentation from the **Callable APIs** list on the application details page.

System-level Header parameters

- (Required) X-Ca-Key: AppKey.
- (Required) X-Ca-Signature: the signature string.
- (Optional) X-Ca-Timestamp: the timestamp passed by the API caller. It is the current time, in milliseconds. A timestamp is valid for 15 minutes.
- (Optional) X-Ca-Nonce: the UUID generated by the API caller. This parameter is used together with the timestamp to prevent replay attacks.
- (Optional) Content-MD5: When the request body is not a form, you can calculate the MD5 value of the body. Then, you can send the value to API Gateway for an MD5 check of the body.
- (Optional) X-Ca-Stage: the stage of the requested API. Values: TEST, PRE, and RELEASE. Default value: RELEASE. If the called API is not in the release environment, you must specify this parameter. If the called API is not in the release environment and you do not specify this parameter, an error is reported.

Signature verification

Organize the strings used for signature calculation

```
String stringToSign=
HTTPMethod + "\n" +
Accept + "\n" +                //Set Accept in Header. If Accept is
 empty, some HTTP clients set the value to */* by default. This can
cause a signature verification failure.
Content-MD5 + "\n"
Content-Type + "\n" +
Date + "\n" +
Headers +
Url
```

The value of HTTPMethod is in all caps, for example, POST.

If Accept, Content-MD5, Content-Type, and Date are empty, add a linefeed `\n`. If Headers is empty, `\n` is not required.

Content-MD5

Content-MD5 indicates the MD5 value of the body. The MD5 value is calculated only when the body is not a form. The calculation formula is as follows:

```
String content-MD5 = Base64.encodeBase64(MD5(bodyStream.getbytes("UTF-8")));
```

bodyStream indicates the byte array.

Headers

It indicates the string constructed by the keys and values of the Header parameters that are used for Headers signature calculation. We recommend that the parameters starting with X-Ca and custom Header parameters be used for signature calculation.

Note:

The following parameters are not used for Headers signature calculation: X-Ca-Signature, X-Ca-Signature-Headers, Accept, Content-MD5, Content-Type, and Date.

Headers organization method:

Sort the keys used for Headers signature calculation **in alphabetical order**. Construct the string based on the following rule: If the value of a Header parameter is empty, use `HeaderKey + ":" + "\n"` for signature calculation. The key and colon (:) must be retained.

```
String headers =
HeaderKey1 + ":" + HeaderValue1 + "\n"\+
HeaderKey2 + ":" + HeaderValue2 + "\n"\+
...
HeaderKeyN + ":" + HeaderValueN + "\n"
```

The keys of Header parameters used for Headers signature calculation are separated by commas (,), and placed in the request header. The key is X-Ca-Signature-Headers.

Url

Url indicates the **`Form`** parameter in **`Path + Query + Body`**. The organization method is as follows: For **`Query + Form`**, sort **`Key`** in alphabetical order and construct the string based on the following rule: If **`Query`** or **`Form`** is empty, `Url = Path` and the question mark (?) is not required. If **`Value`** of a parameter is empty, only **`Key`** is used for signature calculation and the equal sign (=) is not required.

```
String url =
Path +
"?" +
Key1 + "=" + Value1 +
"&" + Key2 + "=" + Value2 +
```

```
...
"&" + KeyN + "=" + ValueN
```

Note:

Note that **`Query`** or **`Form`** may have multiple **`values`**. If there are multiple values, the first value is used for signature calculation.

Calculate the signature

```
Mac hmacSha256 = Mac.getInstance("HmacSHA256");
byte[] keyBytes = secret.getBytes("UTF-8");
hmacSha256.init(new SecretKeySpec(keyBytes, 0, keyBytes.length, "
HmacSHA256"));
String sign = new String(Base64.encodeBase64(Sha256.doFinal(stringToSi
gn.getBytes("UTF-8")),"UTF-8"));
```

`secret` is an AppSecret.

Pass the signature

Place the calculated signature in the request header. The key is X-Ca-Signature.

Signature troubleshooting

If signature verification fails, API Gateway places StringToSign of the server in the HTTP response header and sends the response to the client. The key is X-Ca-Error-Message. Compare StringToSign that is calculated by the client with the one returned by the server.

If the StringToSign values from the client and server are the same, check the AppSecret used for signature calculation.

The HTTP Header does not support linefeeds, so the linefeeds in StringToSign are filtered out. Ignore linefeeds during comparison.

Signature demo

For a detailed demo (Java) of signature calculation, refer to *https://github.com/aliyun/api-gateway-demo-sign-java*.

26.4.6 API call example

You can use a self-compiled HTTP or HTTPS request to call the API.

Part 1: Request

Request address

```
http://e710888d3ccb4638a723ff8d03837095-cn-qingdao.aliapi.com/demo/
post
```

Request method

```
POST
```

Request body

```
FormParam1=FormParamValue1&FormParam2=FormParamValue2
//HTTP Request Body
```

Request header

```
Host: e710888d3ccb4638a723ff8d03837095-cn-qingdao.aliapi.com
Date: Mon, 22 Aug 2016 11:21:04 GMT
User-Agent: Apache-HttpClient/4.1.2 (java 1.6)
Content-Type: application/x-www-form-urlencoded; charset=UTF-8
//The request body type. Set the type based on the actual request
content.
Accept: application/json
//The response body type. Some APIs return data based on the specified
 response body type. We recommend that you set this Header parameter
manually. If you do not set this parameter, some HTTP clients set it
to */* by default. This can cause a signature error.
X-Ca-Request-Mode: debug
//Whether to enable the Debug mode. This parameter is case-insensitive
. The Debug mode is disabled by default. It is usually enabled during
API debugging.
X-Ca-Version: 1
//The API version number. Only version 1 is supported. This parameter
is optional. Default value: 1.
X-Ca-Signature-Headers: X-Ca-Request-Mode,X-Ca-Version,X-Ca-Stage,X-Ca
-Key,X-Ca-Timestamp
//Custom Header parameters that are used for signature calculation
. The server reads Header parameters based on this setting during
signature calculation. Content-Type, Accept, Content-MD5, and Date
are part of the basic signature structure, and are not included in
custom Header parameters. For more information, see request signature
instructions.
X-Ca-Stage: RELEASE
//The stage of the requested API. Values: TEST, PRE, and RELEASE. This
 parameter is case-insensitive. An API provider can decide to which
stage the API is to be published. When you call an API that has not
been published, an error message is returned, indicating an invalid
URL or that the API is not found.
X-Ca-Key: 60022326
//The request AppKey. AppKeys are generated in the API Gateway console
. An application can call an API only when the application has been
authorized to use the API.
```

```
X-Ca-Timestamp: 1471864864235
//The request timestamp. It is the current time, in milliseconds. A
timestamp is valid for 15 minutes.
X-Ca-Nonce: b931bc77-645a-4299-b24b-f3669be577ac
//The unique identifier of the request. The combination of AppKey,
API, and Nonce must be unique for requests within 15 minutes. This
parameter must be used together with the timestamp to prevent replay.
X-Ca-Signature: FJleSrCYPGCU7dMlLTG+UD3Bc5Elh3TV3CWHtSKh1Ys=
//The request signature.
CustomHeader: CustomHeaderValue
//Custom Header parameters. The preceding example serves only as a
reference. You can set several custom Header parameters based on the
API definition.
```

Part 2: Response

Status code

```
400
//The response status code. If the value is greater than or equal to
 200 and smaller than 300, the request is successful. If the value
is greater than or equal to 400 and smaller than 500, a client error
 has occurred. If the value is greater than 500, a server error has
occurred.
```

Response header

```
X-Ca-Request-Id: 7AD052CB-EE8B-4DFD-BBAF-EFB340E0A5AF
//The unique ID of the request. After receiving a request, API Gateway
 generates a request ID and returns the request ID to the client
through the response header. We recommend that the request ID be
recorded by both the client and the backend service for troublesho
oting and tracking.
X-Ca-Error-Message: Invalid Url
//The error message returned by API Gateway. When a request fails, API
 Gateway returns the error message to the client through the response
header.
X-Ca-Debug-Info: {"ServiceLatency":0,"TotalLatency":2}
//The Debug message that is returned when the Debug mode is enabled
. The message may be changed in the future and is used only for
reference during debugging.
```

To call an API, you must attach a signature to the HTTP or HTTPS request. For more information about the signature encryption algorithm, see *Request signature instructions*.

26.5 APIs

26.5.1 Usage limits

Item	Limit
Number of API groups that can be created by a user	50.

Item	Limit
Number of APIs that can be created by a user	10,000. (Each user can create up to 50 API groups, and each group can contain up to 200 APIs.)
Number of independent domain names that can be bound to an API group	5.
Transactions per second (TPS) that can be handled by an API group	500. You can increase this value based on your business requirements.

26.5.2 Manage groups

26.5.2.1 Create a group

You can create API groups in the API Gateway console. You can create up to 50 groups.

Procedure

1. Log on to the API Gateway console.
2. Click the **Groups** tab.
3. On the Groups tab page that appears, click **Create Group**.
4. In the Create Group dialog box that appears, set Department, Project, and other required parameters. Click **Create**.

   Group names must be globally unique. A group name must be 4 to 50 characters in length. It can contain English letters, numbers, and underscores (_) and must start with an English letter.

26.5.2.2 Environment management

To understand environment management, you need to be familiar with two concepts: environment and environment variables.

- An **environment** is a configuration of an API group. You can configure several environments for a group. APIs that are not published are considered as API definitions. After you publish the API to an environment, it starts to provide external services.
- **Environment variables** are environment-specific variables that can be created and managed. For example, you can create an environment variable named `Path` and valued `/stage/release` for the release environment.

In the API definition, you can set `Path` to `#Path#` (which is a variable), and set Parameter Name to `Path`.

When you publish the API to the release environment, the value of `#Path#` in Path is `/stage/release`.

When you publish the API to another environment that does not have the environment variable `#Path#`, the variable in the API fails to obtain the value and the API cannot be called.

You can use environment variables to configure different service addresses for different environments in an API definition. API Gateway calls different backend services based on the environment variable values. Pay attention to the following points:

- Variable names are case-sensitive.
- If you configure a variable in the API definition, you must configure the name and value of the variable for the environment to which the API is published. Otherwise, the variable will not take a value, and the API will not be called.

Create an environment variable

1. Log on to the API Gateway console.
2. Click the **Groups** tab.
3. Locate a group, click the management icon in the Actions column corresponding to the group, and choose View Details from the shortcut menu.
4. On the page that appears, click the **Environment Variables** tab. On the Environment Variables tab page that appears, click Create Variable.
5. In the Create Environment Variable dialog box that appears, set **Variable Name** and **Variable Value**. Click **OK**.

Delete a variable

1. Log on to the API Gateway console.
2. Click the **Groups** tab.
3. Locate a group, click the management icon in the Actions column corresponding to the group, and choose View Details from the shortcut menu.
4. On the page that appears, click the **Environment Variables** tab.
5. On the Environment Variables tab page that appears, select an environment. Locate the variable that you want to delete, click the management icon in the Actions column corresponding to the variable, and choose **Delete** from the shortcut menu.
6. In the message that appears, click **OK**.

26.5.2.3 Delete a group

You can delete an existing group.

Procedure

1. Log on to the API Gateway console.
2. Click the **Groups** tab.
3. Locate the group that you want to delete, click the management icon in the Actions column corresponding to the group, and choose **Delete** from the shortcut menu.
4. In the message that appears, click **OK**.

26.5.3 Create an API

26.5.3.1 Overview

Creating an API is the process of defining the API in the API Gateway console. When creating an API, you need to define the basic information, backend service information, request information, and response information of the API.

- You can configure parameter verification rules. API Gateway pre-verifies API requests based on the verification rules and forwards the requests that contain only valid parameters to the backend service.
- You can configure API Gateway to map a frontend parameter to a backend parameter at any location. For example, you can configure API Gateway to map a **Query** parameter in an API request to a **Header** parameter in a backend service request. In this way, you can package your backend services into standard APIs.
- API Gateway allows you to configure constant parameters and system parameters. These parameters are invisible to users, but API Gateway can attach them to the received requests based on your service requirements before sending the requests to backend services. If you want API Gateway to attach the keyword **apigateway** to each request it forwards to your backend service, you can configure **aligateway** as a constant parameter and specify where it is received.

26.5.3.2 Create an API

Procedure

1. Log on to the API Gateway console.
2. Click the **API** tab.
3. Click **Create API**.

4. Set the basic information of the API and click **Next**.

Parameter	Description
Groups	The group to which the API belongs. APIs are managed by groups. Before you create an API, you must create a group. Select a group from the **Groups** drop-down list.
API Name	The API name.
Authentication Mode	The authentication mode of API requests. Values: **Alibaba Cloud Applications** and **None**. • **Alibaba Cloud Applications**: This authentication mode requires the requester to pass the application authentication to call this API. • **None**: This authentication mode allows any user who knows the request definition of the API to initiate a request. API Gateway directly forwards the requests to your backend service without verifying the identity of the requesters.
Description	The description of the API.

5. Define API requests. Define the settings of the requests that users can send to call the API, including the related protocols, request paths, HTTP methods, and parameters.

Parameter	Description
Network Protocol	The protocol that can be used to call the API. Both HTTP and HTTPS are supported.
Custom Domain Name	The independent domain name that has been bound to the group to which the API belongs.
Second-Level Domain Name	The second-level domain name of the group to which the API belongs.
URL Suffix	The API request path. It corresponds to the service host. The request path can be different from the backend service path. You can provide any valid and semantically -correct path for users. You can configure dynamic parameters in the request path. API Gateway can map these parameters to Query

Parameter	Description
	, Header, or other locations before sending the requests to the backend service.
HTTP Method	The HTTP method supported by the API. Values: PUT, GET, POST, PATCH, DELETE , and HEAD.
Request Parameters	The request parameters of the API. These parameters need to be set by the users. You can define the request parameters in Header, Query, Body, or Path (Parameter Path). If you have defined dynamic parameters in Path, specify how to set these dynamic parameters when your define request parameters. The following parameter types are supported: String, Number, and Boolean.
Parameter verification rules	Click the management icon in the Actions column corresponding to a request parameter, and choose **Configure Advanced Settings** from the shortcut menu. In the Configure Advanced Settings dialog box that appears, you can configure parameter verification rules, such as Maximum Length and Enumeration. API Gateway pre-verifies requests based on the verification rules. The requests with invalid parameters are not sent to your backend service. This greatly reduces the work load of the backend service.

6. Configure the backend service and click **Next**.

   This section defines mappings between frontend and backend parameters, and specifies the API backend service configurations. The backend service configurations include the backend service address, backend service path, backend response timeout period, parameter mappings, constant parameters, and system parameters. After receiving user requests, API Gateway converts the format of the requests to the format required by the backend service based on the backend service configuration. Then, API Gateway forwards the requests to the backend service.

    **Note:**

You can enter the following parameters: dynamic parameters in Path, Header parameters, Query parameters, Body parameters (non-binary), constant parameters, and system parameters. Each parameter name must be globally unique. For example, you are not allowed to enter a parameter named **name** in both Header and Query.

a) Configure basic backend service information.

Parameter	Description
Backend Service URL	The backend service host. It can be a domain name or an address in the format of `http(s)://host:port`.
URL Suffix	The actual request path of your API service on the backend server. If you have configured dynamic parameters in the backend path, you must specify the corresponding request parameters and their locations by declaring the mapping.
Timeout	The maximum amount of time that API Gateway waits for a response from the backend service. API Gateway sends a request to the backend service and waits for a response. The maximum timeout period is 30 seconds. If API Gateway does not receive a response from the backend service within the timeout period, it stops waiting and returns an error.
Mock	You can mock expected responses to return to API callers during the project development process. For more information , see Mock an API.

b) Configure backend service parameters.

API Gateway can set up mappings between frontend and backend parameters, including names and locations. API Gateway can map a request parameter at any location (Path, Header, Query, or Body) to a backend service parameter at a different location. In this way, you can package your backend services into standard APIs. This part declares the frontend-to-backend API mapping.

Note:
The frontend and backend parameters must be globally unique.

c) Configure constant parameters.

To configure API Gateway to add the **`apigateway`** tag to each request it forwards to your backend service, you can configure the tag as a constant parameter. Constant parameters are invisible to users. After a constant parameter is configured, API Gateway automatically adds this parameter to the specified location of the received requests before sending the requests to your backend service.

d) Configure system parameters.

API Gateway does not send its system parameters to you by default. To obtain these parameters, configure their locations and names in the API. The following table lists the system parameters.

Parameter	Description
CaClientIp	The IP address of the client that sends the request
CaDomain	The domain name that is used to send the request
CaRequestHandleTime	The request time (UTC)
CaAppId	The ID of the application that sends the request
CaRequestId	The request ID
CaApiName	The API name
CaHttpSchema	The protocol that is used to call the API, which is HTTP or HTTPS
CaProxy	The proxy (AliCloudApiGateway)

7. Define the response and click **Create**.

You can set Response Content Type, Success Response Example, and Error Response Example, and define error codes. API Gateway does not parse responses, but forwards them to the API requester.

26.5.3.3 Support HTTPS

Hyper Text Transfer Protocol Secure (HTTPS) is based on the Hyper Text Transfer Protocol (HTTP) and Secure Sockets Layer (SSL) protocols. HTTPS is used to encrypt information and data to secure data transmission. HTTPS is widely used today.

API Gateway supports HTTPS-based encryption of API requests. You can configure APIs to support HTTP, HTTPS, or both.

Perform the following steps to configure your APIs to support HTTPS.

Step 1: Make preparations.

Prepare the following items:

- An independent domain name
- An SSL certificate that has been applied for this domain name

  The SSL certificate contains two parts: XXXXX.key and XXXXX.pem, both of which can be opened in text editors.

  Example:

  KEY

```
-----BEGIN RSA PRIVATE KEY-----
MIIEpAIBAAKCAQEA8GjIleJ7rlo86mtbwcDnUfqzTQAm4b3zZEo1aKsfAuwcvCud
....
-----END RSA PRIVATE KEY-----
```

  PEM

```
-----BEGIN CERTIFICATE-----
MIIFtDCCBJygAwIBAgIQRgWF1j00cozRl1pZ+ultKTANBgkqhkiG9w0BAQsFADBP
...
-----END CERTIFICATE-----
```

Step 2: Bind the SSL Certificate to an API group.

After you prepare the preceding items, log on to the API Gateway console and click the **Groups** tab. Locate the API group to which you want to bind the SSL certificate and view the group details.

Bind the independent domain name to the API group before you bind the SSL certificate to the group.

- **Certificate Name**: indicates the name of the certificate.
- **Certificate Content**: indicates the content of the entire certificate. Copy the content in the *XXXXX.pem* file.

- **Private Key**: indicates the private key of the certificate. Copy the content in the *XXXXX.key* file. Click **OK** to bind the SSL certificate to the API group.

Step 3: Adjust the API configuration.

After binding the SSL certificate to the API group, you can configure APIs in the group to support access over HTTP, access over HTTPS, or both. For security considerations, access over HTTPS is recommended.

Locate an API whose configurations you want to adjust on the **API** tab page. Click the management icon in the Actions column and choose **Change** from the shortcut menu. In the Request Basic Settings area on the Define API Request tab page, set Network Protocol.

APIs support the following protocols:

- HTTP: The API supports only access over HTTP.
- HTTPS: The API supports only access over HTTPS.
- HTTP and HTTPS: The API supports both access over both HTTP and HTTPS. If you select HTTPS for Network Protocol, the API supports access over HTTPS.

26.5.3.4 Mock an API

You can mock expected responses to return to API callers during the project development process. This can greatly reduce miscommunication and misunderstanding among team members and significantly improve the development efficiency.

API Gateway supports simple configuration in mock mode.

Configure a mock

On the **Change API** > **Define Backend Service** tab page of an API, configure a mock.

1. Select the Mock type.

   You can set Backend Service Type to Mock and confirm your setting as prompted.

2. Configure the mock response.

   You can enter an actual response in the Mock Response field. Currently, mock responses in the JSON, XML, and text formats are supported. For example:

```
{
"result": {
    "title": " Mock test for API Gateway",
}
```

```
}
```

Save the mock configurations. **Publish** the API to the test or release environment for testing as needed.

26.5.4 API management

26.5.4.1 View and modify an API

You can view APIs and modify them as needed.

> **Note:**
> If you modify an API that has been published to the release environment, you must republish the API for the modifications to take effect in the release environment.

Procedure

1. Log on to the API Gateway console.
2. Click the **API** tab.
3. Locate the API that you want to view.

   View information of the API.
4. Click the management icon in the Actions column corresponding to an API and choose **Change** from the shortcut menu.
5. Set parameters as required and click **Change**.

   The procedure to modify an API is similar to that to create an API. For more information about how to create an API, see *Create an API*.

   If you do not want to continue modifying the API, click **Cancel** in the lower-right corner.

26.5.4.2 Publish an API

After creating an API, you need to debug, test, and publish it.

- When you use a second-level or independent domain name to access an API published to an environment, you must specify the environment to be requested in the request header.
- If you publish an API to the test or release environment where the API already has a version running, the newly published version replaces the running version to take effect in real time. However, all historical versions and definitions are recorded so you can roll the API back to an earlier version.
- You can unpublish an API in the test or release environment. After the API is unpublished , its binding or authorization relationships with the policy, keys, or applications persist and

will automatically take effect when the API is published again. Remove these relationships separately if you no longer need them.

Step 1: Test the API

You can create an application and authorize the application to use this API. Then, use this application to simulate real user requests.

You can write code based on real request scenarios or use the SDK samples provided by API Gateway to call the API.

You can publish an API to the test or release environment. If no independent domain name is bound to the group to which the API belongs, you can test or call the API by using the second-level domain name. Specify an environment in your request. If no environment is specified in your request, the API published to the release environment is called by default.

Step 2: Publish the API

After testing the API, you can publish it.

You can use API Gateway to manage versions of APIs in the test or release environment. You can publish or unpublish an API and change its version. The version change takes effect in real time.

1. Log on to the API Gateway console.
2. Click the **API** tab.
3. Locate the API that you want to publish, click the management icon in the Actions column, and choose **Publish** from the shortcut menu.
4. In the dialog box that appears, select an environment, enter a description, and click **OK**.

26.5.4.3 Authorize applications to call APIs

You need to authorize an application before the application can call your API. After you publish an API to the release environment, you must authorize the applications of users before they can use the API. You can perform an authorization or deauthorization operation to establish or remove an authorization relationship between an API and an application. API Gateway verifies the authorization relationship.

Note:

- You can authorize one or more applications to use one or more APIs.

- If an API has been published to both the test and release environments but only the test environment is selected, applications are authorized to use only the API in the test environment.
- You can locate an application based on its ID or name provided by the user.
- To revoke the authorization from an application under an API, go to the Authorization tab page of the API. Select the application from the application list. Then, click the management icon in the Actions column and choose Deauthorize from the shortcut menu, or click Deauthorize in the upper-right corner.

Applications indicate requesters' identities. When you or your customers test or call an API, you must create an application as the requester's identity and then authorize the application to use the API.

Note:
Authorizations are environment-specific. The same application must be authorized to use the same API in both the test and release environments to avoid errors caused by inconsistency between the authorized environment and requested environment.

Procedure

1. Log on to the API Gateway console.
2. Click the **API** tab.
3. Locate the API that you want to authorize applications to use, click the management icon in the Actions column, and choose **Authorize** from the shortcut menu.
4. In the dialog box that appears, set Environment.
5. Select applications from the left-side list, click **right arrow** to add the selected applications to the right-side list, and click **OK**.

   You can enter an application ID or name to search for applications.

What's next

You can view the authorization information of APIs or revoke the authorization from an application under an API.

In the API list, click the management icon in the **Actions** column corresponding to an API, and choose View Details from the shortcut menu. Click the **Authorization** tab. On the Authorization tab page, you can view the applications that have been authorized to use this API.

You can select one or more application IDs and click **Deauthorize** in the upper-right corner to revoke the authorizations from the selected applications under the API.

26.5.4.4 Delete an API

You can delete existing APIs.

> **Note:**
> Before deleting a published API, you must unpublish it.

Procedure

1. Log on to the API Gateway console.
2. Click the **API** tab.
3. Locate the API that you want to delete, click the management icon in the Actions column, and choose **Delete** from the shortcut menu.
4. In the message that appears, click **OK**.

26.5.4.5 Unpublish an API

You can unpublish a published API.

You can unpublish APIs in the test or release environments. After an API is unpublished, its binding or authorization relationships with policies, keys, or applications still persist. These relationships will take effect if the API is published again. Remove these relationships separately if you no longer need them.

Procedure

1. Log on to the API Gateway console.
2. Click the **API** tab.
3. Locate the API that you want to unpublish, click the management icon in the Actions column, and choose **Take Offline** from the shortcut menu.
4. In the message that appears, click **OK**.

26.5.4.6 View the version history of an API

You can view the version history of an API, including the version number, description, environment, release time, and specific definition of each version.

Procedure

1. Log on to the API Gateway console.

2. Click the **API** tab.
3. Locate the API that you want to view.
4. Click the management icon in the Actions column and choose View Details from the shortcut menu to go to the API details page. Click the **Version History** tab.
5. On the Version History tab page, you can click the management icon in the Actions column corresponding to a version and choose **View** from the shortcut menu to view the API details.

26.5.4.7 Change the version of an API

When viewing the version history of an API, you can select a different version and switch the API to that version. The new version directly replaces the previous one to take effect in real time in the specified environment.

Procedure

1. Log on to the API Gateway console.
2. Click the **API** tab.
3. Locate the API of which you want to change the version.
4. Click the management icon in the Actions column and choose View Details from the shortcut menu to go to the API details page. Click the **Version History** tab.
5. Locate the target version, click the management icon in the Actions column, and choose **Switch to this Version** from the shortcut menu.
6. In the dialog box that appears, enter a description and click **OK**.

26.5.5 Throttling policies

26.5.5.1 Create a throttling policy

You can create throttling policies. Throttling policies are a kind of plug-ins.

Procedure

1. Log on to the API Gateway console.
2. Click the **Plug-ins** tab.
3. Click **Create Plug-in** in the upper-right corner.
4. Set parameters and click **OK**.

26.5.5.2 Bind a throttling policy to APIs

After creating a throttling policy, you need to bind it to an API for the policy to take effect on the bound API.

Context

You can bind a throttling policy to multiple APIs. The limits defined in the throttling policy will be applicable to each bound API. When you bind a throttling policy to an API that is bound with another throttling policy, the new policy takes effect immediately in place of the old policy.

Procedure

1. Log on to the API Gateway console.
2. Click the **Plug-ins** tab.
3. Locate the plug-in that you want to bind to APIs, click the management icon in the Actions column, and choose **Associate API** from the shortcut menu.
4. In the dialog box that appears, set Environment and Groups.
5. Select APIs from the left-side list, click **right arrow** to add the selected APIs to the right-side list, and click**OK**.

26.5.5.3 Delete a throttling policy

You can delete existing throttling policies.

Procedure

1. Log on to the API Gateway console.
2. Click the **Plug-ins** tab.
3. Locate the plug-in that you want to delete, click the management icon in the Actions column, and choose **Delete Plug-in** from the shortcut menu.
4. In the message that appears, click **OK**.

27 Apsara Stack Security

27.1 What is Apsara Stack Security

Apsara Stack Security is a solution that provides Apsara Stack with a full suite of security features, such as network security, server security, application security, data security, and security management.

In today's cloud computing environment, new technologies are developed every day. Border security protection methods that use traditional detection technologies are insufficient to secure cloud services. Apsara Stack Security combines the powerful data analytics capabilities of Alibaba Cloud with the professional expertise of the security operations team. It provides integrated security protection services at the network, application, and server levels.

The Apsara Stack Security Standard Edition provides the following features:

- **Threat Detection Service**: Detects and analyzes network security trends, performs associated tracing and big data analytics on security events, and displays the risks of detected security events.
- **Network security**: Helps the security administrator fully manage the internal and external network security status of the Apsara Stack platform.
- **Application security**: Provides WAF to help the security administrator protect applications on the Apsara Stack platform.
- **Server security**: Provides functions, such as security protection and intrusion detection, to help the security administrator protect servers.
- **Physical machine security**: Provides functions such as file tampering detection, suspicious process detection, suspicious network connection detection, and suspicious port listening detection.
- **Asset management**: Helps the security administrator manage assets in Apsara Stack, including servers and network address translation (NAT) IP addresses.
- **Security audit**: Displays and audits cloud service operation logs, so that the security auditor can promptly discover and eliminate security risks.
- **System management**: provides functions such as account management, rule database synchronization, alert settings, and global settings.

27.2 Restrictions

Before logging on to Apsara Stack Security Center, make sure that your local PC meets the requirements

described in *Table 27-1: Configuration requirements*.

Table 27-1: Configuration requirements

Item	Requirements
Browser	• Internet Explorer: 11 or later • Google Chrome (recommended): 42.0.0 or later • Mozilla Firefox: 30 or later • Safari: 9.0.2 or later
Operating system	• Windows XP, Windows 7, or later • Mac

27.3 Quick start

27.3.1 User permissions

This topic describes the user roles involved in Apsara Stack Security.

All roles in Apsara Stack Security Center are default roles. You cannot add custom roles. Before logging on to Apsara Stack Security Center, make sure that your account has been assigned the corresponding role. For more information about roles in Apsara Stack Security, see *Table 27-2: Default roles in Apsara Stack Security*.

Table 27-2: Default roles in Apsara Stack Security

Role	Description
System administrator of Apsara Stack Security Center	Manages and configures the system settings for Apsara Stack Security Center. The system administrator has the following permissions: Alibaba Cloud account management, rule database synchronization, alert settings, and global settings.
Security administrator of Apsara Stack Security Center	Monitors the security status of the entire Apsara Stack platform and configures security policies for each functional module of Apsara Stack Security. The security administrator has permissions to all functional nodes under Threat Detection, Network Security, Application Security, Server Security, PM Security, and Asset Management.

Role	Description
	Note: The permissions to WAF and Cloud Firewall must be assigned independently.
Department security administrator	Monitors the security status of cloud product resources in the specified department and configures security policies for each functional module of Apsara Stack Security for this department. The department security administrator has permissions to all functional nodes under Threat Detection, Network Security, Application Security, Server Security, PM Security, and Asset Management. In addition, the department security administrator can specify the alert notification method and the alert recipients in this department. **Note:** The permissions to WAF and Cloud Firewall must be assigned independently.
Auditor of Apsara Stack Security Center	Conducts security audits on the entire Apsara Stack platform. The auditor can view audit events and original logs, configure audit policies, and access all functional nodes under Security Audit.

If you do not have an account and a role, contact the administrator to create an account and assign a role to it. For more information, see **Create a user** in *User Guide*.

27.3.2 Log on to Apsara Stack Security Center

This topic describes how to log on to Apsara Stack Security Center.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super . The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.
5. In the top menu bar in the Apsara Stack console, click **Console**.
6. In **Cloud Security Center**, click any service, for example, **Threat Detection**.
7. Specify **Region**, and click **Cloud Security Console**. The **Apsara Stack Security Center** page appears.

What's next

To use an Apsara Stack Security user account to access cloud services such as Cloud Firewall, you must first grant the account permissions in RAM. For more information, see **Grant Permissions in RAM**.

27.3.3 Grant permissions in RAM

This topic describes how to grant permissions to an Apsara Stack Security user in RAM.

Context

To use an Apsara Stack Security user account to access the console of cloud services such as Cloud Firewall, you must first grant the account permissions in RAM.

The Apsara Stack Security services that require authorization in RAM include `yundun-audit:*`, `yundun-sas:*`, `yundun-scs:*`, `yundun-ddos:*`, `yundun-system:*`, `yundun-aegis:*`, `yundun-waf:*`, `yundun-cloudfirewall:*`, `yundun-bastionhost:*`, and `yundun-dbaudit:*`.

Procedure

1. Log on to the Apsara Stack console.

2. Modify the URL of the Apsara Stack console, and then use the modified URL to log on to the RAM console.

   For example, if the URL of the Apsara Stack console is`https://manage.abc.com/xxx`, the URL of the RAM console is `https://manage.abc.com/manage/view/index.html#/system/userCenter/policyMnt`.

   > **Note:**
   > Replace `abc.com` with the actual domain name.

3. Create an authorization policy for accessing the Apsara Stack Security services.

   a) On the **Policy Management** page, click **Custom Authorization Policy**.

   b) Select a department and a region, and click **Search**.

   c) Click **Create Authorization Policy** to create an authorization policy for accessing the Apsara Stack Security services.

   Specify **Policy Details** as follows:

```
{
  "Statement":
  [
    {
      "Action":
      [
        "yundun-cloudfirewall:*"
      ],
      "Effect":"Allow",
      "Resource": "*",
    },
    {
      "Action":
      [
        "yundun-waf:*"
      ],
      "Effect":"Allow",
      "Resource": "*",
    },
    {
      "Action":
      [
        "yundun-aegis:*"
      ],
      "Effect":"Allow",
      "Resource": "*",
    },
    {
      "Action":
      [
        "yundun-audit:*"
      ],
      "Effect":"Allow",
```

```
      "Resource": "*",
    },
    {
      "Action":
      [
        "yundun-sas:*"
      ],
      "Effect":"Allow",
      "Resource": "*",
    },
    {
      "Action":
      [
        "yundun-scs:*"
      ],
      "Effect":"Allow",
      "Resource": "*",
    },
    {
      "Action":
      [
        "yundun-ddos:*"
      ],
      "Effect":"Allow",
      "Resource": "*",
    },
    {
      "Action":
      [
        "yundun-system:*"
      ],
      "Effect":"Allow",
      "Resource": "*",
    },
    {
      "Action":
      [
        "yundun-bastionhost:*"
      ],
      "Effect":"Allow",
      "Resource": "*",
    },
    {
      "Action":
      [
        "yundun-dbaudit:*"
      ],
      "Effect":"Allow",
      "Resource": "*",
    }
  ],
  "Version": "1"
}
```

4. Apply the authorization policy to the Apsara Stack Security user account.

   If the authorization policy is created after the Apsara Stack Security user account has been created, you must manually apply this policy to the account.

Note:
If the authorization policy is created before the Apsara Stack Security user account is created, you only need to specify the department of the account as the authorized department to complete the authorization.

a) Go to the **RAM Users** page.

You can modify the URL of the Apsara Stack console, and then use the modified URL to access the RAM Users page. For example, if the URL of the Apsara Stack console is `https://manage.abc.com/xxx`, the URL of the **RAM Users** page is `https://manage.abc.com/manage/view/index.html#/system/userCenter/accessControl`.

Note:
Replace `abc.com` with the actual domain name.

b) Select a department and a region, enter a username, and click **Search**.

c) Click the More icon in the Actions column for the target RAM user, and select **Authorize**.

d) Select the authorization policy created for accessing the Apsara Stack Security services in **Existing Authorization Policies**, move this policy to **Selected Authorization Policies**, and click **OK**.

27.3.4 Switch regions

This topic describes how to switch regions managed by Apsara Stack Security.

Context

When you log on to Apsara Stack Security Center, you have already selected a region. To manage the server or network security of another region, follow these steps:

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Select the target region from the Region drop-down list in the upper-left corner.

The region of Apsara Stack Security Center will be switched to the selected one.

27.3.5 User interface of Apsara Stack Security Center

This topic describes the user interface of Apsara Stack Security Center.

The user interface of Apsara Stack Security Center is divided into three areas, as shown in *Figure 27-1: User interface of Apsara Stack Security Center*.

Figure 27-1: User interface of Apsara Stack Security Center

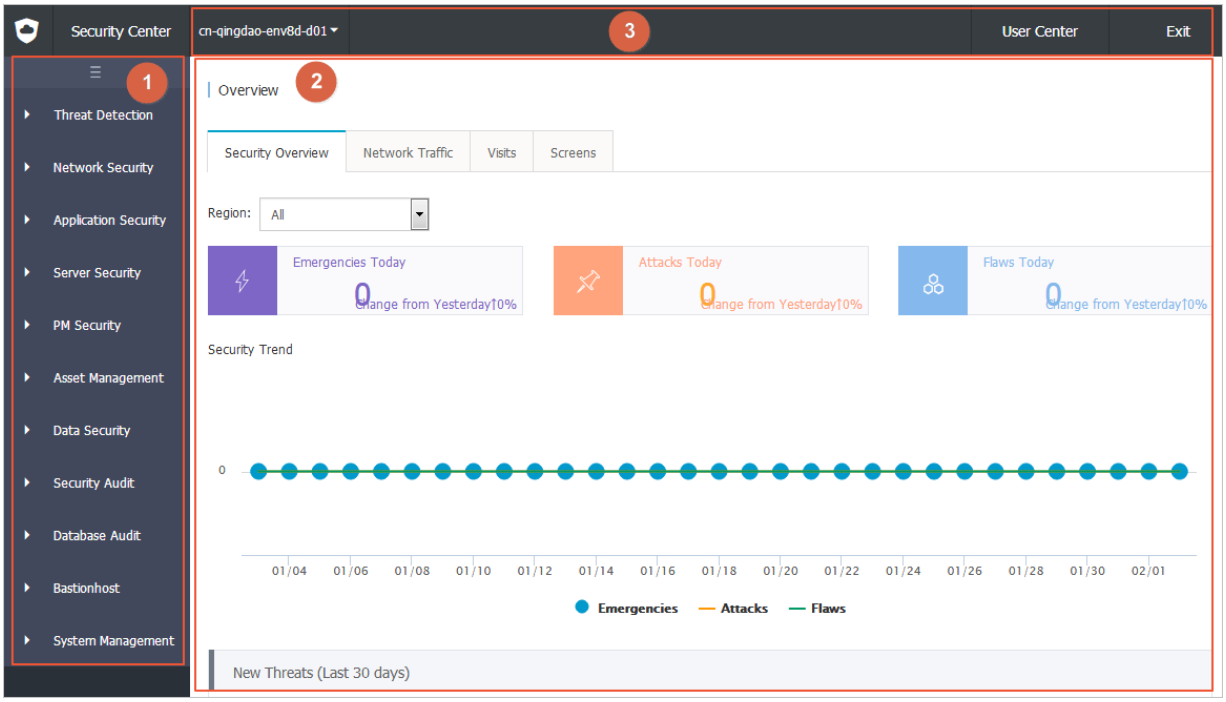


Table 27-3: Areas on the user interface of Apsara Stack Security Center

No.	Area	Description
1	Navigation pane	Apsara Stack Security Center Standard Edition provides the following features: • **Threat Detection Service**: Detects and analyzes network security trends, performs associated tracing and big data analytics on security events, and displays the risks of detected security events. • **Network Security**: Helps the security administrator fully manage the internal and external network security status of the Apsara Stack platform. • **Application Security**: Provides WAF to help the security administrator ensure the security of applications on the Apsara Stack platform.

No.	Area	Description
		• **Server Security**: Provides functions, such as threat protection and intrusion detection, to help the security administrator protect servers. • **Physical Machine Protection**: Provides file tampering detection, suspicious process detection, suspicious network connection detection, and suspicious port listening detection. • **Asset Management**: Helps the security administrator manage assets in Apsara Stack, including servers and network address translation (NAT) IP addresses. • **Security Audit**: Displays and audits cloud service operation logs, so that the security auditor can promptly discover and eliminate security risks. • **System Management**: Provides account management, rule database synchronization, alert settings, and global settings.
2	Operation area	After you select a menu item from the left-side navigation pane, its configuration page is displayed in the right-side operation area.
3	Operation bar	• Region drop-down list: Select a region protected by Apsara Stack Security. • **User Center**: Click this button to go to the personal information page. You can view the basic information about your account or change the logon password. • **Exit**: Click this button to log off.

27.4 Threat Detection Service

27.4.1 Threat Detection Service Overview

This topic describes the basic concepts of Threat Detection Service (TDS).

TDS incorporates a full range of capabilities to monitor enterprise vulnerabilities, hacker intrusions, Web attacks, DDoS attacks, threat intelligence, enterprise security reputation, and other security threats. Through modeling and analysis, TDS obtains key information from traffic features, server behavior, and server operation logs to identify intrusions that cannot be detected simply by inspecting traffic or scanning files. By combining the output from cloud-based analytics models with intelligence data, TDS identifies threat sources and attack behavior, and assesses the level of threat.

TDS provides the following functions:

- **Overview**: Displays the overall security situation, network traffic, access analysis, and information related to the security screens.
- **Event analysis**: Displays security events detected in the system and their development trends.
- **Threat analysis**: Displays security risks of the current system.
- **Security reports**: Allows you to configure Apsara Stack security report tasks.
- **Vulnerability scan**: Displays vulnerabilities and risks in the current system.

27.4.2 Overview

27.4.2.1 View security overview information

This topic describes how to view security trends, latest threats, and asset information on the Apsara Stack platform.

Context

The **Security Overview** page presents an overview of the detected security events, latest threats, and inherent vulnerabilities and flaws. The security administrator can view the information on the **Security Overview** page to have a comprehensive understanding of the system security situation.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose**Threat Detection** > **Overview**and click the **Security Overview** tab.

   View the current security situation on the Apsara Stack platform, as shown in *Figure 27-2: Security Overview tab page*.

Figure 27-2: Security Overview tab page

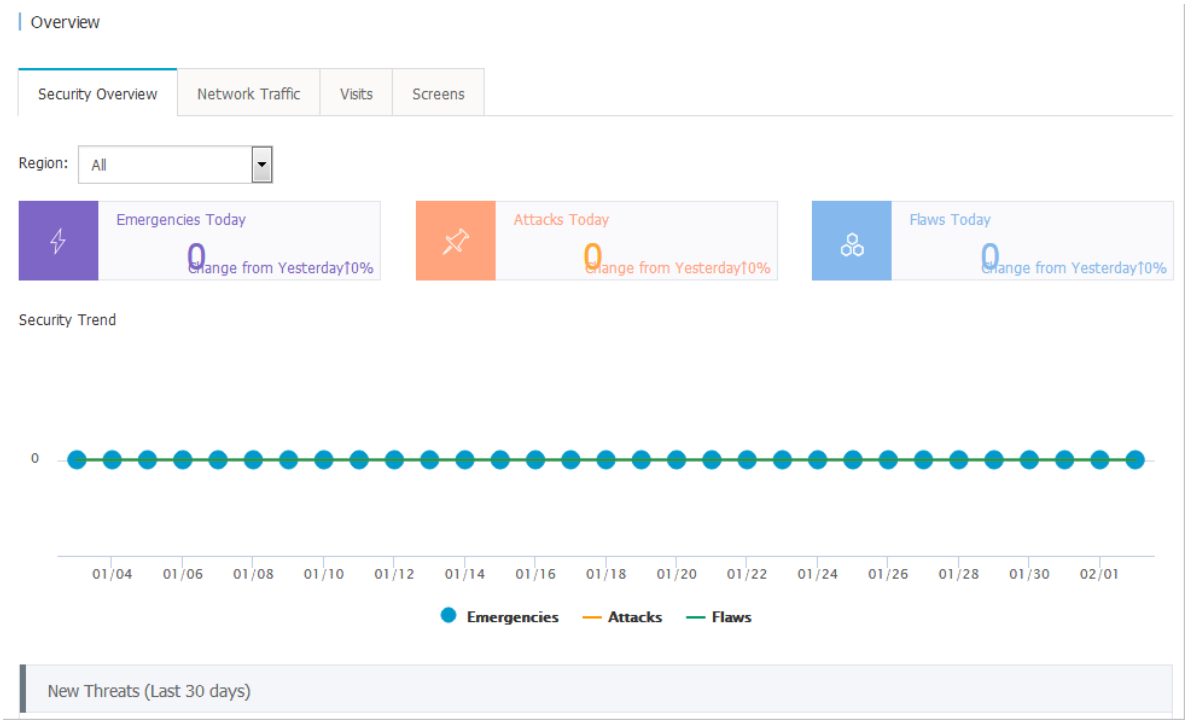


Table 27-4: Areas on the Security Overview tab page

Area	Description
Security Trend	Displays the detected security events and attacks, system vulnerabilities and flaws, and system security trends by time.
New Threats	Displays the existing security threats in the system. These threats require immediate attention. **Note:** These security threats are identified by the core scanner of Apsara Stack Security and analyzed by the big data analytics model of Apsara Stack.
Asset Overview	Displays the information about your most important assets so that you can learn the real-time asset status.

3. Click **Emergencies Today**, **Attacks Today**, or **Flaws Today** to view details on the corresponding page. You can also click View in each threat in the New Threats area to view the details.

   For example, you can click Emergencies to go to the **Event Analysis** page.

27.4.2.2 View network traffic information

This topic describes how to view network traffic information.

Context

Network traffic information over a period of time in the past is displayed in the form of a line chart. By checking the traffic at different time periods, in different regions, and from a specific IP address, you can identify the traffic peak and trough periods and view traffic distribution by rate or region. You can also check the top five IP addresses that generate the most traffic to effectively block access from malicious IP addresses.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Overview** and click the **Network Traffic** tab.
3. Optional: Set the **Region** parameter, enter the IP address of an ECS instance, and then click **Search**.

   You can query the traffic information by region and IP address.
4. Optional: Click **Today**, **Last 30 Days**, or **Last 90 Days**.

   The traffic information is displayed for different time periods.
5. View traffic information at a specific time point.
   - In the Outbound/Inbound Traffic diagram, hover over a time point on the traffic curve. You can view detailed information about the outbound or inbound traffic at the specified time point and the top five IP addresses that generate the most traffic.
   - In the QPS (Average) diagram, hover over a time point on the traffic curve. You can view the detailed QPS information at the specified time point.

27.4.2.3 View visit analysis results

This topic describes how to view visit analysis results.

Context

Based on big data analytics on visits from different sources: normal IP addresses, malicious IP addresses, and crawler IP addresses. Based on visits from malicious and crawler IP addresses, the security administrator can identify the causes of possible security issues in the system.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Overview** and click the **Visits** tab.

The Visits tab page appears.

Table 27-5: Visits tab page

Area	Description
All Visits Yesterday	Shows the top 10 most-visited domain names on the previous day and the number of IP addresses that visited each domain name.
Visitors Detected	Shows the number of normal IP addresses, the number of malicious IP addresses, and the number of crawler IP addresses on the previous day.
History Details	Shows the history details about visits from malicious and crawler IP addresses.

3. In the **History Details** area, click **All**, **Malicious IP**, or **Crawler IP**.

   The detailed visit information is displayed.

Table 27-6: Detailed visit information

Parameter	Description
Visitor IP	The IP address of a visitor.
Detection Method	Visits can be from malicious or crawler IP addresses.
Time	The time of the visit.
UserAgent	The User-Agent information contained in the HTTP request.
Target Application	The URL of the application that is visited.
Visited Pages	The number of pages that have been visited.
Maximum Visits per Second	The maximum number of visits per second.
Web Attack Detected	Indicates whether the visit involves Web attacks.

27.4.2.4 View information on visualization screens

This topic describes how to view the information on visualization screens.

Context

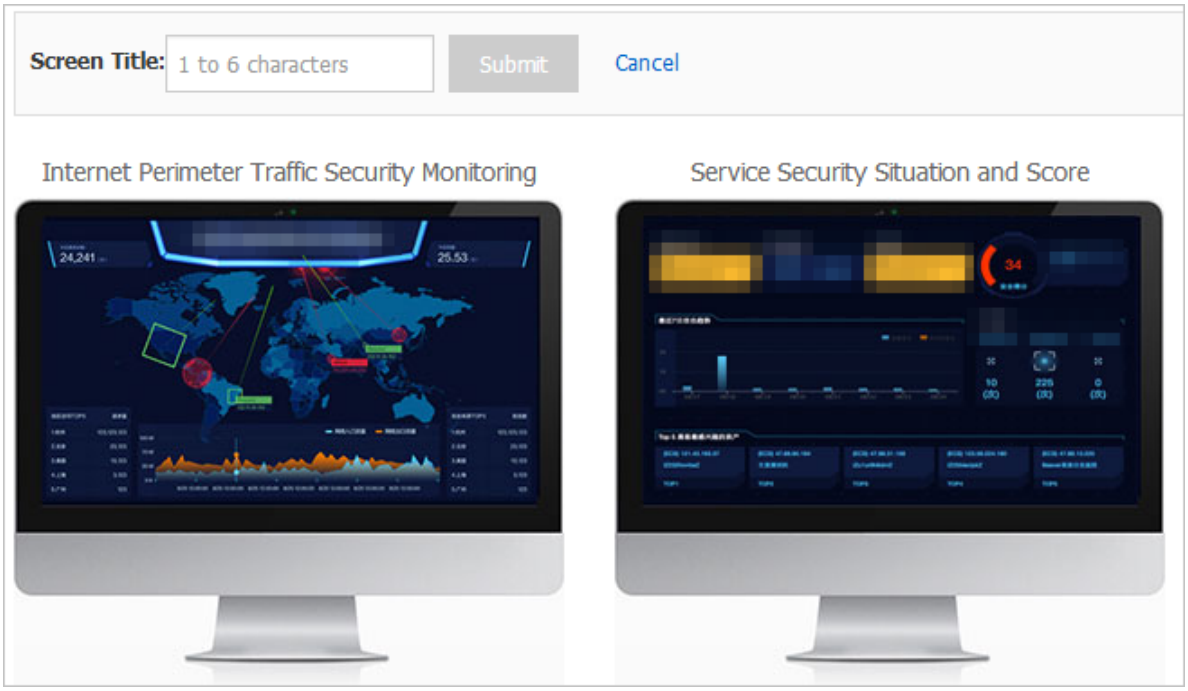
Visualization screens use animations to present key security event metrics. This provides the security administrator with a general picture of the security situation, effectively supporting security decisions.

The visualization screens include the screen for monitoring the security of the Internet perimeter traffic and the service security situation and scoring screen.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Overview** and click the **Screens** tab.

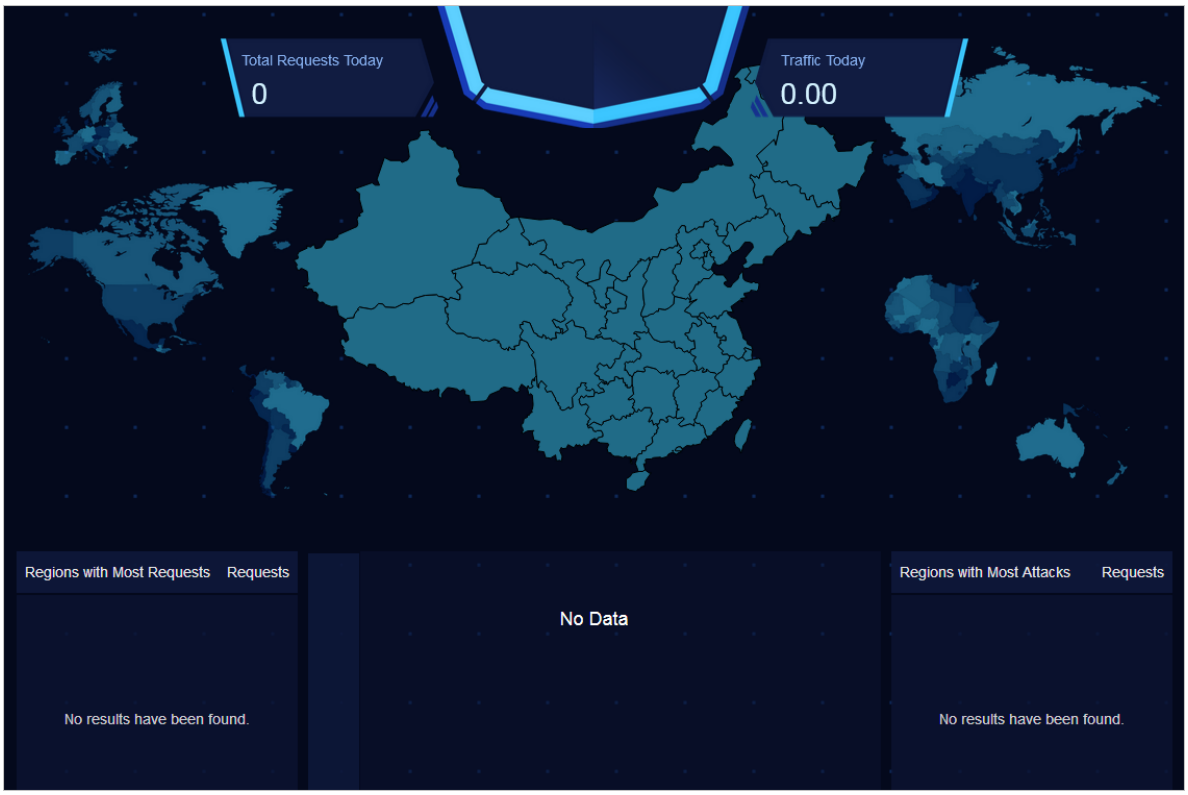
   Figure 27-3: Screens tab page



   Click **Modify** on the Screens tab page to modify the displayed title for **Internet Perimeter Traffic Security Monitoring**.

3. Click the screen of **Internet Perimeter Traffic Security Monitoring**.

   You can view the information displayed on the screen.



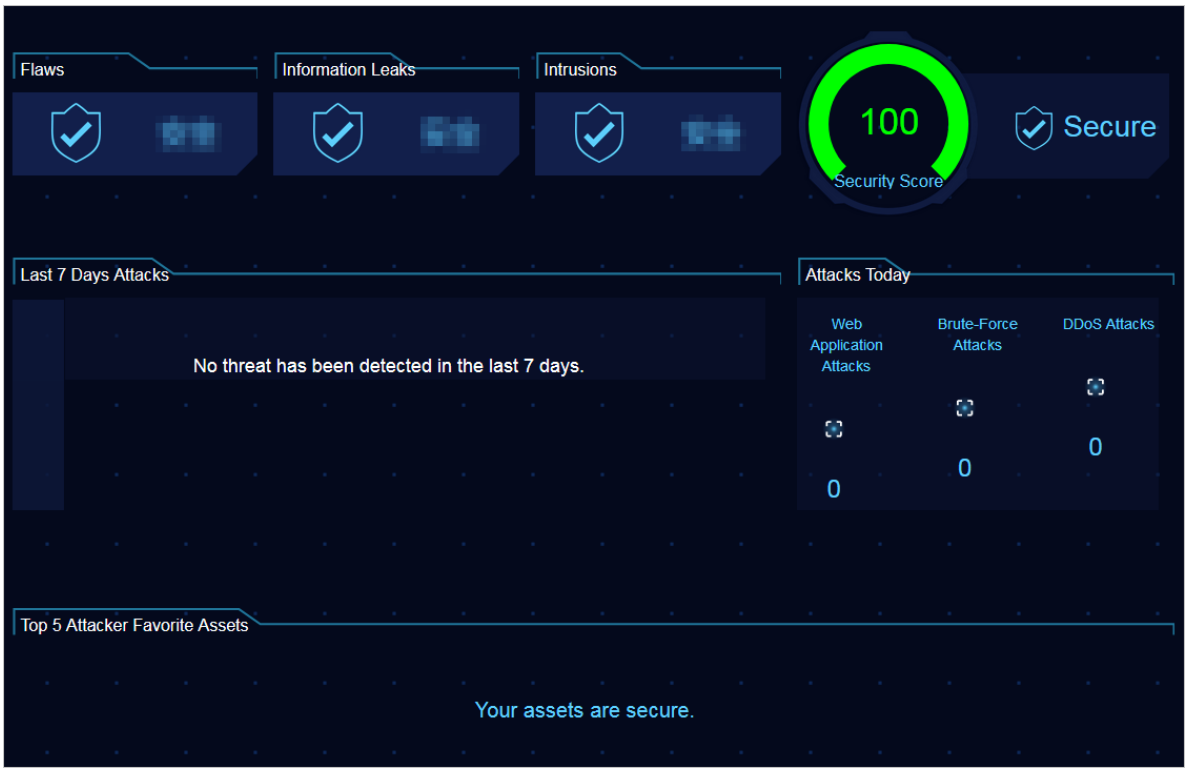
The **Internet Perimeter Traffic Security Monitoring** screen displays statistics on the source areas and number of current requests and attacks. It also displays a general picture of the system traffic. It lists the top five attack sources and top five attacked areas, providing the security administrator with an accurate understanding of the area distribution of requests and attacks.

Table 27-7: Access traffic sources

Type	Implementation mechanism
Request analysis	The assets that interest users are pushed to the traffic security monitoring module, which then reports access information for these assets.
Attack analysis	The traffic security monitoring module of Apsara Stack Security detects, reports, and displays events similar to Web attacks.
Traffic display	The traffic security monitoring module collects and reports traffic informatio n to the console for recording.

4. Click the **Service Security Situation and Score** screen.

   You can view the information displayed on the dashboard.



The **Service Security Situation and Score** screen displays detailed information about the current security events in the system. By analyzing the system flaws and the assets that have been attacked or that interest hackers, this screen evaluates the system's security situation and displays the security score.

The data shown on this screen is derived mainly from reports and scanned by modules such as the traffic security monitoring module, Server Guard, and flaw analysis of Apsara Stack Security. The top five assets that interest hackers are analyzed by the big data engine through modeling.

27.4.3 Event analysis

27.4.3.1 View emergencies

This topic describes how to view emergencies.

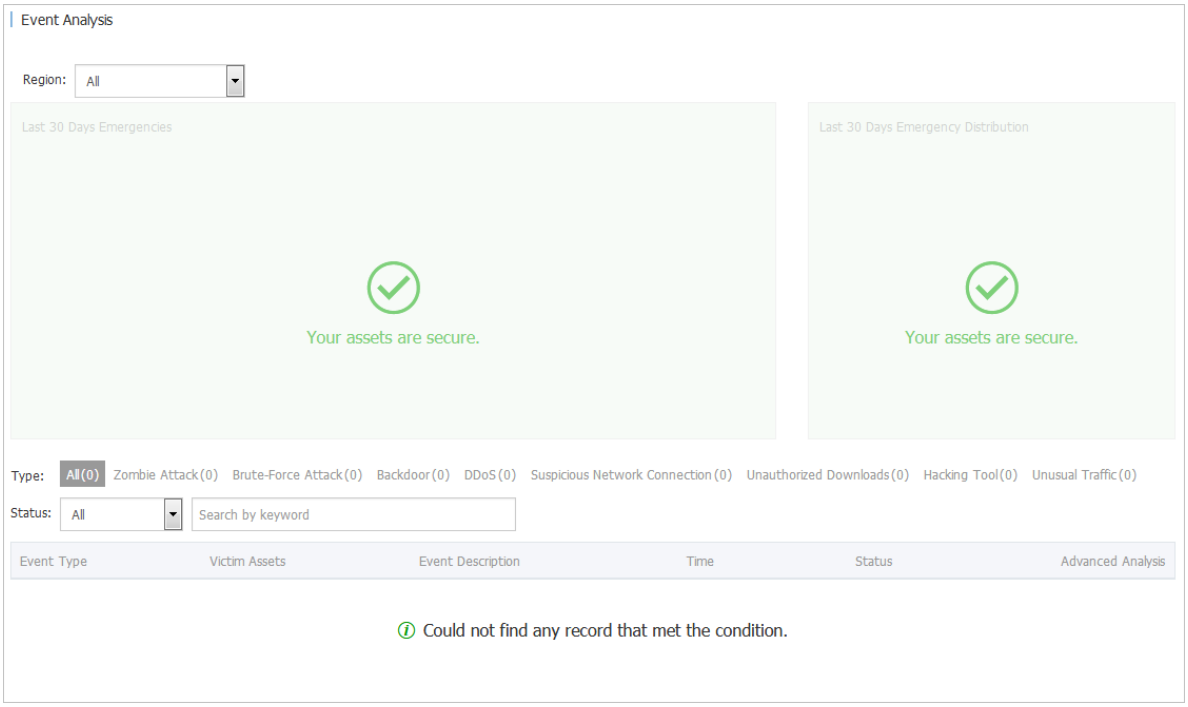
Context

Emergencies are security events that have occurred or are currently occurring in the system. Emergencies are detected and reported by the Apsara Stack Security modules, such as traffic security monitoring, Server Guard, and flaw analysis. When emergencies occur, the security administrator must pay immediate attention and take appropriate security measures.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Event Analysis**.

Figure 27-4: Event Analysis page



3. Select a specific emergency type for **Type**.

   Emergencies of the specified type are displayed in a list.

Table 27-8: Emergency event types

Emergency event type	Description
Zombie attack	A user server is controlled by hackers and used as a bot to launch external attacks.
Brute-force attack	The Server Guard agent reports both brute-force attack attempts and successful brute-force attacks. A successful force-force attack will be displayed in the emergency list. It must be immediately handled by the security administrator.
Backdoor	The Server Guard agent detects webshell files in the system. The big data analytics module analyzes traffic imported from the traffic security monitoring module to detect one-line trojans and complex trojans.
DDoS	DDoS attacks are detected by the traffic security monitoring module.

Emergency event type	Description
Suspicious network connection	Apsara Stack Security uses big data analytics models to analyze the information reported by security modules, and detects suspicious behavior such as suspicious external connections, malicious program downloads, and malicious file downloads.
Unauthorized download	Distinctive responses within the specified quantity range (between 1 and 20) are selected from the output traffic of the traffic security monitoring module. This allows the big data analytics module to detect unauthorized downloads.
Hacking tool	The residual hacker tools or hacker attack behaviors on the servers can be detected based on the information reported by Server Guard.
Unusual traffic	Attacks such as miner programs can be detected based on the information reported by the traffic security monitoring module and Server Guard.

4. View detailed event information in the list.

27.4.4 Threat analysis

27.4.4.1 View threat analysis results

This topic describes how to view threat analysis results.

Context

Apsara Stack Security analyzes traffic information by using the big data model to discover attack features, integrates attack information by attack type, and presents the current security threats in the system.

Threat analysis covers the following information:

- Last 7 Days Attacks/Last 30 Days Attacks: Displays normal attack and targeted attack trends for the last 7 days and attack analysis for the last 30 days.
- Attacker’s Top 5 Favorite Assets: Analyzes traffic information by using the big data model and grading each asset by threat. The five most risky assets are displayed for attention and protection by the security administrator.
- Targeted Attacks: Analyzes the traffic information provided by the traffic security monitoring module by using the big data model, to detect targeted attacks.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Threat Analysis** and click the **Threats** tab.

   Threat analysis results are displayed.

   Table 27-9: Threat analysis results

Item	Description
Last 7 Days Attacks	Displays attacks to servers and applications on the Apsara Stack platform in the last 7 days.
Last 30 Days Attacks	Displays attacks to servers and applications on the Apsara Stack platform in the last 30 days.
Attacker's Top 5 Favorite Assets	Displays the top five assets that interest hackers. These asset IP addresses are obtained by Apsara Stack Security Center by using the big data computing model based on detected attack and threat information. We recommend that the security administrator enhance the protection on these assets.
Targeted Attacks	Displays the targeted attacks of a specific type in Apsara Stack Security Center.

3. Select a targeted attack type for **Type**.

   The targeted attack events of the specified type are displayed in a list. *Table 27-10: Targeted attack types* describes targeted attack types.

   Table 27-10: Targeted attack types

Targeted attack type	Description
Targeted Web attack	When the system discovers a targeted Web attack, this means that hackers are more interested in a website than others. The hackers have performed SQL injection, command execution, directory scans, or some other malicious operations on this website.
Server-targeted password cracking	Server-targeted password cracking attacks aim at cracking users' logon passwords. Hackers generally launch untargeted cracking attacks on server passwords. A targeted attack generally implies that hackers are interested in specific servers.
User enumeration	The system can analyze unusual logon activities to detect logons resembling user enumeration attacks. Such attacks

Targeted attack type	Description
	indicate that hackers may be using username and password combinations leaked on the Internet in an attempt to forcibly log on to your website. This may harm user interests.
CMS unusual logon	The system can detect unusual logon events for the application administration console. If a logon attempt is not made by an authorized user, the hacker may have already stolen the console password. In this case, we recommend that the user check the password strength and change the password as soon as possible.
Scanner attack	The system can detect hackers' behavior of using a dedicated vulnerability scanner to scan servers on the Apsara Stack platform. After detecting a vulnerability on a server, hackers may launch a targeted attack to the server.
Pingback exploit	The system can detect targeted attacks that are launched by hackers by exploiting the vulnerability of pingback.
Logon with multiple accounts	The system can detect attackers that are using a large number of low-quality accounts to log on. Such accounts are most likely bot accounts.

4. In the **Targeted Attacks** list, click **View** next to a specific attack.

   The detailed information and solution for the targeted attack are displayed.

27.4.4.2 View threat attack information

This topic describes how to view attack information.

Context

Attacks include application attacks and brute-force attacks:

- **Application attacks**: The traffic security monitoring module of Apsara Stack Security monitors all traffic to Web servers and extracts attack information.
- **Brute-force attacks**: The Server Guard agent installed on a server detects hackers' brute-force attacks to this server and reports the attacks to the console.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Threat Analysis** and click the **Attacks** tab.
3. Select a **Region** and click **Application Attack**.

   Application attacks are displayed.

Table 27-11: Application attacks

Item	Description
Last 7 Days Attacks	Displays attacks to applications on the Apsara Stack platform in the last 7 days.
Attack Types of Last 7 Days	Displays attacks to applications on the Apsara Stack platform in the last 7 days.
Application Attack List	Displays application attack events.

Select an application attack type for **Type**.

Attack events of the corresponding type are displayed in a list. *Table 27-12: Application attack types* describes the application attack types.

Table 27-12: Application attack types

Application attack type	Description
SQL injection	A Web application does not check the validity of the data that users provide. An attacker creates SQL statements to insert special characters and commands in an input area such as a URL or form on the Web page, and interacts with the database to obtain private information or tamper with the database data.
XSS attack	A Web application does not filter or restrict the statements and variables that users provide. An attacker submits malicious code to the database or HTML page from an input area on the Web page. When users click the link or open a page that contains malicious code, malicious code automatically runs on the browser.
Code or command execution	An attacker issues requests using URLs and runs unauthorized code or commands on the Web server.
Local or remote file inclusion	An attacker adds invalid parameters to URLs when issuing requests to the Web server. The Web server fails to filter variables and uses these invalid parameters. These invalid parameters may be the names of local files or remote malicious files. This vulnerability is caused by the failure to strictly filter PHP variables . Therefore, only PHP-based Web applications may have the file inclusion vulnerability.
Trojan script	A Trojan script is a command execution environment in the form of Web files such as ASP, PHP, and JSP. It is also known as a webshell. After intruding a website, an attacker usually mixes

Application attack type	Description
	ASP or PHP webshell files with normal Web page files in the Web directory of the website server. Then, the attacker can access the webshell files from a browser to obtain the command execution environment for controlling the website server.
Upload vulnerability	When processing a file uploaded by a user, a Web application stores a file on the server without checking the validity of the file name extension or the validity of the file content. This file may be a webshell that can control the Web server directly.
Path traversal	When issuing requests to a Web server, an attacker adds `../` and its variant to a URL or a special directory. The attacker can then access the unauthorized directory and run commands in directories except for the root directory of the Web server.
Denial-of-service (DoS)	An attacker uses DoS to exhaust the network or system resources on a server and interrupt or stop services on this server. This prevents authorized users from accessing the server.
Unauthorized access	An application has vulnerabilities in the authentication process. An attacker exploits these vulnerabilities to bypass authentication and access or operate unauthorized code.
Others	Other application attack types.

4. Select a **Region** and click **Brute-Force Attack**.

   Brute-force attack events are displayed.

   Select a brute-force attack event and click **Show** to view the details of the event.

27.4.5 Security reports

27.4.5.1 Add a report task

This topic describes how to add a report task. A report task regularly sends security reports of the Apsara Stack platform to the specified email address, informing the security administrator of the current security situation.

Context

The following table describes the information that a security report can contain.

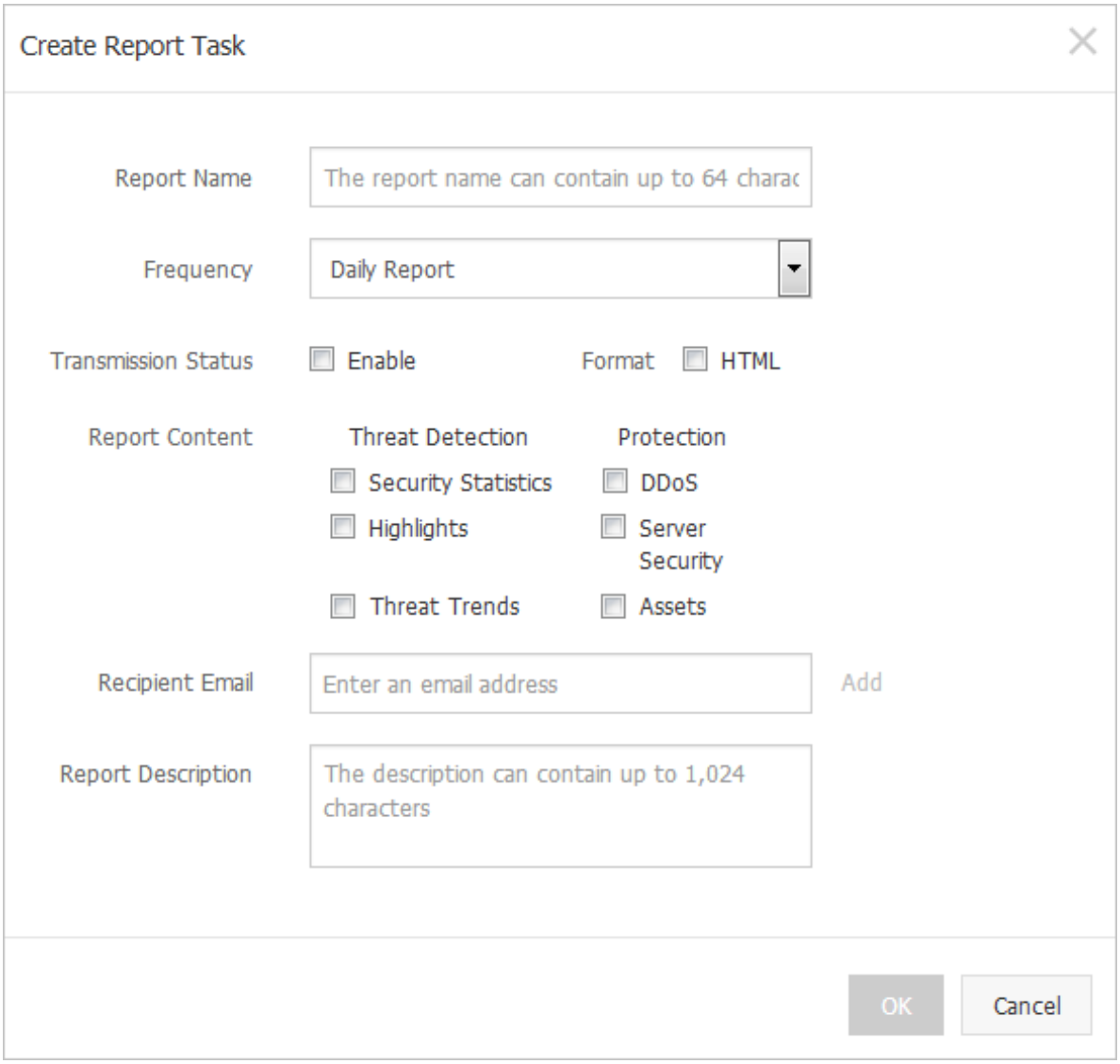
Category	Item	Description
Threat Detection Service	Security statistics	Security overview information on the Overview page of Threat Detection Service.

Category	Item	Description
	Highlights	Important emergency information displayed on the Event Analysis page of Threat Detection Service.
	Threat trends	Attack trend and analysis information displayed on the Threat Analysis page of Threat Detection Service.
Protection	DDoS	DDoS attack events detected by Apsara Stack Security Center.
	Server security	Server security vulnerabilities, unusual logons, brute-force attacks, and configuration risks discovered by Apsara Stack Security Center.
	Protected assets	Assets protected by Apsara Stack Security Center, including server assets and NAT assets.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Security Reports**.
3. On the **Security Reports** page, click **Create Task**.
4. In the **Create Report Task** dialog box, set relevant parameters.

Figure 27-5: Create a report task



Parameter	Description
Report Name	The name of a report task.
Frequency	The interval at which reports are sent. Value options are as follows: • `Daily Report`: Indicates that security reports are sent on a daily basis. • `Weekly Report`: Indicates that security reports are sent on a weekly basis. • `Monthly Report`: Indicates that security reports are sent on a monthly basis.
Transmission Status	Indicates whether to enable this report task.

Parameter	Description
Format	Reports are output in HTML format.
Report Content	Select the items to be contained in a security report.
Recipient Email	The email address that receives security reports. **Note:** Click **Add** next to the text box to add up to 10 email addresses.
Report Description	The report description.

5. Click **OK**.

Result

After the report task has been created, the specified email addresses receive security reports at the set intervals.

27.4.5.2 Manage a report task

This topic describes how to view, modify, and delete a report task.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose**Threat Detection** > **Security Reports**.
3. On the **Security Reports** page, manage existing report tasks.
   - Select a report task and click **Details** to view details of this task.
   - Select a report task and click **Modify** to modify this task.
   - Select a report task and click **Delete** to delete this task.

27.4.6 Vulnerability scan

27.4.6.1 Manage application vulnerabilities

This topic describes how to view and manage application vulnerabilities.

Context

Threat Detection Service can scan applications installed on the server to detect and report vulnerabilities.

Procedure

1. *Log on to Apsara Stack Security Center*.

2. Choose **Threat Detection** > **Vulnerability Scan** and click the **Vulnerabilities** tab.
3. Select the target data center from the **Region** drop-down list.
4. Click **Application Vulnerability** to view application vulnerabilities.

   Analysis of application vulnerabilities for the last 7 days and detailed information about the application vulnerabilities are displayed.

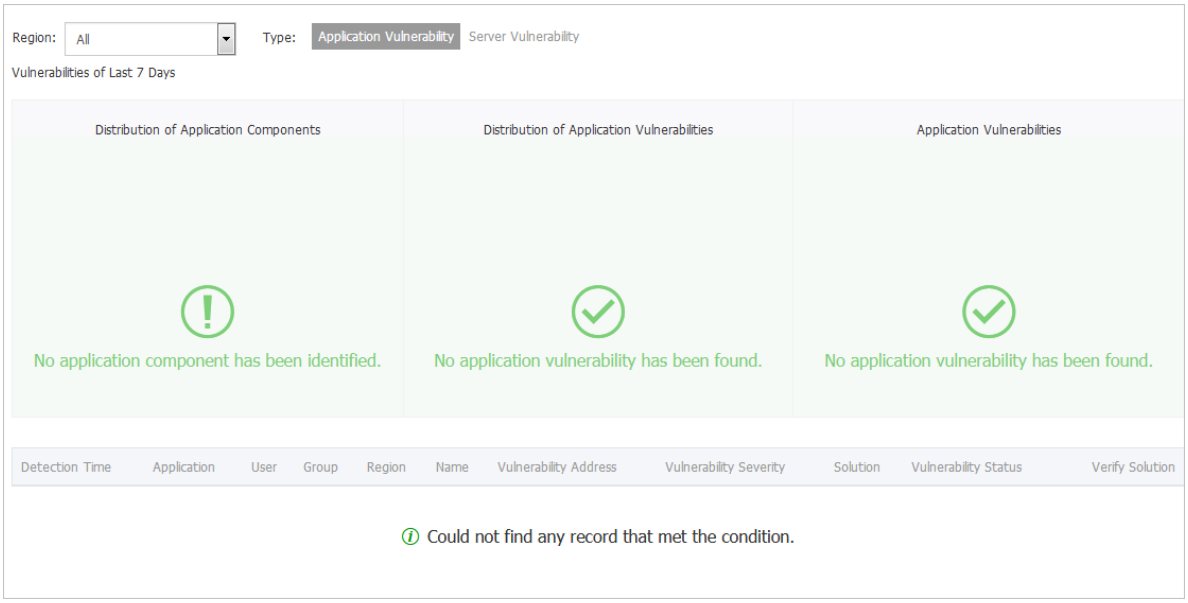


Table 27-13: Parameters in the vulnerability list

Parameter	Description
Name	The name of the application vulnerability.
Vulnerability Address	The server, port, Web URL, and other information about the vulnerability.
Vulnerability Severity	The assessed severity of the vulnerability.
Solution	Click the link to view the solution for fixing the vulnerability.
Vulnerability Status	The status of the vulnerability, including `Unfixed`, `Verifying`, and `Fixed`.
Verify Solution	Click **Verify Now** to check whether the application vulnerability has been fixed.

5. Click **Verify Now** next to an application vulnerability to verify that the application vulnerability has been fixed.

27.4.6.2 View server vulnerabilities

This topic describes how to view server vulnerabilities.

Context

Note:

The server vulnerabilities are detected by Server Guard, which scans the servers and reports the detected server vulnerabilities.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Vulnerability Scan** and click the **Vulnerabilities** tab.
3. Select the target data center from the **Region** drop-down list.
4. Click **Server Vulnerability** to view server vulnerabilities.

   To export the information about detected server vulnerabilities for local use, click **Export** in the upper-right corner.

27.4.6.3 View weak password information

This topic describes how to view weak password information.

Context

The system detects simple asset passwords that are easy to guess or crack. It will then remind the security administrator to change weak usernames and passwords for stronger ones.

Weak passwords pose huge security risks on your system. We recommend that you take the following measures when setting system and application accounts:

- Increase the password complexity, for example, include numbers, characters, and special characters in the password, or increase the password length.
- Manage passwords in a hierarchical manner.
- For important passwords (for example, the SSH password), increasing the complexity is not enough. You also need to regularly change the password.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Vulnerability Scan**.
3. Click the **Weak Passwords** tab to check the weak passwords that have been detected.

4. Select a weak password record and click **Show** to view the details.
5. Select a weak password record and click **Ignore** to ignore the weak password alert, which will no longer be triggered for this weak password.

   You can select an **ignored** weak password record and click **Restore** to enable check for this weak password again.

27.4.6.4 Add custom weak passwords

This topic describes how to add custom weak passwords.

Context

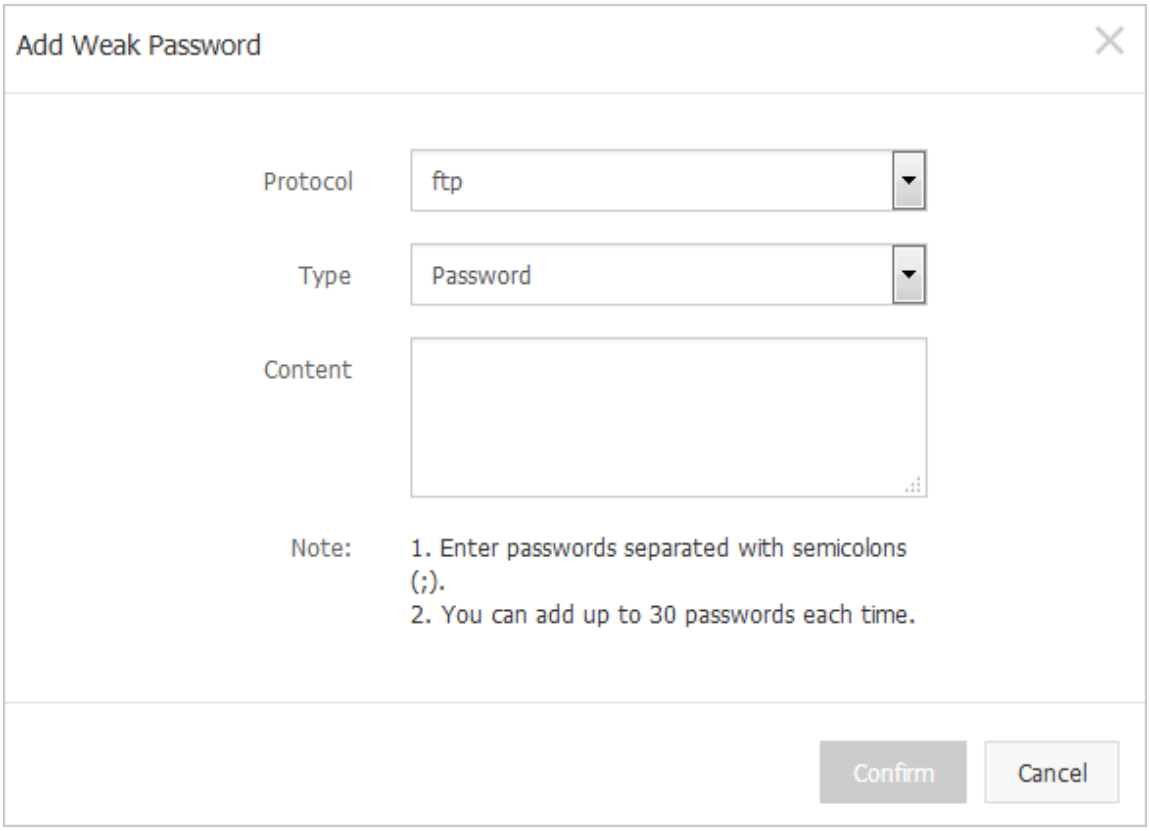
Apart from the default weak passwords, Apsara Stack Security can also scan for the custom weak passwords. Apart from common usernames and passwords, users may use their names, birth dates, or other personal information in their usernames and passwords. This makes their usernames and passwords weak and easy to be guessed. You can add frequently used weak usernames and passwords to Apsara Stack Security so that Apsara Stack Security can scan for them.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Vulnerability Scan** and click the **Weak Passwords** tab.
3. Click **Custom Weak Password**.

   The **Custom Weak Password** page appears. You can view the weak usernames and passwords that have been added and taken effect.
4. Add a custom weak password checking rule.
   a) Click **Add**.
   b) In the **Add Weak Password** dialog box, add a weak password checking rule.

Figure 27-6: Add Weak Password dialog box



c) Click **Confirm**.

d) In the message that appears, click **Confirm**.

5. Check the added weak password checking rule.

   - Click **Modify** to modify the weak password or username.
   - Click **Delete** to delete the weak password checking rule.

6. Click **Export** to generate and download a new weak password configuration file.

   Note:
   The file is downloaded as a .zip package to the default path `C:\Users\username\Downloads`. You do not need to decompress the package.

7. Upload the package to a directory, for example, `/root/war/` in the system that contains the master project of Cactus-keeper for vulnerability analysis. Run the `cactusConfig.sh` script, as shown in *Figure 27-7: Script running page*.

Figure 27-7: Script running page

```
[root@spark2 cactusConfig]# ./cactusConfig.sh
975550d3af6d
/root/war
Archive:  cactus_config_2016-10-28.zip
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_ftp_user.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_ftp_psw.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_mysql_user.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_mysql_psw.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_ssh_user.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_ssh_psw.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_mssql_user.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_mssql_psw.txt
65c5671c8259
cactus_config_2016-10-28.zip
Archive:  /home/datal/yundun/cactus-keeper/cactus_config.zip
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_ftp_user.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_ftp_psw.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_mysql_user.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_mysql_psw.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_ssh_user.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_ssh_psw.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_mssql_user.txt
  inflating: /home/datal/yundun/cactus-keeper/cactus_config/dic_mssql_psw.txt
commad.sh
done!
[root@spark2 cactusConfig]#
```

After the script is run, the new weak password checking rule takes effect.

27.4.6.5 View configuration risks

This topic describes how to view configuration risks.

Context

If your system configuration files or other sensitive files are stored in an improper location, attackers may be able to obtain them without authorization, resulting in the leakage of key information. The configuration inspection function scans the system configuration files and reports configuration files that may be accessed without authorization. Based on the inspection results, the security administrator must promptly add permission protection to the directories that store configuration files, or move sensitive files to a secure directory.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Threat Detection** > **Vulnerability Scan**.
3. Click the **Configuration Risks** tab to view the configuration risks that have been detected.
4. Select a configuration risk and click **Show** to view the details.

5. Select a configuration risk and click **Ignore** to ignore the configuration risk alert, which will no longer be triggered for this configuration item.

   You can select an **ignored** configuration item record and click **Restore** to enable inspection for this configuration item.

27.5 Network security

27.5.1 Enable network security blocking

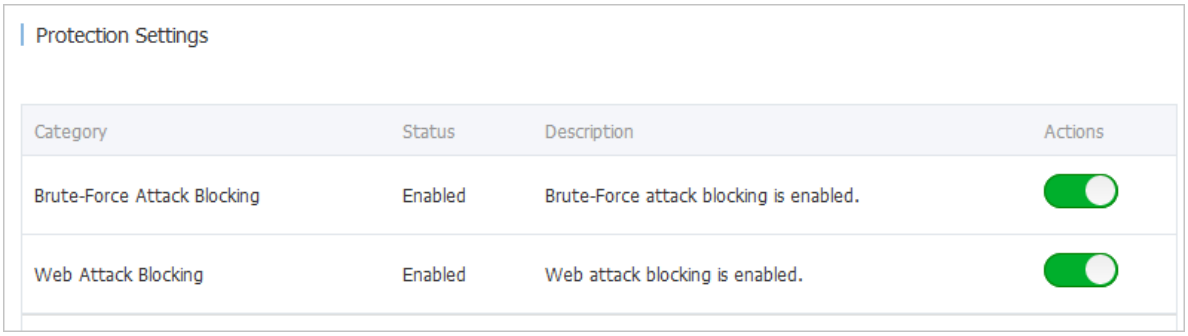
This topic describes how to enable network security blocking.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Network Security** > **Protection Settings**.
3. Click the **Web Attack Blocking** or **Brute-Force Attack Blocking** toggle switch to enable or disable the corresponding function.

   Note:
   After you disable a blocking function, only the warning function is provided.

   Figure 27-8: Configure blocking functions



27.6 Application security

27.6.1 WAF overview

This topic describes the functions of Web Application Firewall (WAF).

Functions

As a website security protection service, WAF protects website applications. WAF protects traffic of HTTP and HTTPS website services. In the WAF console, you can import certificates and private

keys to implement end-to-end encryption for businesses, and prevent data from being intercepted over connections. WAF also meets the security protection requirements of HTTPS services.

WAF provides protection against the following Web attacks:

- Common Web application attacks: such as SQL injections and Cross Site Scripting (XSS) attacks
- DDoS Layer-7 attacks: such as HTTP flood attacks

WAF also allows you to customize accurate protection rules based on the specific businesses of your website and use these rules to block malicious Web requests targeting at your website.

Restrictions

Follow these restrictions when you use WAF:

- WAF supports access protection for a maximum of 100 domain names. It supports wildcard domain names and does not impose any restrictions on top-level or second-level domain names.
- WAF supports domain name protection only for HTTP port 80 and HTTPS port 443.

27.6.2 Connect domain names to WAF

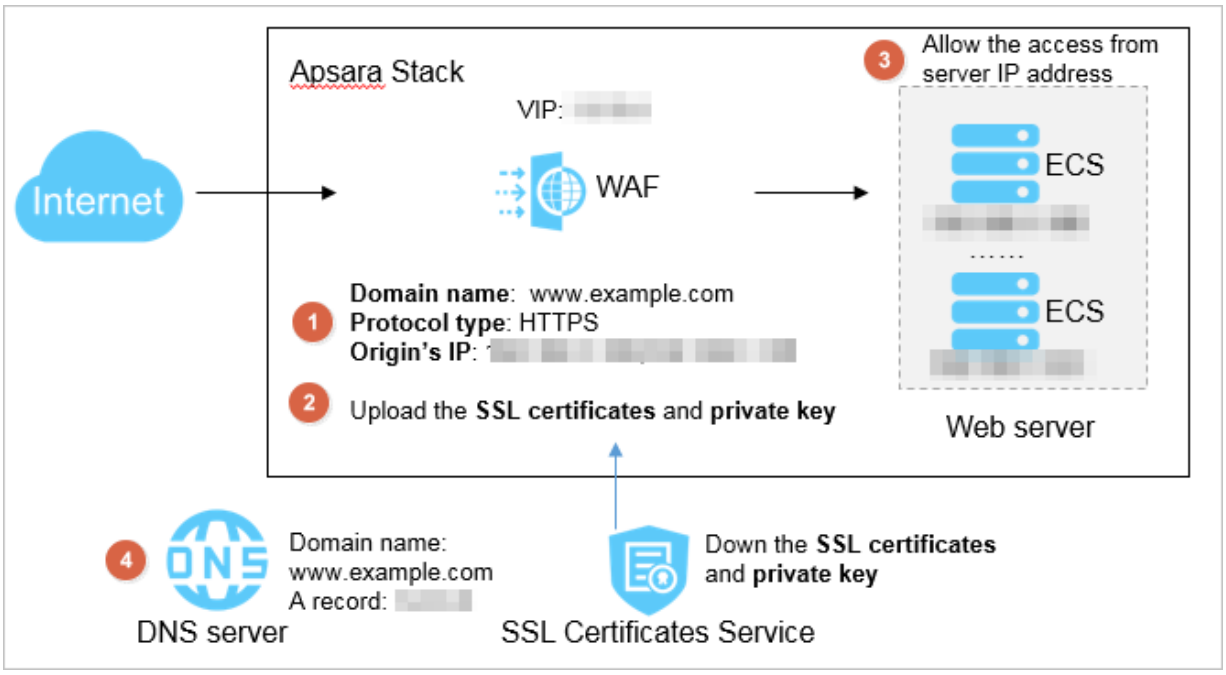
27.6.2.1 Before you start

Before using WAF, you must prepare the following information:

- Information about the domain names you want to protect. You cannot simply use IP addresses.
- Origin IP addresses, which is typically the IP addresses of real servers.

  Note:
  WAF allows you to configure up to 20 origin IP addresses for a single domain name.
- If you use HTTPS, you must also prepare certificates and private keys of servers.

Use the workflow in the following figure to configure the domain names that need to be protected by WAF.

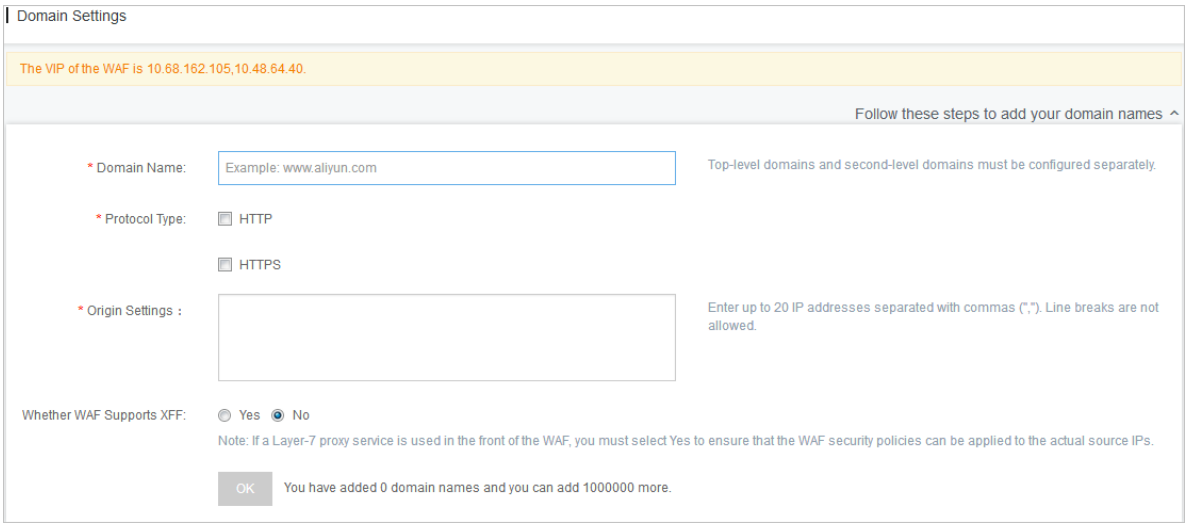


27.6.2.2 Add a protected domain name

This topic describes how to add a protected domain name.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Application Security** > **Domain Settings**.



3. Enter a **Domain Name** that you want to protect.

   Both exact-match domain names and wildcard domain names are supported. A wildcard domain name (such as *.aliyundemo.cn) can match related second-level domain names. When both a wildcard domain name and an exact-match domain name are configured, the forwarding and protection policies take effect for the exact-match domain name first.

4. Set the **Protocol Type**.

   You can select a protocol type, HTTP (port 80) or HTTPS (port 443), based on your site requirements.

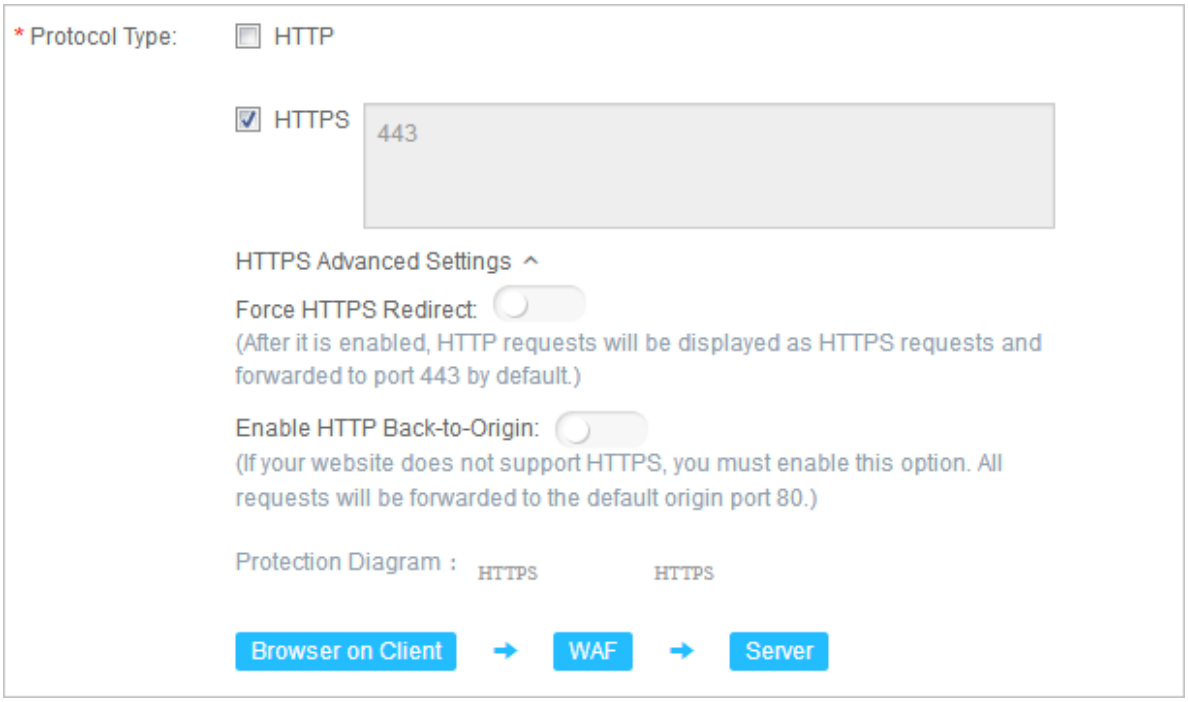
   - For HTTP sites:
     - If you only need HTTP access, select HTTP.
     - If you only need HTTPS access, select HTTPS.
     - If you need both HTTP and HTTPS accesses, select both HTTP and HTTPS.
   - For HTTPS sites:

     HTTPS is required and HTTP is optional. We recommend that you also select HTTP to ensure smooth access in the case of HTTP redirection.

5. Optional: Click **HTTPS Advanced Settings** to configure HTTPS properties.

   Note:
   You need to set **HTTPS Advanced Settings** only when you have selected HTTPS for **Protocol Type**.

   Figure 27-9: HTTPS advanced settings



   - For HTTP sites:

Turn on the **Enable HTTP Back-to-Origin** switch to enable WAF to support HTTPS. After enabling this option, clients can access the site over HTTP and HTTPS.

> **Note:**
> With HTTP back-to-origin, you do not need to change any settings of the origin site server or configure HTTPS, provided that you have uploaded correct certificates and private keys to WAF.

- For HTTPS sites:

  By default, **Force HTTPS Redirect** is disabled. If you want to force clients to access your site using HTTPS, enable **Force HTTPS Redirect**.

  > **Note:**
  > After enabling force HTTPS redirect, you can enable or disable HTTP back-to-origin as needed. If you enable HTTP back-to-origin at the same time, WAF redirects the HTTP requests of clients to HTTPS and sets HSTS for clients for one day. Then clients that support HSTS access the site over HTTPS directly, and clients that do not support HSTS access the site by means of redirection.

6. Enter origin IP addresses in **Origin Settings**.

   An origin IP address refers to the IP address to which WAF forwards requests. You can use either the origin server IP address, such as the IP address of your ECS instance, or the IP address of your SLB instance.

   A maximum of 20 origin IP addresses are supported. WAF supports load balancing and health check.

7. Set **Whether WAF Supports XFF** to Yes or No.

   The X-Forwarded-For (XFF) HTTP header field reveals the real IP address of an originating client. It is used in forwarding services such as HTTP proxy and load balancing.

   If you have already configured a Layer-7 proxy for WAF, set this parameter to `Yes`, so that the WAF security policies can take effect on the originating IP addresses.

8. Click **OK**.

27.6.2.3 Upload HTTPS certificates and private keys for domain names of HTTPS websites

This topic describes how to upload HTTPS certificates and private keys for domain names of HTTPS websites.

Prerequisites

You have applied for an HTTPS certificate and a private key.

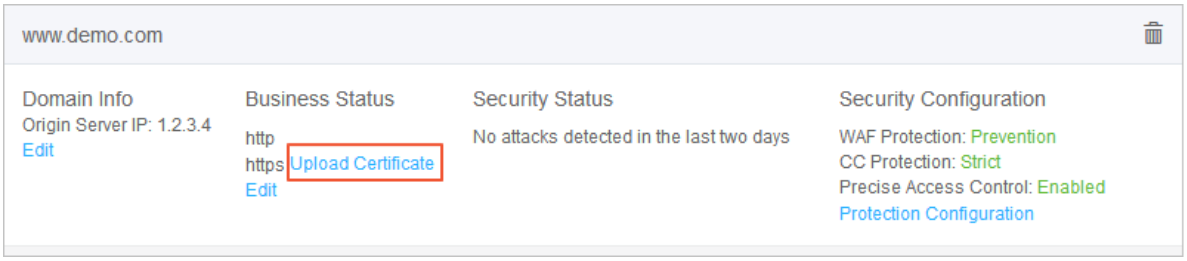
Context

To protect the domain name of an HTTPS website, you must upload the HTTPS certificate and private key to WAF. Otherwise, users will fail to access the HTTPS website.

If a domain name does not support access over HTTPS, ignore this task.

Procedure

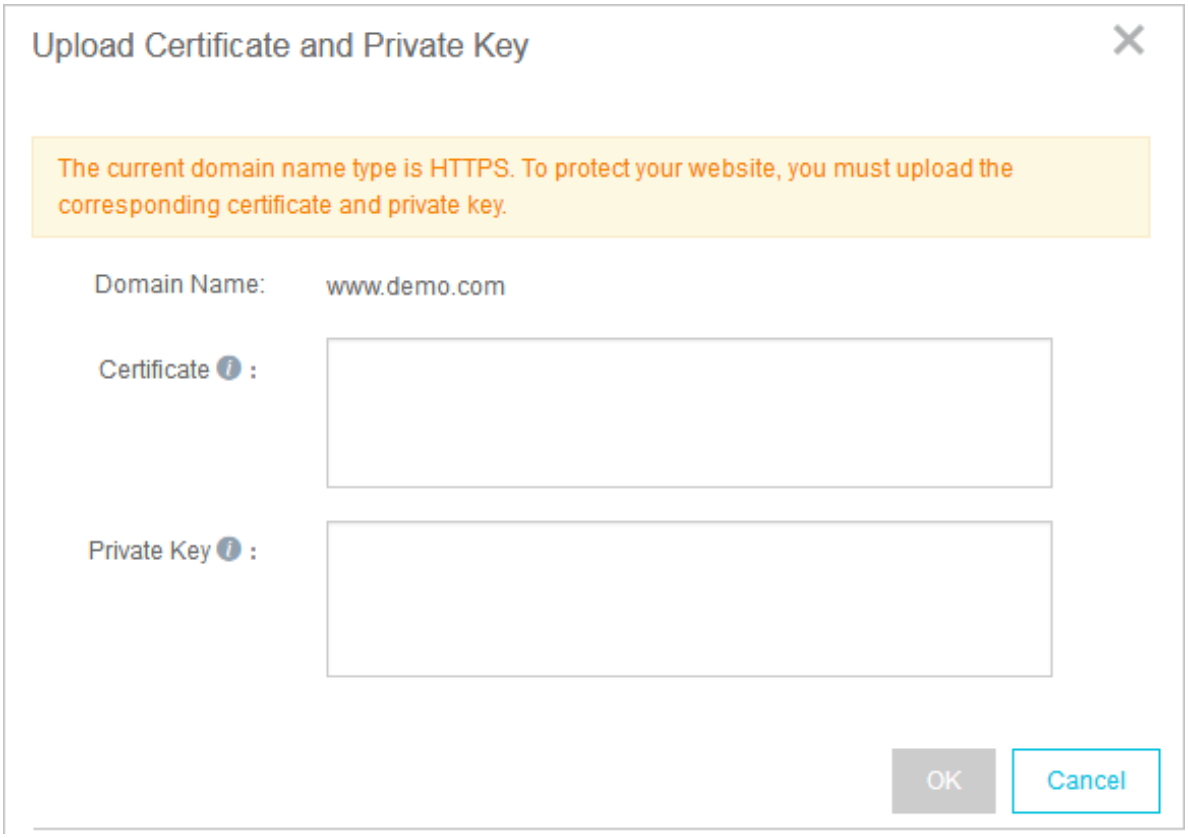
1. *Log on to Apsara Stack Security Center*.
2. Choose **Application Security** > **Domain Settings**.
3. Select an HTTPS domain name in the domain name list and click **Upload Certificate**.



4. Upload the certificate and private key.

   Copy and paste the certificate file content and private key file content to the corresponding text boxes.

Figure 27-10: Upload the certificate and private key



You can directly open certificate files in common formats, such as PEM, CER, and CRT, in a text editor and copy the content. Certificate files in other formats, such as PFX and P7B, must first be converted to common formats. If multiple certificate files exist, for example, a certificate chain, you can splice and upload them together.

- The following is an example of the certificate file content that can be recognized by WAF:

```
-----BEGIN CERTIFICATE----- 62EcYPWd2Oy1vs6MTXcJSfN9Z7rZ9f
mxWr2BFN2XbahgnsSXM48ixZJ4krc+1M+j2kcubVpsE2 cgHdj4v8H6jUz9Ji4mr7
vMNS6dXv8PUkl/qoDeNGCNdyTS5NIL5ir+g92cL8IGOkjgvhlqt9vc 65Cgb4mL+
n5+DV9uOyTZTW/MojmlgfUekC2xiXa54nxJf17Y1TADGSbyJbsC0Q9nIrHsPl8YKk
 vRWvIAqYxXZ7wRwWWmv4TMxFhWRiNY7yZIo2ZUhl02SIDNggIEeg== -----END
CERTIFICATE-----
```

- The following is an example of the private key file content that can be recognized by WAF:

```
-----BEGIN RSA PRIVATE KEY----- DADTPZoOHd9WtZ3UKHJTRgNQmioPQn
2bqdKHop+B/dn/4VZL7Jt8zSDGM9sTMThLyvsmLQKBgQ Cr+ujntC1kN6pGBj2Fw2l
/EA/W3rYEce2tyhjgmG7rZ+A/jVE9fld5sQra6ZdwBcQJaiygoIYo aMF2EjRwc0
qwHaluq0C15f6ujSoHh2e+D5zdmkTg/3NKNjqNv6xA2gYpinVDzFdZ9Zujxvuh9o
4Vqf0YF8bv5UK5G04RtKadOw== -----END RSA PRIVATE KEY-----
```

5. Click **OK**.

27.6.2.4 Allow the access from a server IP address in the WAF cluster

This topic describes how to configure the security group of the origin server to allow the access from a server IP address in the WAF cluster.

Context

WAF acts as a proxy for a server and uses VIP to receive traffic and requests from clients. WAF uses the IP address of the server in the WAF cluster to send the filtered traffic to the server.

Therefore, you must configure the security group of the origin website to allow the access from the server IP address in the WAF cluster. Otherwise, the website denies the access or responds very slowly.

Procedure

1. Obtain the IP address of a server in the WAF cluster.

   Contact the administrator to obtain the server IP address in the WAF cluster. The administrator can follow these steps to obtain the server IP address:

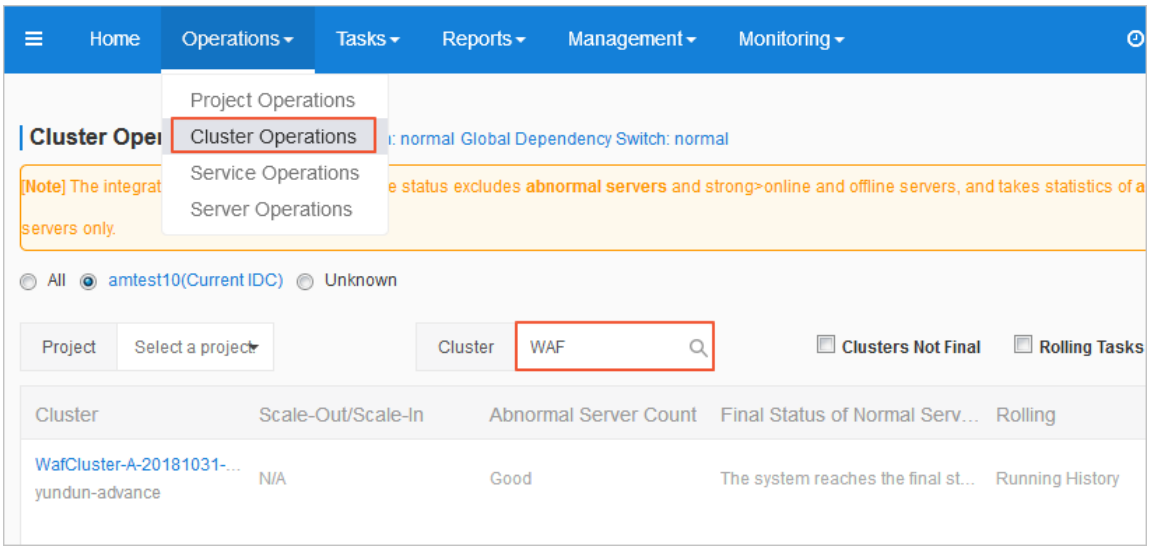
   a) In the address bar of a browser, enter `https://address of Apsara Service Operation platform` and press Enter.

   b) On the logon page, enter the username and password and click **Log On**.

   c) In the left-side navigation pane, click **Products**.

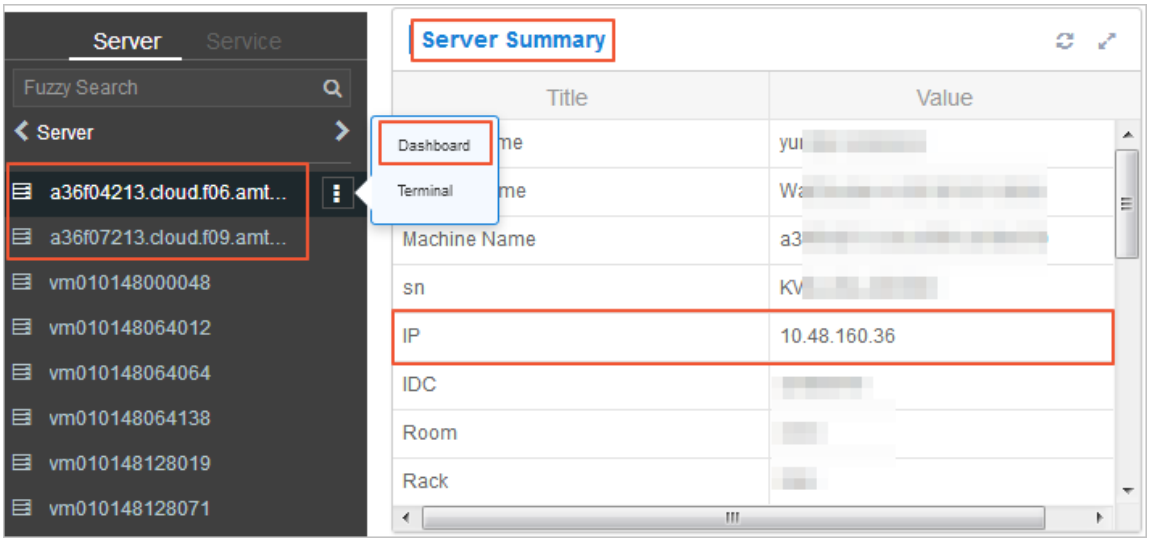
   d) In the product list, click **Apsara Infrastructure Management Framework** to go to the Apsara Infrastructure Management Framework console.

   e) Choose **Operations** > **Cluster Operations**. On the Cluster Operations page, enter WAF in the **Cluster** parameter.



f) Click the WAF cluster in the list. On the **Cluster Dashboard** page, view the server list on the left.

g) Hover over ⋮ of the target server and click **Dashboard**. In the **Server Summary** area, check the IP address of the server in the WAF cluster.



2. Log on to the Apsara Stack console and configure the security group of the origin server to allow the access from the server IP address in the WAF cluster.

   For more information about how to configure a security group, see **Manage security groups** in *User Guide*.

27.6.2.5 Verify the WAF connection configuration for a domain name locally

This topic describes how to verify the WAF connection configuration for a domain name by accessing the domain name from a local PC.

Context

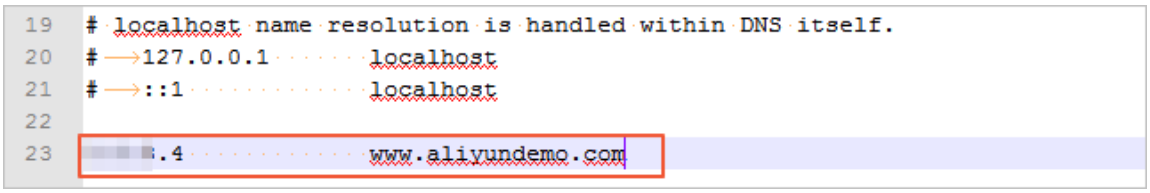
Before you switch business traffic to WAF, we recommend that you perform a local verification to ensure that the domain name has been connected to WAF and that WAF can forward traffic correctly. After you have added the VIP of WAF and the domain name of a website to the local hosts file, the request to access the domain name from a local browser passes through WAF first.

Procedure

1. *Log on to Apsara Stack Security Center*
2. Add the VIP and domain name to the `hosts` file on your local PC.
   In Windows 7, the hosts file path is `C:\Windows\System32\drivers\etc\hosts`.

a) Open the hosts file by using a text editor such as Notepad.

b) Add `<WAF VIP> <Protected domain name>` to the last line.



Note:

The IP address in front of the domain name is the VIP assigned by WAF.

3. Ping the protected domain name from the local PC.

   The resolved IP address must be the WAF VIP that is bound in the hosts file. If the resolved IP address is still the IP address of the origin website, refresh the local DNS cache.

4. Enter the domain name in the address bar of a browser and press Enter.

   If the domain name has been correctly connected to WAF, you can visit the website properly.

5. Verify the WAF protection configuration.

   Simulate a Web attack request to check whether WAF can block the access.

   For example, add `/? alert(xss)` to the URL of the domain name, for example, `www.aliyundemo.cn/? alert(xss)`. The block page will appear.

27.6.2.6 Modify DNS resolution settings

This topic describes how to connect your businesses to WAF by modifying the DNS resolution settings.

Context

Before modifying the DNS resolution settings and switching business traffic to WAF, make sure that you have passed local verification.

A protected domain name of a website may not be resolved by a DNS provider, for example, the website uses a Server Load Balancer (SLB) instance to connect to the Internet. To connect such a domain name to WAF, use the following procedure to specify the WAF VIP address as the origin IP address of the SLB instance:

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Application Security** > **Domain Settings**.

3. Record the VIP address that is assigned by WAF to the protected domain name.
4. Log on to the console provided by the DNS provider and find the domain name resolution settings for the relevant domain name. Then, change the A record value to the WAF origin IP address.

> **Note:**
> We recommend that you set the TTL value to 600 seconds in DNS resolution settings. The greater the TTL value is, the longer it takes to synchronize and update DNS records.

27.6.3 Configure protection functions

27.6.3.1 Configure Web application protection

This topic describes how to enable Web application protection, and how to set the prevention mode and protection rule.

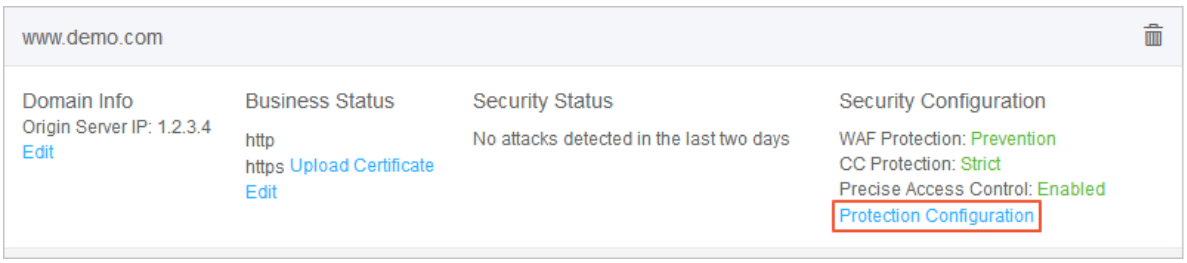
Context

This function protects your website against common Web application attacks, such as SQL injections, Cross-Site Scripting (XSS) attacks, unrestricted file uploads, file inclusion attacks, common directory traversal attacks, common CMS vulnerabilities, code injection attacks, webshell attacks, and attacks based on scanners.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Application Security** > **Domain Settings**.
3. Select a domain name that is already protected by WAF, and click **Protection Configuration**.

Figure 27-11: Configure domain protection



4. In the **Web Application Protection** area, turn on the switch next to **Status**, select Prevention or Detection, and select a protection mode.

Figure 27-12: Web Application Protection

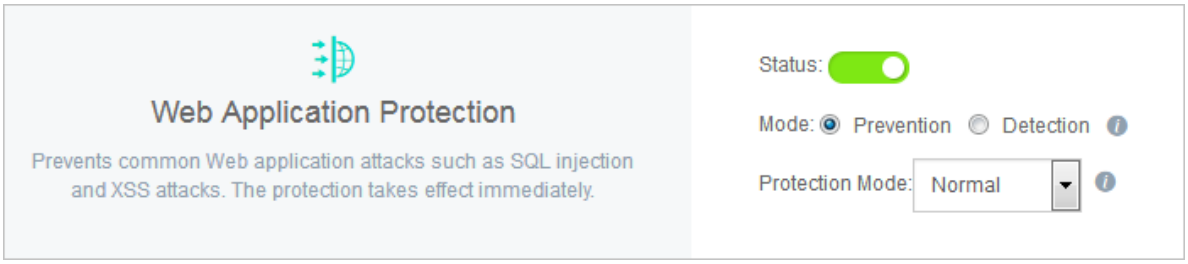


Table 27-14: Protection configurations

Protection configuration		Description
Mode	Prevention	WAF automatically blocks attacks.
	Detection	For suspicious attacks, WAF sends alerts instead of blocking the attacks immediately. This allows you to detect false positives.
Protection Mode	Normal	By default, this mode is selected.
	Loose	If a protection rule is frequently triggered to intercept many normal requests by mistake in normal mode or your business has a relatively high amount of uncontrollable user input (such as rich text editors and technical forums), we recommend you select loose mode.
	Strict	If you need more rigorous protection rules to protect against path traversal, SQL injection, command execution, and other attacks, we recommend that you select strict mode.

27.6.3.2 Configure malicious IP blocking

This topic describes how to block malicious IP addresses. When multiple Web attacks are launched from a specific IP address in a short period of time, you can use this function to block all access requests from this IP address for a certain period.

Prerequisites

You can enable malicious IP blocking only after you have enabled both Web application protection and HTTP flood protection.

Context

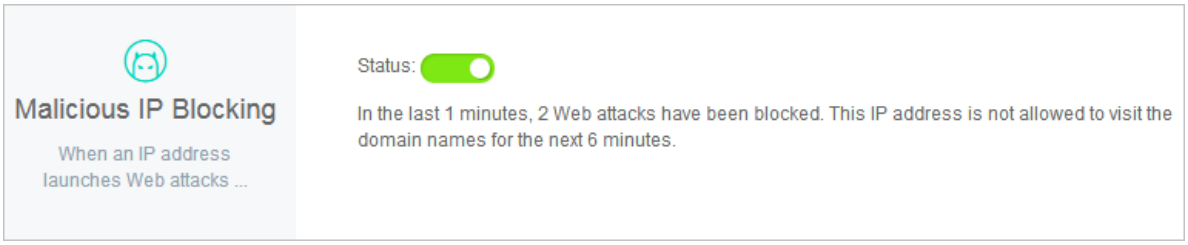
Traditional WAF products block requests based on IP addresses and URLs. After determining that a request is an attack, they only block this request once. However, malicious attackers may scan and attack your website repeatedly. These attackers may explore the vulnerabilities of your website all night long, study your protection policies, and attempt to bypass them.

Apsara Stack Security WAF allows you to configure malicious IP blocking. Backed by the machine learning function, Apsara Stack Security WAF generates judging rules for malicious IP address by using a database with a large number of malicious IP addresses. Apsara Stack Security WAF keeps studying and analyzing the attacks and attack frequencies of the malicious IP addresses. When an IP address initiates continuous attacks, WAF automatically blocks all access requests from this IP address.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Application Security** > **Domain Settings**.
3. Select a domain name that is already protected by WAF, and click **Protection Configuration**.
4. In the **Malicious IP Blocking** area, turn on the switch next to **Status** to enable malicious IP blocking.

Figure 27-13: Malicious IP Blocking



27.6.3.3 Configure HTTP flood protection

This topic describes how to block HTTP flood attacks by using Apsara Stack protection engines and big data.

Context

An HTTP flood attack is a type of DDoS attacks that are targeted at Web server applications. An attacker uses a proxy server or zombie to initiate massive HTTP requests to the target server, which exhausts the server resources.

WAF provides the following protection modes to protect servers against HTTP flood attacks:

- **Strict**: The default mode. This mode is used to prevent false positives by blocking only suspicious requests.
- **Loose**: This mode provides stronger HTTP flood attack blocking performance but has a higher rate of false positives. If you find that your website is under an HTTP flood attack and the attack cannot be blocked by using the strict mode, you can use the loose mode.

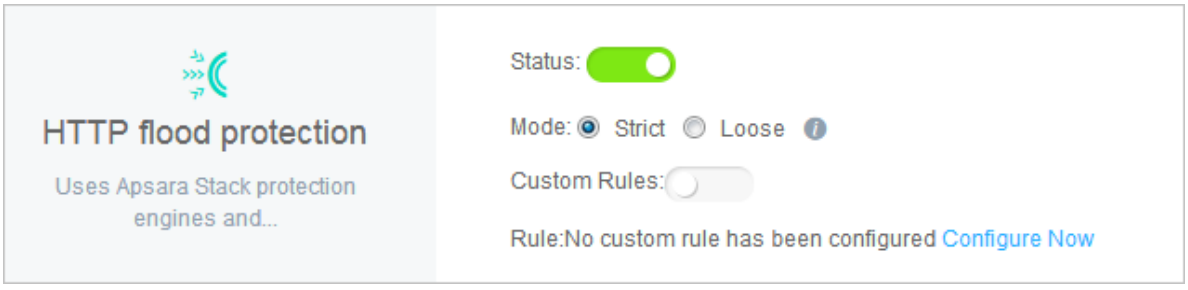
  Note:
  The loose mode applies only to common Web pages and HTML5 pages but is not applicable to APIs or native apps because it may block a large number of normal requests by mistake. To protect APIs and native apps against HTTP flood attacks, customize protection rules for them.

WAF allows you to customize HTTP flood attack protection rules. You can customize the access frequency limit for specific URL paths in the console.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Application Security** > **Domain Settings**.
3. Select a domain name that is already protected by WAF, and click **Protection Configuration**.
4. In the **HTTP Flood Protection** area, turn on the switch next to **Status** and select a protection mode.

   Figure 27-14: HTTP Flood Protection

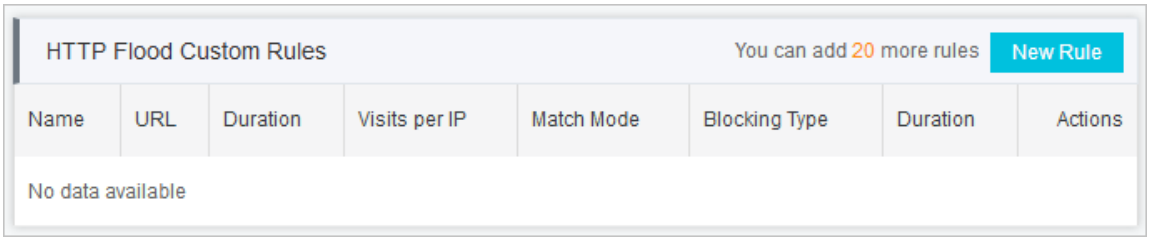
   

5. Turn on the switch next to **Custom Rules** to enable custom protection rules.

   To configure custom protection rules, follow these steps:

   a) Click **Configure Now** to configure custom protection rules.

Figure 27-15: Custom HTTP flood protection rules



b) Click **New Rule** to add a rule.

For example, you can customize a protection rule that blocks an IP address for an hour if this IP address is used to access www.example.com/login.html for more than 20 times in 10 seconds.

Figure 27-16: Create a custom rule

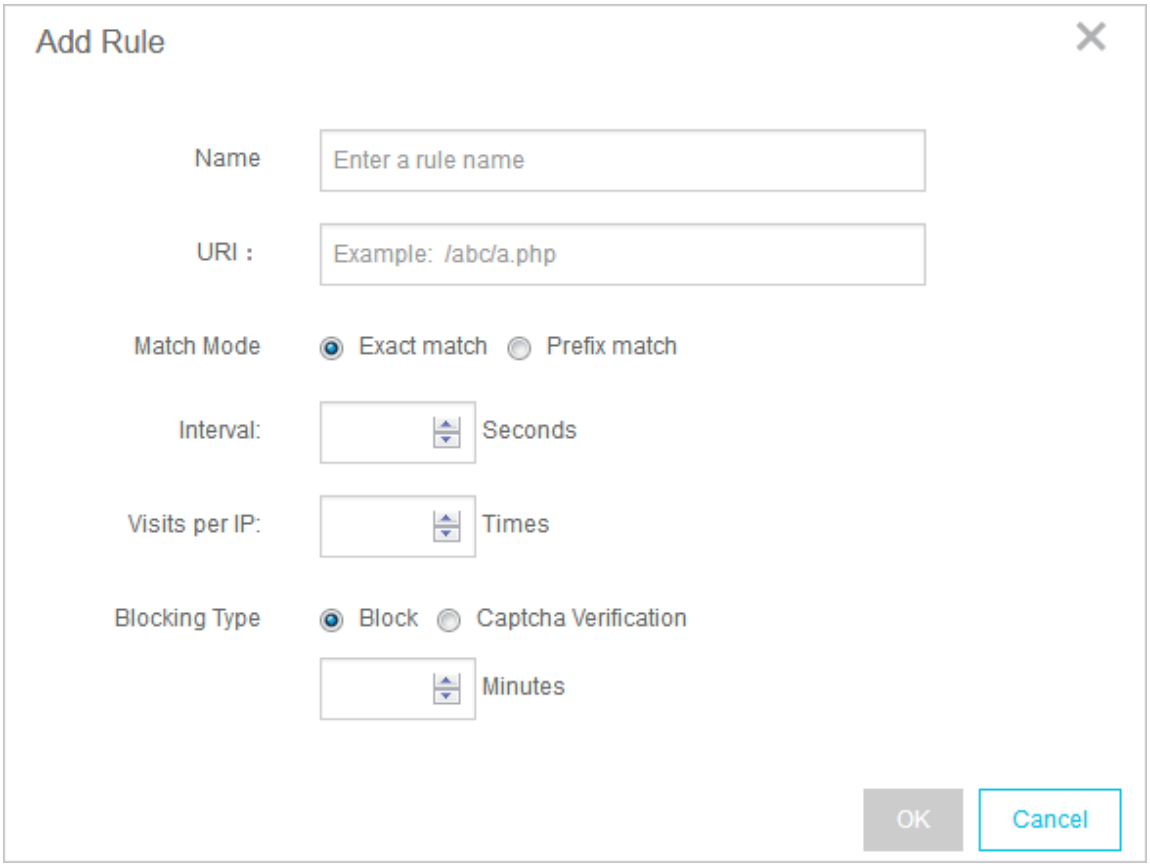


Table 27-15: Rule parameters

Parameter	Description
URI	The URI to be protected, for example, `/register`. You can specify parameters in the URI, for example, `/user? action= login`.
Match Mode	• **Exact Match**: This rule applies only when the request exactly matches the configured URI. • **Prefix Match**: This rule applies when the requested URI starts with the configured URI, for example, `/register.html`.
Interval	The period during which access attempts are counted. This parameter is used together with the Visits per IP parameter.
Visits per IP	The number of visits from a single source IP address to the URL during the statistics period.
Blocking Type	• **Block**: When this rule is triggered, the connection is interrupted immediately. • **Captcha Verification**: When this rule is triggered, the client is redirected to a human/machine identification test. Only verified requests are allowed to pass.
Duration	The time during which the IP address is blocked.

c) Click **OK**.

27.6.3.4 Configure precise access control

This topic describes how to configure precise access control. You can customize access rules to filter access requests by criterion such as client IP address or request URL.

Context

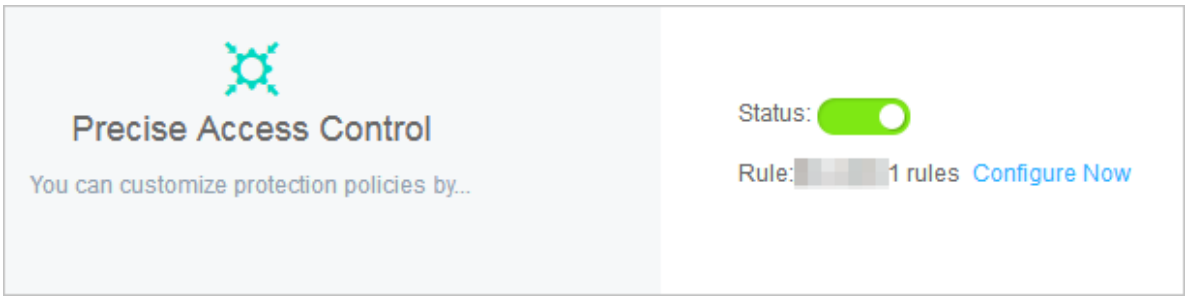
With the precise access control function, you can configure access control rules to filter access requests based on match conditions. A matching condition is a combination of commonly used HTTP fields, such as IP, URL, Referer, User-Agent, and Params. For requests that meet the matching conditions, you can allow them to pass, block them, or report alerts for them. This function applies to different business scenarios, such as hotlinking protection and website administration console protection.

Precise access control rules are matched in a specific order. You can adjust the order of the rules to achieve optimal protection effects.

Procedure

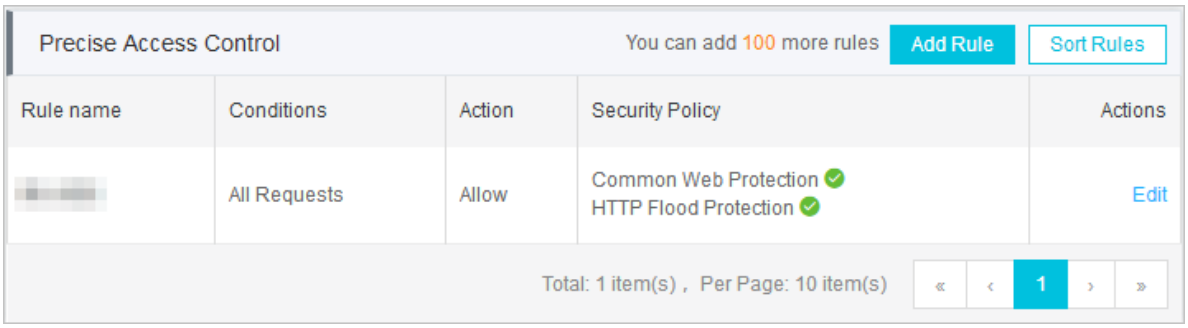
1. *Log on to Apsara Stack Security Center*.
2. Choose **Application Security** > **Domain Settings**.
3. Select a domain name that is already protected by WAF, and click **Protection Configuration**.
4. In the **Precise Access Control** area, turn on the switch next to **Status** to enable the precise access control function.

   Figure 27-17: Precise Access Control

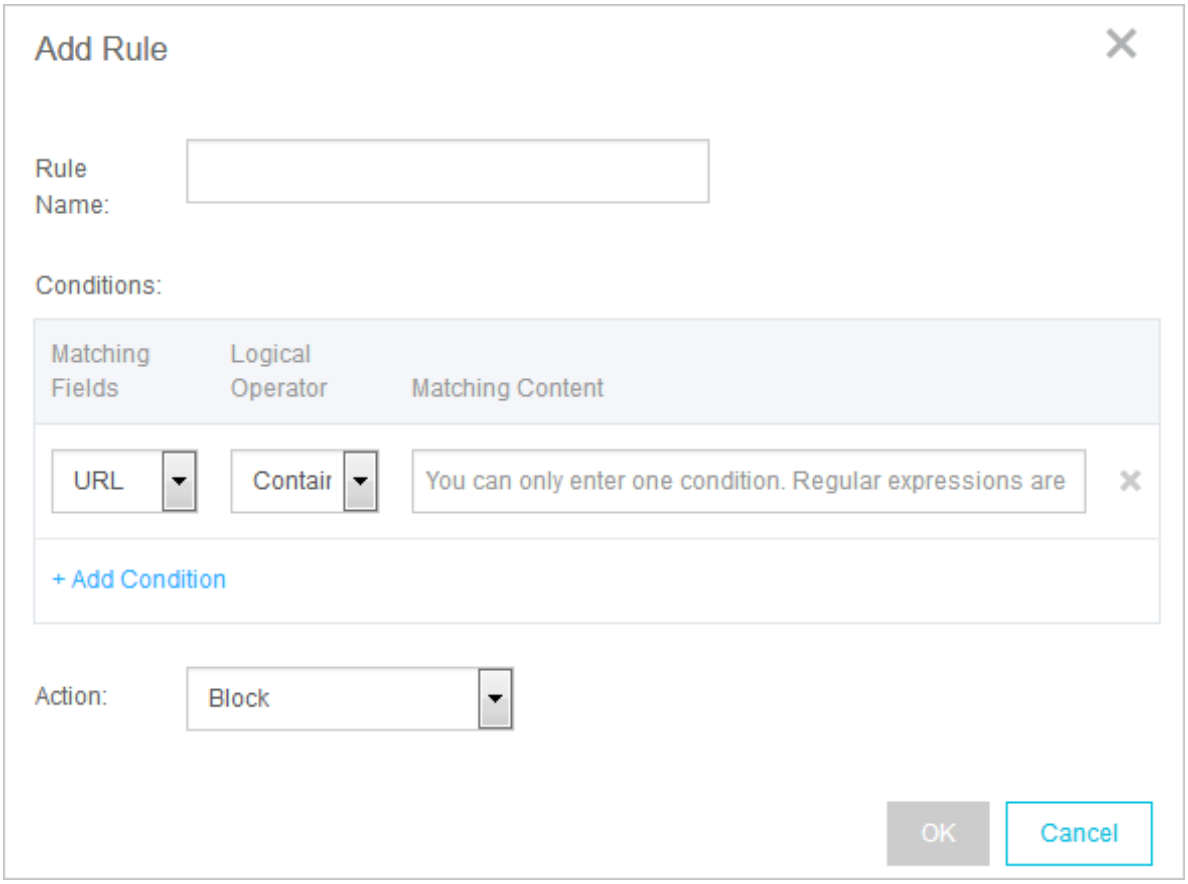
5. Click **Configure Now**.

   Figure 27-18: Configure precise access control rules

6. Click **Add Rule**, configure a precise access control rule, and click **OK**.

Figure 27-19: Add a precise access control rule



Each precise access control rule consists of matching conditions and a matching action. When creating a rule, you can define matching conditions and set an action to be triggered for the access requests that meet the matching conditions.

- **Conditions**:

  A matching condition includes a matching field, logical operator, and matching content. You can set a maximum of three matching conditions. The logical operator between these conditions is **AND**. That is, an access request must meet all matching conditions to hit the rule.

  The supported matching fields are described as follows.

Matching field	Description
IP	The source IP address of the access request.
URL	The URL of the access request.

Matching field	Description
Referer	The source website address of the access request. This address represents the Web page from which you are redirected.
User-Agent	The browser ID, rendering engine ID, version information, and other browser-related information of the client that launches the access request.
Params	The parameters in the access request URL, which start after the question mark (?).

- **Action**:

  The following actions can be taken after a rule is matched:

  - Block: Blocks the access request that meets the matching conditions.
  - Allow: Allows the access request that meets the matching conditions to pass.
  - Alert: Allows the access request that meets the matching conditions to pass and reports an alert for the request.

  Note:
  After selecting `Allow` or Alert, you can further set whether the request needs to be detected and filtered by other WAF protection functions.

7. Click **Sort Rules** to adjust the order of the existing precise access control rules.

   If you configure multiple rules, they follow a specific matching order. Access requests are matched based on the order of the precise access control rules. The rule with the higher ranking is matched first.

   You can set the order of precise access control rules to achieve optimal protection effects.

27.6.3.5 Configure blocked areas

This topic describes how to configure blocked areas to block access requests from IP addresses in these areas.

Context

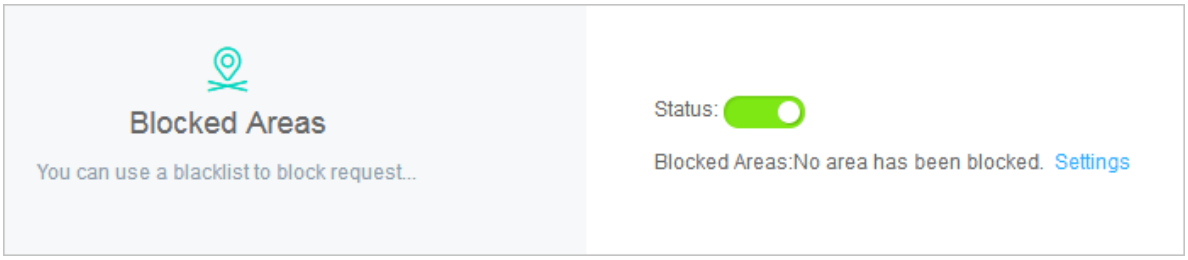
WAF can block source IP addresses in specific areas. Based on the IP geolocation database, this function can block IP addresses both in and outside China.

Procedure

1. *Log on to Apsara Stack Security Center*.

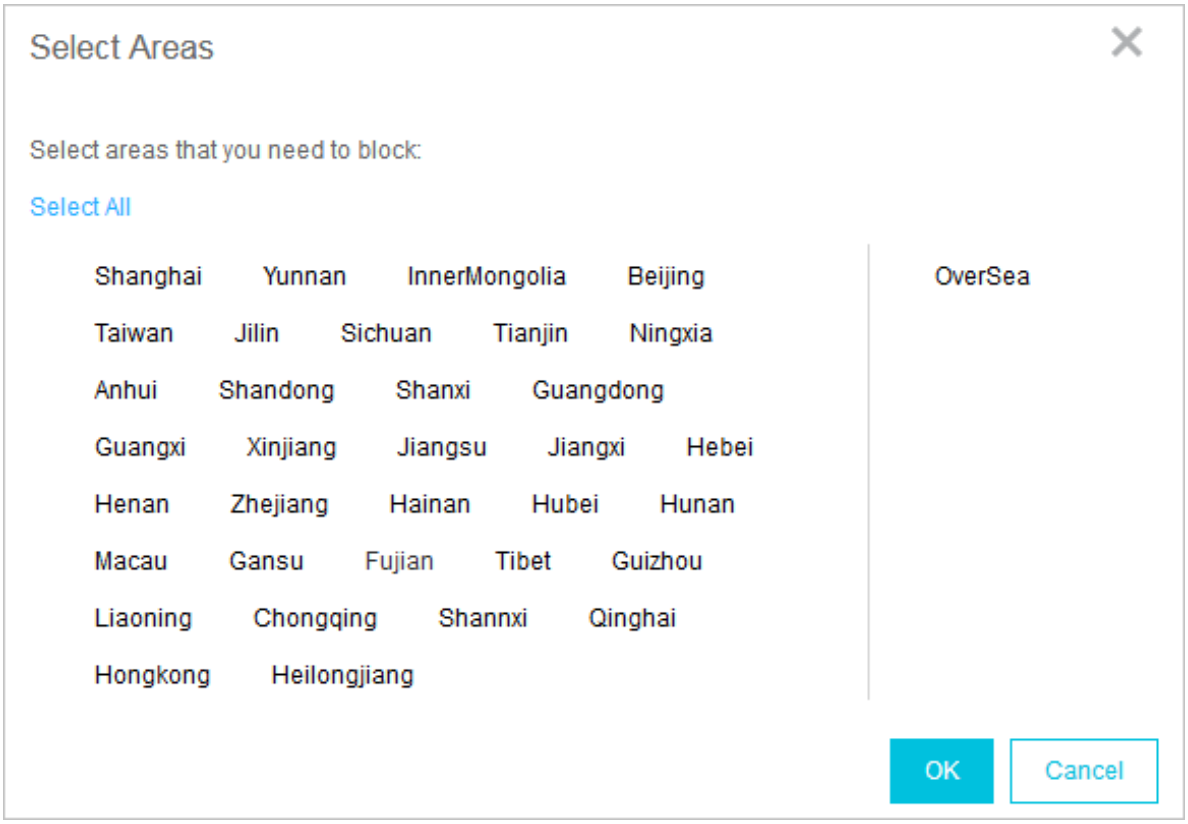
2. Choose **Application Security** > **Domain Settings**.
3. Select a domain name that is already protected by WAF, and click **Protection Configuration**.
4. In the **Blocked Areas** area, turn on the switch next to **Status** to block access requests from IP addresses in specific areas.

Figure 27-20: Blocked Areas



5. Click **Settings** to configure the areas in which IP addresses are prevented from accessing the domain name.

Figure 27-21: Configure blocked areas



27.6.4 View security reports

27.6.4.1 View security overview

This topic describes how to view security overview. The Overview page of Application Security displays attack protection reports and messages related to WAF protection rules.

Context

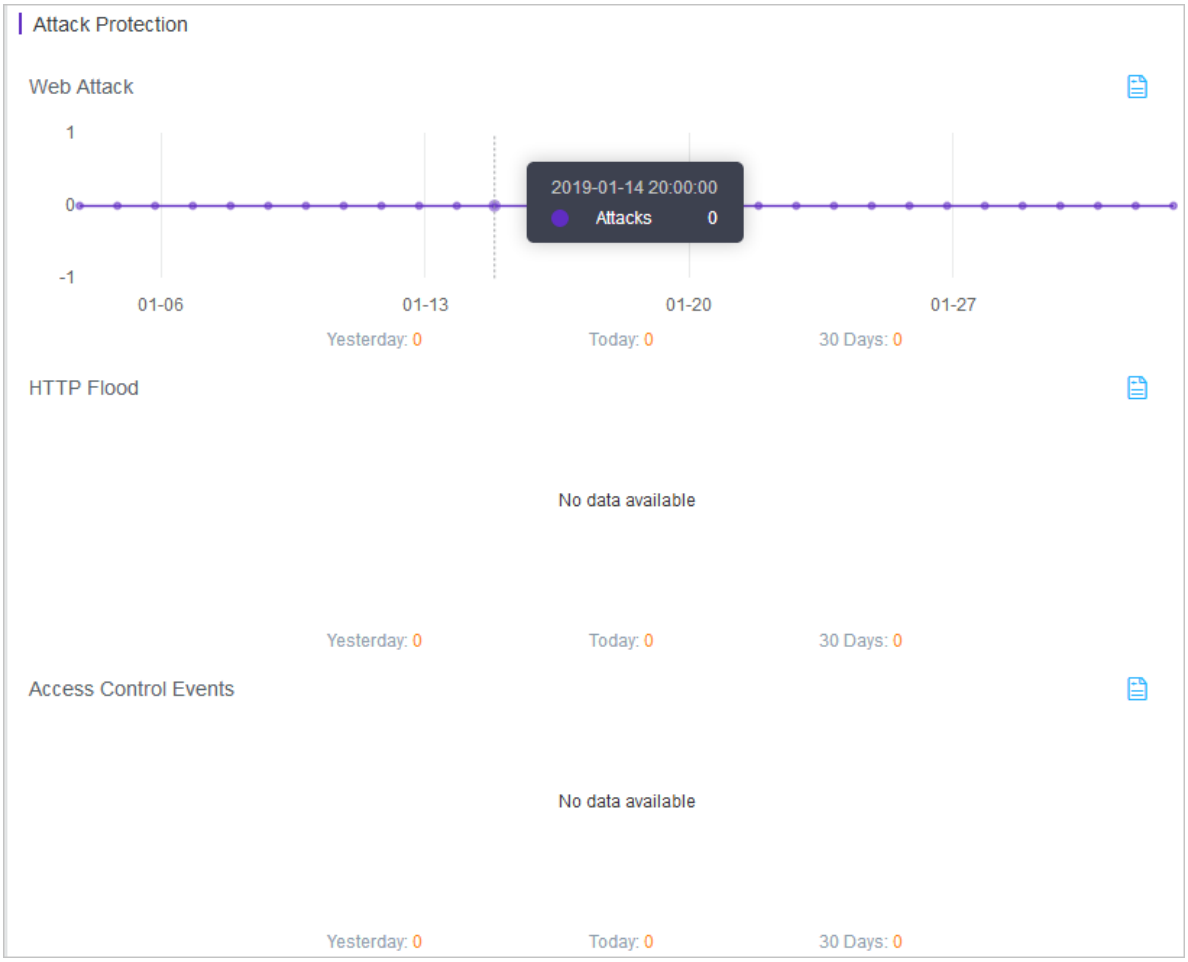
The Overview page of Application Security displays attack protection reports and messages related to WAF protection rules.

- In the attack protection reports, you can view the protection information of Web attacks, HTTP flood attacks, and access control events, and the number of attacks that have been prevented yesterday, today, or in the last 30 days. This helps the security administrator quickly learn the overall security status of protected domain names.
- In the Messages area, you can view the update messages of WAF protection rules.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Application Security** > **Overview**.
3. In the **Attack Protection** area, view the WAF attack protection information.

Figure 27-22: Overview page



In the **Attack Protection** area, you can view the line charts for the number of Web attacks, HTTP flood attacks, and access control events. Click the View Details icon in the upper-right corner of each report to go to the **Security Report** page. The detailed attack protection information is displayed.

4. In the **Messages** area, you can view the update messages of WAF protection rules.

27.6.4.2 View security reports

This topic describes how to view detailed protection information for the domain names connected to WAF.

Context

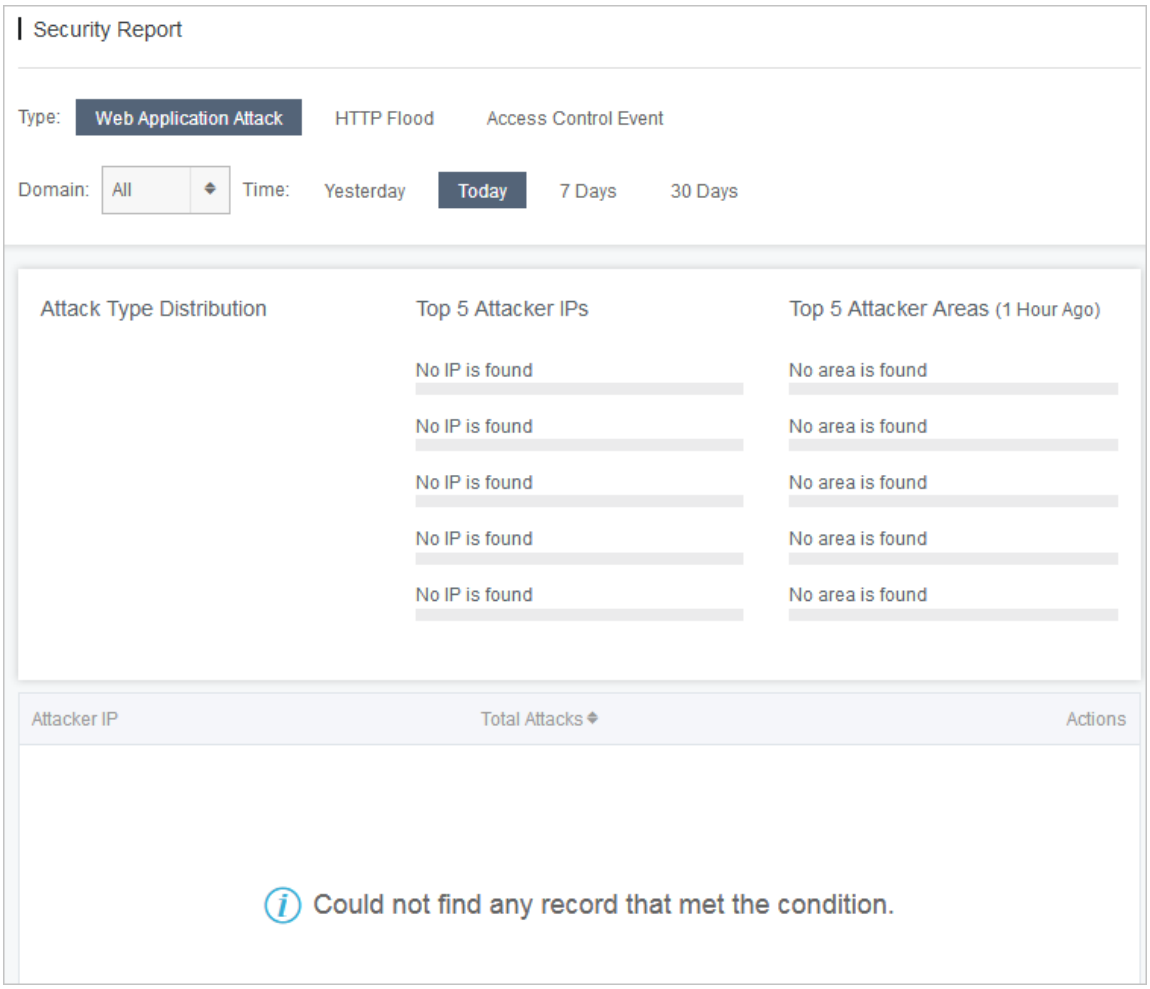
On the Security Report page, you can view detailed protection information about the domain names protected by WAF.

- Click **Web Application Attack** to view the attack type distribution, attacker IP addresses, attacker areas, and detailed attack records.
- Click **HTTP Flood** to view queries per second (QPS) information for servers, including the total QPS, attack QPS, and detailed records about malicious HTTP flood attack events.
- Click **Access Control Event** to view the hits of the configured precise access control rules and the corresponding actions.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Application Security** > **Security Report**.
3. View the security report on Web application attacks.
   a) Click **Web Application Attack** and set the domain name and time period, as shown in *Figure 27-23: Security report on Web application attacks*.

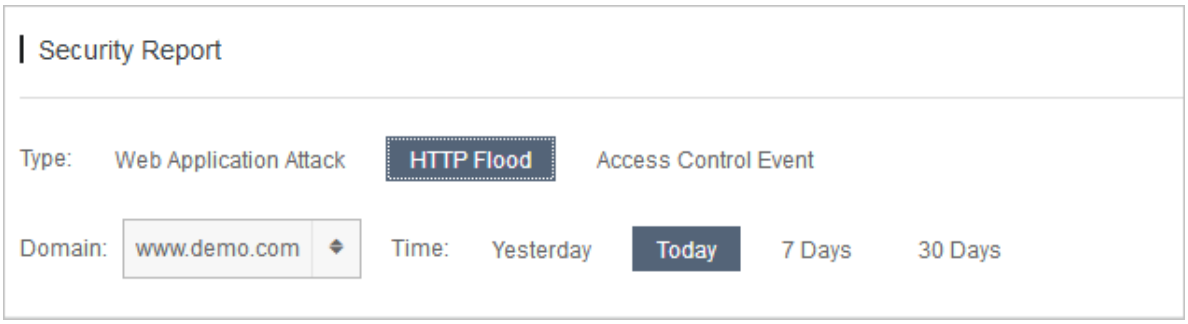
Figure 27-23: Security report on Web application attacks



b) Select an attack record and click **View Attack Details**. The attack details and the triggered protection rule are displayed.

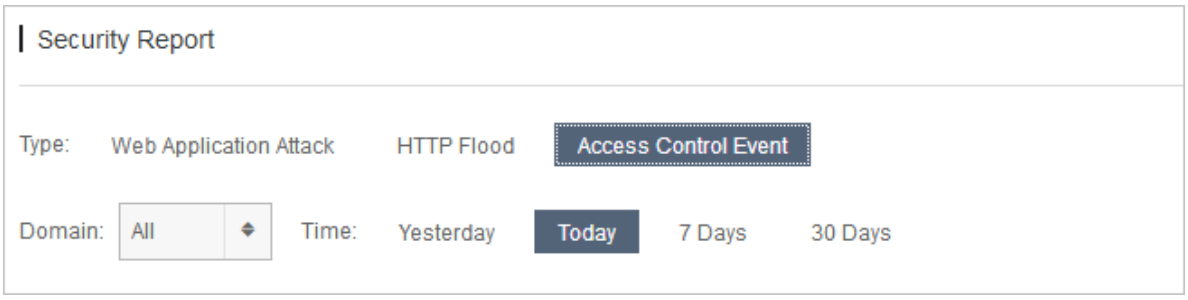
4. View the security report on HTTP flood attacks, as shown in *Figure 27-24: Security report on HTTP flood attacks*.

Figure 27-24: Security report on HTTP flood attacks



5. View the security report on access control events, as shown in *Figure 27-25: Security report on access control events*.

Figure 27-25: Security report on access control events



27.6.4.3 View business analysis results

This topic describes how to view business analysis results.

Prerequisites

Make sure that MaxCompute has been deployed in Apsara Stack.

The business analysis function depends on MaxCompute (formerly known as ODPS) for data analysis. The business analysis function cannot be used without MaxCompute.

Context

Based on the attack blocking information and access traffic information provided by WAF, the business analysis function uses the big data engine to analyze business access information for

protected domain names. This helps the security administrator detect vulnerabilities in a timely manner and improves the defense capabilities.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Application Security** > **Business Analysis**.
3. Select a domain name and time period to view the business analysis results.

27.7 Server security

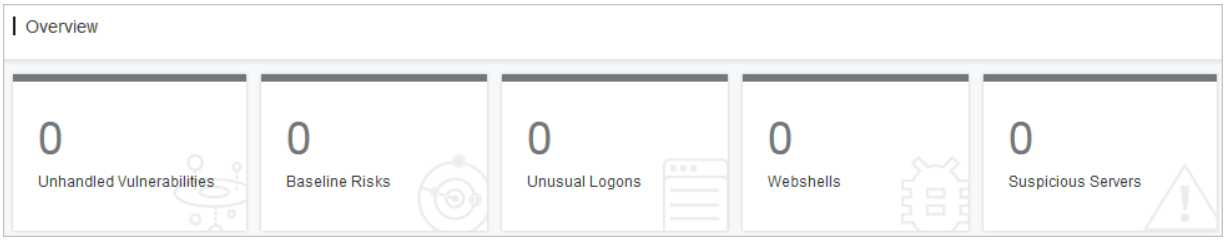
27.7.1 Server security overview

The security administrator can view the current security status of all servers on the server security overview page of Apsara Stack Security Center.

The server security overview page contains the following areas: Overview, Flaws, Events, Agent Status, and Key Flaws and Events.

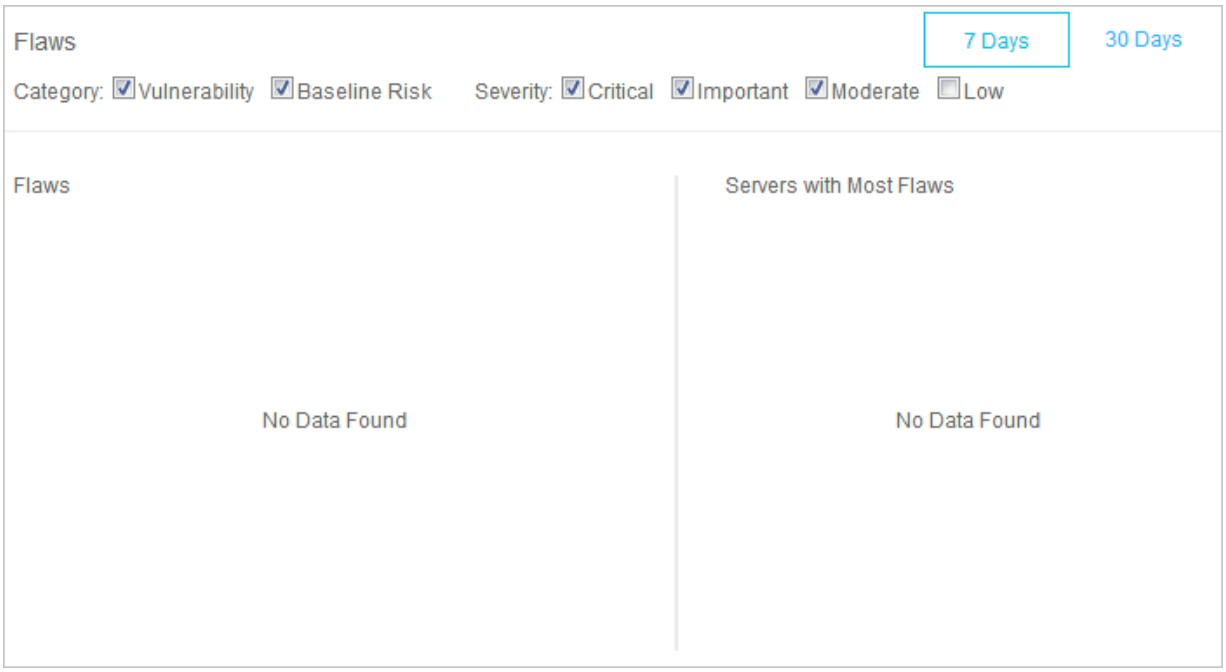
Overview area

This area displays the number of security flaws of each type, such as unhandled vulnerabilities and baseline risks, and the number of security events of each type, such as unusual logons and suspicious servers.



Flaws area

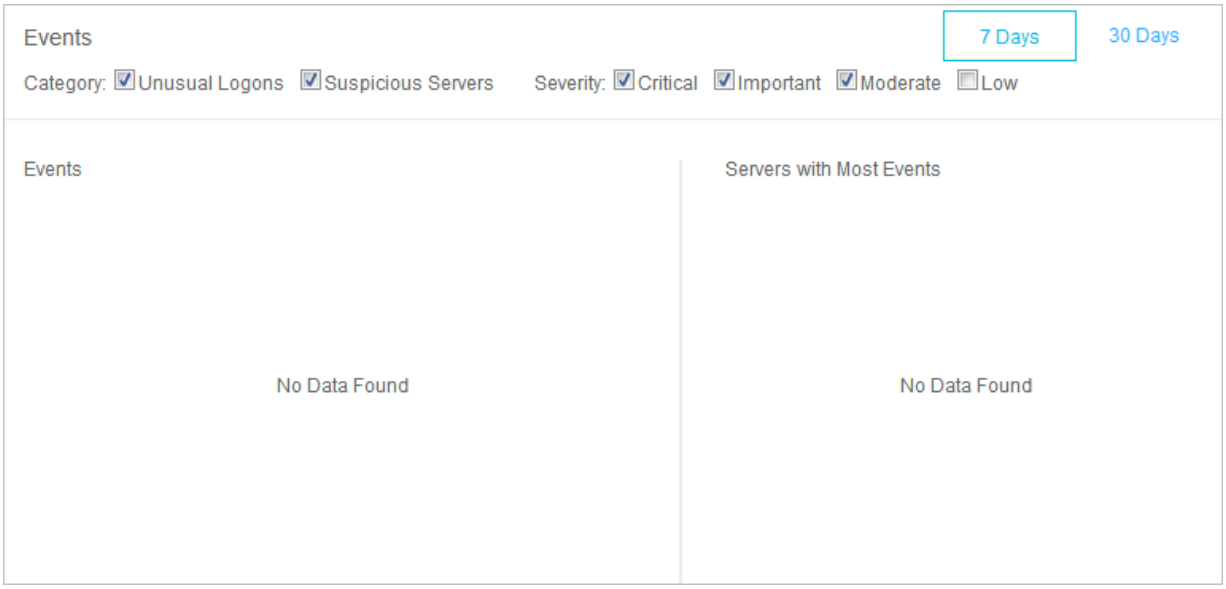
This area displays the trend of security flaws on servers. Security flaws may cause risks if they are not fixed.



- The filter criteria are displayed at the top of this area. You can filter security flaws by type, severity, and time period.
- The server flaws in a specific period is displayed on the left.
- The servers with the most flaws are displayed on the right.

Events area

This area displays the trend of security events on servers. Security events are security intrusions that have been detected on servers.

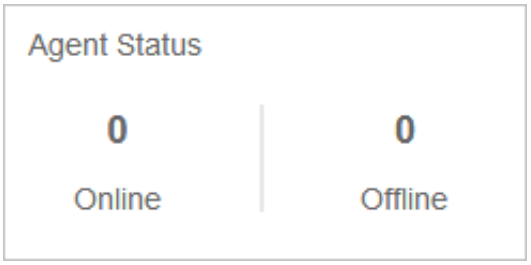


- The filter criteria are displayed at the top of this area. You can filter security events by type, severity, and time period.

- The intrusions detected on servers in a specific period are displayed on the left.
- The servers that suffered from the most intrusions are displayed on the right.

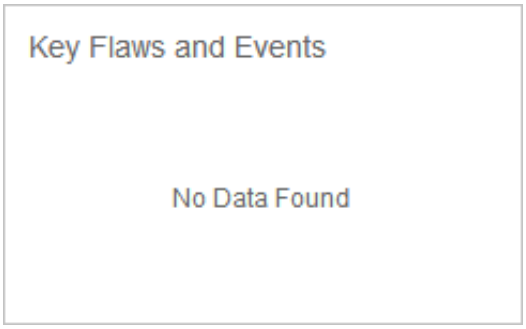
Agent Status area

This area displays the number of servers being protected and the number of offline servers.



Key Flaws and Events area

This area displays the recent key server flaws and events on servers. You can click a flaw or event to view the details.



27.7.2 Server list

27.7.2.1 Manage the server list

On the Servers page, you can view the status of servers protected by Server Guard.

Context

The protection status of a server can be:

- **Online**: Server Guard provides complete security protection for this server.
- **Offline**: Server Guard cannot provide security protection for this server because it cannot connect to the Server Guard agent on this server.
- **Disable Protection**: Security protection is temporarily disabled for this server. For more information, see *Disable protection*.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Servers**.
3. Optional: Search for a server.

   To view the protection status of a server, enter the server IP address in the search box and click **Search**. The detailed server information, including security information, is displayed.
4. View the protection status and detailed security information of the server.

   Click  in the upper-right corner of the page to select the information columns to be displayed. The following table lists the information categories.

Category	Information
Basic information	• Server IP/name • Tag • Operating system • Region
Agent status	Agent status
Security protection	• Vulnerabilities • Baseline risks
Intrusion detection	• Unusual logons • Webshells • Suspicious servers
Server fingerprints	• Number of processes • Number of ports • Root account or all accounts

5. Manage servers.

Function	Actions
Change Group	Select servers and click **Change Group** to assign a new group to the selected servers. For more information, see *Manage groups*.
Modify Tag	Select servers and click **Modify Tag** to modify tags for the servers.
Security Inspection	Select servers and click **Security Inspection** to check the security of the servers in multiple dimensions.
Delete External Servers	Select an **External** server and choose **More** > **Delete External Servers**.

Function	Actions
Disable Protection	Select servers in the **Online** status and choose **More** > **Disable Protection**to temporarily disable security protection for the servers. This action reduces the resource consumption of the servers.
Enable Protection	Select servers in the **Disable Protection** status and choose **More** > **Enable Protection**to enable security protection for the servers.

27.7.2.2 Manage groups

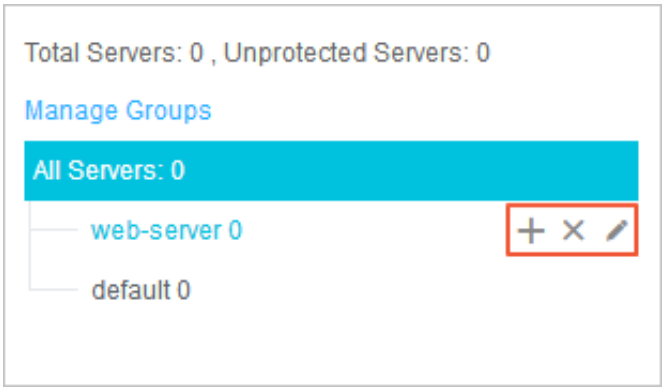
To facilitate the security control of servers, you can add servers to different groups and view security events by group.

Context

Servers that have not been added to any group are assigned to the **default** group. If you delete a group, all servers in the group are moved to the **default** group automatically.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Servers**.
3. Manage groups.



- Create a group.

  Click the Add Subgroup icon next to **All Servers** or a specific group, enter the name of the group to be added, and click **OK**.

  > **Note:**
  > The system supports three levels of groups.

- Modify a group.

Click the Modify Group Name icon next to the target group, enter a new name, and click **OK**.

- Delete a group.

  Click the Delete icon next to the target group. In the message that appears, click **OK**.

  > **Note:**
  > After you delete the group, servers in the group are moved to the **default** group automatically.

4. Group servers.
   a) Select servers from the list on the right.
   b) Click **Change Group**.
   c) In the Change Group dialog box, select a group from the drop-down list.
   d) Click **OK**.
5. Sort groups.

   Click **Manage Groups** to sort groups in descending order by priority.

27.7.3 Threat protection

27.7.3.1 Vulnerability management

27.7.3.1.1 Manage Linux software vulnerabilities

Apsara Stack Security scans the software that has been installed on your servers against the Common Vulnerabilities and Exposures (CVE) list to discover matching vulnerabilities in your software and send alerts. Apsara Stack Security also provides commands for you to fix vulnerabilities that have been detected and allows you to verify these vulnerability fixes.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Threat Prevention** > **Vulnerabilities** and click the **Linux Software Vulnerabilities** tab.
3. View all vulnerabilities.

   You can quickly locate a vulnerability by using the search and filter functions.
4. Click a vulnerability name to go to the vulnerability details page. You can view detailed vulnerability information and affected assets on this page.

You can quickly locate specific affected assets by using the search and filter functions.

5. Select an action according to the impact of the vulnerability. *Table 27-16: Actions on vulnerabilities* describes the actions.

Table 27-16: Actions on vulnerabilities

Action	Description
Generate Fix Command	Select this option to generate the commands for fixing the vulnerability. You can then log on to the server to run these commands.
Fix Now	Select this option to fix the vulnerability directly.
Restarted and Verified	If a vulnerability fix requires a server reboot to take effect, reboot the server only after the status of the vulnerability changes to **Fixed (To Be Restarted)**. After the reboot, click **Restarted and Verified**.
Ignore	Select this option to ignore the vulnerability. The system will no longer alert you about this vulnerability.
Verify	Select this option to **verify** the vulnerability fix. If you do not perform a manual verification, the system automatically verifies the fix 48 hours after the fix is applied.

You can fix the vulnerability for a single affected asset or multiple affected assets at one time.

- For a single asset: Select an action from the **Actions** column of an affected asset to fix the vulnerability.
- For multiple assets: Select one or more affected assets, and select an action in the lower-left corner to fix the vulnerability.

27.7.3.1.2 Manage Windows vulnerabilities

Apsara Stack Security automatically checks if your servers have the latest Microsoft updates installed, and notifies you of any detected vulnerabilities. Apsara Stack Security can also automatically detect and fix major vulnerabilities on your servers.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Servers** > **Threat Prevention** > **Vulnerabilities** and click the **Windows Vulnerabilities** tab.
3. View all vulnerabilities.

   You can quickly locate a vulnerability by using the search and filter functions.

4. Click a vulnerability to go to the vulnerability details page. You can view detailed vulnerability information and affected assets on this page.

   You can quickly locate specific affected assets by using the search and filter functions.

5. Select an action according to the impact of the vulnerability. *Table 27-17: Actions on vulnerabilities* describes the actions.

   Table 27-17: Actions on vulnerabilities

Action	Description
Fix Now	Select this option to fix the vulnerability directly. The system will cache an official Windows patch in the cloud for your server to download and update.
Ignore	Select this option to ignore the vulnerability. The system will no longer alert you about this vulnerability.
Verify	Select this option to **verify** the vulnerability fix.
Restarted and Verified	If a vulnerability fix requires a server reboot to take effect, reboot the server only after the status of the vulnerability changes to **Fixed (To Be Restarted)**. After the reboot, click **Restarted and Verified**.

   You can fix the vulnerability for a single affected asset or multiple affected assets at one time.

   - For a single asset: Select an action from the **Actions** column of an affected asset to fix the vulnerability.
   - For multiple assets: Select one or more affected servers, and select an action in the lower-left corner to fix the vulnerability.

27.7.3.1.3 Manage WCMS vulnerabilities

Apsara Stack Security obtains the latest Web content management system (WCMS) vulnerability alerts and patches, issues patches, and fixes the vulnerabilities in a timely manner.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Servers** > **Threat Prevention** > **Vulnerabilities** and click the **Web CMS Vulnerabilities** tab.
3. View all vulnerabilities.

   You can quickly locate a vulnerability by using the search and filter functions.

4. Click a vulnerability to go to the vulnerability details page. You can view detailed vulnerability information and affected assets on this page.

   You can quickly locate specific affected assets by using the search and filter functions.

5. Select an action according to the impact of the vulnerability. *Table 27-18: Actions on vulnerabilities* describes the actions.

   Table 27-18: Actions on vulnerabilities

Action	Description
Fix Now	Select this option to fix the WCMS vulnerability by replacing the Web files that contain the vulnerability on your server. **Note:** Before fixing the vulnerability, we recommend that you back up the Web files affected by the vulnerability. For more information about the paths of the Web files, see the path that is specified in the remarks.
Ignore	Select this option to ignore the vulnerability. The system will no longer alert you about the vulnerability.
Verify	Select this option to **verify** the vulnerability fix. If you do not perform a manual verification, the system automatically verifies the fix 48 hours after the fix is applied.
Roll Back	For vulnerabilities that have been fixed, click **Roll Back** to restore the Web files that have been replaced.

   You can fix the vulnerability for a single affected asset or multiple affected assets at one time.

   - For a single asset: Select an action from the **Actions** column of an affected server to fix the vulnerability.
   - For multiple assets: Select one or more affected servers, and select an action in the lower-left corner to fix the vulnerability.

27.7.3.1.4 Manage other vulnerabilities

Apsara Stack Security automatically detects vulnerabilities on servers, such as the Redis unauthorized access vulnerability and Struts S2-052 vulnerability, and sends vulnerability alerts. After you fix a vulnerability, you can also verify that your fix is successful.

Procedure

1. *Log on to Apsara Stack Security Center*.

2. Choose **Server Security** > **Threat Prevention** > **Vulnerabilities** and click the **Others** tab.
3. View all vulnerabilities.

   You can quickly locate specific vulnerabilities by using the search and filter functions.
4. Click a vulnerability to go to the vulnerability details page. You can view detailed vulnerability information and affected assets on this page.

   You can quickly locate specific affected assets by using the search and filter functions.
5. Select an action according to the impact of the vulnerability. *Table 27-19: Actions on vulnerabilities* describes the actions.

   Follow the instructions to manually fix the vulnerability that is displayed on the **Others** tab page.

   Table 27-19: Actions on vulnerabilities

Action	Description
Ignore	Select this option to ignore the vulnerability. The system will no longer alert you about this vulnerability.
Verify	Select this option to **verify** the fix after you have manually fixed the vulnerability. If you do not perform a manual verification, the system automatically verifies the fix 48 hours after the fix is applied.

   You can fix the vulnerability for a single affected asset or multiple affected assets at one time.

   - For a single asset: Select an action from the **Actions** column of an affected server.
   - For multiple assets: Select one or more affected servers, and select an action in the lower-left corner.

27.7.3.1.5 Configure vulnerability management

You can enable or disable automatic detection for different types of vulnerabilities, and enable vulnerability detection for specific servers. You can also set a time duration for which invalid vulnerabilities are retained, and set a vulnerability whitelist.

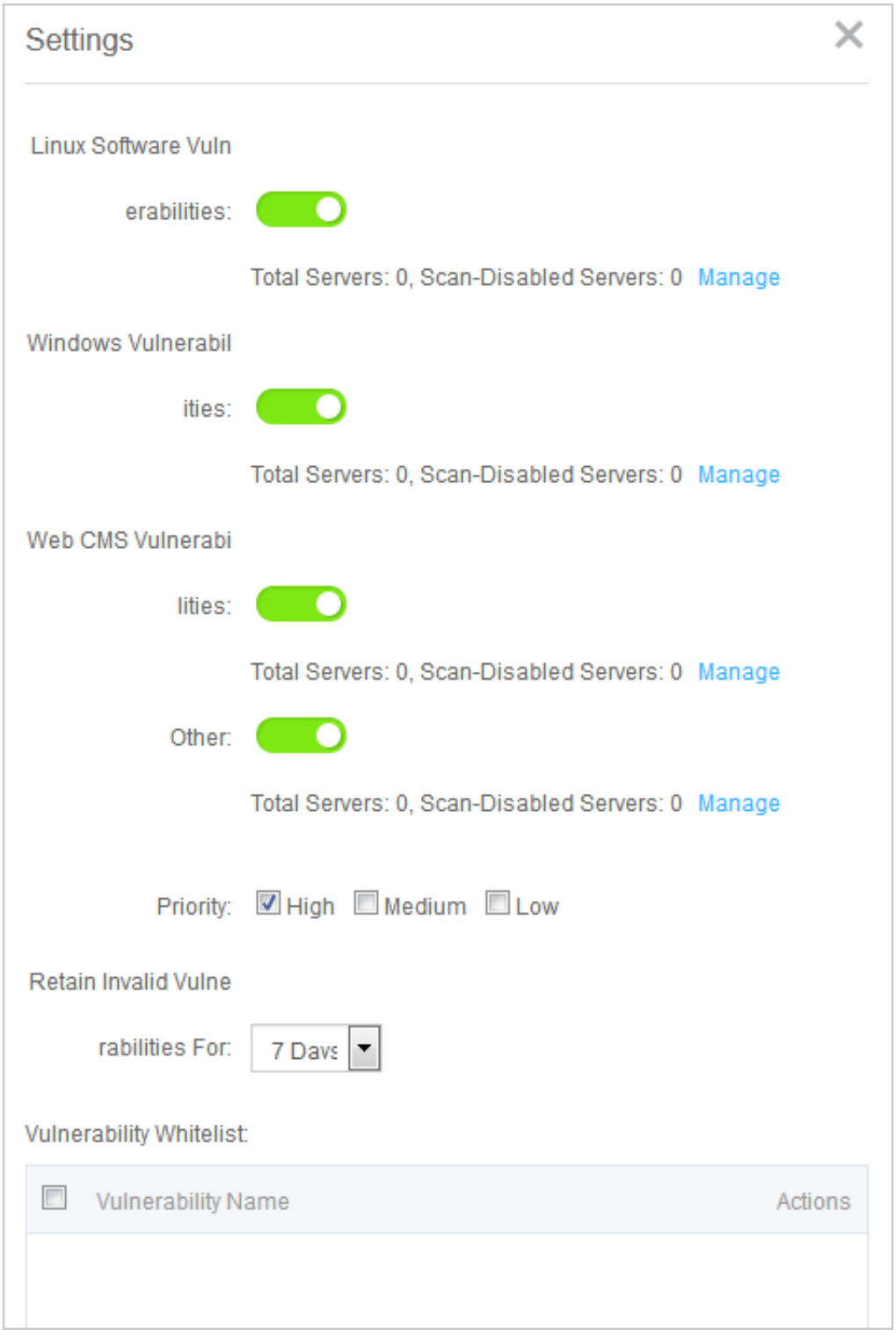
Context

A vulnerability whitelist allows you to completely exclude vulnerabilities from the detection list. You can add multiple vulnerabilities in the vulnerability list to the whitelist. After the vulnerabilities have been added to the whitelist, the system no longer detects these vulnerabilities. You can also maintain the vulnerability whitelist.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Threat Prevention** > **Vulnerabilities**.
3. Click **Settings** in the upper-right corner to configure vulnerability management, as shown in *Figure 27-26: Settings dialog box*.

Figure 27-26: Settings dialog box



- Select a vulnerability type, and enable or disable detection for vulnerabilities of this type.
- Click **Manage** next to a vulnerability type and specify the servers on which vulnerabilities of this type are detected.

- Select priorities from High, Medium, and Low.
- Select a time duration for which invalid vulnerabilities are retained: 7 days, 30 days, or 90 days.

> **Note:**
> If you do not take any action on a detected vulnerability, the system determines that the alert settings for the vulnerability are invalid. The system removes the vulnerability after the specified duration.

- Select vulnerabilities in the whitelist, and click **Remove** to enable the system to detect these vulnerabilities and send alerts again.

27.7.3.2 Baseline check

27.7.3.2.1 Baseline check overview

The baseline check function of Server Guard automatically detects system, database, and account configuration risks on servers and provides fix solutions for the detected issues.

How baseline check works

The baseline check function automatically detects system, permission, account, and database configuration risks on servers and provides fix solutions.

Check interval

By default, a complete check is automatically performed between 00:00 AM and 06:00 AM every 3 days. You can set the check interval and time on the security settings page.

Notes

When checking some items, such as weak passwords in MySQL and SQL Server services, Server Guard may use server resources for logon attempts and generate some logon failure records. By default, baseline check is disabled for such check items. You must be aware of these risks before enabling baseline check for such check items on the baseline check settings page.

Check items

Table 27-20: Check items

Category	Check item
CIS baseline	CentOS 7 baseline
	Tomcat 7 baseline

Category	Check item
Weak password	PostgreSQL weak password
	SSH weak password
	Anonymous FTP logon
	SQL Server logon password
	MySQL weak password
	RDP weak password
	FTP weak password
System	Group policy
	File monitoring
	Baseline policy
	Registry
Account	System account security
Database	Redis compliance

27.7.3.2.2 Configure baseline check

This topic describes how to use the baseline check function to check and fix configuration risks on servers.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Threat Prevention** > **Baseline Check** to go to the Baseline Check page.
3. View all configuration risks.

   You can locate a risk quickly by using the search and filter functions.
4. Click a risk to view the risk details and related fix solutions.

   If the risk has no impact on your servers, and does not need to be checked in the future, you can click **Add to Whitelist** in the upper-right corner.
5. Fix the risk on the affected servers based on the solutions.

   - After you have fixed the risk for a server, click **Verify** to verify that the fix is successful. If you do not perform a manual verification, Server Guard automatically verifies the fix 72 hours after the fix is applied.

- If no fix is required for a server, click **Ignore** to ignore the risk, so that Server Guard will stop reporting and sending alerts about this risk on the server.

27.7.3.2.3 Set a baseline check policy

This topic describes how to set baseline check items, check interval, and risk level based on the actual business situations.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Threat Prevention** > **Baseline Check** to go to the Baseline Check page.
3. Click **Settings**. You can create or modify a policy.
4. Click **Create** to create a policy.

   Select a policy and click **Modify** to modify this policy or click **Delete** to delete this policy.
5. Manage the whitelist.

   If you need to recheck a baseline item in the whitelist, select the item in **Baseline Risk Whitelist** and click **Remove**.

27.7.4 Intrusion detection

27.7.4.1 Unusual logons

27.7.4.1.1 How unusual logon detection works

On the **Unusual Logons** page of the Server Guard console, you can view the IP address, account name, and time of each unusual logon. You can also view the alerts for unusual logons, disapproved IP addresses, disapproved logon time, and disapproved accounts.

The Server Guard agent regularly collects logon logs of your server, and uploads them to the Server Guard server where the logs are analyzed and matched. An alert is reported when Server Guard detects a successful logon from a disapproved location, using a disapproved IP address or account, or at a disapproved time.

Note:

To enable SMS notification, choose **System Settings** > **Alert Settings**, and then choose **Logon Security** > **Unusual Logons** to set your preferred notification methods. Value options include mobile number and email. By default, both methods are selected.

You can also set approved logon IP addresses, logon time period, and accounts for specific servers. All logon attempts, except for those using the approved logon IP addresses and accounts during the approved logon time period, will trigger alerts. These logon security settings have a higher priority than the unusual logon alert policy.

27.7.4.1.2 Check unusual logon alerts

This topic describes how to check the alerts for unusual logons, including logons from disapproved locations, brute-force attacks, logons using disapproved IP addresses, logons using disapproved accounts, and logons at a disapproved time.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Intrusion Detection** > **Unusual Logons**.
3. Check all unusual logon alerts.

   You can quickly locate a specific unusual logon alert by using the search and filter functions.
4. Handle unusual logon alerts.

   Select an unusual logon alert to check whether it is a false positive.

   - If this alert is a false positive, click **Label as Handled**.
   - If the logon is an intrusion, improve the security of the related server. For example, use a more complex password, fix vulnerabilities on the server, remove risks that are detected during baseline check, or specify a blacklist and whitelist. Then, click **Label as Handled**.

27.7.4.1.3 Configure logon security

This topic describes how to configure logon security. You can set approved logon IP addresses, approved logon time periods, and approved accounts.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Intrusion Detection** > **Unusual Logons**.
3. In the upper-right corner of the **Unusual Logons** tab page, click **Logon Security**.
4. Set approved IP addresses.

   a) Click **Add** to add an approved CIDR block.

   b) Enter a CIDR block.

   c) Specify the servers on which the specified CIDR block takes effect.

- Click **All Servers** to select specific servers.
- Click **Server Groups** to select servers by group.

d) Click **OK**.

e) Select an approved CIDR block and click **Modify** to modify the settings.

f) Select an approved CIDR block and click **Delete** to delete the approved CIDR block.

5. Set approved logon time periods.
6. Set approved accounts.

27.7.4.2 Webshells

27.7.4.2.1 Manage webshell files

This topic describes how to view and quarantine webshell files.

Context

Server Guard inspects the files in the Web directory on your server to check whether any webshell files exist. If a webshell file is detected, an alert is triggered.

Server Guard inspects webshell files in PHP, JSP, or other common formats in real time or at scheduled time locally or in the cloud. Server Guard also provides the function of quickly quarantining the detected webshell files.

Server Guard uses dynamic inspection or scheduled inspection to inspect webshell files.

- **Dynamic inspection**: When any file in the Web directory is modified, Server Guard inspects the modified content instantly.
- **Scheduled inspection**: Server Guard scans the entire Web directory every early morning.

> **Note:**
> By default, scheduled inspection is enabled for all servers protected by Server Guard. You can also enable scheduled inspection for specific servers. Choose**Settings** > **Security Settings.**In the **Trojan Scan** area, click **Manage** next to **Web Directory Periodic Scan** and enable periodic trojan scan for the specified servers.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Intrusion Detection** > **Webshells**.
3. Specify a server and check the detected webshell files.

4. Process the webshell files.
   - **Quarantine**: Quarantine the file. You can select and quarantine multiple files at one time.
   - **Restore**: Click **Restore** to restore a file that has been quarantined by mistake.
   - **Ignore**: Ignore the file. Server Guard will no longer generate alerts for this file.

   Note:
   Server Guard does not delete webshell files on your server. It simply quarantines the files. You can restore a quarantined file if you determine that the file can be trusted. After a webshell file is restored, Server Guard will no longer generate alerts for the file.

27.7.4.3 Suspicious servers

27.7.4.3.1 Manage server exceptions

This topic describes how to view alerts for server exceptions and handle the exceptions.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Intrusion Detection** > **Suspicious Servers**.
3. Select a server to view the detected exceptions.
4. Select an action to handle each exception based on its impact. *Table 27-21: Handle server exceptions* describes the actions.

Table 27-21: Handle server exceptions

Action	Description
Fix	Select this option to fix the exception.
Ignore Once	Select this option to ignore the alert if the exception does not have any impact on the server security.
Confirm	Select this option to confirm the exception.
Label as False Positive	Select this option if the alert is a false positive.
View	Select this option to view the alert details.

27.7.5 Server fingerprints

27.7.5.1 Manage listening ports

You must regularly collect information about listening ports on a server.

Context

This task is applicable to the following scenarios:

- Check servers that listen to the specified port.
- Check enabled ports of a server.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Server Fingerprints** and click the **Listening Ports** tab.
3. View all enabled ports, their corresponding network protocols, and the number of servers on which these ports are enabled.

   You can search for a port by the port number or process name.
4. Click a port number to view the details, such as the corresponding asset and protocol.

27.7.5.2 Manage processes

You must regularly collect information about processes on a server.

Context

This task is applicable to the following scenarios:

- Check the servers that run the specified process.
- Check the processes running on a server.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Server Fingerprints** and click the **Processes** tab.
3. View all running processes and the number of servers that run these processes.

   You can search for a process by process name or user.
4. Click a process name to view the details, such as the corresponding assets, path, and startup parameters.

27.7.5.3 Manage account information

You must regularly collect account information on a server.

Context

This task is applicable to the following scenarios:

- Check the servers where the specified account is created.
- Check the accounts created on a server.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Server Fingerprints** and click the **Accounts** tab.
3. View all accounts that have logged on and the number of servers that use these accounts.

   You can search for an account by account name.
4. Click an account name to view the details, such as the corresponding assets, root permissions, and user group.

27.7.5.4 Manage software versions

You must regularly collect software version information of a server to check the software assets.

Context

This task is applicable to the following scenarios:

- Check unauthorized software assets, which are software that has been installed without authorization.
- Check software assets that are outdated.
- Quickly locate the affected assets when vulnerabilities occur.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Server Fingerprints** and click the **Software** tab.
3. View all software in use and the number of servers that use such software.

   You can search by software name, version, or installation directory.
4. Click a software name to view the corresponding assets, software version, and other information.

27.7.5.5 Set the server fingerprint refresh frequency

You can set the frequency at which data of running processes, system accounts, enabled ports, and software versions is collected and refreshed.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Server Fingerprints** and click **Settings**.
3. Select the refresh frequency from the corresponding drop-down list.
4. Click **OK**.

27.7.6 Log retrieval

27.7.6.1 Log retrieval overview

The log retrieval function provided by Server Security allows you to manage logs scattered in various systems of Apsara Stack in a centralized manner, so that you can easily identify the causes of issues that occur on your servers.

The log retrieval function supports storage of logs for 180 days and query of logs generated within 30 days.

Benefits

The log retrieval function provides the following benefits:

- **End-to-end log retrieval platform**: Allows you to retrieve logs of various Apsara Stack services in a centralized manner and trace issues easily.
- **Cloud-based SaaS service**: Allows you to query logs on all servers in Apsara Stack without additional installment and deployment.
- Supports TB-level data retrieval. It also allows you to add a maximum of 50 inference rules (Boolean expressions) in a search condition and obtain full-text search results within several seconds.
- Supports a wide range of log sources.
- Supports log shipping, which allows you to import security logs to Log Service for further analysis.

Scenarios

You can use log retrieval to meet the following requirements:

- **Security event analysis**: When a security event is detected on a server, you can retrieve the logs to identify the cause and assess the damage and affected assets.
- **Operation audit**: You can audit the operation logs on a server to identify high-risk operations and serious issues in a meticulous way.

Supported log types

Table 27-22: Log types

Log type	Description
Logon history	Log entries about successful system logons
Brute-force attack	Log entries about system logon failures that are generated during brute-force attacks
Process snapshot	Log entries about processes on a server at a specific time
Listening port snapshot	Log entries about listening ports on a server at a specific time
Account snapshot	Log entries about account logon information on a server at a specific time
Process initiation	Log entries about process initiation on a server
Network connection	Log entries about active connections from a server to external networks

27.7.6.2 Search for logs

This topic describes how to search for and view server logs.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Log Retrieval**.
3. Set search conditions.

Table 27-23: Search condition parameters

Parameter	Description
Log source	Select a supported log source. For more information, see *Table 27-24: Log sources*.

Parameter	Description
Field	Select a field that is supported by the specified log source. For more information, see *Table 27-25: Supported fields in logon logs*.
Keyword	Enter the keyword of the field to be searched for.
Logical operator	Select a logical operator from value options: AND, OR, and NOT. For more information, see *Table 27-32: Logical operators*.
+	Add inference rules in a search condition for a log source.
Add conditions	Add search conditions for different log sources.

4. Click **Search** and view the search result.

   - **Reset**: Click **Reset** to clear the search condition configuration.
   - **Save Search**: Click **Save Search** to save the search condition configuration for future reuse.
   - **Saved Searches**: Click **Saved Searches** to select and apply a search condition configuration that has been previously saved.

27.7.6.3 Supported log sources and fields

This topic describes the types of logs and fields that are supported by the log retrieval function.

The log retrieval function allows you to query the following types of logs. You can click a log source to view the fields that can be retrieved.

Table 27-24: Log sources

Log source	Description
Logon log	Log entries about successful system logons.
Brute-force attack log	Log entries about system logon failures that are generated during brute-force attacks.
Process snapshot log	Log entries about processes on a server at a specific time.
Listening port snapshot log	Log entries about listening ports on a server at a specific time.
Account snapshot log	Log entries about account logon information on a server at a specific time.
Process initiation log	Log entries about process initiation on a server.

Log source	Description
Network connection log	Log entries about active connections from a server to external networks.

Logon log

The following fields are supported in logon log queries:

Table 27-25: Supported fields in logon logs

Field	Data type	Description
uuid	string	The agent ID.
IP	string	The server IP address.
warn_ip	string	The source IP address for the logon.
warn_port	string	The logon port.
warn_user	string	The logon username.
warn_type	string	The logon type.
warn_count	string	The number of logon attempts.
time	datetime	The logon time.

Brute-force attack log

The following fields are supported in brute-force attack log queries:

Table 27-26: Supported fields in brute-force attack logs

Field	Data type	Description
uuid	string	The agent ID.
IP	string	The server IP address.
warn_ip	string	The attacker IP address.
warn_port	string	The target port number.
warn_user	string	The target username.
warn_type	string	The type.
warn_count	string	The number of brute-force attack attempts.

Field	Data type	Description
time	datetime	The attack time.

Process initiation log

The following fields are supported in process initiation log queries.

Table 27-27: Supported fields in process initiation logs

Field	Data type	Description
uuid	string	The agent ID.
IP	string	The server IP address.
pid	string	The process ID.
groupname	string	The name of the user group.
ppid	string	The parent process ID.
uid	string	The user ID.
username	string	The username.
filename	string	The file name.
pfilename	string	The file name of the parent process.
cmdline	string	The command line.
filepath	string	The process path.
pfilepath	string	The parent process path.
time	datetime	The time when the process was started.

Listening port snapshot log

The following fields are supported in listening port snapshot log queries:

Table 27-28: Supported fields in listening port snapshot logs

Field	Data type	Description
uuid	string	The agent ID.
IP address	string	The server IP address.
src_port	string	The listening port.

Field	Data type	Description
src_ip	string	The listening IP address.
proc_path	string	The process path.
PID	string	The process ID.
proc_name	string	The process name.
proto	string	The protocol.
time	datetime	The time when data was collected.

Account snapshot log

The following fields are supported in account snapshot log queries:

Table 27-29: Supported fields in account snapshot logs

Field	Data type	Description
uuid	string	The agent ID.
IP address	string	The server IP address.
perm	string	Indicates whether the client has root permissions.
home_dir	string	The home directory.
warn_time	string	The password expiration notification time.
groups	string	The group to which the user belongs.
login_ip	string	The IP address of the last logon.
last_chg	string	The last time when the password was changed.
shell	string	The Linux shell command.
domain	string	The Windows domain.
tty	string	The logon terminal.
account_expire	string	The account expiration time.
passwd_expire	string	The password expiration time.
last_logon	string	The last logon time.

Field	Data type	Description
user	string	The user.
status	string	The user status. Value options include: • 0: disabled. • 1: normal.
time	datetime	The time when data was collected.

Process snapshot log

The following fields are supported in process snapshot log queries.

Table 27-30: Supported fields in process snapshot logs

Field	Data type	Description
uuid	string	The agent ID.
IP address	string	The server IP address.
path	string	The process path.
start_time	string	The time when the process was started.
uid	string	The user ID.
cmdline	string	The command line.
pname	string	The parent process name.
name	string	The process name.
pid	string	The process ID.
user	string	The username.
md5	string	The MD5 value of the process file. This value is not calculated if the file size exceeds 1 MB.
time	datetime	The time when data was collected.

Network connection log

The following fields are supported in network connection log queries.

Table 27-31: Supported fields in network connection logs

Field	Data type	Description
uuid	string	The agent ID.
IP	string	The server IP address.
src_ip	string	The source IP address.
src_port	string	The source port.
proc_path	string	The process path.
dst_port	string	The destination port.
proc_name	string	The process name.
dst_ip	string	The destination IP address.
status	string	The status.
proto	string	The protocol.
time	datetime	The connection time.

27.7.6.4 Inference rules and logical operators

The log retrieval function supports multiple search conditions. You can add multiple inference rules in one search condition for one log source, or combine multiple conditions for several log sources by using different logical operators. This topic describes the inference rules and logical operators that are supported in log queries. Some examples are provided to help you understand them.

The following table describes the logical operators that are supported in log queries.

Table 27-32: Logical operators

Logical operator	Description
AND	Binary operator. This operator is in the format of `query1 and query2`, indicating the intersection of the query results of `query1` and `query2`. **Note:** If no logical operator is specified for multiple keywords, the default operator is AND.
OR	Binary operator.

Logical operator	Description
	This operator is in the format of `query1 or query2`, indicating the combination of the query results of `query1` and `query2`.
NOT	Binary operator. This operator is in the format of `query1 not query2`, indicating the results that match `query1` but not `query2`, which is equivalent to `query1 - query2`. **Note:** If only `not query1` is specified, the query returns all records that do not match `query1`.

27.7.7 Settings

27.7.7.1 Manage security settings

This topic describes how to manage the security settings of servers. You can enable or disable periodic trojan scan and set the resource usage mode of the Server Guard agent for servers.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Settings**.
3. Enable periodic trojan scan for servers.
   a) Click **Manage**.
   b) Select the servers that require periodic trojan scan.
   c) Click **OK**.
4. Specify the resource usage mode of the Server Guard agent for servers.
   - **Business First Mode**: The peak CPU usage is less than 10% and the peak memory usage is less than 50 MB.
   - **Protection First Mode**: The peak CPU usage is less than 20% and the peak memory usage is less than 80 MB.

   a) Click **Manage**.
   b) Specify the resource usage mode of the Server Guard agent for servers.
   c) Click **OK**.

27.7.7.2 Install the Server Guard agent

This topic describes how to manually install the Server Guard agent on a Windows or Linux server.

Prerequisites

If you have installed security software on your server, the system may fail to install the Server Guard agent correctly. We recommend that you disable or uninstall security software, if any, before you install the Server Guard agent.

Context

The Server Guard agent has been integrated in public images. If you select the public image when you create an ECS instance, the Server Guard agent is automatically integrated in the ECS instance.

For an external server that runs Windows, you must use the Server Guard agent installation package to install the agent. For an external server that runs Linux, you must run the relevant command to install the agent.

To ensure that the agent can run correctly in the following situations, you must delete the Server Guard agent directory and follow the preceding steps to manually reinstall the agent:

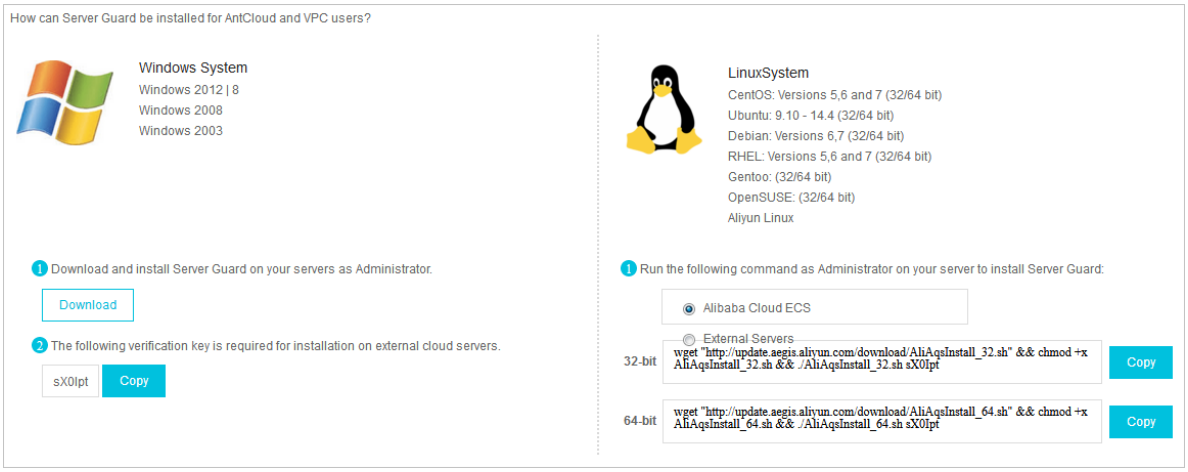
- Use an image that includes the Server Guard agent to install the agent on multiple external servers at one time.
- Directly copy the Server Guard agent files from a server that has been installed with the Server Guard agent to your external servers.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Settings** > **Install/Uninstall**.

   The Server Guard agent installation page appears, as shown in *Figure 27-27: Install the agent*.

Figure 27-27: Install the agent



3. Obtain and install the Server Guard agent based on the operating system type of your server.
   - **Windows operating system**
     a. On the left side of the page, click **Download** to download the installation package to your local PC.
     b. Upload the installation package to your server. For example, you can use an FTP client to upload the package to the server.
     c. Run the installation package on your server as an administrator.

     > **Note:**
     > When installing the agent on an external server, you will be prompted to enter the installation verification key. You can find the installation verification key on the Server Guard agent installation page.

   - **Linux operating system**
     a. On the right side of the page, select **`Alibaba Cloud ECS`** or **`External Servers`**.
     b. Select the installation command for 32-bit or 64-bit according to your operating system, and click **Copy** to copy the command.
     c. Log on to your Linux server as an administrator.
     d. Run the installation command on your Linux server to download and install the Server Guard agent.
4. View the agent status of your server.

You can view the agent status of your server in the Server Guard console 5 minutes after installing the Server Guard agent.

- If your server is an ECS instance, the status of the server changes from offline to online.
- If your server is an external server, the server is added to the server list.

27.7.7.3 Uninstall the Server Guard agent from a server

If you decide not to use any of the Server Guard functions on your server, you can use the following procedure to uninstall the Server Guard agent.

Context

When uninstalling the Server Guard agent from a specific server in the console, make sure that the agent status of the server is online. If the status is offline, the server cannot receive the command for uninstalling the agent.

If you need to reinstall the Server Guard agent within 24 hours (the protection period) after the uninstallation, install it manually and ignore the error messages. You must repeat the install operation at least three times before it can be successfully reinstalled.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Server Security** > **Settings** > **Install/Uninstall**.
3. Click **Uninstall** in the upper-right corner.
4. In the **Uninstall Server Guard** dialog box, select the server from which you want to uninstall the Server Guard agent.
5. Click **Uninstall**. Then, the system automatically uninstalls the Server Guard agent.

27.8 Physical machine security

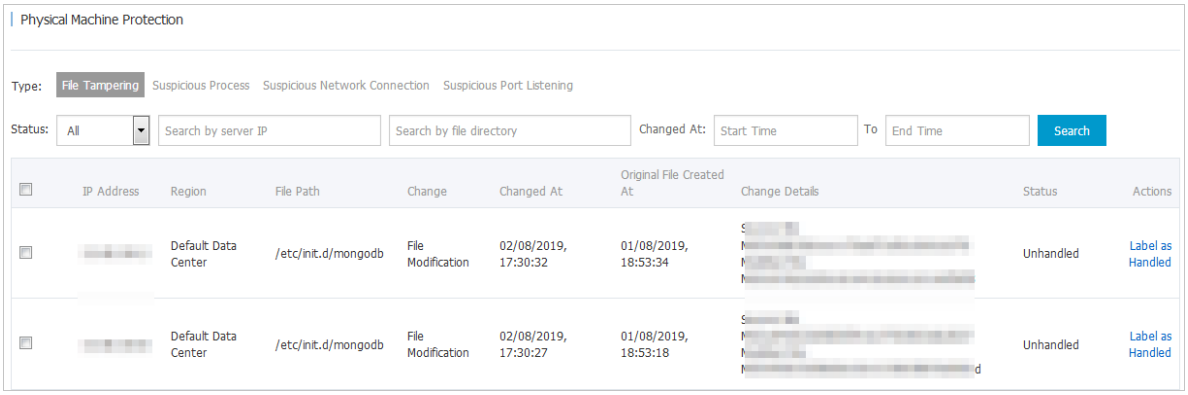
27.8.1 View and handle file tampering events

This topic describes how to check the integrity of files in specific directories of the server system, detect tampering events, and generate tampering alerts in a timely manner.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **PM Security** > **PM Protection** and click **File Tampering**.
3. View file tampering events, as shown in *Figure 27-28: File tampering events*.

Figure 27-28: File tampering events



4. Further troubleshoot a file tampering event.
   - If you confirm that the event is an intrusion, take security hardening measures immediately for the server, and further investigate and analyze the causes.
   - If you confirm that the event is normal or is an intrusion that has been handled, click **Label as Handled**. In the message that appears, click **Confirm** to change the event status to Handled.

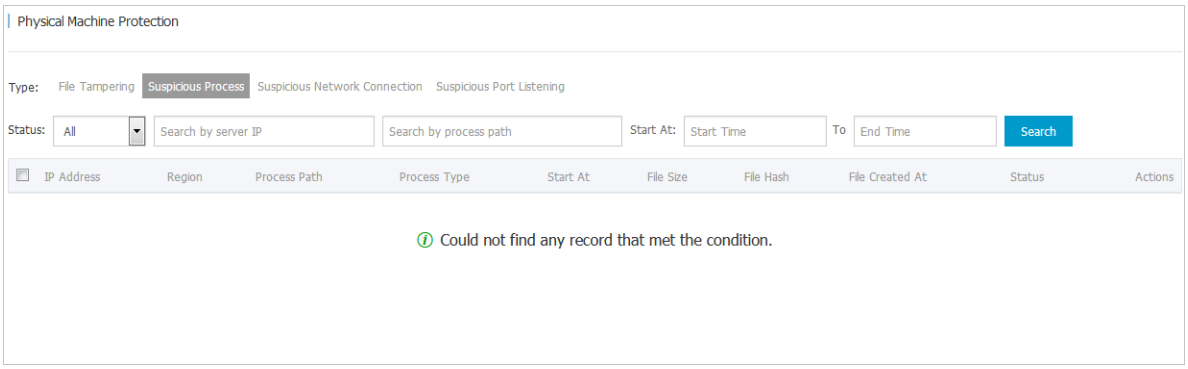
27.8.2 View and handle suspicious processes

This topic describes how to detect suspicious running processes in a timely manner and generate alerts accordingly.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **PM Security** > **PM Protection** and click **Suspicious Process**.
3. View suspicious processes, as shown in *Figure 27-29: Suspicious processes*.

Figure 27-29: Suspicious processes



4. Further troubleshoot a suspicious process.
   - If you confirm that the process is suspicious, take security hardening measures immediately for the server, and further investigate and analyze the causes.
   - If you confirm that the process is normal or is a suspicious process that has been handled, click **Label as Handled**. In the message that appears, click **Confirm** to change the status to Handled.

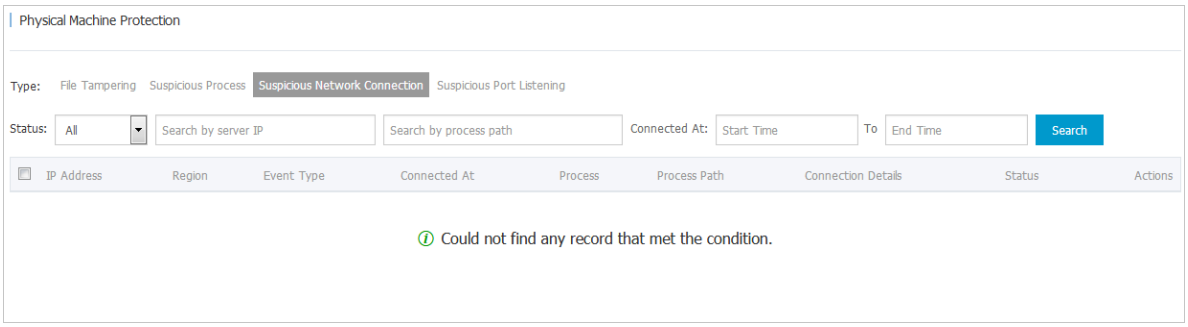
27.8.3 View and handle suspicious network connections

This topic describes how to detect active connections to external networks in a timely manner, and generate alerts accordingly.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **PM Security** > **PM Protection** and click **Suspicious Network Connection**.
3. View suspicious network connections, as shown in *Figure 27-30: Suspicious network connections*.

Figure 27-30: Suspicious network connections



4. Further troubleshoot a suspicious network connection.
   - If you confirm that the connection is suspicious, take security hardening measures immediately for the server, and further investigate and analyze the causes.
   - If you confirm that the connection is normal or is a suspicious connection that has been handled, click **Label as Handled**. In the message that appears, click **Confirm** to change the event status to Handled.

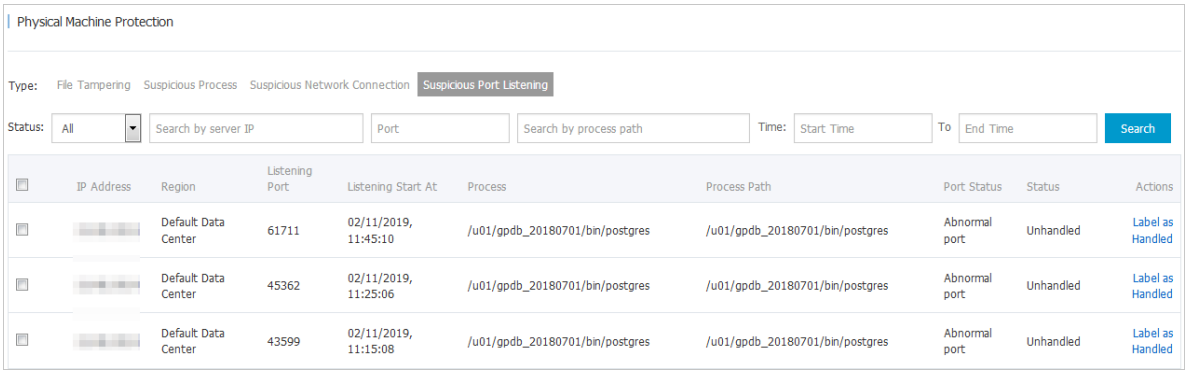
27.8.4 View and handle suspicious port listening events

This topic describes how to detect suspicious port listening events and generate alerts accordingly.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **PM Security** > **PM Protection** and click **Suspicious Port Listening**.
3. View suspicious port listening events, as shown in *Figure 27-31: Suspicious port listening events*.

   Figure 27-31: Suspicious port listening events



4. Further troubleshoot a suspicious port listening event.
   - If you confirm that the port listening event is suspicious, take security hardening measures immediately for the server, and further investigate and analyze the causes.
   - If you confirm that the port listening event is normal or is a suspicious event that has been handled, click **Label as Handled**. In the message that appears, click **Confirm** to change the event status to Handled.

27.9 Asset overview

27.9.1 Overview

Apsara Stack Security Center presents statistical information about your assets in charts, for example, your server assets and NAT assets, frequency of increase or decrease in the assets, and regional distribution. The security administrator can query asset information by group or type, so that they can better understand the general asset information for better asset management.

On the **Asset Overview** page, the security administrator can view the overall asset information in a direct and clear way, including the total number of assets, number of new assets in the current

month, number of groups, number of regions, and asset distributions by report time, group, and region. This helps users better manage their assets.

Figure 27-32: Asset Overview page

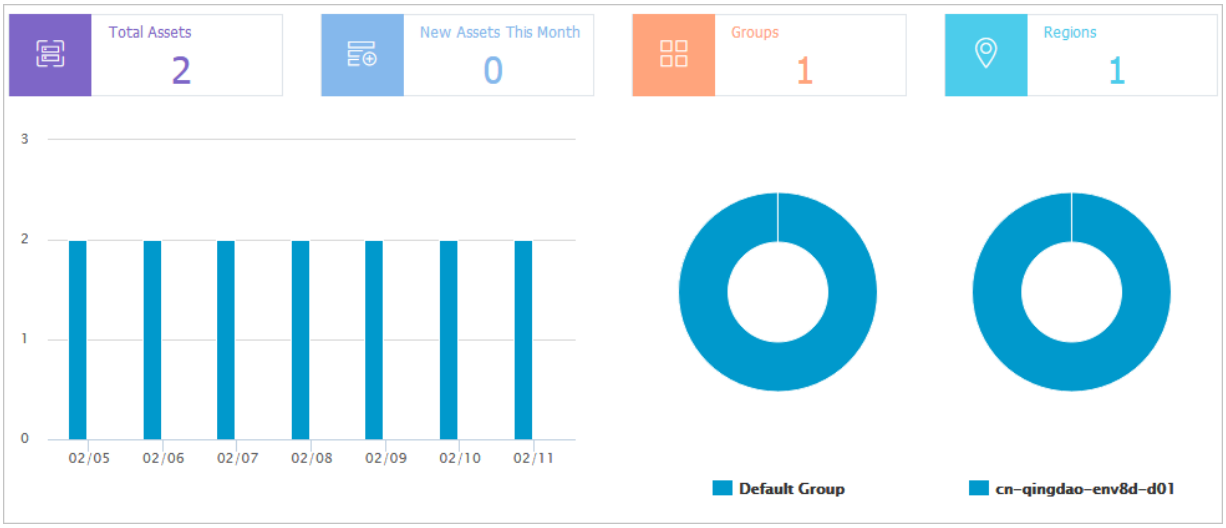


Table 27-33: Parameters on the Asset Overview page

Parameter	Description
Total Assets	The total number of assets reported by the Server Guard agent, including server assets and NAT assets.
New Assets This Month	The total number of new assets in this month, including server assets and NAT assets.
Asset Distribution by Report Time	The change in the number of server assets and that of NAT assets over the last 7 days.
Groups	The number of existing groups.
Asset Distribution by Group	The pie chart that shows the proportion of assets in each group to the total assets.
Regions	The number of configured regions.
Asset Distribution by Region	The pie chart that shows the proportion of assets in each region to the total assets.

27.9.2 Manage groups

27.9.2.1 Add a group

This topic describes how to add a group. Asset groups are used to classify assets, allowing you to query and modify asset information more easily.

Context

Note:
A maximum of 10 asset groups are supported.

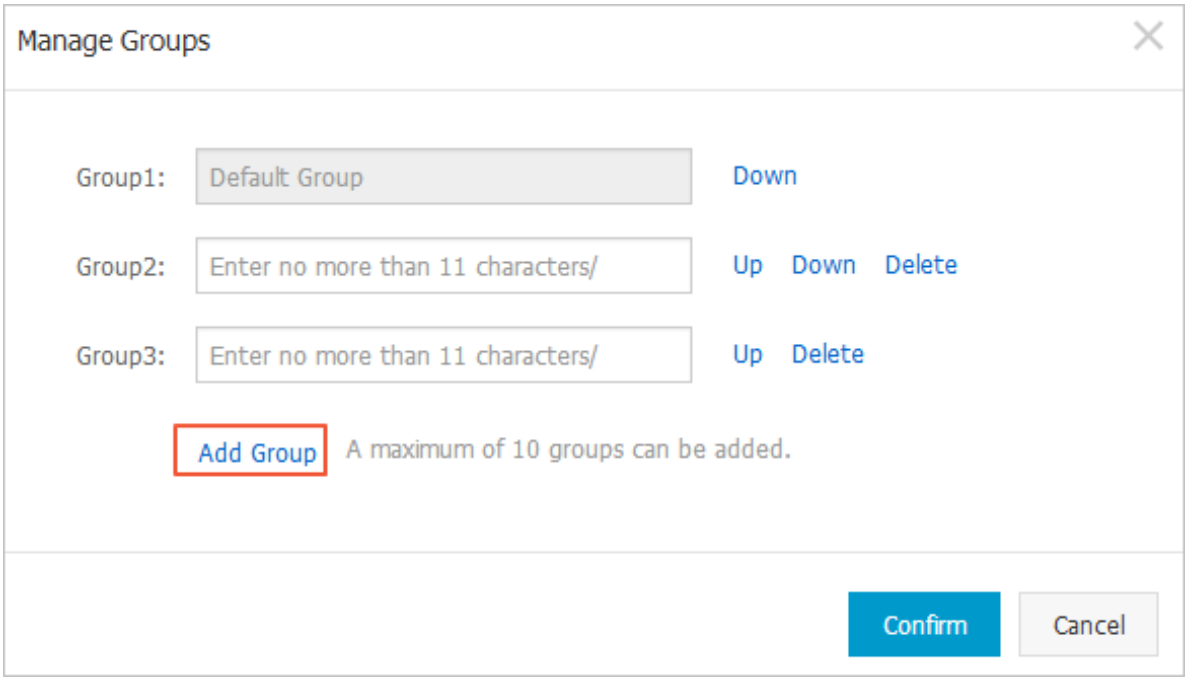
Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Asset Management** > **Asset Overview** and click **Manage Groups**.

Figure 27-33: Manage groups

3. In the **Manage Groups** dialog box, click **Add Group**, as shown in *Figure 27-34: Manage Groups dialog box*.

Figure 27-34: Manage Groups dialog box



4. Enter a group name.
5. Click **Confirm**.

27.9.2.2 Delete a group

This topic describes how to delete unnecessary groups to facilitate asset information query and modification.

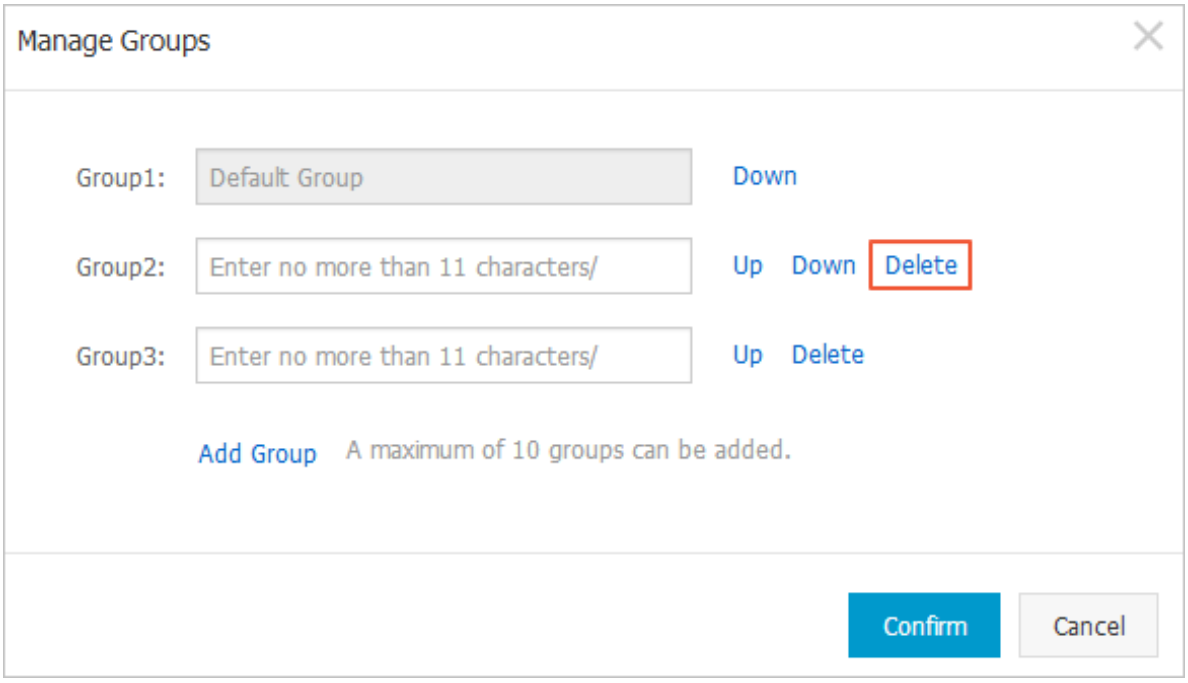
Context

- The default group cannot be deleted or renamed.
- Groups that contain assets cannot be deleted.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Asset Management** > **Asset Overview** and click **Manage Groups**.
3. In the **Manage Groups** dialog box, click **Delete** next to a group, as shown in *Figure 27-35: Delete a business group*.

Figure 27-35: Delete a business group



4. Click **Confirm**.

27.9.2.3 Sort groups

This topic describes how to sort groups. You can move frequently used groups to the top to facilitate asset information query and modification.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Asset Management** > **Asset Overview** and click **Manage Groups**.
3. In the **Manage Groups** dialog box, click **Up** or **Down** next to a group to change the group order.
4. Click **Confirm** to save the new group order.

27.9.3 Asset information

27.9.3.1 Manage server assets

A server asset refers to a server where a Server Guard agent has been installed and has connected to the Server Guard server.

Context

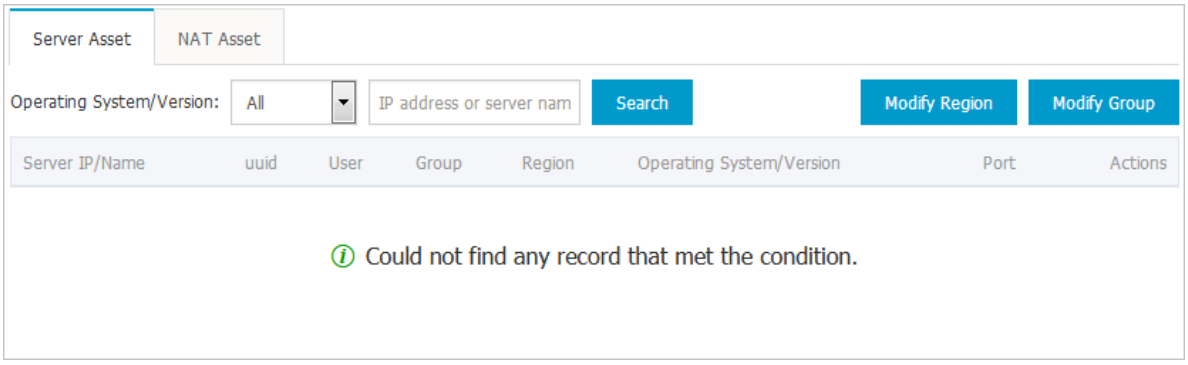
The security administrator can search for server assets to view their general information such as the operating systems, enabled ports, and installed common software. The security administrator can also change the region and group for each server asset.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Asset Management** > **Asset Overview** and click the **Server Asset** tab.
3. Set search criteria and click **Search** to view server asset information, as shown in *Figure 27-36: Server Asset tab page*.

> **Note:**
> You can filter server assets by operating system, region, and group. You can also enter the server IP address or server name for a fuzzy search. By default, the Server Asset tab page displays servers in all regions, and the servers are sorted by IP address.

Figure 27-36: Server Asset tab page



4. Select a server to view its details.
   - Click **Details** to view the ports that are enabled on the server.
   - Click **Show Application Information** to view the server applications that can be monitored.
5. Maintain information for a server asset.
   - Click **Modify**. In the **Modify Asset** dialog box, change the asset group and region and click **Confirm**.
   - Click **Delete**. In the **Delete Asset** message, click **Confirm** to delete the asset.

   Note:
   If the Server Guard agent on a server is uninstalled or an ECS instance is removed from Apsara Stack, you must manually delete the corresponding asset.

27.9.3.2 Manage NAT assets

NAT assets are public IP addresses that are converted from private IP addresses through NAT, namely, IP addresses exposed to the Internet. Multiple servers can share a public IP address but use different ports for receiving Internet requests. After an IP address is set as a NAT asset, Threat Detection Service analyzes the asset to detect attack events.

Context

The security administrator can search for a NAT asset protected by Apsara Stack Security to view the basic information about the asset or change the asset group and region. The security administrator can also add a NAT asset or add multiple NAT assets by CIDR block.

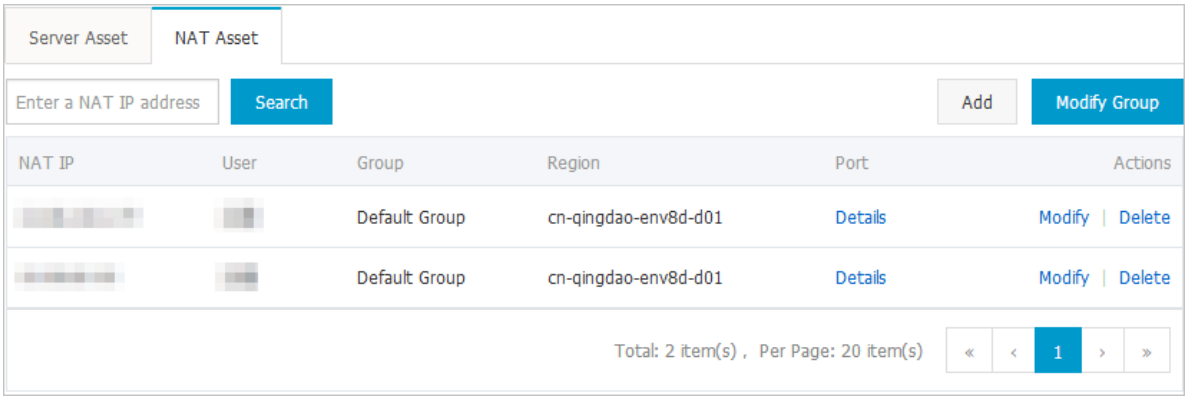
Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Asset Management** > **Asset Overview** and click the **NAT Asset** tab.

3. Set search criteria and click **Search** to view NAT asset information, as shown in *Figure 27-37: NAT Asset tab page*.

> **Note:**
> You can search for NAT assets by region or group, or by IP address in a fuzzy mode. By default, the NAT Asset tab page displays assets in all regions, which are sorted by IP address.

Figure 27-37: NAT Asset tab page



4. Add a NAT asset.

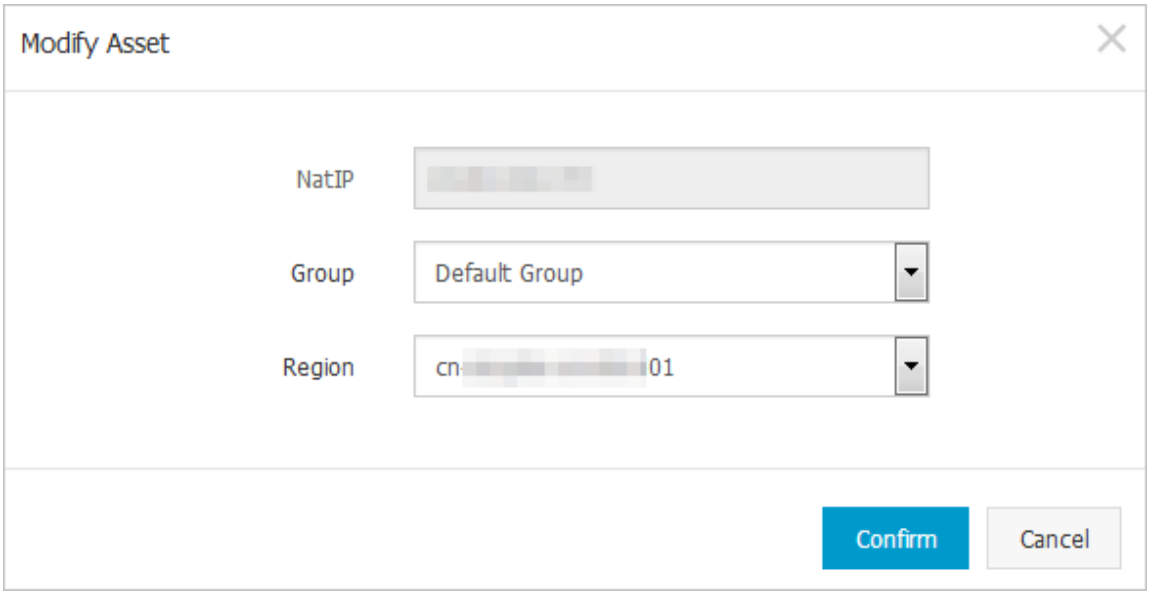
   The IP address of the asset to be added cannot conflict with an existing IP address. The NAT IP parameter must be set to a valid IP address or CIDR block.

   a) On the **NAT Asset** tab page, click **Add** in the upper-right corner.
   b) In the **Add Asset** dialog box, enter an IP address or a CIDR block and select the business group and region.
   c) Click **Confirm**.

5. Maintain NAT asset information.

   - Click **Details** to view the ports that are enabled for the NAT asset.
   - Click **Modify**. In the Modify Asset dialog box, change the business group and region of the asset and click **Confirm**, as shown in *Figure 27-38: Modify Asset dialog box*.

Figure 27-38: Modify Asset dialog box



- Click **Delete**. In the **Delete Asset** message, click **Confirm** to delete the NAT asset.

27.9.3.3 Modify attributes for multiple assets

This topic describes how to modify the group and region attributes for multiple assets.

Context

You can modify the attributes for multiple assets individually or at one time.

- For a single asset:

  This method applies when you modify only one asset or when you modify multiple assets but they are not in the same CIDR block and their server names do not follow any rules. For more information about how to modify a single asset, see *Manage server assets* and *Manage NAT assets*.

- For multiple assets:

  This method applies when you modify multiple assets that belong to the same CIDR block or have similar server names.

Note:
The server IP address, server name, operating system type, and operating system version are fixed information for an asset. These information cannot be modified.

Procedure

1. *Log on to Apsara Stack Security Center*.

2. Choose **Asset Management** > **Asset Overview**.
3. Change the group or region for multiple assets.
   - Click **Modify Group** to change the group for multiple assets at one time.
   - Click **Modify Region** to change the region for multiple assets at one time.
4. In the **Modify Group** or **Modify Region** dialog box, specify the assets to be modified, select a new group or region for these assets, and click **Confirm**.
   - Select **`CIDR Block`** from the Type drop-down list, and enter the CIDR block for the specified server assets or NAT assets.
   - Select **`Server Name`** from the Type drop-down list, and enter the common part of the server names for the specified server assets.

   Note:

   - If a CIDR block is specified, all server assets and NAT assets in the specified CIDR block are modified.
   - If the common part of server names is specified, all server assets whose names have the specified common part are modified.

27.10 Security audits

27.10.1 Overview

A security audit refers to the systemic and independent inspection and verification of activities and behavior in the computer network environment. Delegated by property owners and authorized by management authorities, professional auditors give their assessments according to relevant laws and regulations. When the administrator needs to backtrack system operations, the administrator can perform a security audit.

Security audits are long-term security management activities throughout the lifecycle of cloud services. The security audit feature of Apsara Stack Security can collect system security data , analyze weaknesses in system operations, report audit events, and classify audit events into important, moderate, and low risk levels. The security administrator views and analyzes audit events to continuously improve the system and ensure the security and reliability of cloud services .

27.10.2 View audit overview

The **Overview** page provides four types of reports: raw log trend, audit event trend, audit risk distribution, and security issue distribution. The reports are displayed in the form of a run chart or pie chart, helping the security administrator analyze the trend of risks facing Apsara Stack services.

Context

- Data in the **Trends of Raw Log** report includes the number of logs generated by physical servers, network devices, RDS instances, ECS instances, and APIs in the last week. Based on the log trend of the Apsara Stack platform, the security administrator can check whether the number of logs generated by the system is normal.
- Data in the **Audit Events** report includes the number of audit events generated by physical servers, network devices, RDS instances, ECS instances, and APIs in the last week. Based on the audit event trend, the security administrator can check whether the number of audit events generated by the system is normal.
- Data in the **Audit Risk Distribution** report includes the number of events at important, moderate, low, and urgent risk levels in the last week. Based on the audit risk distribution, the security administrator can check whether the number of audit events generated by the system at each risk level is normal.
- Data in the **Security Issue Distribution** report includes the proportion of events of each type among all events in the last week. Based on security issue distribution, the security administrator can check the type of audit event that accounts for the largest proportion of the events, identify important-risk events, and take preventive measures.

In addition, on the **Overview** page, the security administrator can check the online and offline log storage and log storage for each audit type in the specified time range.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Security Audit** > **Overview** to go to the **Overview** page.
3. Set the **End Time** parameter and click **View**. The overall audit information collected one week before the specified time is displayed.

Note:

You can view the specific time range of the displayed audit logs in the **Duration** parameter.

4. You can select or clear an audit type for the **Audit Type** parameter to determine whether to display audit logs of this type.

27.10.3 Query audit events

On the **Audit Query** page, you can view details of audit events, for example, log creation time, audit type, audit target, operation type, risk level, and log content.

Context

The system matches the logs collected by Security Audit against the audit policies. If the log content matches any regular expression in the audit policies, an audit event is reported. For more information about audit policies, see *Add an audit policy*.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Security Audit** > **Audit Query** to go to the **Audit Query** page.
3. Set the search criteria and time, and click **Search**. Audit events generated during the specified period of time are displayed.

   Basic search criteria include **Audit Type**, **Audit Target**, **Action Type**, and **Risk Level**.

   You can click **Advanced Search** to set more detailed search criteria, such as **User**, **Target**, **Action**, **Result**, and **Cause**.
4. Click **Export** to create an export task.

   For more information about how to download the exported file to your local disk for analysis, see *Manage export tasks*.

27.10.4 View raw logs

On the **Raw Logs** page, you can view the raw logs generated during the running of the audit targets. As necessary information for debugging, raw logs can help the security administrator locate system faults.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Security Audit** > **Raw Logs**.
3. Set the **Audit Type** and **Audit Target**, set the query time, and click **Search**. Raw log information of the specified audit target within the time range is displayed.

Figure 27-39: Raw log information

4. Click **Export** to create an export task.

   For more information about how to download the exported file to your local disk for analysis, see *Manage export tasks*.

27.10.5 Policy settings

27.10.5.1 Add an audit policy

When a log record matches an audit policy, an audit event is reported.

Context

Audit policies support regular expression matching. A regular expression defines a character string matching pattern, which can be used to check whether a specific string exists. The following table provides some examples.

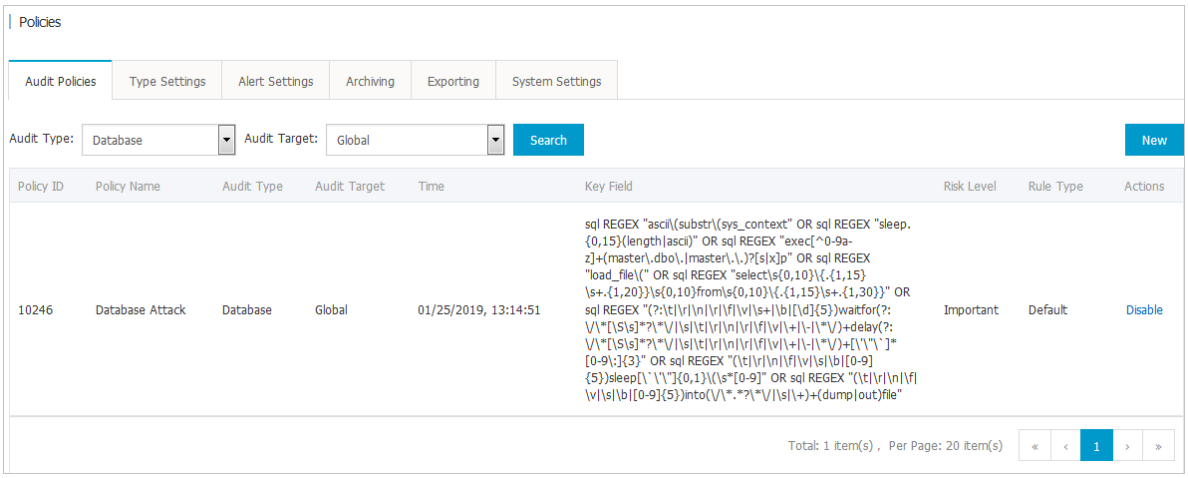
Regular expression	Description
`^\d{5,12}$`	Matches the consecutive numbers from the 5th number to the 12th number
`load_file\(`	Matches the string: load_file(

Security Audit defines the default audit policy based on the string generated in the log when an audit event is reported. The security administrator can also customize an audit policy based on the string generated in the log when the system encounters an attack.

Procedure

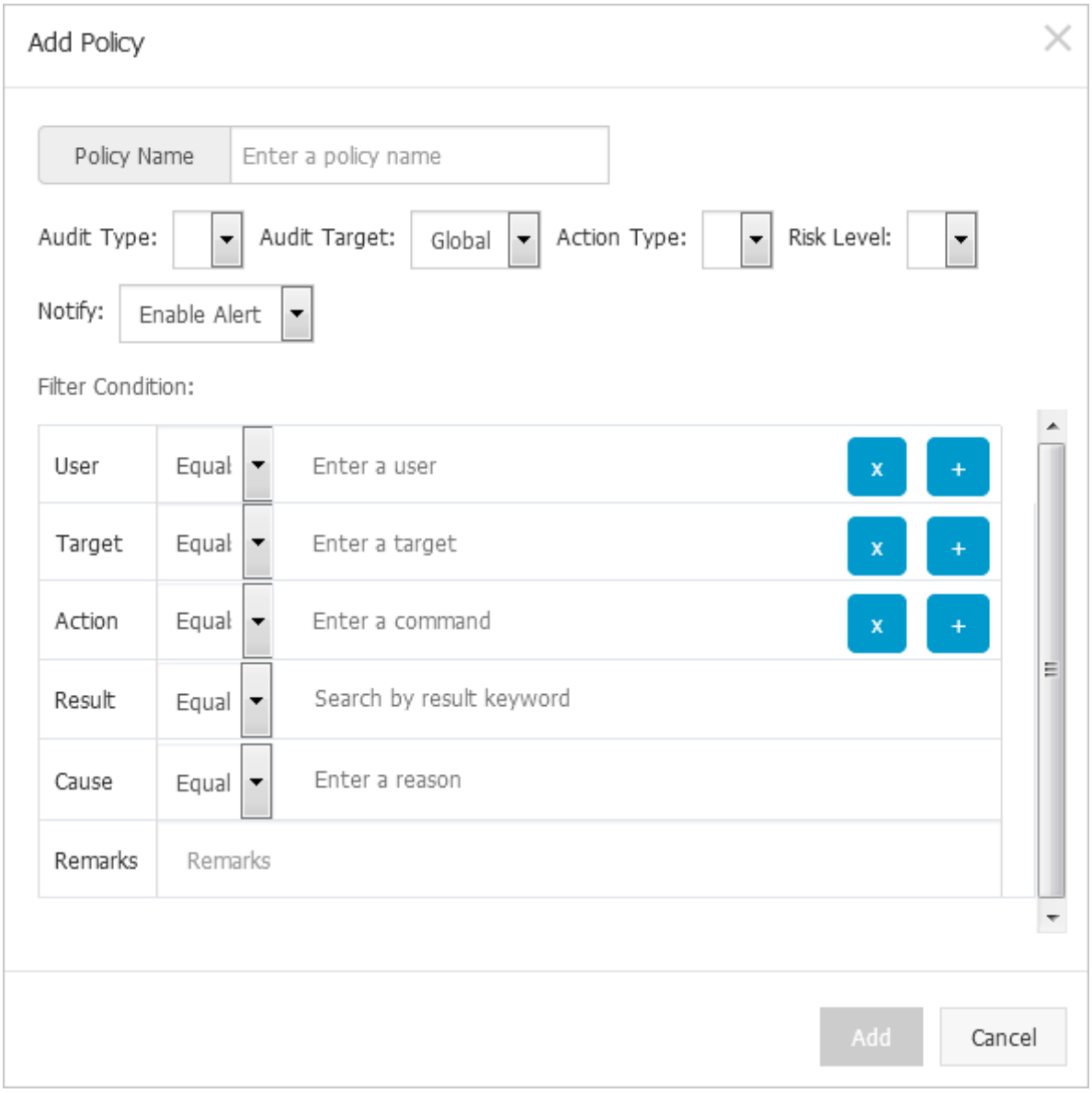
1. *Log on to Apsara Stack Security Center*.
2. Choose **Security Audit** > **Policies**.
3. Click the **Audit Policies** tab.

Figure 27-40: Audit Policies tab page



4. Click **New**.
5. In the **Add Policy** dialog box, configure the audit policy.

Figure 27-41: Add Policy dialog box



6. Click **Add**.

   After an audit policy is added, if any string in an audit log of the specified audit type, audit target , or risk level matches the regular expression of the audit policy, an alert email is sent to the specified recipient.

27.10.5.2 Manage action types

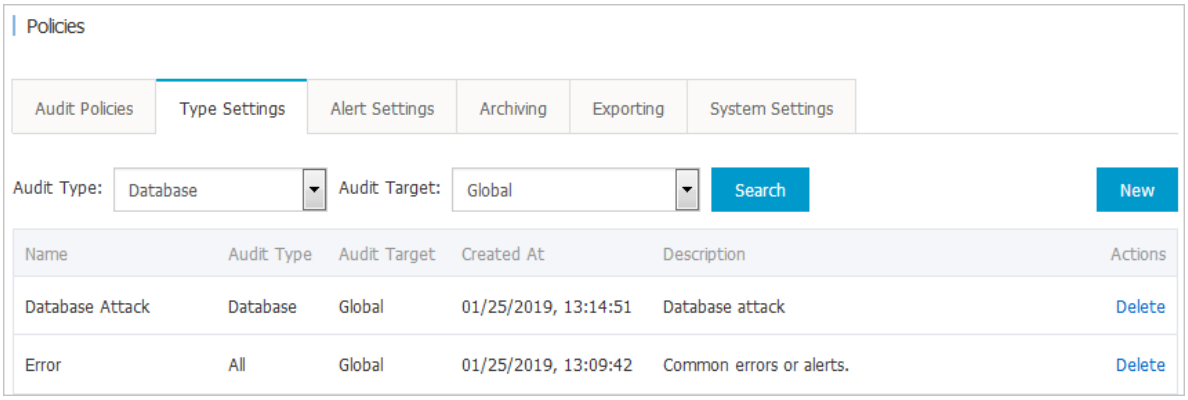
This topic describes how to add, query, and delete action types for audit policies.

Procedure

1. *Log on to Apsara Stack Security Center*.

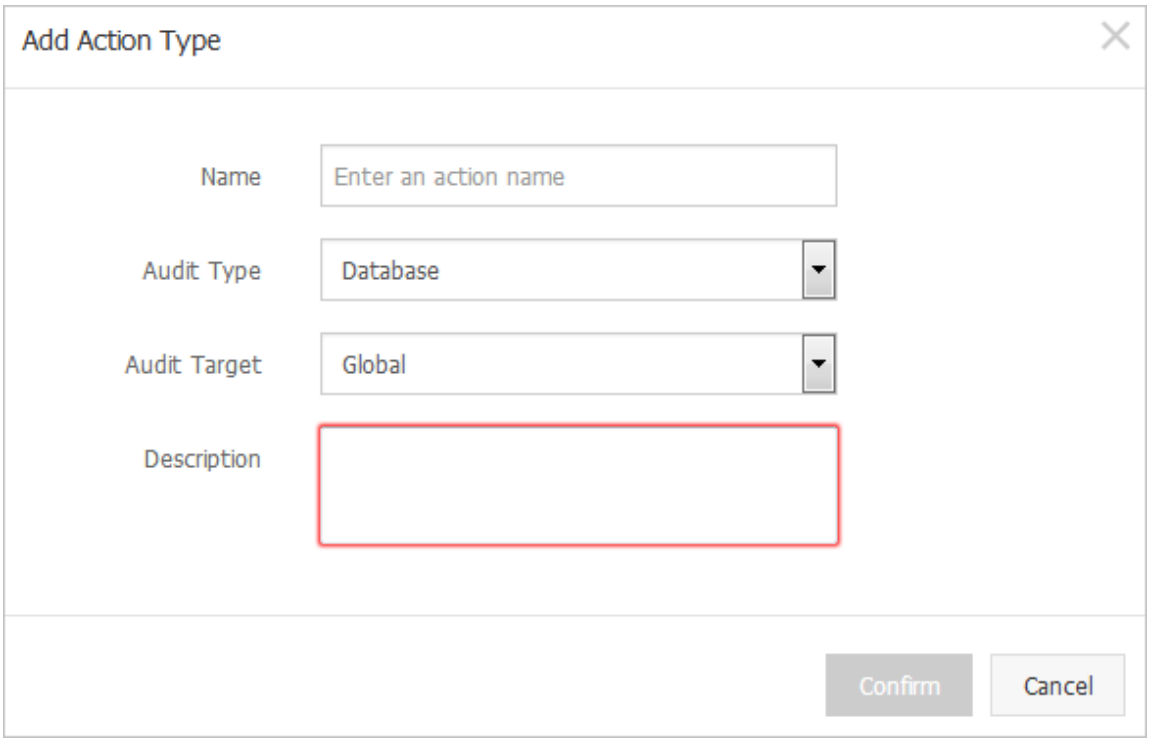
2. Choose **Security Audit** > **Policies**.
3. Click the **Type Settings** tab.

Figure 27-42: Type Settings tab page



4. Add an action type.
   a) Click **New**.
   b) In the **Add Action Type** dialog box, configure the action type.

Figure 27-43: Add an action type



   c) Click **Confirm**.
5. Search for action types

Set the **Audit Type** and **Audit Target** parameters, and click **Search**. The matching action types are displayed.

6. Delete an action type.

   Select a record and click **Delete** to delete the action type.

> **Note:**
> The default action types cannot be deleted.

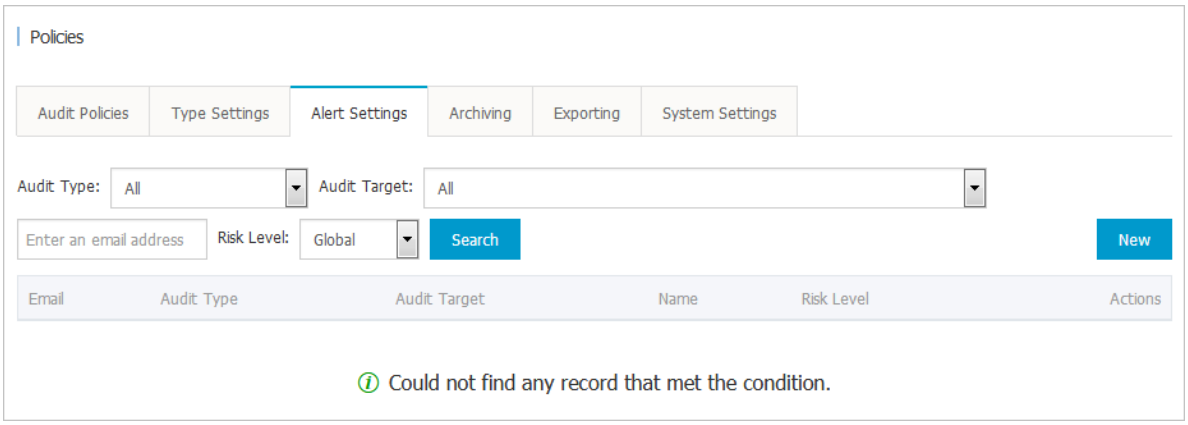
27.10.5.3 Set an alert recipient

This topic describes how to set the email address of the alert recipient. When an audit event occurs, the event is reported to the alert recipient by email.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Security Audit** > **Policies**.
3. Click the **Alert Settings** tab.

Figure 27-44: Alert Settings tab page

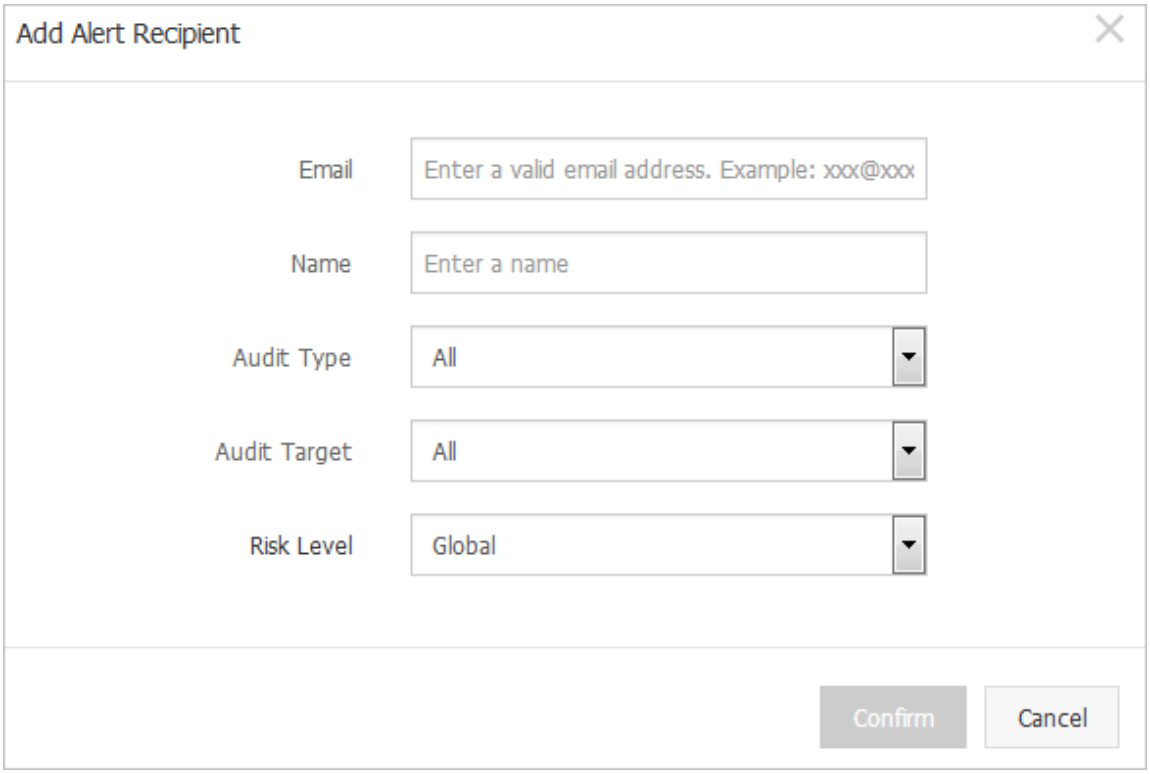


4. Add an alert recipient.

   a) Click **New**.

   b) In the **Add Alert Recipient** dialog box, configure the alert recipient information.

Figure 27-45: Add Alert Recipient dialog box



c) Click **Confirm**.

5. Search for alert recipients.

   Set the **Audit Type**, **Audit Target**, and Risk Level parameters, and click **Search**. The matching alert recipients are displayed.

6. Delete an alert recipient.

   Select a record and click **Delete** to delete the alert recipient.

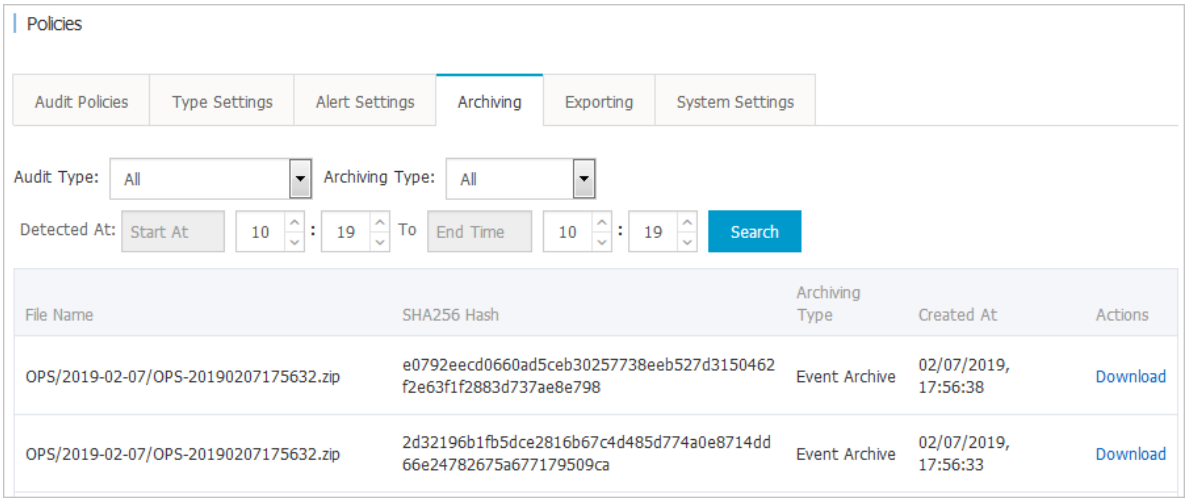
27.10.5.4 Manage event log archives

You can manage archives of audit events and raw logs.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Security Audit** > **Policies**.
3. Click the **Archiving** tab.

Figure 27-46: Archiving tab page



4. Set the search criteria, such as **Audit Type**, **Archiving Type**, and **Detected At**, and click **Search** to query the archive information.
5. Select an archived file that you want to download and click **Download** to download the file to your local disk.

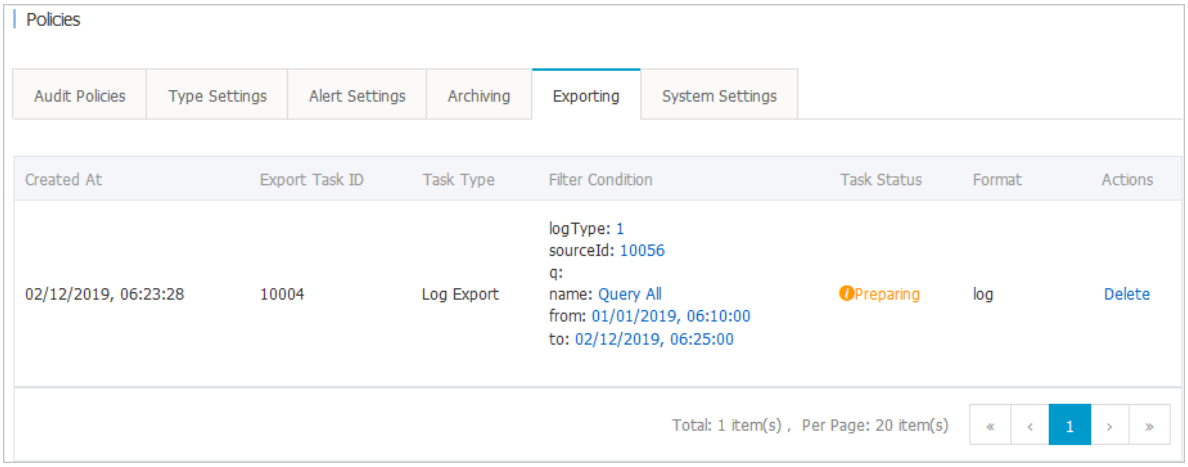
27.10.5.5 Manage export tasks

After exporting audit events or logs on the **Audit Query** or **Raw Logs** page, you can manage the export tasks on the Exporting page.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Security Audit** > **Policies**.
3. Click the **Exporting** tab.
4. View the export tasks that have been created.

Figure 27-47: Exporting tab page



5. Click **Download** to download the specified audit event or log file to your local disk.
6. Click **Delete** to delete the specified export task.

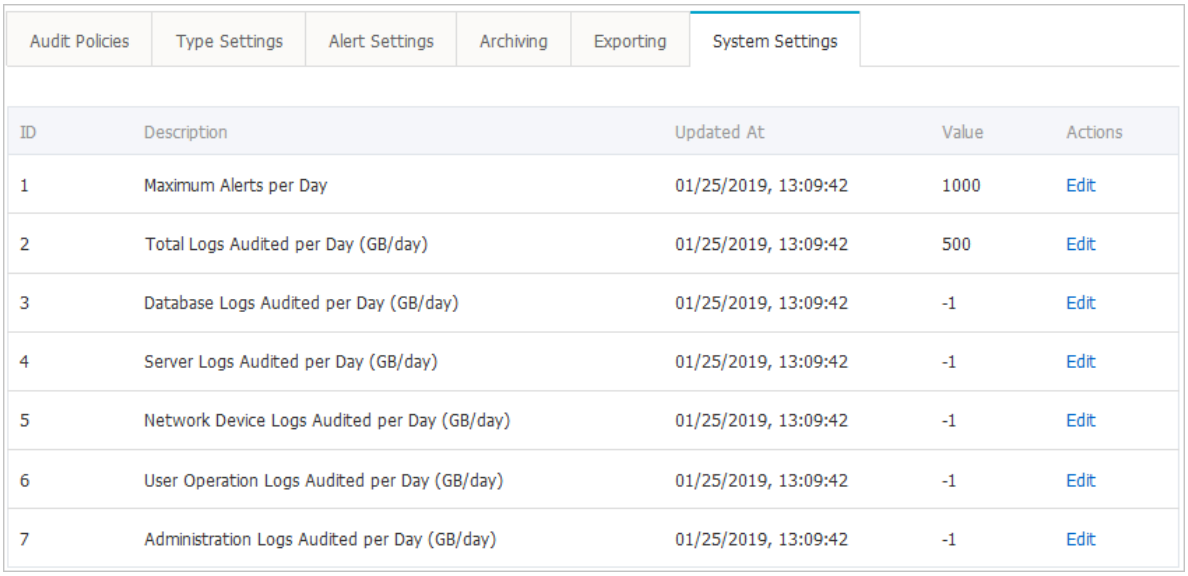
27.10.5.6 Modify system settings

By setting system parameters, you can configure the maximum number of system alerts per day and the maximum number of audits per day for various raw logs.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Security Audit** > **Policies**.
3. Click the **System Settings** tab.
4. Select a configuration item and click **Edit**.

Figure 27-48: System Settings tab page



5. Set the corresponding parameter and click **Confirm**.

27.11 System management

27.11.1 Manage your Alibaba Cloud account

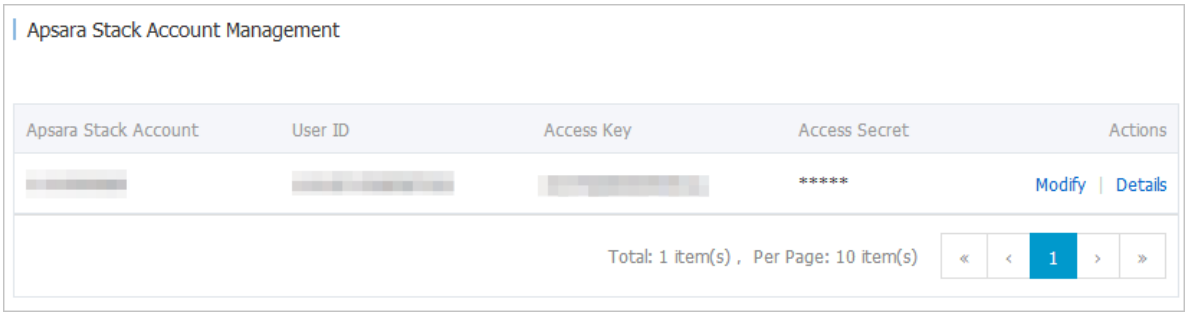
This topic describes how to manage your Alibaba Cloud account bound to Apsara Stack Security.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Accounts** to view and modify your Alibaba Cloud account bound to the system, as shown in *Figure 27-49: Apsara Stack Account Management page*.

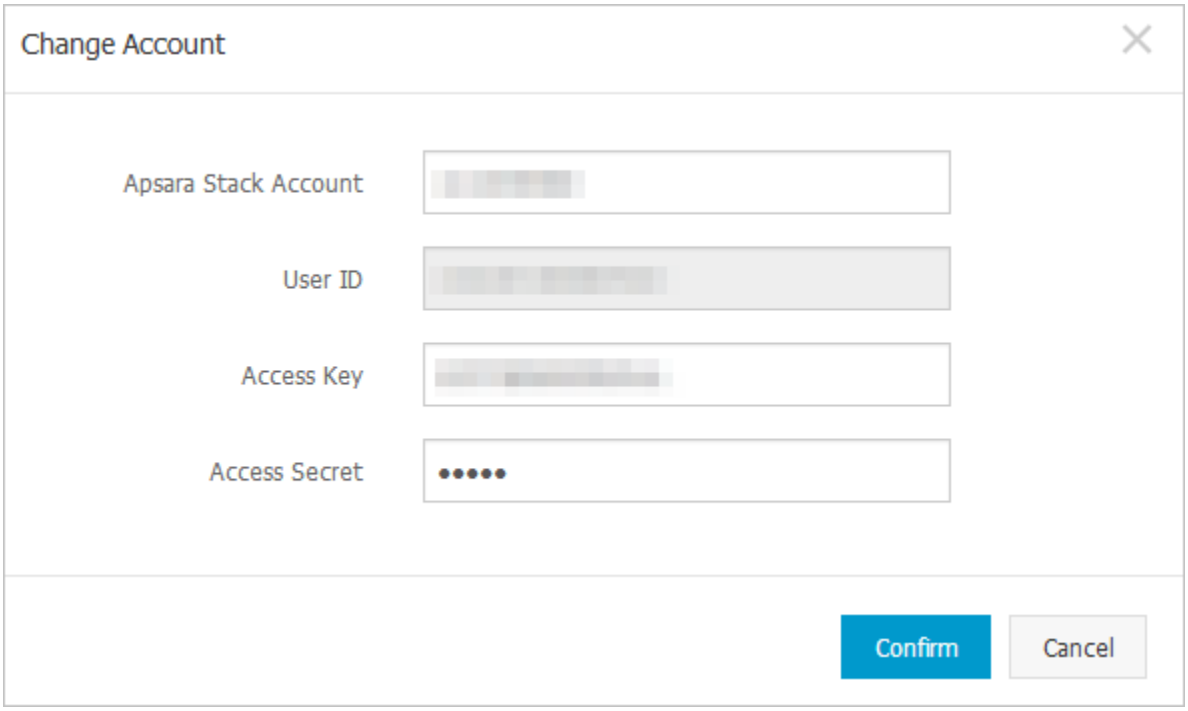
   In Apsara Stack Security, all assets are bound to your Alibaba Cloud account. Use caution when you modify the account.

Figure 27-49: Apsara Stack Account Management page



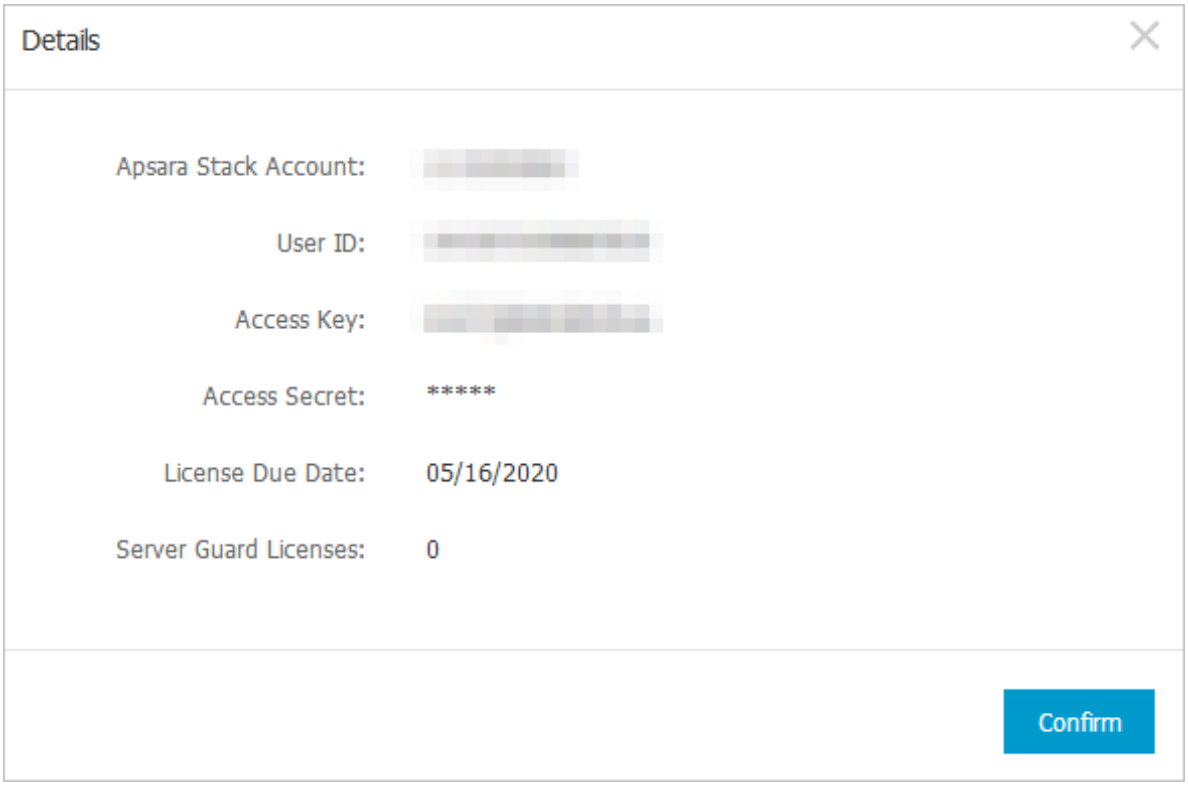
3. Click **Modify**. In the Change Account dialog box, modify the account information and click **Confirm**, as shown in *Figure 27-50: Change Account dialog box*.

Figure 27-50: Change Account dialog box



4. Click **Details** to view detailed information about your Alibaba Cloud account, including the license expiration date and number of Server Guard licenses, as shown in *Figure 27-51: Account details*. The details are obtained based on the configured user ID and access key.

Figure 27-51: Account details



27.11.2 Rule database synchronization

27.11.2.1 Synchronization overview

This function synchronizes specific security information such as rule databases and vulnerabilities from Alibaba Cloud to Apsara Stack.

Rule databases are in the initial status before being synchronized from Alibaba Cloud. Only rule databases on Alibaba Cloud can be synchronized to Apsara Stack. The synchronized rules are used in functional modules of Apsara Stack Security to ensure that the security capabilities of Alibaba Cloud are provided in Apsara Stack.

The rule database information and rule database synchronization frequency and time can be set by the administrator. The synchronization can also be triggered automatically. If the administrator do not set the synchronization frequency and time, the default settings will apply.

Table 27-34: Synchronization status description

Status	Description
To Be Upgraded	Indicates that a new version of the rule database is available for upgrade.
Upgrading	Indicates that the rule database is being downloaded from Alibaba Cloud or is being upgraded.
Upgraded	Indicates that the rule database has been upgraded.
Upgrade Failed	Indicates that the rule database failed to be upgraded.

Rule databases can be synchronized from Alibaba Cloud through online upgrade or offline upgrade by importing upgrade package, depending on the deployment mode of Apsara Stack.

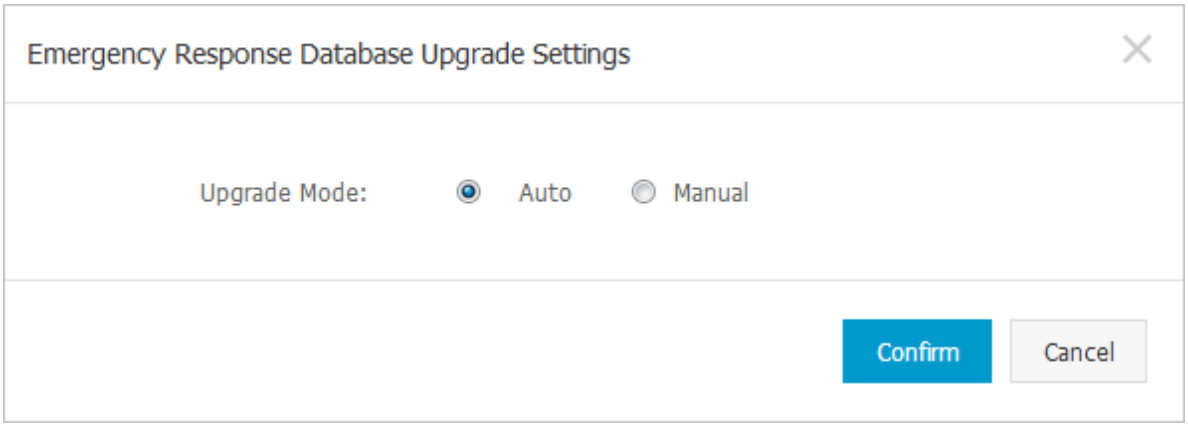
27.11.2.2 Specify the upgrade mode for rule databases

Apsara Stack Security supports both automatic and manual upgrade for rule databases. This topic describes how to specify the upgrade mode for rule databases.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Synchronizations**.
3. Click **Settings** in the **Actions** column of a rule database.
4. In the **Upgrade Settings** dialog box, select **Auto** or **Manual**.

Figure 27-52: Specify the upgrade mode



5. Click **Confirm**.

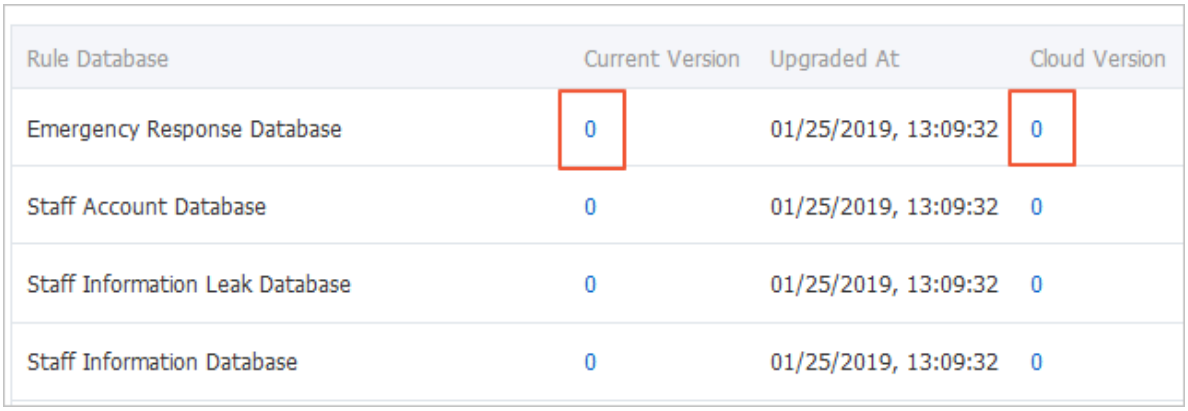
27.11.2.3 Refresh the cloud synchronization list

This topic describes how to refresh the rule databases synchronized from Alibaba Cloud to view the rule databases with updates.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Synchronizations**.
3. Click **Refresh** in the upper-right corner.
4. View detailed information about rule databases.

   After the rule databases are refreshed, you can view the current version, cloud version, upgrade mode, upgrade frequency, and status for each rule database in Apsara Stack.
5. Click the current version or cloud version of a rule database to view version details of the rule database.



Rule Database	Current Version	Upgraded At	Cloud Version
Emergency Response Database	0	01/25/2019, 13:09:32	0
Staff Account Database	0	01/25/2019, 13:09:32	0
Staff Information Leak Database	0	01/25/2019, 13:09:32	0
Staff Information Database	0	01/25/2019, 13:09:32	0

27.11.2.4 Manually upgrade a rule database

This topic describes how to manually upgrade a rule database. You can upgrade a single rule database or all rule databases at one time.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Synchronizations**.
3. Manually upgrade a rule database.

   You can upgrade a single rule database or all rule databases at one time.

   - Upgrade a single rule database

     Click **Upgrade** in the **Actions** column of a rule database. Download the rule database information from the cloud to the version database.

- Upgrade all rule databases.

  Click **Upgrade** in the upper-right corner of the page to manually upgrade all rule databases.

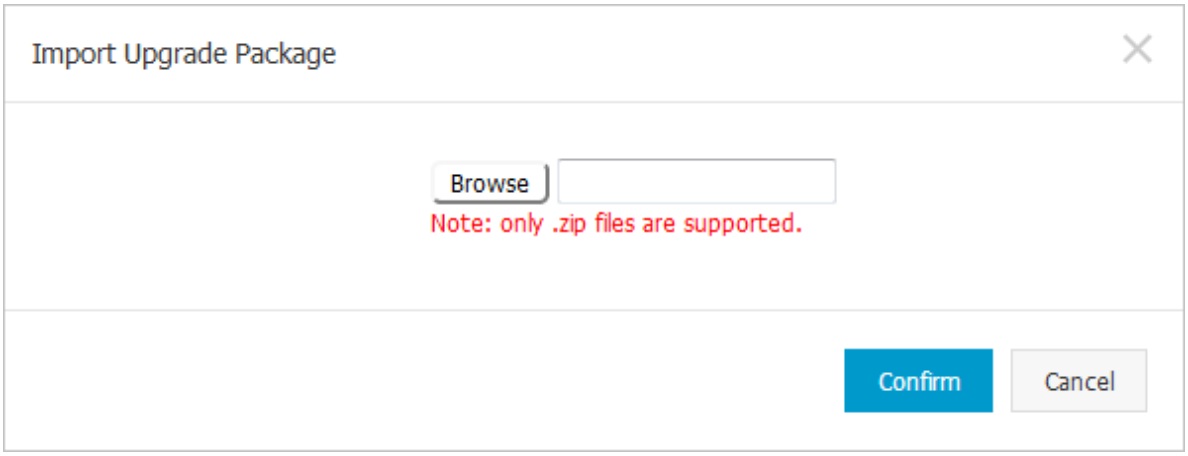
27.11.2.5 Import an offline upgrade package

If the Apsara Stack environment cannot connect to Alibaba Cloud, you can upgrade a rule database by importing an offline upgrade package.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Synchronizations**.
3. Click **Import Upgrade Package** in the upper-right corner.
4. In the **Import Upgrade Package** dialog box, click **Browse** to select an offline upgrade package that has been downloaded to your local device.

   Figure 27-53: Import an offline upgrade package

   Apsara Stack Security upgrades a rule database based on the corresponding offline upgrade package, and updates the status of the rule database on the **Synchronizations** page.
5. Click **Confirm**.

27.11.2.6 Roll back a rule database

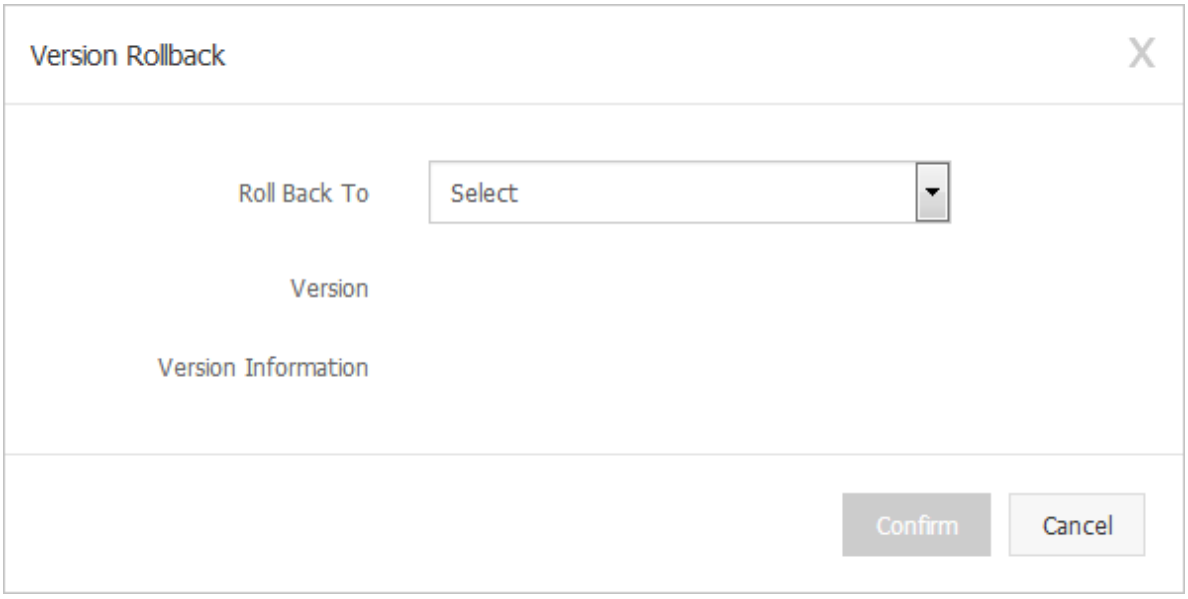
If there are any problems with an upgraded rule database, you can roll the rule database back to a previous version.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Synchronizations**.

3. In the Actions column of the rule database that needs to be rolled back, click **Roll Back**.
4. In the **Version Rollback** dialog box, select the target rule database version for the rollback from the **Roll Back To** drop-down list.

   Figure 27-54: Version Rollback dialog box



5. Click **OK**.

27.11.2.7 View version history of a rule database

This topic describes how to view upgrade history of a rule database.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Synchronizations**.
3. Click **History** in the **Actions** column of a rule database.

   In the **History** dialog box, you can view the upgrade history of the rule database.

27.11.3 Alert settings

27.11.3.1 Set alert recipients

Alert recipients are those who receive alert messages. Alerts can be sent by SMS or email. When the monitored data meets an alert rule, an alert message is sent to the alert recipient.

Procedure

1. *Log on to Apsara Stack Security Center*.

2. Choose **System Management** > **Alert Settings** > **Alert Recipient**.
3. Click **Add Recipient**.
4. Enter the recipient information, and click **OK** to add an alert recipient.

   After adding an alert recipient, you can click **Edit** or **Delete** to edit or delete the recipient information.

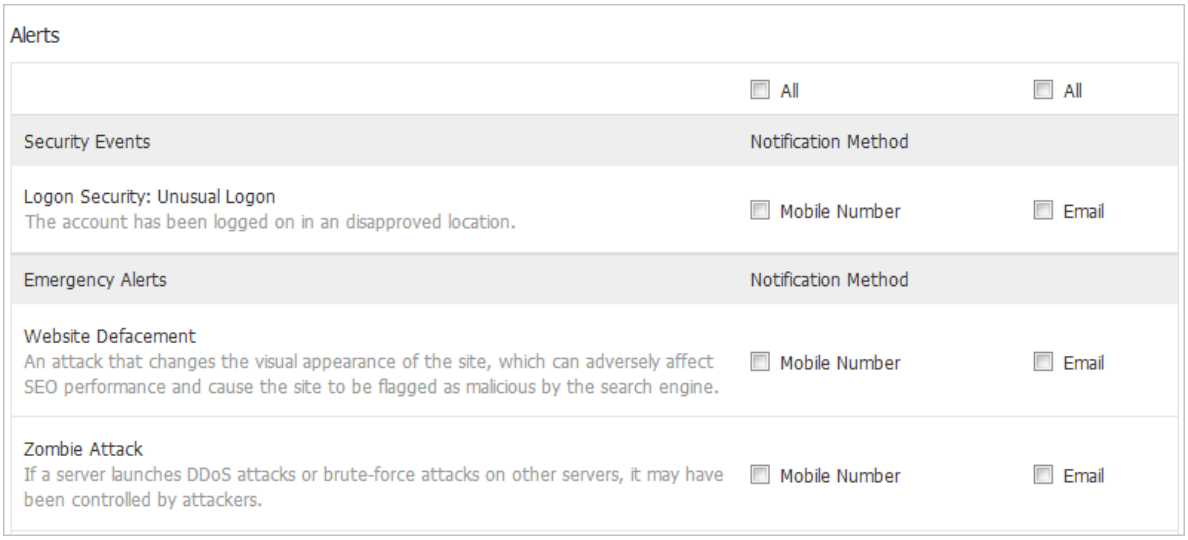
27.11.3.2 Set alert information

You can set alerts for various security events. These alerts can be sent by SMS or email.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Alert Settings** > **Alert Settings**.
3. In the **Alerts** area, select a notification method for each security event, as shown in *Figure 27-55: Alert Settings tab page*.

   Figure 27-55: Alert Settings tab page



4. Click **Confirm**.

27.11.4 Global settings

27.11.4.1 Set CIDR blocks for traffic monitoring

27.11.4.1.1 Add a CIDR block for traffic monitoring

This topic describes how to add a traffic monitoring CIDR block for the traffic security monitoring module of Apsara Stack Security.

Context

CIDR blocks are configured for the traffic security monitoring module. The security administrator can change the CIDR blocks for monitoring as needed. The settings of CIDR blocks for traffic monitoring only apply to data centers in the corresponding regions.

Note:

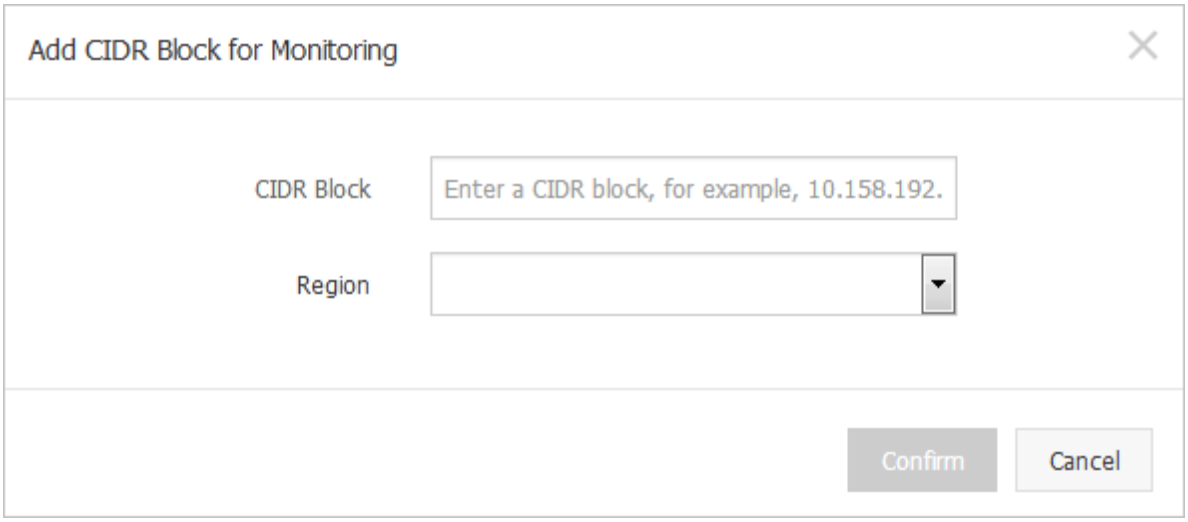
Changes to CIDR block settings take effect immediately without security administrators' interventi on.

If the same CIDR block is configured for the traffic security monitoring module and region detection, make sure that the same region is specified for the CIDR block.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Global Settings** > **Traffic Collection IP Range**.
3. Click **Add**.
4. In the **Add CIDR Block for Monitoring** dialog box, configure a monitored CIDR block.

Figure 27-56: Add CIDR Block for Monitoring dialog box



- Enter a CIDR block.

> **Note:**
> The CIDR block must be valid and unique.

- Select the region to which the CIDR block belongs.

5. Click **Confirm**.

27.11.4.1.2 Manage CIDR blocks for traffic monitoring

This topic describes how to modify and delete the CIDR blocks for traffic monitoring.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Global Settings** > **Traffic Collection IP Range**.
3. Select a region, enter a CIDR block, and click **Search**.

   You can view the CIDR block for traffic monitoring and region information in the query results.
4. In the **Actions** column of a CIDR block, manage the CIDR block for traffic monitoring.
   - Modify the CIDR block for traffic monitoring:

     Click **Modify** to modify the region of the CIDR block.
   - Delete the CIDR block for traffic monitoring

     Click **Delete** to delete the CIDR block.

27.11.4.2 Region settings

27.11.4.2.1 Add a CIDR block for a region

This topic describes how to add CIDR blocks for regions that are detected and reported by Server Guard.

Context

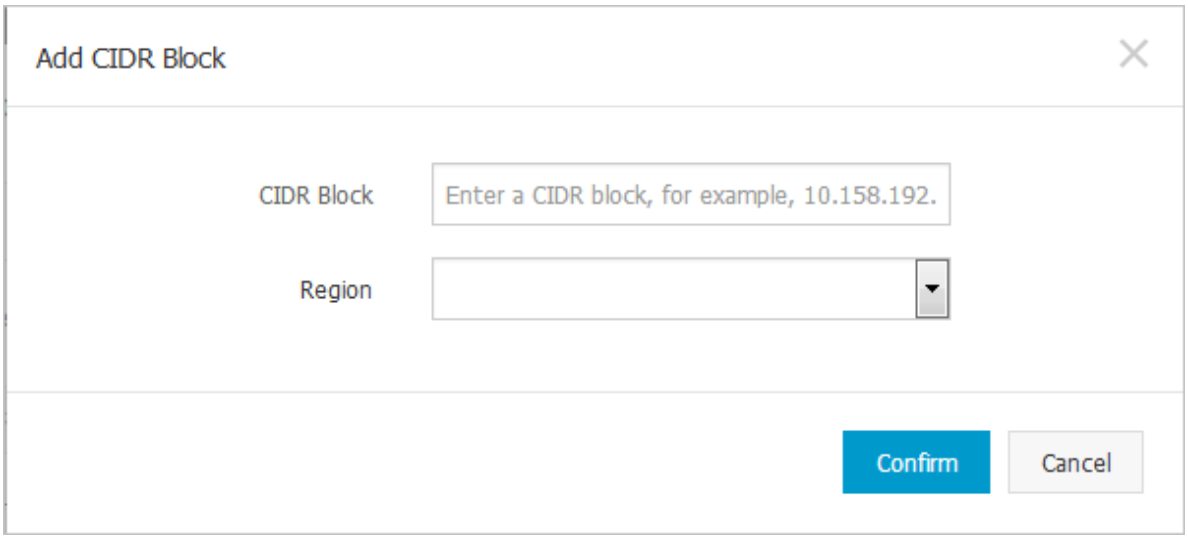
Region settings are used for region detection of Server Guard agents in different data centers. Server Guard servers automatically detects and matches the regions of servers based on the IP address information reported by Server Guard agents.

> **Note:**
> You can change the region of a CIDR block. After modification, you must modify the region for all assets in the CIDR block on the Asset Overview page.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Global Settings** > **Region**.
3. Click **Add**.
4. In the **Add CIDR Block** dialog box, set the CIDR block.

   Figure 27-57: Add CIDR Block dialog box

   - Enter a CIDR block.

     Note:
     The CIDR block must be valid and unique.
   - Select the region to which the CIDR block belongs.
5. Click **Confirm**.

27.11.4.2.2 Manage CIDR blocks for region detection

This topic describes how to modify and delete CIDR blocks for region detection.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Global Settings** > **Region**.
3. Select a region, enter a CIDR block, and click **Search**.

   View information about the CIDR block for region detection in the query result.
4. In the **Actions** column of the region, manage the CIDR block for region detection.
   - Modify the CIDR block for region detection:

Click **Modify** to modify the CIDR block.

- Delete the CIDR block for region detection:

  Click **Delete** to delete the CIDR block.

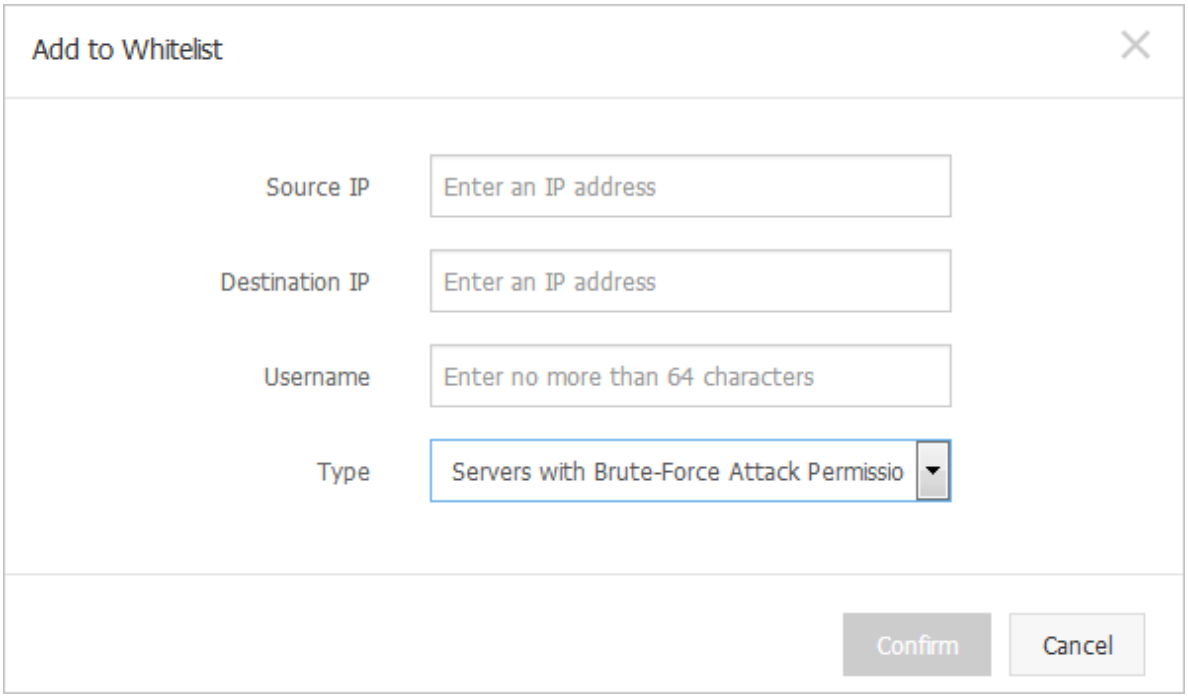
27.11.4.3 Configure whitelists

This topic describes how to configure the following types of whitelists: Beaver WAF whitelist, servers with brute-force attack permissions, and IP addresses with application attack permissions.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **System Management** > **Global Settings** > **Whitelist**.
3. Click **Add**.
4. In the **Add to Whitelist** dialog box, set parameters for the whitelist entry.

Figure 27-58: Add to Whitelist dialog box



Parameter	Description
Source IP	The source IP address or CIDR block.
Destination IP	The destination IP address or CIDR block.
Username	The name of the user who adds the whitelist entry.

Parameter	Description
Type	• Beaver WAF Whitelist: Traffic defined in this whitelist is not inspected. • Servers with Brute-Force Attack Permissions: Brute-force server attacks defined in this whitelist are not inspected. • IPs with Application Attack Permissions: Suspicious application attacks defined in this whitelist are not inspected.

5. Click **Confirm**.

   You can click **Delete** to delete an unnecessary whitelist entry.

27.12 Optional security products

27.12.1 Anti-DDoS settings

27.12.1.1 Overview

Distributed Denial of Service (DDoS) exploits client/server technology to combine multiple computers and form a platform on which an attack is initiated against one or more targets. In this way, the threat posed by the denial-of-service (DoS) attack is increased exponentially.

Common DDoS attack types include:

- **Network-layer attacks**: Typically UDP reflection attacks, such as NTP flood. These attacks usually congest the network bandwidth of the victim by using heavy traffic, making the victim unable to normally respond to customer access requests.
- **Transport-layer attacks**: Typically SYN flood and connections flood. These attacks occupy connection pool resources of a server to achieve the purpose of DoS.
- **Session-layer attacks**: Typically SSL flood. These attacks occupy SSL session resources of a server to achieve the purpose of DoS.
- **Application-layer attacks**: Typically DNS flood, HTTP flood, and dummy attacks. These attacks occupy application processing resources and significantly consume the processing resources of a server to achieve the purpose of DoS.

Apsara Stack Security can redirect, scrub, and re-inject attack traffic to defend your server against DDoS attacks and ensure normal business operations.

27.12.1.2 View DDoS events

During or after the traffic scrubbing process, Apsara Stack Security reports security events to Apsara Stack Security Center. This topic describes how to view DDoS events.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Network Security** > **Network Protection** and click the **Anti-DDoS Events** tab.
3. Set the search criteria and click **Search**.

   The system returns a list of DDoS events that meet the search criteria.

Parameter	Description
Trigger	The DDoS attack traffic metric that exceeds the configured alert threshold.
External Service IP	The IP address that was under a DDoS attack.
Status	• Scrubbing: Indicates that traffic scrubbing is in progress. • Scrubbed: Indicates that traffic scrubbing is complete.
Actions	• Stop Scrubbing: Allows you to stop traffic scrubbing. • Traffic Analysis: Allows you to view the traffic protocol and top 10 attacked servers in the Anti-DDoS event. • View Traffic: Allows you to view the alert thresholds and traffic diagram.

4. View and analyze an Anti-DDoS event.
   - Click **View Traffic** to view the alert thresholds and traffic diagram of the corresponding IP address.
   - Click **Traffic Analysis** to view the traffic protocol and top 10 attacked servers in the Anti-DDoS event.

27.12.1.3 Anti-DDoS rules

27.12.1.3.1 Set alert thresholds

If an alert threshold is set for a CIDR block or an IP address, DDoS detection is triggered depending on this threshold. Otherwise, DDoS detection is triggered depending on the global threshold.

Context

After a DDoS traffic alert threshold is set for an IP address, an alert is triggered when the traffic to the IP address reaches the threshold. The alert thresholds of the IP address must be set based on the traffic volume. Excessive traffic volume indicates a possible DDoS attack. We recommend that you set a alert threshold to a value slightly higher than the peak traffic volume.

Apsara Stack Security supports global alert thresholds or alert thresholds for a specific CIDR block or IP address.

- **Global alert threshold**: You cannot add a global threshold. The default value is imported during service initialization.
- **Alert threshold for a specific CIDR block**: You can set an alert threshold for a specific CIDR block based on its traffic volume. The CIDR block-specific threshold is more precise than the global threshold for a CIDR block.
- **Alert threshold for a specific IP address**: You can set an alert threshold for a specific IP address based on its traffic volume. The IP address-specific threshold is more precise than the CIDR block-specific threshold for an IP address.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Network Security** > **Network Protection**, and click the **Anti-DDoS Settings** tab.
3. Click **Create Anti-DDoS Rule**.
4. In the **Create Anti-DDoS Rule** dialog box, set relevant parameters.

Parameter	Description
IP Address	The IP address or CIDR block to which the alert thresholds are applied. **Note:** Before setting the alert thresholds, make sure that the corresponding CIDR block has been added in *Add a CIDR block for traffic monitoring*.
Bandwidth Threshold	The alert threshold for bandwidth usage in a data center . When the inbound or outbound traffic rate reaches this threshold, DDoS detection is triggered. Generally, this value is set to be slightly higher than the traffic peak. We recommend that you set this parameter to 100 Mbps or higher . Bandwidth is measured in Mbps (megabits per second).

Parameter	Description
Packets Threshold	The alert threshold for the packet transmission rate in a data center. When the inbound or outbound packet transmission rate reaches this threshold, DDoS detection is triggered. Generally, this value is set to be slightly higher than the traffic peak. We recommend that you set this parameter to 20,000 pps or higher. The packet transmission rate is measured in packets per second (pps).
HTTP Requests Threshold	The alert threshold for the rate at which the servers in a data center receive HTTP requests. When the inbound or outbound HTTP request rate reaches this threshold, DDoS detection is triggered. Generally, this value is set to be slightly higher than the traffic peak. We recommend that you set this parameter to 100,000 QPS or higher. The HTTP request rate is measured in queries per second (QPS).

5. Click **Confirm**.

27.12.1.3.2 Manage anti-DDoS rules

This topic describes how to modify or delete anti-DDoS rules.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose **Network Security** > **Network Prevention** and click **Anti-DDoS Settings**.
3. In the **Actions** column of an anti-DDoS rule, manage the anti-DDoS rule.
   - Modify the anti-DDoS rule:

     Click **Modify**, enter the desired threshold values in the **Modify Anti-DDoS Rule** dialog box, and click **Confirm**.
   - Delete the anti-DDoS rule:

     Click **Delete** to delete the anti-DDoS rule.

     Note:
     The default anti-DDoS rules cannot be deleted.

28 Key Management Service (KMS)

28.1 What is KMS

Key Management Service (KMS) is a secure and easy-to-use key management service provided by Apsara Stack. KMS allows you to create and manage CMKs with ease and use a DEKs to encrypt your data.

KMS integrates many Alibaba Cloud products and services to help protect your data in the cloud.

Table 28-1: KMS solutions describes how KMS provides solutions for a variety of concerns and issues.

Table 28-1: KMS solutions

Role	Requirement	Solution
Application or website developer	• My program needs keys or certificates for encryption or signature, and I want secure and independent key management services. • I want to securely access keys regardless of where my application is deployed , and cannot take the risk of deploying plaintext keys elsewhere.	KMS provides envelope encryption, allowing you to store the Customer Master Key (CMK) in KMS and deploy only the EDKs. You can simply call a KMS API to decrypt DEKs only when necessary.
Service developer	• I do not want to be responsible for securing users keys and data. • I want users to manage their own keys. I want to use specified keys to encrypt their data after obtaining their authorization . In this way, I can focus on developing service features.	Envelop encryption and KMS APIs allow service developers to use specified CMKs to encrypt and decrypt DEKs. **`Plaintexts are not directly stored in a storage device.`** This method helps service developers manage CMKs.
Chief security officer (CSO)	• There are compliance requirements that I expect	KMS can connect to RAM for unified authorization management.

Role	Requirement	Solution
	our key management activities to meet. • I need to ensure that keys are reasonably authorized and that the use of any keys is audited.	

28.2 Log on to the KMS console

This topic describes how to log on to the KMS console.

Prerequisites

- Before logging on to the Apsara Stack console, make sure that you obtain the IP address or domain name address of the Apsara Stack console from the deployment personnel. The access address of the Apsara Stack console is http://*IP address or domain name address of the Apsara Stack console*/manage.
- We recommend that you use the Chrome browser.

Procedure

1. Open your browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.
3. Enter the correct username and password.
   - The system has a default super administrator with the username super and password super. The super administrator can create system administrators, and system administrators can create other system users and notify them of their default passwords by SMS or email.
   - You must change the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet the minimum complexity requirements, that is to be 8 to 20 characters in length and contain at least two types of the following characters: English uppercase/lowercase letters (A to Z or a to z), numbers (0 to 9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **LOGIN** to go to the **Dashboard** page.

5. In the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Key Management Service**.

28.3 Create a CMK

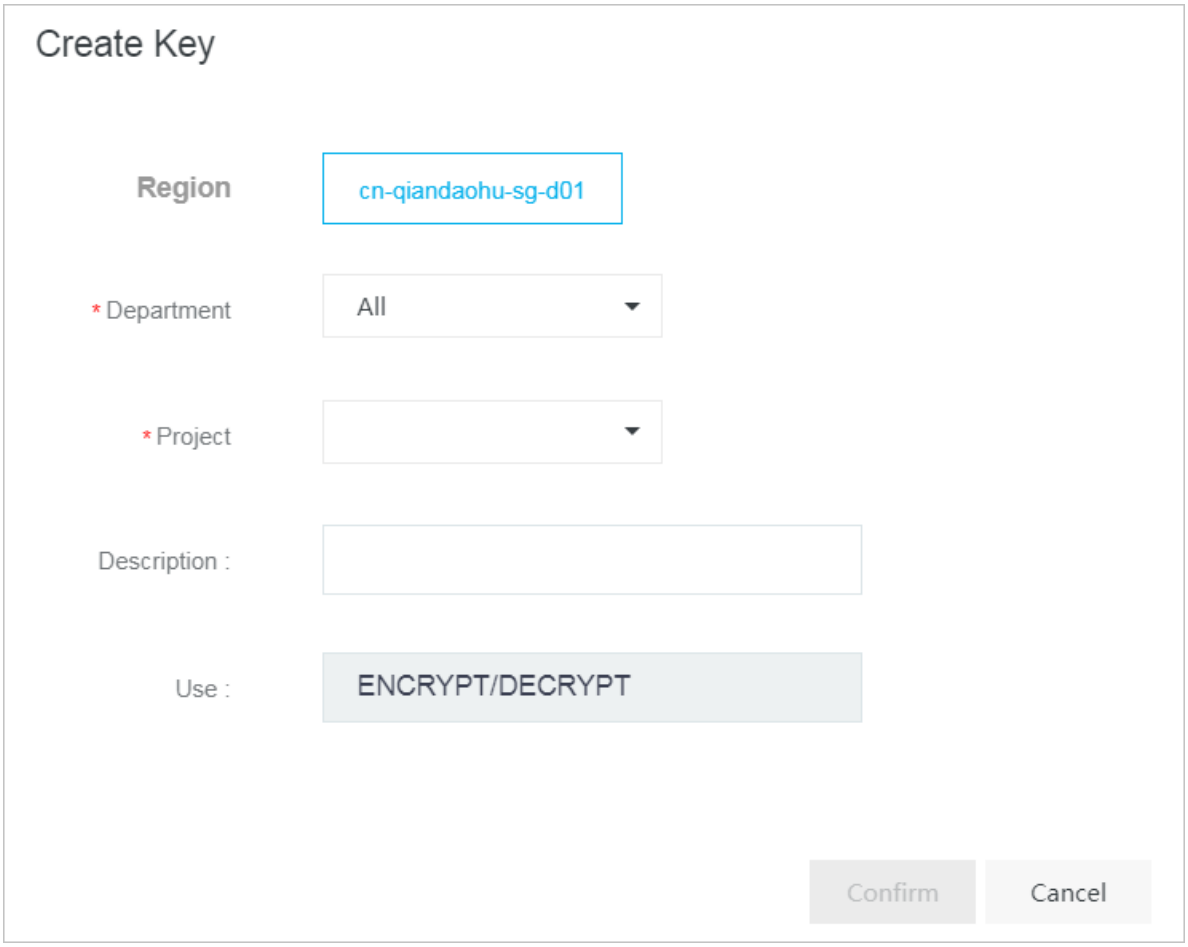
This topic describes how to create a CMK in Apsara Stack Management Console for subsequent encryption and decryption operations.

Procedure

1. *Log on to the KMS console*.
2. Click **Create Key**.

   The **Create Key** dialog box appears, as shown in *Figure 28-1: Create a CMK*.

   Figure 28-1: Create a CMK



3. Select a region, department, and project. Enter descriptive information. Then, click **OK**.

   After the CMK is created, call the KMS API by programming based on Scenarios and the *KMS Development Guide*.

28.4 View CMK details

After a CMK is created, you can view the key ID, key status, key purpose, and creator information.

Procedure

1. *Log on to the KMS console*.
2. In the CMK list, select a CMK that you want to view. Click the link of the key ID, or click the  icon and choose **Details** from the shortcut menu.

   The **Key Details** page appears.
3. In **Basic Information**, you can view the key ID, key status, key purpose, and creator information.

28.5 Enable a CMK

This topic describes how to enable a CMK.

Procedure

1. *Log on to the KMS console*.
2. Select a CMK that is in the **`Disabling`** state, click the  icon in the Actions column, and choose **Enable Key** from the shortcut menu.

   After the CMK is enabled, the CMK status changes from **`Disabling`** to **`Enabling`**.

28.6 Disable a CMK

If a CMK is disabled, it cannot be used for encryption or decryption. The ciphertext encrypted by using the CMK cannot be decrypted until the CMK is enabled again.

Context

After a CMK is created, it is in the **`Enabling`** state by default.

Procedure

1. *Log on to the KMS console*.
2. Select a CMK that is in the **`Enabling`** state, click the  icon, and choose **Disable Key** from the shortcut menu.

   The **Disable Key** message appears.
3. Click **OK** to disable the CMK.

   After the CMK is disabled, the CMK status changes from **`Enabling`** to **`Disabling`**.

28.7 Schedule a CMK to be deleted

You can schedule a CMK to be deleted after a specified period from 7 to 30 days.

Context

To delete a CMK, you must specify a scheduled period. The period ranges from 7 to 30 days.

You can set **`Cancel Key Deletion`** to cancel the CMK deletion before the scheduled period expires.

Note:

- Within the CMK scheduled deletion period, the CMK is in the Pending Deletion state and cannot be used for operations such as encryption, decryption, and DEK generation.
- Deleting a CMK has a severe impact on data availability. Typically, we recommend that you select *Disable a CMK*.
- A CMK cannot be recovered after it is deleted. The encrypted content along with the DEK generated by using the CMK cannot be decrypted. Therefore, you can schedule a CMK to be deleted instead of directly deleting it.
- KMS deletes the CMK within 24 hours after the scheduled period.

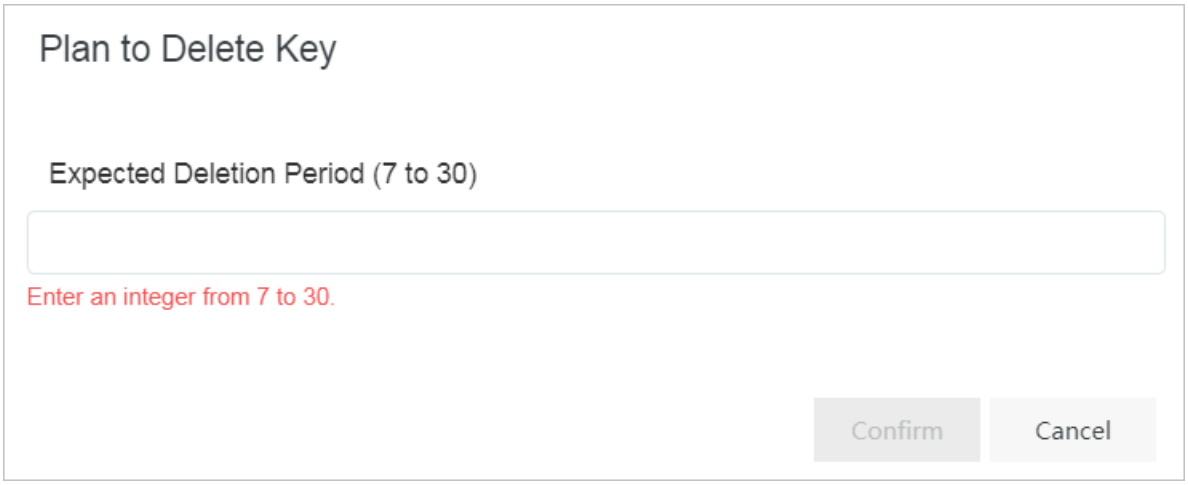
  For example, if you schedule a CMK to be deleted at 14:00, September 10, 2017, and the scheduled period is seven days, KMS deletes the CMK within 24 hours after 14:00, September 17, 2017.

Procedure

1. *Log on to the KMS console*.
2. Locate a CMK, click the  icon in the Actions column, and choose **Plan to Delete Key** from the shortcut menu.

   The **Plan to Delete Key** dialog box appears, as shown in *Figure 28-2: Schedule a CMK to be deleted*.

Figure 28-2: Schedule a CMK to be deleted



3. Enter the scheduled period (in days) in the text box, and click **OK**.

   Then the CMK status becomes **`Pending Deletion`**.

   If you want to cancel the scheduled deletion before the scheduled period expires, click the  icon in the Actions column and choose **Cancel Key Deletion** from the shortcut menu.