Alibaba Cloud
Apsara Stack Enterprise

User Guide - Cloud Essentials and Security

Version: 1911, Internal: V3.10.0

Issue: 20200317

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, consequent**

ial, 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.**

Document conventions

Style	Description	Example
	A danger notice 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.
	A warning notice 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 restart an instance.
	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	 **Notice:** If the weight is set to 0, the server no longer receives new requests.
	A note indicates supplemental instructions, best practices, tips, and other content.	 **Note:** You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click Settings > Network > Set network type.
Bold	Bold formatting is used for buttons, menus, page names, and other UI elements.	Click **OK**.
`Courier font`	Courier font is used for commands.	Run the `cd /d C:/window` command to enter the Windows system folder.
Italic	Italic formatting is used for parameters and variables.	`bae log list --instanceid` *`Instance_ID`*
[] or [a|b]	This format is used for an optional value, where only one item can be selected.	`ipconfig [-all|-t]`

Style	Description	Example
{} or {a|b}	**This format is used for a required value, where only one item can be selected.**	`switch {active|stand}`

Contents

1 ASCM console

1.1 What is the ASCM console?

The Apsara Stack Cloud Management (ASCM) console is a service capability platform based on the Alibaba Cloud Apsara Stack platform and designed for government and enterprise customers. This platform improves IT management and troubleshooting and is dedicated to providing a leading service capability platform of the cloud computing industry. It provides large-scale and cost-efficient end-to-end cloud computing and big data services for customers in industries such as government, education, healthcare, finance, and enterprise.

Overview

The ASCM console simplifies the management and deployment of physical and virtual resources by building an Apsara Stack platform that supports various business types of government and enterprise customers. The console helps you build your business systems in a simple and quick manner, fully improve resource utilization, and reduce O&M costs, allowing you to shift your focus from O&M to business. The console brings the Internet economy model to government and enterprise customers, and builds a new ecosystem chain based on cloud computing .

Workflow

ASCM console operations are divided into the following parts:

1. **System initialization: This part is designed to complete basic system configurat ions, such as creating organizations, resource sets, and users, creating basic resources such as VPCs, and creating contacts and contact groups in CloudMonit or.**
2. **Cloud resource creation: This part is designed to create resources as needed.**
3. **Cloud resource management: This part is designed to complete resource management operations, such as starting, using, and releasing resources and changing resource configurations.**

1.2 User roles and permissions

This topic describes roles and their permissions.

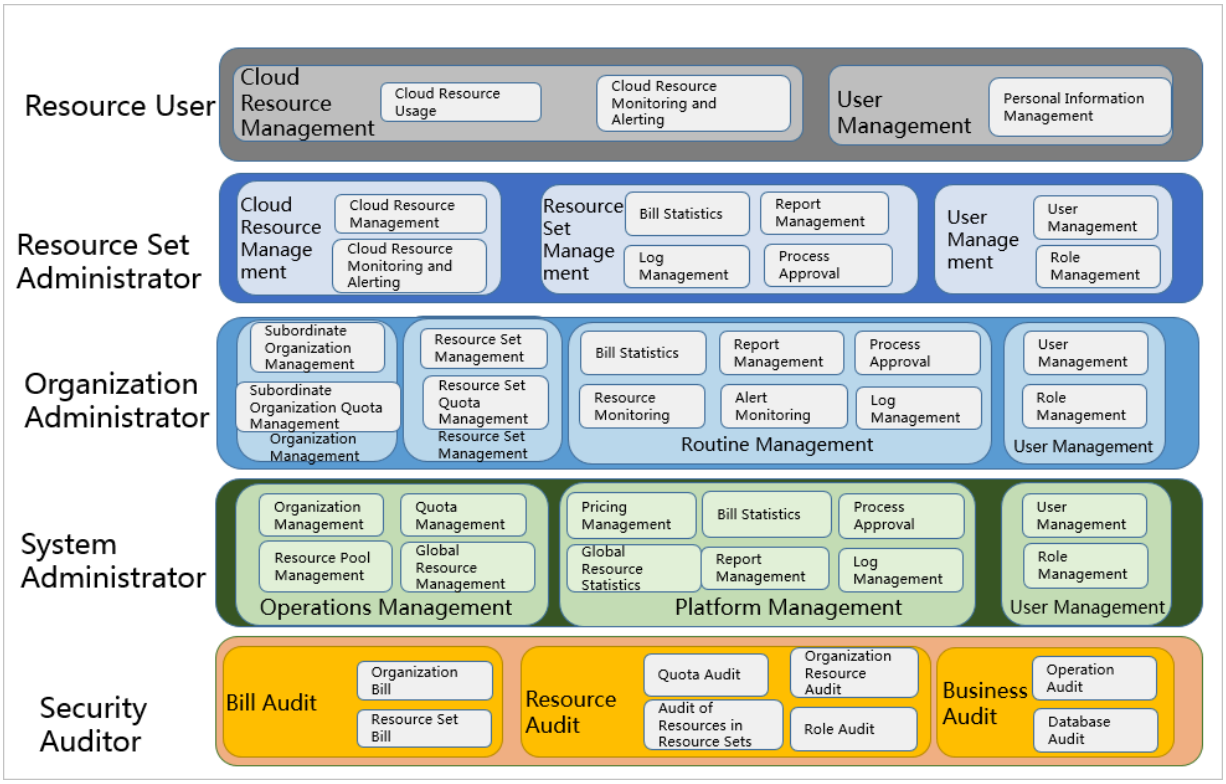


Table 1-1: Roles and permissions

Role name	Role permission
Resource user	This role has the permissions to view and modify resources in a resource set and create alarm rules.
Resource set administrator	This role has the permissions to create , modify, and delete resources in a resource set and manage the users of the resource set.
Organization administrator	This role has the permissions to manage an organization and its subordinate organizations, create, modify, and delete the resources of the organizations, create and view alarm rules for the resources, and export reports.
System administrator	This role has read and write permissions on all resources.

Role name	Role permission
Security auditor	This role has read-only permissions on all resources.
Super administrator	This role has the permissions to initialize the system and create system administrators.

1.3 Log on to the ASCM console

This topic describes how to log on to the Apsara Stack Cloud Management (ASCM) console.

Prerequisites

- Before logging on to the ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel. The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.
- We recommend that you use the Google Chrome browser.

Procedure

1. In the address bar, enter the URL used to access the ASCM console. Press Enter.
2. Enter your username and password.

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

   > **Note:**
   > When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).

3. Click Login to go to the ASCM console homepage.

1.4 Webpage introduction

The webpage of the ASCM console consists of the main menu bar, information area of the currently logged-on user, and operation area.

ASCM console page

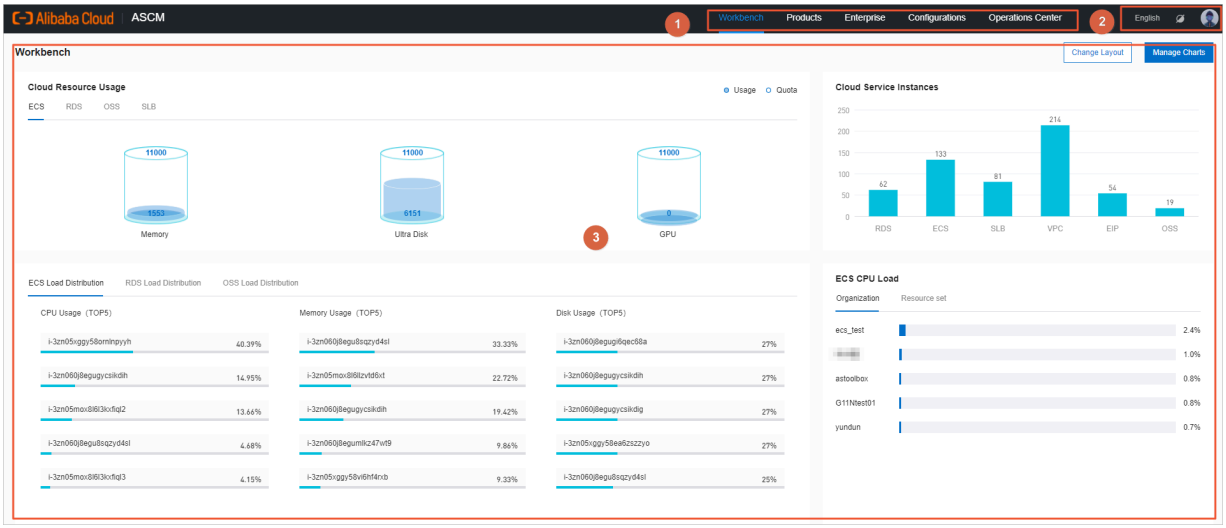
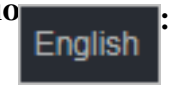
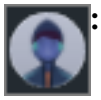


Table 1-2: Functional areas of the webpage

Area		Description
1	Main menu bar	It includes the following modules: · Home: uses charts to display the usage and monitoring data of existing system resources in each region. · Product: manages all types of basic cloud products and resources. · Enterprise: manages organizations, resource sets, roles, users, and logon policies for enterprises. · Configurations: manages system configurations, password policies, specifications, and message center. · Operations: manages the daily operations of cloud resources, including usage statistics, quotas, and . · Security: provides operation logs and system logs.

Area		Description
2	Information area of the currently logged-on user	•  : allows you to switch between English, simplified Chinese, and traditional Chinese. •  : allows you to switch between styles. •  : Move the pointer over the icon of the currently logged-on user. The User Information and Exit menu items are displayed. On the User Information page, you can perform the following operations: - View basic information. - Modify personal information. - Change the logon password. - View the AccessKey pair of your Apsara Stack tenant account.
3	Operation area	Operation area: the information display and operation area.

1.5 Initial configuration

1.5.1 Configuration description

Before using the Apsara Stack Cloud Management (ASCM) console, you must complete a series of basic configuration operations as an administrator, such as creating organizations, resource sets, users, and roles and initializing resources. This is the initial system configuration.

Based on the service-oriented principle, the ASCM console manages the organizations, resource sets, users, and roles of cloud data centers in a centralized manner to grant different resource access permissions to different users.

- **Organization**

  After the ASCM console is deployed, a root organization is automatically generated. You can create other organizations under the root organization.

  Organizations are displayed in a hierarchical structure. You can create subordinate organizations under each organization level.
- **Resource set**

  A resource set is a container used to store resources. Each resource must belong to a resource set.
- **User**

  A user is a resource manager and user.
- **Role**

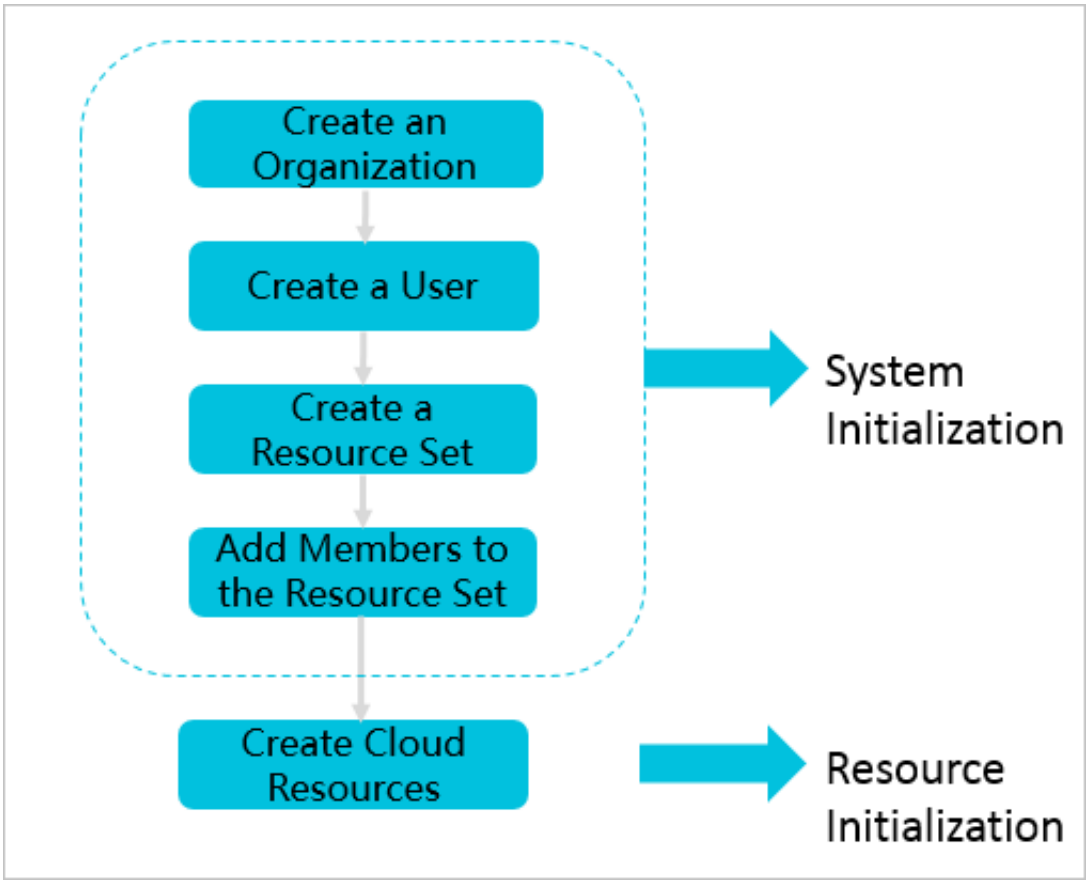
  A role is a set of access permissions. You can assign different roles to different users to meet different requirements for system access control.

The following table describes the relationships among organizations, resource sets, users, roles, and cloud resources.

1.5.2 Configuration process

This topic describes the initial configuration process.

Before using the Apsara Stack Cloud Management (ASCM) console, you must complete the initial system configurations as an administrator according to the process shown in the following figure.



1. *Create an organization*

   You can create organizations to store resource sets and their resources.

2. *Create a user*

   You can create users and assign different roles to different users to meet different requirements for system access control.

3. *Create a resource set*

   Before applying for resources, you must create a resource set.

4. *Add a member to a resource set*

   Add members to the resource set.

5. **Create cloud resources**

   You can create instances in each service console based on project requirements. For more information about how to create cloud service instances, see the detailed introduction of each cloud service.

1.6 Monitoring

1.6.1 View the workbench

The ASCM console uses charts to keep you updated on the current usage of resources.

Context

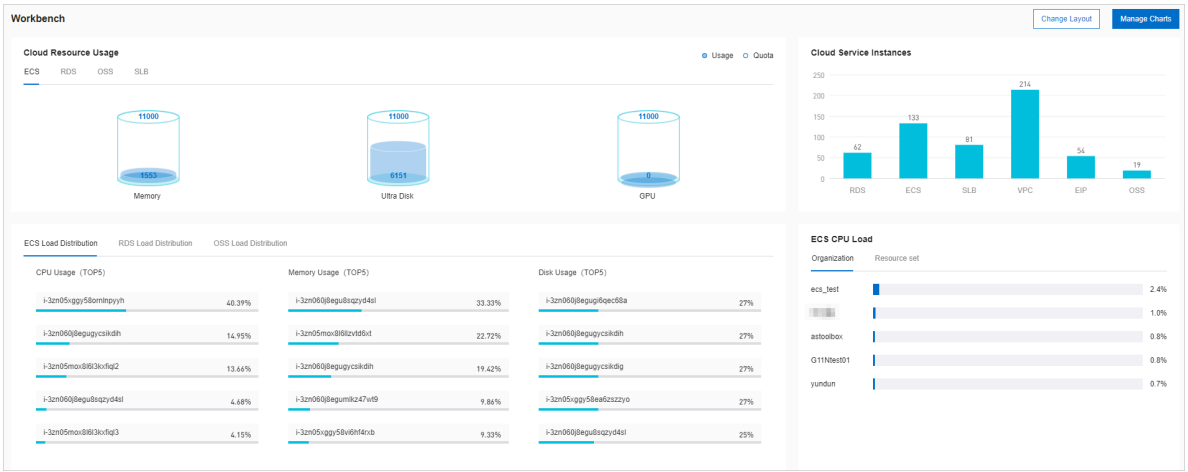
Note:

The resource types displayed vary with region types. See the actual page when using this document.

Procedure

1. *Log on to the ASCM console*.

   By default, the Workbench page appears when you log on to the ASCM console. To return to the Workbench page from other pages, click Home in the top navigation bar.



2. **On the Workbench page, you can view the instance summary information for all regions of the Apsara Stack environment.**

   The Workbench page consists of four modules. You can click Manage Charts in the upper-right corner of the page to select all or some modules to view relevant

information. You can also click Change Layout in the upper-right corner of the page and drag a specific module to the target location.

- Cloud Resource Usage

  Shows the usage and quota of ECS, ApsaraDB for RDS, OSS, and SLB resources for top-level organizations.
- Cloud Service Instances

  Performs statistics on the total number of instances created in all regions.
- Cloud Resource Load Distribution

  Shows the top five ECS, ApsaraDB for RDS, and OSS resources in terms of CPU, memory, and disk usage.
- ECS Load Distribution

  Performs statistics of the top five ECS instances in terms of memory and disk usage.

1.6.2 CloudMonitor

1.6.2.1 CloudMonitor overview

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

CloudMonitor can monitor metrics for ECS, ApsaraDB for RDS, SLB, OSS, and KVStore for Redis.

You can use the monitoring metrics of Apsara Stack services to configure alarm rules and notification policies. This way, you can keep up-to-date on the running status and performance of your service instances, and scale resources in a timely manner when resources are insufficient.

1.6.2.2 Metrics

This topic describes the metrics available for each service.

CloudMonitor checks the availability of a service based on the metrics for the service. You can configure alarm rules and notification policies for these metrics to stay up-to-date on the running status and performance of monitored service instances.

CloudMonitor can monitor resources of ECS, SLB, ApsaraDB for RDS, OSS, and KVStore for Redis. The following table lists the metrics for each service.

Metric	Description	Apsara Stack service	Calculation formula	Remarks
CPU utilization	Measures the CPU utilization of an ECS instance. Unit: %.	ECS	CPU utilization of an ECS instance/ Total CPU cores of the ECS instance	None
Memory usage	Measures the memory usage of an ECS instance. Unit: %.	ECS	Memory usage of an ECS instance /Total memory of the ECS instance	When you use the `free` or `top` command to query memory usage of a Linux ECS instance, you may find that the memory usage is inconsistent with the actual memory usage displayed in the Apsara Stack Cloud Management (ASCM) console. This is because when CloudMonitor is calculating the memory usage, it does not take the amount of memory used as cache into account.
Disk I/O read	Measures the volume of data read from an ECS instance disk per second. Unit: KB/s.	ECS	Total bytes read from an ECS instance disk/ Statistical period	For a Linux ECS instance, the disk I/O data can be obtained by using the iostat tool. If you find that the Linux ECS instance does not have any disk I/O data, check whether iostat has been installed in your ECS instance. If not, Redhat and CentOS users can use yum to install the tool , while Ubuntu and Debian users can use apt-get to install the tool.

Metric	Description	Apsara Stack service	Calculation formula	Remarks
Disk I/O write	Measures the volume of data written to an ECS instance disk per second. Unit: KB/s.	ECS	Total bytes written to an ECS instance disk/ Statistical period	None
Disk usage	Measures the disk usage of an ECS instance. Unit: %.	ECS	Used capacity of an ECS instance disk/Total capacity of the ECS instance disk	None
Inbound traffic	Measures the inbound network traffic to an ECS instance per second. Unit: Kbit/s.	ECS	None	None
Outbound traffic	Measures the outbound network traffic from an ECS instance per second. Unit: Kbit/s.	ECS	None	When the purchased bandwidth is used up, access will fail or requests will slow down. On the monitoring chart, eth0 indicates the internal NIC of the ECS instance, and eth1 indicates the public NIC of the ECS instance.

Metric	Description	Apsara Stack service	Calculation formula	Remarks
Number of TCP connections	Measures the total number of TCP connections established by an ECS instance.	ECS	None	None
Number of processes	While an alarm rule is configured with this metric, the number of specified processes is counted and the total number displayed.	ECS	None	To monitor the running conditions of processes in an ECS instance, configure an alarm rule with this metric to trigger the alarm when the number of running processes does not equal the actual number of processes.
Average load	Measures the average load value of a Linux ECS instance.	ECS	None	The average load value cannot be greater than 1. If your ECS instance has a multi-core processor, the average load value must be divided by the number of CPU cores and the obtained value must be smaller than 1. If the average load value is greater than 1, processes are queued up and the ECS instance slows down.

Note:

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

Installation method: In the ECS instance list on the CloudMonitor page, select the instance to be monitored, and click Batch Install below the instance list.

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

Table 1-3: Metrics for ApsaraDB for RDS

Metric	Description	Apsara Stack service	Calculation formula
CPU utilization	Measures the CPU utilization of an RDS instance. Unit: %.	ApsaraDB for RDS	CPU utilization of an RDS instance/Total CPU cores of the RDS instance
Memory usage	Measures the memory usage of an RDS instance. Unit: %.	ApsaraDB for RDS	Used memory of an RDS instance/Total memory of the RDS instance
Disk usage	Measures the disk usage of an RDS instance. Unit: %.	ApsaraDB for RDS	None
IOPS utilization	Measures the number of I/O requests for an RDS instance per second. Unit: %.	ApsaraDB for RDS	Number of I/O requests for an RDS instance/Statistical period
Connection utilization	Measures the number of connections between an application and an RDS instance per second. Unit: %.	ApsaraDB for RDS	Number of connections between an application and an RDS instance per second/Statistical period

Table 1-4: Metrics for SLB

Metric	Description	Apsara Stack service	Remarks
Number of outbound packets sent over a port per second	Measures the number of packets sent by an SLB instance per second.	SLB	None

Metric	Description	Apsara Stack service	Remarks
Number of inbound packets received over a port per second	Measures the number of packets received by an SLB instance per second.	SLB	None
Amount of data received over a port per second	Measures the traffic consumed to access an SLB instance from the Internet. Unit: Kbit/s.	SLB	None
Amount of data sent over a port per second.	Measures the traffic consumed by an SLB instance to access the Internet. Unit: Kbit/s.	SLB	None
Number of active connections on a port	Measures the number of connections in the ESTABLISHED state.	SLB	It can be interpreted as, but cannot be equivalent to, the concurrent connections. If persistent connections are used, a connection can transfer multiple file requests at one time.
Number of inactive connections on a port	Measures the number of all TCP connections except connections in the ESTABLISHED state.	SLB	None

Metric	Description	Apsara Stack service	Remarks
Number of new connections on a port	Measures the average number of times that the status is SYN_SENT at first for a TCP three-way handshake in the statistical period.	SLB	Number of New Port Connections, Number of Active Port Connections, and Number of Inactive Port Connections are all used to measure the number of requests for connecting a client to an SLB instance.

Table 1-5: Metrics for OSS

Metric	Description	Apsara Stack service
Number of reads	Measures the number of reads of an OSS bucket.	OSS
Number of internal errors	Measures the number of errors that occur on an OSS bucket.	OSS
Inbound traffic from the Internet	Measures the inbound traffic from the Internet to an OSS bucket per second. Unit: bytes.	OSS
Outbound traffic to the Internet	Measures the outbound traffic from an OSS bucket to the Internet per second. Unit: bytes.	OSS
Inbound traffic from the internal network	Measures the inbound traffic from the internal network to an OSS bucket per second. Unit: bytes.	OSS
Outbound traffic to the internal network	Measures the outbound traffic from an OSS bucket to the internal network per second. Unit: bytes.	OSS
Number of writes	Measures the number of writes of an OSS bucket.	OSS
Used storage space	Measures the used storage space of an OSS bucket. Unit: bytes.	OSS

Metric	Description	Apsara Stack service
Number of successful requests	Measures the number of requests with an HTTP status code of 200.	OSS
Valid request percentage	Measures the percentage of valid requests to all requests.	OSS
Number of valid requests	Measures the number of valid requests.	OSS
Inbound and outbound network traffic	Measures the inbound and outbound traffic on the Internet and internal network.	OSS
Number of requests with server-side errors	Measures the number of requests with server-side errors.	OSS
Percentage of requests with server-side errors	Measures the percentage of total requests whose status code is 4xx.	OSS
Percentage of requests with network errors	Measures the percentage of total requests with network errors.	OSS
Number of requests with client-side errors	Measures the number of requests with client-side errors.	OSS
Number of GET requests	Measures the number of HTTP GET requests.	OSS
Number of PUT requests	Measures the number of HTTP PUT requests.	OSS
Maximum latency of GetObject requests	Measures the maximum allowable latency of GetObject requests.	OSS
Maximum latency of HeadObject requests	Measures the maximum allowable latency of HeadObject requests.	OSS

Metric	Description	Apsara Stack service
Maximum latency of PutObject requests	Measures the maximum allowable latency of PutObject requests.	OSS
Maximum latency of PostObject requests	Measures the maximum allowable latency of PostObject requests.	OSS

Table 1-6: Metrics for KVStore for Redis

Metric	Description	Apsara Stack service	Unit
CPU utilization	Measures the CPU utilization of a KVStore for Redis instance.	KVStore for Redis	%
Memory usage	Measures the percentage of used memory to total memory.	KVStore for Redis	%
Used memory	Measures the amount of memory currently in use.	KVStore for Redis	Bytes
Number of used connections	Measures the total number of client connections.	KVStore for Redis	N/A
Percentage of used connections	Measures the percentage of total connections that are used.	KVStore for Redis	%
Write bandwidth	Measures the write traffic per second.	KVStore for Redis	Bytes/s
Read bandwidth	Measures the read traffic per second.	KVStore for Redis	Bytes/s
Number of failed operations per second	Measures the number of failed operations on a KVStore for Redis instance per second.	KVStore for Redis	N/A
Write bandwidth usage	Measures the percentage of total bandwidth used by write operations.	KVStore for Redis	%

Metric	Description	Apsara Stack service	Unit
Read bandwidth usage	Measures the percentage of total bandwidth used by read operations.	KVStore for Redis	%
Used QPS	Measures the number of queries per second (QPS).	KVStore for Redis	N/A
QPS usage	Measures the QPS utilization rate.	KVStore for Redis	%
Average response time	Measures the average response time.	KVStore for Redis	Milliseconds
Maximum response time	Measures the maximum response time.	KVStore for Redis	Milliseconds
Number of failed commands	Measures the number of failed commands.	KVStore for Redis	N/A
Hit rate	Measures the current hit rate.	KVStore for Redis	%
Inbound traffic	Measures the inbound traffic to a KVStore for Redis instance.	KVStore for Redis	Bytes
Inbound bandwidth usage	Measures the inbound bandwidth usage of a KVStore for Redis instance.	KVStore for Redis	%
Outbound traffic	Measures the outbound traffic from a KVStore for Redis instance.	KVStore for Redis	Bytes
Outbound bandwidth usage	Measures the outbound bandwidth usage of a KVStore for Redis instance.	KVStore for Redis	%

1.6.2.3 View monitoring charts

You can view monitoring charts to obtain up-to-date information about each instance.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, choose Products > Monitoring and O&M > CloudMonitor.**

3. In the left-side navigation pane of the CloudMonitor page, click Cloud Service Monitoring.
4. Click a cloud service.
5. Click Monitoring Charts in the Actions column corresponding to an instance.

   On the Monitoring Charts page that appears, you can select a date and time to view the monitoring data of each metric.

1.6.3 Alerts

1.6.3.1 View alarm overview

On the Overview page in CloudMonitor, you can view the alarm status statistics and alarm logs.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, choose Products > Monitoring and O&M > CloudMonitor.
3. In the left-side navigation pane of the CloudMonitor page, click Overview.
4. On the Overview page, view the alarm status statistics and alarm logs generated in the last 24 hours.

1.6.3.2 View alarm logs

You can view alarm information to stay up-to-date on the running status of ECS, ApsaraDB for RDS, SLB, OSS, and KVStore for Redis.

Context

Alarm information contains information for all of the items that do not comply with your configured alarm rules.

Note:

- **A maximum of one million alarm items generated within the last three months can be retained.**
- **This topic describes how to view alarm information for ECS. You can view the alarm information for other cloud resources in a similar manner.**

Procedure

1. *Log on to the ASCM console* as an administrator.

2. In the top navigation bar, choose Products > Monitoring and O&M > CloudMonitor.
3. In the left-side navigation pane of the CloudMonitor page, choose Alarms > Alarm Logs.
4. On the Alarm Rule History List page, you can filter alarm information by rule ID, rule name, product, metric, and date.

   The following table describes the fields in the query result.

   Table 1-7: Alarm information fields

Field	Description
Product	The product for which the alarm was triggered.
Fault Instance	The instance for which the alarm was triggered.
Occurred At	The time when the alarm was triggered.
Rule Name	The name of the alarm rule.
Status	The status of the alarm rule.
Notification Contact	The recipient of the alarm notification.

1.6.3.3 Alarm rules

1.6.3.3.1 Query alarm rules

After creating alarm rules, you can view your alarm rules on the Alarm Rules page.

Context

The system provides alarm rules for ECS, ApsaraDB for RDS, SLB, OSS, and KVStore for Redis.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, choose Products > Monitoring and O&M > CloudMonitor.
3. In the left-side navigation pane of the CloudMonitor page, click Cloud Service Monitoring.
4. Click a cloud service.

5. Click Alarm Rules in the Actions column corresponding to an instance to go to its Alarm Rules page.

   On the Alarm Rules page that appears, view the detailed alarm rule information.

1.6.3.3.2 Create an alarm rule

You can create an alarm rule to monitor an instance.

Prerequisites

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

The installation method is as follows:

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, choose Products > Monitoring and O&M > CloudMonitor.
3. In the left-side navigation pane, choose Cloud Service Monitoring > ECS.
4. In the ECS instance list, select the instances that you want to monitor, and click Batch Install.

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

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, choose Products > Monitoring and O&M > CloudMonitor.
3. In the left-side navigation pane of the CloudMonitor page, click Cloud Service Monitoring.
4. Click a cloud service.
5. Click Alarm Rules in the Actions column corresponding to an instance to go to its Alarm Rules page.

   > **Note:**
   > You can also use the search function to query specific instances for which you want to create alarm rules.

6. Click Create Alarm Rule.

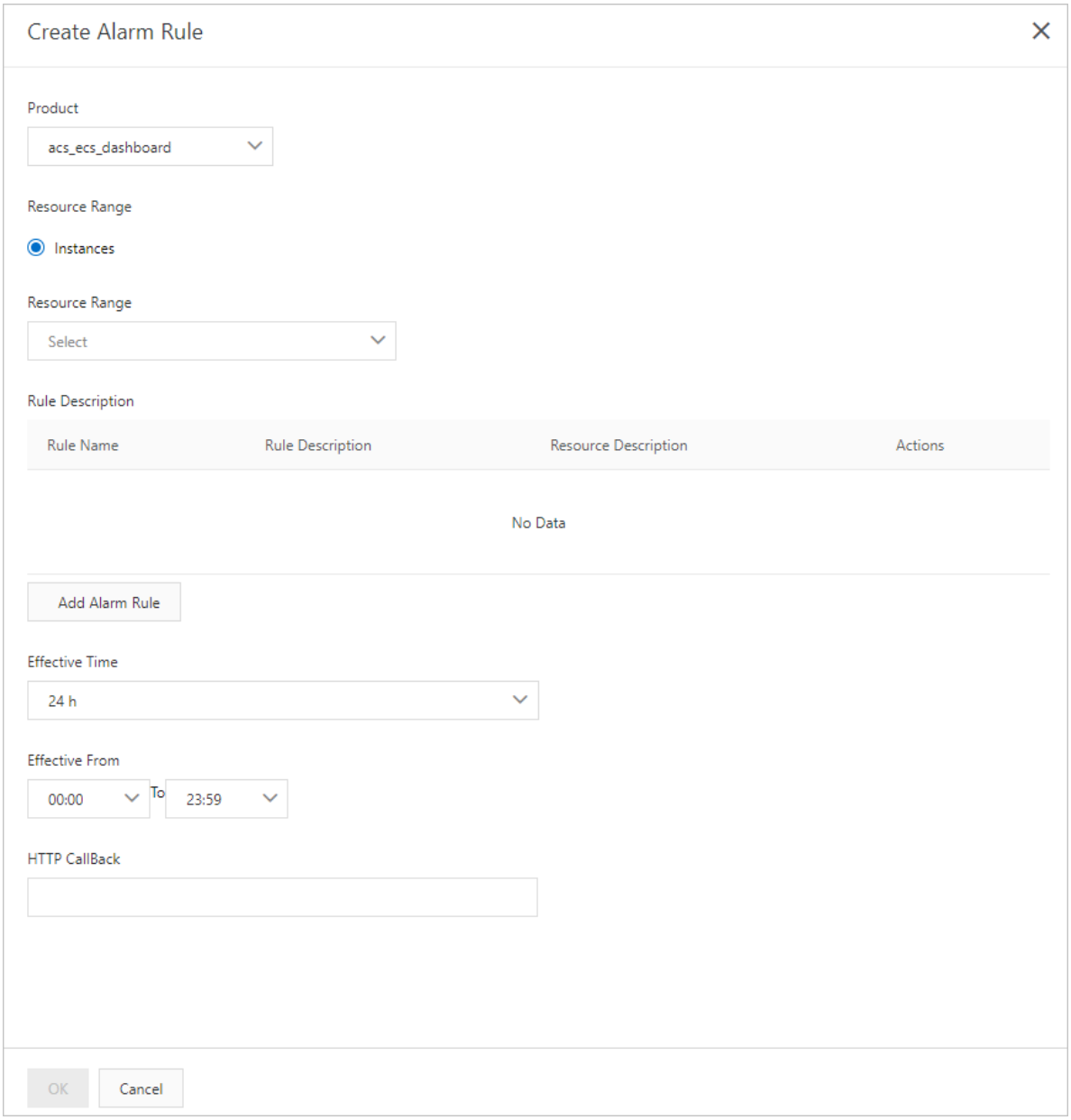


Table 1-8: Alarm rule creation parameters

Parameter	Description
Product	The monitored cloud service. Currently monitored cloud services include ECS, ApsaraDB for RDS, SLB, and KVStore for Redis.
Resource Range	The range of resources associated with the alarm rule. Set Resource Range to Instances. Application groups and users will be supported in the future.
Description	The description of the alarm rule.

Parameter	Description
Effective Time	Only a single alarm is sent during each mute duration, even if the metric value exceeds the alarm rule threshold several consecutive times. Unit: seconds. Default value: 86400 (one day). Minimum value: 3600 (one hour).
Effective From	An alarm is sent only when the threshold is met during the effective period.
HTTP Callback	The callback URL when the alarm conditions are met.

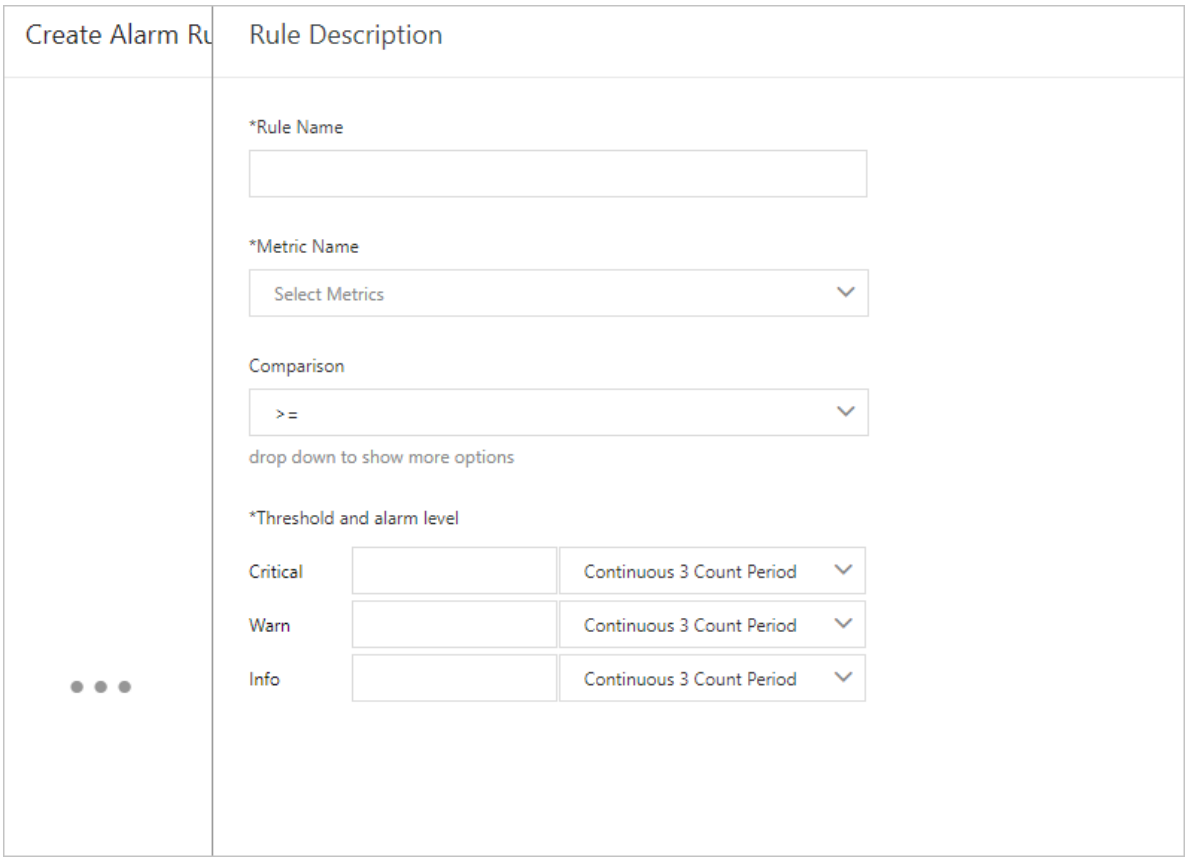


Table 1-9: Rule description configuration

Parameter	Description
Rule Name	The name of the alarm rule, which must be easy to distinguish. The name must be 1 to 64 characters in length and can contain letters and digits.
Metric Name	Different products have different monitoring metrics. For more information, see *Metrics*.

Parameter	Description
Comparison	The comparison between thresholds and observed values. The comparison operators include >, > =, <, and <=. When a comparison rule is met, an alarm rule is triggered.
Threshold and alarm level	Different metrics have different reference thresholds.

7. Click OK.

1.6.3.3.3 Modify an alarm rule

You can modify alarm rules as needed.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, choose Products > Monitoring and O&M > CloudMonitor.
3. In the left-side navigation pane of the CloudMonitor page, click Cloud Service Monitoring.
4. Click a cloud service.
5. Click Alarm Rules in the Actions column corresponding to an instance to go to its Alarm Rules page.
6. Select the alarm rule that you want to modify, and click Modify in the Actions column.

   For more information about how to create alarm rules, see *Create an alarm rule*.

1.6.3.3.4 Disable an alarm rule

You can disable one or more alarm rules as needed.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, choose Products > Monitoring and O&M > CloudMonitor.
3. In the left-side navigation pane of the CloudMonitor page, click Cloud Service Monitoring.
4. Click a cloud service.
5. Click Alarm Rules in the Actions column corresponding to an instance to go to its Alarm Rules page.

6. Select the alarm rule that you want to disable, and click Disable below the alarm rule list.
7. In the message that appears, click OK.

1.6.3.3.5 Enable an alarm rule

After an alarm rule is disabled, it can be re-enabled as needed.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, choose Products > Monitoring and O&M > CloudMonitor.
3. In the left-side navigation pane of the CloudMonitor page, click Cloud Service Monitoring.
4. Click a cloud service.
5. Click Alarm Rules in the Actions column corresponding to an instance to go to its Alarm Rules page.
6. Select the alarm rule that you want to enable, and click Enable below the alarm rule list.
7. In the message that appears, click OK.

1.6.3.3.6 Delete an alarm rule

You can delete alarm rules that are no longer needed.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, choose Products > Monitoring and O&M > CloudMonitor.
3. In the left-side navigation pane of the CloudMonitor page, click Cloud Service Monitoring.
4. Click a cloud service.
5. Click Alarm Rules in the Actions column corresponding to an instance to go to its Alarm Rules page.
6. Select the alarm rule that you want to delete and click Delete in the Actions column.
7. In the message that appears, click OK.

1.7 Enterprise

1.7.1 Organizations

1.7.1.1 Create an organization

You can create organizations to store resource sets and their resources.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Organizations.**
4. **In the organization navigation tree, move the pointer over the name of a parent organization, and click on the right.**
5. **Choose Add Organization from the shortcut menu.**
6. **In the dialog box that appears, enter an organization name and click OK.**

1.7.1.2 Query an organization

You can query an organization by name to view its resource sets, users, and user groups.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Organizations.**
4. **In the search box below Organizations, enter an organization name to query information about the corresponding organization.**

1.7.1.3 Modify organization information

An administrator can modify the information of an organization.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Organizations.**

4. In the organization navigation tree, move the pointer over the name of an organization, and click on the right.
5. Choose Edit Organization from the shortcut menu.
6. In the dialog box that appears, modify the name of the organization and click OK.

1.7.1.4 Delete an organization

Administrators can delete organizations that are no longer needed.

Prerequisites

> **Note:**
> Before deleting an organization, make sure that the organization does not contain any users, resource sets, or subordinate organizations. Otherwise, the organization cannot be deleted.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Organizations.
4. In the organization navigation tree, move the pointer over the name of an organization, and click on the right.
5. Choose Delete Organization from the shortcut menu.
6. In the message that appears, click OK.

1.7.1.5 Obtain the AccessKey pair of an organization

An administrator can obtain the AccessKey pair of an organization.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Organizations.
4. In the organization navigation tree, move the pointer over the name of the target organization and click on the right.
5. Choose AccessKey from the shortcut menu.
6. In the message that appears, view the AccessKey information of the organization.

1.7.2 Resource sets

1.7.2.1 Create a resource set

Before applying for resources, you must create a resource set.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Resource Sets.**
4. **In the upper-right corner of the page, click Create.**
5. **In the Create Resource Set dialog box that appears, set Name and Organization.**
6. **Click OK.**

1.7.2.2 View the details of a resource set

When you need to use a cloud resource in your organization, you can view the details of the resource set that contains the resource, including all resource instances and members of the resource set.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Resource Sets.**
4. **Select an organization from the Organization drop-down list, or enter a resource set name in the search bar.**
5. **Click Search.**
6. **Click the name of the resource set that you want to view to go to the Resource Set Details page.**

   Click the Resources and Members tabs to view information about all resource instances and members of the resource set.

1.7.2.3 Modify the name of a resource set

An administrator can modify the name of a resource set to keep it up-to-date.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**

3. In the left-side navigation pane of the Enterprise page, click Resource Sets.
4. Click More in the Actions column corresponding to a resource set, and choose Edit Name from the shortcut menu.
5. In the dialog box that appears, enter the new name.
6. Click OK.

1.7.2.4 Add a member to a resource set

You can add a member to a resource set so that the member can use the resources in the resource set.

Prerequisites

Before adding a member, make sure that the following prerequisites are met:

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

Context

Members of a resource set have the permissions to use resources in the resource set .

Deleting resources from a resource set does not affect the members of the resource set. Similarly, deleting members from a resource set does not affect the resources in the resource set.

You can delete a member that is no longer in use in a resource set. After the member is deleted, it will no longer be able to access the resource set.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Resource Sets.
4. Click More in the Actions column corresponding to a resource set, and choose Add Member from the shortcut menu.
5. In the dialog box that appears, select a username.
6. Click OK.

1.7.2.5 Delete a resource set

Administrators can delete resource sets that are not needed.

Context

Ensure that the resource set to be deleted does not contain any members or resources.

Notice:

A resource set cannot be deleted if it contains resources or members.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Resource Sets.**
4. **Click More in the Actions column corresponding to a resource set, and choose Delete from the shortcut menu.**
5. **In the message that appears, click OK.**

1.7.3 Roles

1.7.3.1 Create a custom role

You can add custom roles in the Apsara Stack Cloud Management (ASCM) console to more efficiently grant permissions to users so that different personnel can work with different functions.

Context

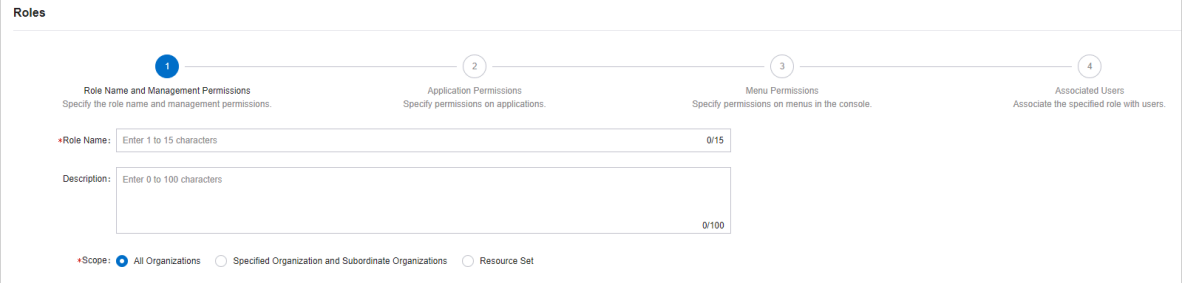
A role is a set of access permissions. Each role has a range of permissions. A user can have multiple roles, which means that the user is granted all the permissions defined for these roles. A role can be used to grant the same permissions to a group of users.

Before adding a custom role, note that the total number of custom and default roles cannot exceed 20.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**

3. In the left-side navigation pane of the Enterprise page, click Roles.
4. In the upper-right corner of the page, click Create Custom Role.
5. On the Roles page that appears, set the role name and management permissions.



The following table describes the role creation parameters.

Table 1-10: Role creation parameters

Parameter	Description
Role Name	The name of the role. The name can be up to 15 characters in length and can contain only letters and digits.
Description	Optional. The description of the role. The description can be up to 100 characters in length and can contain letters, digits, commas (,), semicolons (;), and underscores (_).
Management Permissions	All Organizations The permissions apply to all organizations involved. · Specified Organization and Subordinate Organizations The permissions apply to the organization to which the user belongs and its subordinate organizations. · Resource Set The permissions apply to the resource sets assigned to the user.

6. Set the parameters in the Application Permissions, Menu Permissions, and Associated Users steps in sequence.

Note:
The system automatically selects the required permissions that the specified operation depends on. Removing the dependency may cause the operation to fail.

1.7.3.2 View the details of a role

If you are not certain about the specific permissions of a role, go to the Roles page to view the role permissions.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Roles.**
4. **Click the name of the role that you want to view. On the Roles page, view information about the role.**

1.7.3.3 Modify custom role information

An administrator can modify the name and permissions of a custom role.

Context

> **Note:**
> **Information about preset roles cannot be modified.**

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Roles.**
4. **In the role name list, click More in the Actions column corresponding to a custom role, and choose Modify from the shortcut menu to go to the Roles page.**
5. **You can modify the name, permissions, and associated users of a custom role.**
   - **Procedure of modifying a role name: Move the pointer over the role name and click to enter a new role name.**
   - **Procedure of modifying permissions: Click the Management Permissions, Application Permissions, or Menu Permissions tab, select or remove related permissions from the corresponding tab, and then click Update.**
   - **Procedure of binding a user to a role: Click the Associated Users tab, and select a user from the Select one or more users drop-down list to add the user. To unbind the user from the role, click Remove.**

1.7.3.4 Delete a custom role

You can delete custom roles that are no longer needed.

Context

> **Note:**
> **Default or preset roles cannot be deleted.**

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Roles.**
4. **Click More in the Actions column corresponding to a role, and choose Delete from the shortcut menu.**
5. **In the message that appears, click OK.**

1.7.4 Users

1.7.4.1 System users

1.7.4.1.1 Create a user

An administrator can create a user and assign the user different roles to meet different requirements for system access control.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Users.**
4. **Click Create.**
5. **In the dialog box that appears, configure the parameters.**

Parameter	Description
Username	The Apsara Stack account name of the user. The name must be 3 to 30 characters in length and can contain letters, digits, hyphens (-), underscores (_), periods (.), and at signs (@). It must start with a letter or digit.

Parameter	Description
Display Name	The display name of the user. The name must be 2 to 30 characters in length and can contain letters, digits, hyphens (-), underscores (_), periods (.), and at signs (@).
Roles	The roles for the user.
Organization	The organization to which the user belongs.
Logon Policy	The logon policy that restricts the logon time and IP addresses of the user. By default, the default policy is bound to new 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 modify the user's logon policy or create a logon policy for the user. For more information, see *Create a logon policy*.
Mobile Number	The mobile number of the user. The mobile number is used by the system to notify users of resource application and usage. Make sure that the entered mobile number is correct.  Note: If the mobile number is changed, be sure to update it on the system in a timely manner.
Landline Number	Optional. The landline number of the user. It must be 4 to 20 characters in length and can contain only digits (0 to 9) and hyphens (-).
Email	The email address of the user. The email address is used by the system to notify users of resource application and usage. Make sure that the entered email address is correct.  Note: If the email address is changed, be sure to update it on the system in a timely manner.

6. Click OK.

1.7.4.1.2 Query a user

You can view user information such as name, organization, mobile number, email address, role, logon time, and initial password.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Users.**
4. **Click the System Users tab.**
5. **Set Username, Organization, or Role, and then click Search.**
6. **Click More in the Actions column corresponding to a user, and choose User Information from the shortcut menu to view basic information about the user.**

1.7.4.1.3 Modify user information

You can modify user information such as display name, mobile number, and email address to keep it up-to-date.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Users.**
4. **Click the System Users tab.**
5. **Click More in the Actions column corresponding to a user, and choose Edit from the shortcut menu.**
6. **In the Modify User Information dialog box, enter the relevant information and click OK.**

1.7.4.1.4 Change user roles

You can add, change, and delete roles for a user.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Users.**
4. **Click the System Users tab.**

5. **Click More in the Actions column corresponding to a user, and choose Authorize from the shortcut menu.**
6. **In the `Role` field, add, delete, or change user roles as needed.**
7. **Click OK.**

1.7.4.1.5 Modify a user logon policy

An administrator can modify a user's logon policy to restrict the permitted logon time and IP addresses of the user.

Prerequisites

A new logon policy is created. For more information about how to create a logon policy, see *Create a logon policy*.

Modify a user logon policy

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Users.**
4. **Click the System Users tab.**
5. **Click More in the Actions column corresponding to a user, and choose Logon Policy from the shortcut menu.**
6. **In the Assign Logon Policy dialog box, select a logon policy and click OK.**

Modify multiple user logon policies at a time

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Users.**
4. **Click the System Users tab.**
5. **Select multiple users.**
6. **In the upper-right corner of the page, click Logon Policy.**
7. **In the Assign Logon Policies dialog box, select a logon policy.**

1.7.4.1.6 View the initial password of a user

After a user is created, the system automatically generates an initial password for the user.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Users.
4. Click the System Users tab.
5. Select the user for which you want to view the initial password.
6. You can use one of the following methods to view the initial password of a user:
   - Click View Initial Password in the upper-right corner of the Users page to view the initial password.
   - Click More in the Actions column corresponding to the user, and choose User Information from the shortcut menu. On the user information page, click View Password to view the initial password.

1.7.4.1.7 Reset the password of a user

If users forget their logon passwords, the system administrator can reset the logon passwords for them.

Prerequisites

The logon password of a user can be reset by only the user and those who have the permissions to create resource sets for the user.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Users.
4. Click the System Users tab.
5. Select the user for which you want to reset the password.
6. Click More in the Actions column corresponding to the user, and choose User Information from the shortcut menu.
7. Click Reset Password.

   After the password is reset, a message is displayed, indicating that the password has been reset. If you want to view the initial password after password reset, click View Password.

1.7.4.1.8 Disable and enable a user

You can disable a user to prevent the user from logging on to the Apsara Stack Cloud Management (ASCM) console. Disabled users must be re-enabled before they can log on to the ASCM console again.

Context

By default, users are enabled when they are created.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Users.**
4. **Click the System Users tab.**
5. **You can perform the following operations on the current tab:**
   - **Select a user whose Status is Enabled, click More in the Actions column, and choose Disable from the shortcut menu to disable the user.**
   - **Select a user whose Status is Disabled, click More in the Actions column, and choose Enable from the shortcut menu to enable the user.**

1.7.4.1.9 Delete a user

An administrator can delete a specified user as needed.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Users.**
4. **Click the System Users tab.**
5. **Click More in the Actions column corresponding to a user, and choose Delete from the shortcut menu.**

1.7.4.2 Historical users

1.7.4.2.1 Query historical users

You can check whether a user has been deleted, or quickly find and restore the user.

Procedure

1. *Log on to the ASCM console* **as an administrator.**

2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Users.
4. Click the Historical Users tab.
5. Enter the username that you want to query in the Username search box.
6. Click Search.

1.7.4.2.2 Restore historical users

An administrator can restore a deleted user account from the Historical Users tab.

Context

The basic information such as logon password of a restored user will be the same as it was before the user was deleted, except for the organization and role.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Users.
4. Click the Historical Users tab.
5. Find the user to be restored and click Restore in the Actions column.
6. In the Restore User dialog box that appears, select an organization and a role.
7. Click OK.

1.7.5 Logon policies

1.7.5.1 Create a logon policy

To improve the security of the Apsara Stack Cloud Management (ASCM) console, an administrator can create a logon policy to control the logon time and IP addresses of a user.

Context

Logon policies are used to control the time period and IP addresses for users to log on. After a user is bound to a logon policy, the user's logons will be restricted based on the logon time and IP addresses specified in the policy.

When providing services, the ASCM console automatically generates a default policy without restrictions on the logon time and IP addresses. The default policy cannot be deleted.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Logon Policies.
4. In the upper-right corner of the page, click Create.
5. In the Create Logon Policy dialog box that appears, set the Name, Policy Properties, Time Period, and IP Address fields.

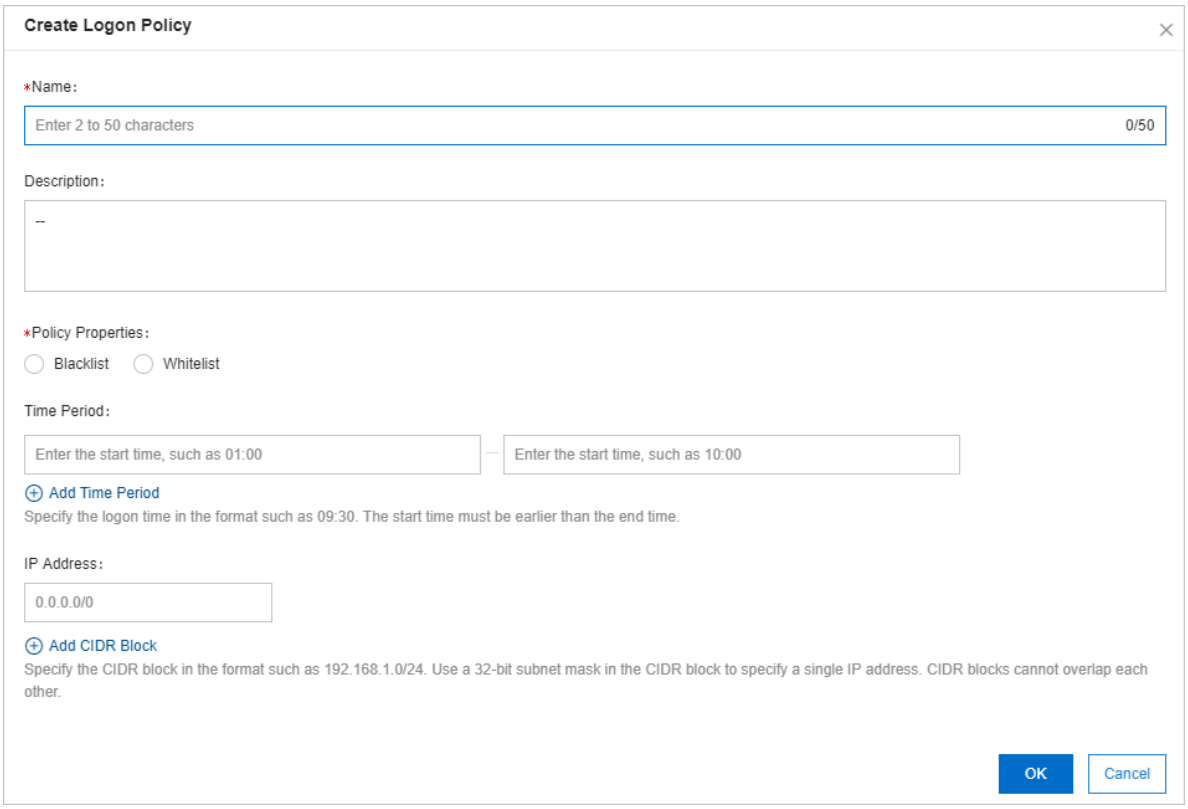


Table 1-11: Logon policy parameters

Parameter	Description
Name	The name of the logon policy. The name must be 2 to 50 characters in length and can contain only letters and digits. The name must be unique in the system.
Description	The description of the logon policy.
Policy Properties	The authentication method for filtering user logons. · Whitelist: Logon is allowed if the following parameters are met. · Blacklist: Logon is denied if the following parameters are met.

Parameter	Description
Time Period	The permitted logon time period. When this policy is configured, users can log on to the ASCM console only during the configured period.
IP Address	The permitted CIDR block. When this policy is configured, users can log on to the ASCM console only from IP addresses within the specified CIDR block.

1.7.5.2 Query a logon policy

When providing services, the Apsara Stack Cloud Management (ASCM) console automatically generates a default policy without restrictions on the logon time and IP addresses.

Context

When providing services, the Apsara Stack Cloud Management (ASCM) console automatically generates a default policy without restrictions on the logon time and IP addresses.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Logon Policies.**
4. **Enter the name of the policy that you want to view and click Search.**
5. **View the logon policy, including the permitted logon time and IP addresses.**

1.7.5.3 Modify a logon policy

You can modify the policy name, policy properties, permitted logon time period, and IP addresses of a logon policy.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Logon Policies.**
4. **Click More in the Actions column corresponding to a policy, and choose Modify from the shortcut menu.**
5. **In the Modify Logon Policy dialog box that appears, modify the logon policy information.**

6. Click OK.

1.7.5.4 Delete a logon policy

You can delete logon policies that are no longer needed.

Context

> **Note:**
> The default policy cannot be deleted.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Logon Policies.
4. Click More in the Actions column corresponding to a policy, and choose Delete from the shortcut menu.
5. In the message that appears, click OK.

1.7.6 User groups

1.7.6.1 Create a user group

You can create a user group in a selected organization and grant batch authorizations to users in the group.

Prerequisites

Before creating a user group, you must create an organization. For more information, see *Create an organization*.

Context

Relationship between user groups and users:

- A user group can contain zero or more users.
- You can add users to user groups as needed.
- You can add a user to multiple user groups.

Relationship between user groups and organizations:

- A user group can only belong to a single organization.
- You can create multiple user groups in an organization.

Relationship between user groups and roles:

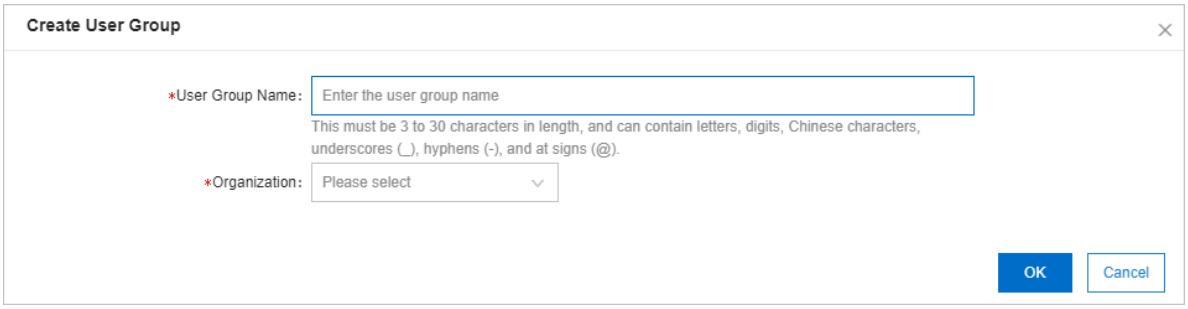
- A user group can only be bound to a single role.
- A role can be associated with multiple user groups.
- When a role is associated with a user group, the role permissions are automatically granted to users in the user group.

Relationship between user groups and resource sets:

- You can add zero or more user groups to a resource set.
- A user group can be added to multiple resource sets.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click User Groups.
4. In the upper-right corner of the page, click Create User Group.
5. In the dialog box that appears, set User Group Name and Organization.



6. Click OK.

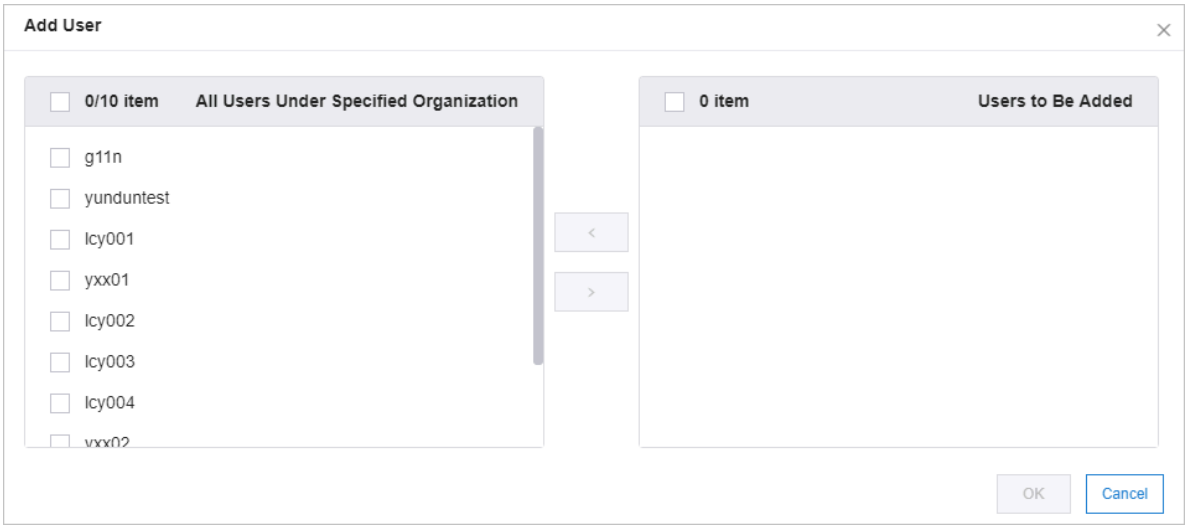
1.7.6.2 Add users to a user group

You can add users to a user group.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click User Groups.
4. Click Add User in the Actions column corresponding to a user group.

5. Select the names of users to be added from the left list, and click the right arrow to move them to the right list.



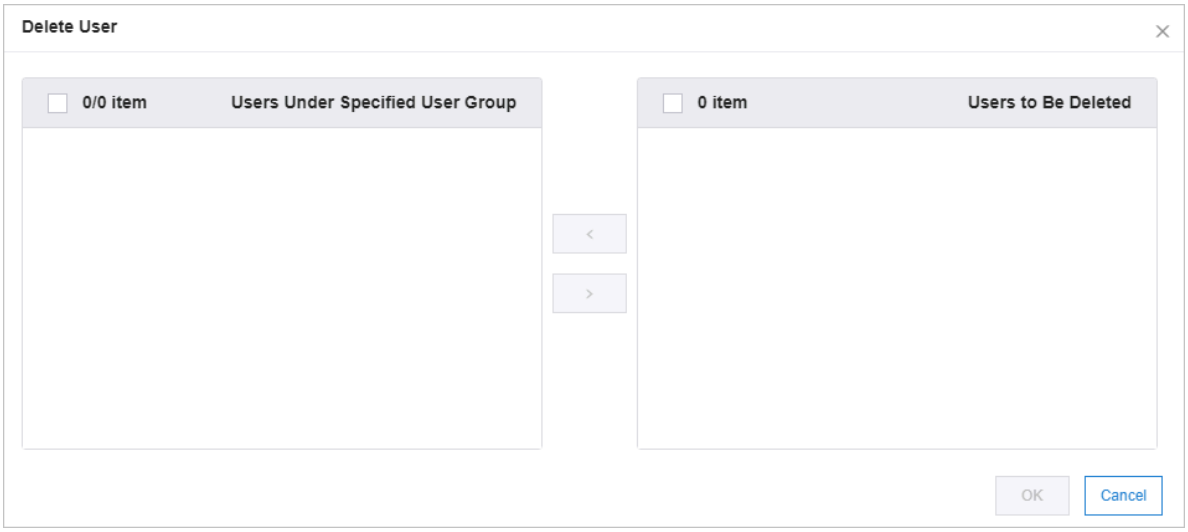
6. Click OK.

1.7.6.3 Delete users from a user group

You can delete users from a user group.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click User Groups.
4. Click Delete User in the Actions column corresponding to a user group.
5. Select the names of users to be deleted from the Users Under Specified User Group list, and click the right arrow to move them to the Users to Be Deleted list.



6. Click OK.

1.7.6.4 Add a role

You can add a role to a user group and assign the role to all users in the group.

Context

Note:

You can add only one role to a user group.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click User Groups.**
4. **Click Add Role in the Actions column corresponding to a user group.**
5. **In the dialog box that appears, select a role.**
6. **Click OK.**

1.7.6.5 Delete a role

You can delete existing roles.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click User Groups.**
4. **Click Delete Role in the Actions column corresponding to a user group.**
5. **In the Confirm message that appears, click OK.**

1.7.6.6 Modify the name of a user group

You can modify the names of user groups.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click User Groups.**
4. **Click Edit User Group in the Actions column corresponding to a user group.**
5. **In the dialog box that appears, enter the new name.**
6. **Click OK.**

1.7.6.7 Delete a user group

You can delete user groups that are no longer needed.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click User Groups.**
4. **Click Delete User Group in the Actions column corresponding to a user group.**
5. **In the Confirm message that appears, click OK.**

1.7.7 Resource pools

1.7.7.1 Update associations

The Apsara Stack Cloud Management (ASCM) console can be deployed in multiple regions. You can update the associations between organizations and regions.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Resource Pool Management.**
4. **In the left-side organization tree, click the name of an organization.**
5. **In the corresponding region list, select the names of regions to be associated.**
6. **Click Update Association.**

1.8 Configurations

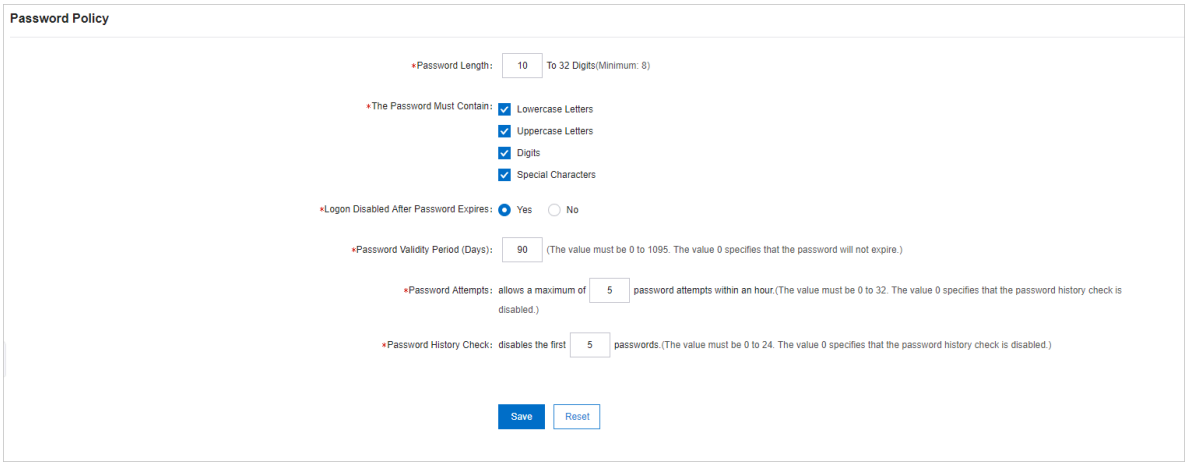
1.8.1 Password policies

You can configure password policies for user logons.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Configurations.**
3. **In the left-side navigation pane of the Configurations page, click Password Policies.**

4. **On the Password Policy page, set the password policy parameters.**



To restore to the default password policy, click Reset.

1.8.2 Menus

1.8.2.1 Create a menu

You can create a menu and add its URL to the ASCM console for quick access.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Configurations.**
3. **In the left-side navigation pane of the Configurations page, click Menu Settings.**
4. **On the Main Menu page, click Create in the upper-right corner.**

5. In the Create dialog box that appears, set the menu parameters.

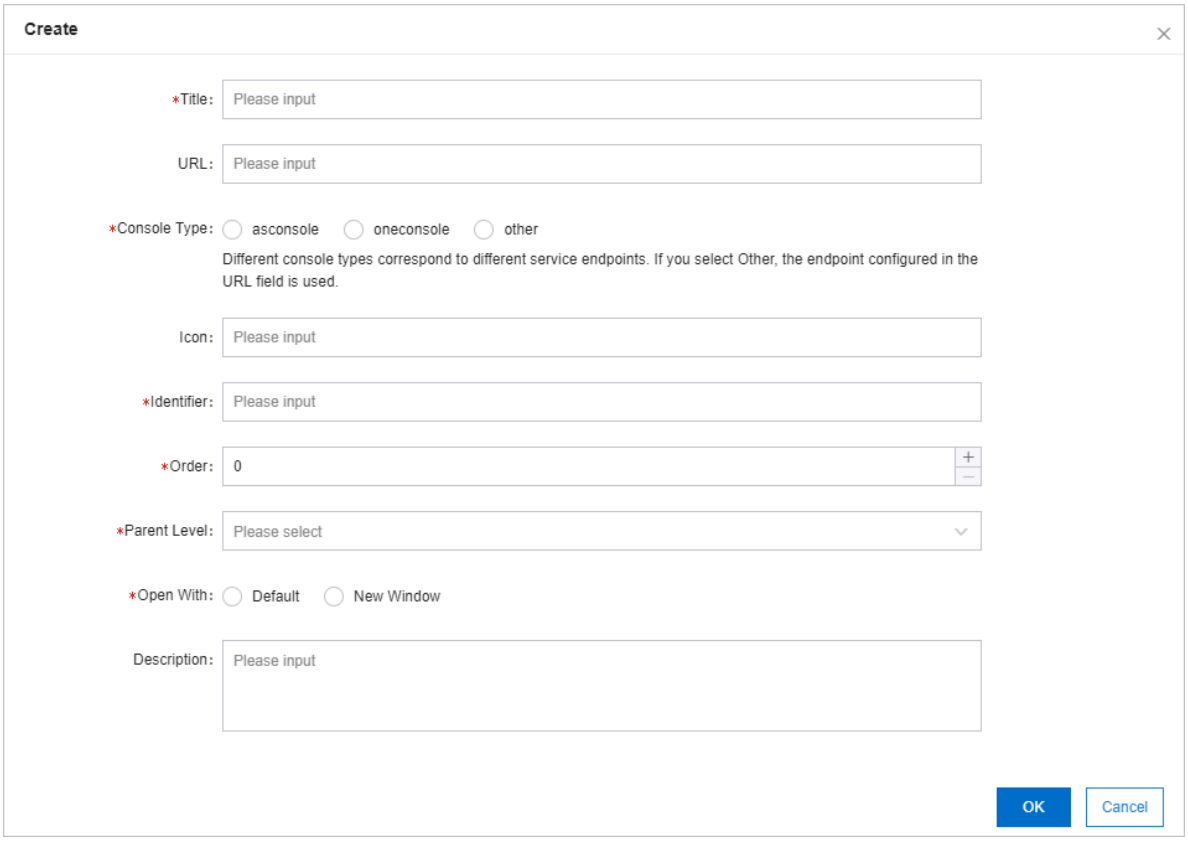


Table 1-12: Menu parameters

Parameter	Description
Title	The display name of the menu.
URL	The URL of the menu.
Console Type	Different console types correspond to different domain names. · oneconsole: You only need to enter the path in the URL field. The domain name is automatically matched. · asconsole: You only need to enter the path in the URL field. The domain name is automatically matched. · other: You must enter the domain name in the URL field.
Icon	The icon displayed in the left-side navigation pane. The icon cannot be changed.
Identifier	The unique identifier of the menu in the system. This identifier can be used to indicate whether the menu is selected in the navigation bar. The identifier cannot be changed.

Parameter	Description
Order	The display order among the same-level menus. The larger the value, the lower the display order. Leave the Order field empty.
Parent Level	The displayed tree structure.
Open With	Specifies whether to open the menu in the current window or in a new window.
Description	The description of the menu.

1.8.2.2 Modify a menu

You can modify an existing menu, including the menu name, URL, icon, and menu order.

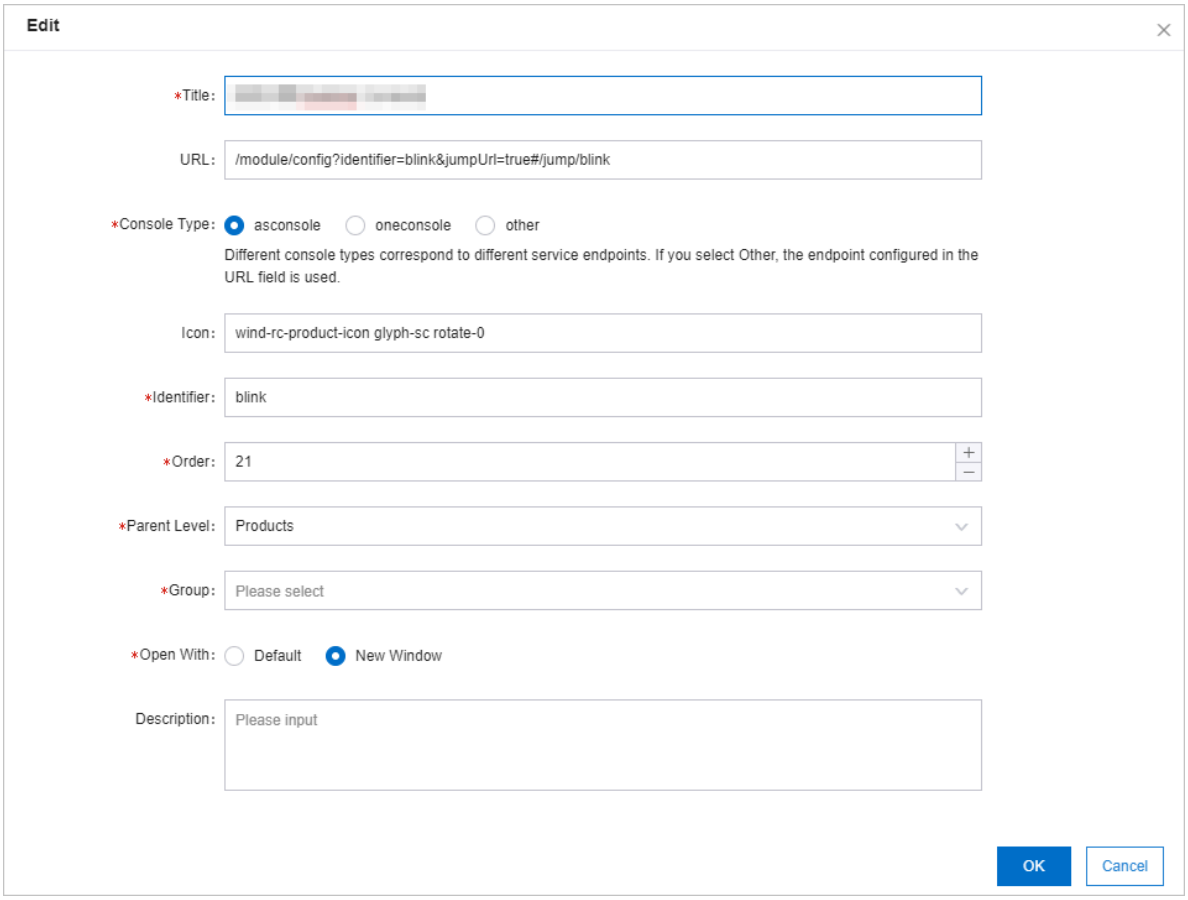
Prerequisites

Default menus cannot be modified.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Configurations.**
3. **In the left-side navigation pane of the Configurations page, click Menu Settings.**
4. **Click Edit in the Actions column corresponding to a menu.**

5. In the Edit dialog box that appears, modify relevant information about the menu.



1.8.2.3 Delete a menu

You can delete menus that are no longer needed.

Prerequisites

Default menus cannot be deleted.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Configurations.**
3. **In the left-side navigation pane of the Configurations page, click Menu Settings.**
4. **Click Delete in the Actions column corresponding to a menu.**
5. **In the message that appears, click OK.**

1.8.2.4 Display or hide menus

You can display or hide menus as follows:

Procedure

1. *Log on to the ASCM console* **as an administrator.**

2. **In the top navigation bar, click Configurations.**
3. **In the left-side navigation pane of the Configurations page, click Menu Settings.**
4. **Select or clear the check box in the Displayed column corresponding to a menu.**

1.8.3 Specifications

1.8.3.1 Create specifications

You can customize specifications for each resource type.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Configurations.**
3. **In the left-side navigation pane of the Configurations page, click Specifications.**
4. **Click the resource type for which you want to create specifications.**
5. **In the upper-right corner of the page, click Create Specifications.**
6. **In the dialog box that appears, set the parameters.**
7. **Click OK.**

1.8.3.2 View specifications

You can view the specifications of each resource type.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Configurations.**
3. **In the left-side navigation pane of the Configurations page, click Specifications.**
4. **Click the resource type for which you want to view specifications.**
5. **In the list of specifications, view information about the specifications.**

1.8.3.3 Disable specifications

By default, the status of newly created specifications is Enabled.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Configurations.**
3. **In the left-side navigation pane of the Configurations page, click Specifications.**
4. **Select the resource type for which you want to disable specifications.**

5. Click Disable in the Actions column corresponding to the target specifications.
6. In the message that appears, click OK.

1.8.3.4 Export specifications

You can export specifications that you want to view and share.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Configurations.
3. In the left-side navigation pane of the Configurations page, click Specifications.
4. Click the resource type for which you want to create specifications.
5. In the upper-right corner of the page, click Export.
6. Save the specifications file to the target path.

1.9 Operations

1.9.1 Quotas

1.9.1.1 Quota parameters

This topic briefly describes the quota parameters of each service.

A department administrator can set resource quotas and create resources within the allowed quotas for the department. When the quotas for the department are used up, the system does not allow the department administrator to create more resources for the department. To create more resources, you must first increase the quotas for the department.

If no quotas are set, you can create an unlimited amount of resources.

ECS

Parameter	Description
CPU Quota	The total number of CPU cores that you can configure for ECS and the number of used cores.
Memory Quota (GB)	The total memory size that you can configure for ECS.
GPU Quota	The total number of GPU cores that you can configure for ECS.

Parameter	Description
SSD Quota (GB)	The total SSD capacity that you can configure for ECS.
Ultra Disk Quota (GB)	The total number of cloud disks that you can configure for an ECS instance.

VPC

Parameter	Description
VPC Quota	The maximum number of VPCs that you can configure.

OSS

Parameter	Description
OSS Quota	The maximum capacity that you can allocate for OSS.

ApsaraDB for RDS (including primary and read-only instances)

Parameter	Description
CPU Quota	The total number of CPU cores that you can configure for ApsaraDB RDS for MySQL, SQL Server, PPAS, or PostgreSQL, and the number of used cores.
Memory Quota (GB)	The total memory size that you can configure for ApsaraDB RDS for MySQL, SQL Server, PPAS, or PostgreSQL.
Disk Quota (GB)	The total storage size that you can configure for ApsaraDB RDS for MySQL, SQL Server, PPAS, or PostgreSQL.

SLB

Parameter	Description
Virtual IP Quota	The maximum number of private IP addresses that you can configure for SLB.
Public Virtual IP Quota	The maximum number of public IP addresses that you can configure for SLB.

EIP

Parameter	Description
EIP Quota	The maximum number of Elastic IP (EIP) addresses that you can configure.

MaxCompute

Parameter	Description
CU Quota	The total number of CUs that you can configure for MaxCompute.
Disk Quota (GB)	The total storage size that you can configure for MaxCompute.

1.9.1.2 Set quotas for a cloud service

The Apsara Stack Cloud Management (ASCM) console allows you to set quotas to properly allocate resources among organizations.

Prerequisites

You must set quotas for a parent organization before you can set quotas for its subordinate organizations.

Context

If the parent organization has quotas (except when the parent organization is a level-1 organization), the available quotas for a subordinate organization are equal to the quotas for the parent organization minus the quotas for other subordinate organizations.

This topic describes how to set quotas for ECS. You can set quotas for other cloud resources in a similar manner.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Operations.**
3. **In the left-side navigation pane of the Operations page, click Quotas.**
4. **In the left-side navigation tree, click the name of the organization for which you want to create cloud resources.**
5. **Select the cloud service for which you want to set quotas. For this example, ECS is selected.**

6. In the upper-right corner of the quota section, click Set.
7. Set the total quotas and click Save.

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

1.9.1.3 Modify quotas

Administrators can adjust quotas for cloud resources based on organizational requirements.

Context

This topic describes how to modify quotas for ECS. You can modify quotas for other cloud resources in a similar manner.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Operations.**
3. **In the left-side navigation pane of the Operations page, click Quotas.**
4. **In the left-side navigation tree, click the name of the organization for which you want to create cloud resources.**
5. **Select the Apsara Stack service for which you want to modify quotas. For this example, ECS is selected.**
6. **In the upper-right corner of the quota area, click Modify.**
7. **Set the total quotas and click Save.**

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

1.9.1.4 Reset quotas

Administrators can reset quotas as needed.

Prerequisites

Before deleting a quota for an organization, make sure that no subordinate organizations have any quotas.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Operations.**
3. **In the left-side navigation pane of the Operations page, click Quotas.**

4. In the left-side organization navigation tree, click the name of the target organization.
5. Select the cloud service for which you want to reset quotas. For this example, ECS is selected.
6. In the upper-right corner of the quota section, click Reset.
7. In the message that appears, click OK.

1.9.2 Usage statistics

1.9.2.1 View the usage statistics of cloud resources

The Apsara Stack Cloud Management (ASCM) console displays statistics for the number of resource instances running in the Apsara Stack environment by time, organization, resource set, or region, and allows you to export statistical reports.

Context

Currently, the cloud resources that can be measured include ECS, VPC, SLB, OSS, ApsaraDB for RDS, and Elastic IP Address (EIP).

This topic describes how to set quotas for ECS. You can set quotas for other cloud resources in a similar manner.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Operations.
3. In the left-side navigation pane of the Operations page, click Usage Statistics.
4. In the Resource Type section, click Elastic Compute Service ECS.
5. In the Search Conditions section, you can filter resources by time, organization, resource set, region, and instance ID.

   You can directly view the export result or click Export to export the currently displayed information to your personal computer in .xls format.

   > **Note:**
   > In the console, you can view or export up to 1,000 statistical records to an Excel file. Use the statistics query API to obtain more statistical data.

   The exported file is named `Resource type name.xls`. Find the downloaded file from the download path in the browser setting.

1.10 Security

1.10.1 View operation logs

You can view operation logs to obtain up-to-date information about various resources and functional modules in the Apsara Stack Cloud Management (ASCM) console. You can also export operation logs to your personal computer.

Procedure

1. *Log on to the ASCM console* **as a security administrator.**
2. **In the top navigation bar, click Security.**
3. **You can filter logs by username, object, level, source IP address, details, start time, and end time.**

   The following table describes the fields in the query result.

   Table 1-13: Fields in the query result

Log field	Description
Username	The name of the operator.
Object	The Apsara Stack service on which operations are performed. The operations include creating, modifying, deleting, querying , updating, binding, unbinding, enabling, and disabling service instances, applying for and releasing service instances, and changing the ownership of service instances.
Level	The operation level. Valid values: INFO, DEBUG, and ERROR.
Source IP	The IP address of the operator.
Details	The brief introduction of the operation.
Start Time	The time when the operation started.
End Time	The time when the operation ended.

4. **Optional: Click Export to export the logs displayed on the current page to your personal computer in.xls format.**

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

1.11 RAM

1.11.1 RAM introduction

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

You can use RAM to manage users and control which resources are accessible to employees, systems, and applications.

RAM provides the following features:

- **RAM role**

  To authorize a cloud service in a level-1 organization to use other resources in the organization, you must create a RAM role. This role contains the operations that the cloud service can perform on resources.

  Only system administrators and level-1 organization administrators can create RAM roles.

- **User group**

  You can create multiple users within an organization and grant them different operation permissions on cloud resources.

  You can create RAM user groups to classify and authorize RAM users within your Apsara Stack tenant account. This simplifies the management of RAM users and their permissions.

  You can create RAM permission policies to grant different operation permissions to different user groups.

1.11.2 Permission policy structure and syntax

This topic describes the structure and syntax used to create or update permission policies in Resource Access Management (RAM).

Policy characters and usage rules

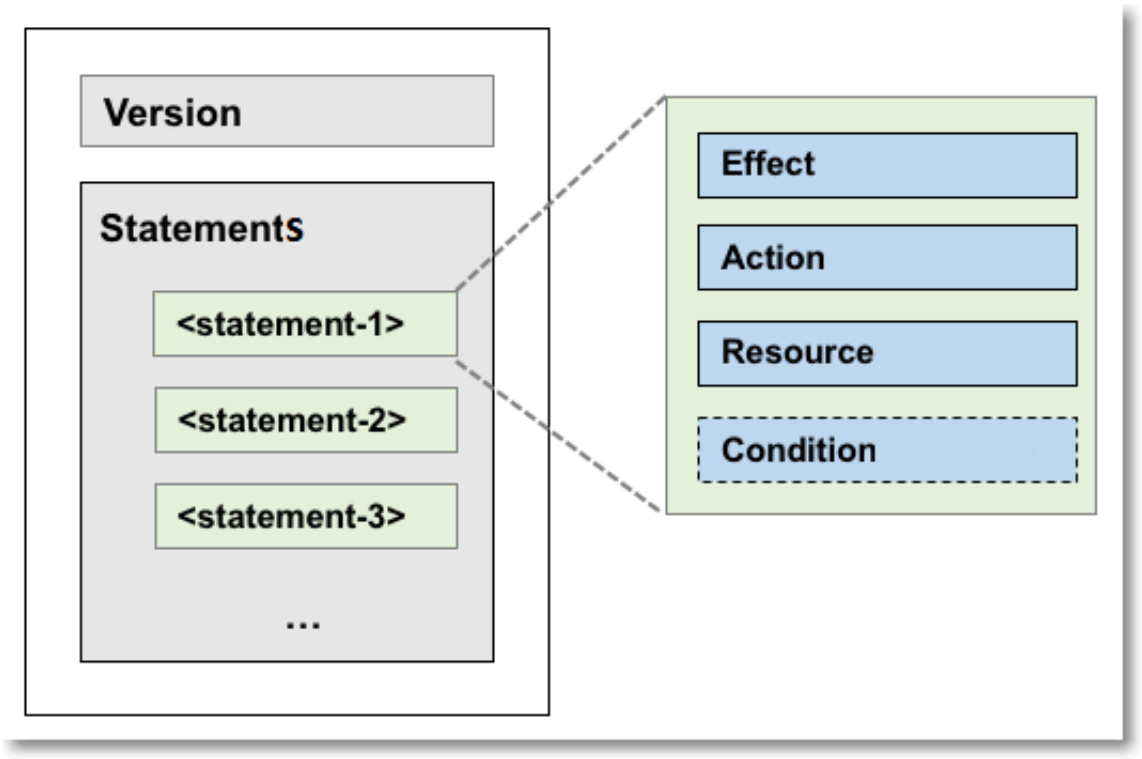
- **Characters in a policy**
  - **The following characters are JSON tokens and are included in policies: `{ } [ ] " , :`.**
  - **The following characters are special characters in the syntax and are not included in policies: `= < > ( ) |`.**

- **Use of characters**
  - **If an element can have more than one value, you can perform the following operations:**
    - **Separate multiple values by using commas (,) as delimiters between each value and use an ellipsis (...) to describe the remaining values. Example:** `[ <action_string>, <action_string>, ...]`.
    - **Include only one value. Examples:** `"Action": [<action_string>]` **and** `"Action": <action_string>`.
  - **A question mark (?) following an element indicates that the element is optional. Example:** `<condition_block? >`.
  - **A vertical bar (|) between elements indicates multiple options. Example:** `("Allow" | "Deny")`.
  - **Elements that must be text strings are enclosed in double quotation marks (""). Example:** `<version_block> = "Version" : ("1")`.

Policy structure

The policy structure includes the following components:

- **The version number.**
- **A list of statements. Each statement contains the following elements: Effect, Action, Resource, and Condition. The Condition element is optional.**



Policy syntax

```
policy  = {
     <version_block>,
     <statement_block>
}
<version_block> = "Version" : ("1")
<statement_block> = "Statement" : [ <statement>, <statement>, ... ]
<statement> = {
    <effect_block>,
    <action_block>,
    <resource_block>,
    <condition_block? >
}
<effect_block> = "Effect" : ("Allow" | "Deny")
<action_block> = ("Action" | "NotAction") :
    ("*" | [<action_string>, <action_string>, ...])
<resource_block> = ("Resource" | "NotResource") :
    ("*" | [<resource_string>, <resource_string>, ...])
<condition_block> = "Condition" : <condition_map>
<condition_map> = {
  <condition_type_string> : {
      <condition_key_string> : <condition_value_list>,
      <condition_key_string> : <condition_value_list>,
      ...
  },
  <condition_type_string> : {
      <condition_key_string> : <condition_value_list>,
      <condition_key_string> : <condition_value_list>,
      ...
  }, ...
}
<condition_value_list> = [<condition_value>, <condition_value>, ...]
```

```
<condition_value> = ("String" | "Number" | "Boolean")
```

Description:

- **The current policy version is 1.**
- **The policy can have multiple statements.**
  - **The effect of each statement can be either `Allow` or `Deny`.**

    Note:
    In a statement, both the Action and Resource elements can have multiple values.
  - **Each statement can have its own conditions.**

    Note:
    A condition block can contain multiple conditions with different operators and logical combinations of these conditions.
- **You can attach multiple policies to a RAM user. If policies that apply to a request include an `Allow` statement and a `Deny` statement, the Deny statement overrides the Allow statement.**
- **Element value:**
  - **If an element value is a number or Boolean value, it must be enclosed in double quotation marks ("") in the same way as strings.**
  - **If an element value is a string, characters such as the asterisk (`*`) and question mark (`?`) can be used for fuzzy matching.**
    - **The asterisk (`*`) indicates any number (including zero) of allowed characters. For example, `ecs:Describe*` indicates all ECS API operations that start with `Describe`.**
    - **`The question mark (?)` indicates an allowed character.**

Policy format check

Policies are stored in RAM as JSON documents. When you create or update a policy, RAM first checks whether the JSON format is valid.

- **For more information about JSON syntax standards, see *RFC 7159*.**
- **We recommend that you use tools such as JSON validators and editors to check whether the policies meet JSON syntax standards.**

1.11.3 RAM roles

1.11.3.1 View basic information about a RAM role

You can view basic information about a RAM role, including its user groups and existing permission policies.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Roles.**
4. **On the Roles page, click the name of the target RAM role.**
5. **In the basic information section, click the User Groups and Permissions tabs to view relevant information.**

1.11.3.2 Create a RAM role

To authorize a cloud service in a level-1 organization to use other resources in the organization, you must create a RAM role. This role contains the operations that the cloud service can perform on resources.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Roles.**
4. **In the upper-right corner of the page, click Create RAM Role.**
5. **On the Roles - Create RAM Role page that appears, set Role Name and Description.**
6. **Click Create.**

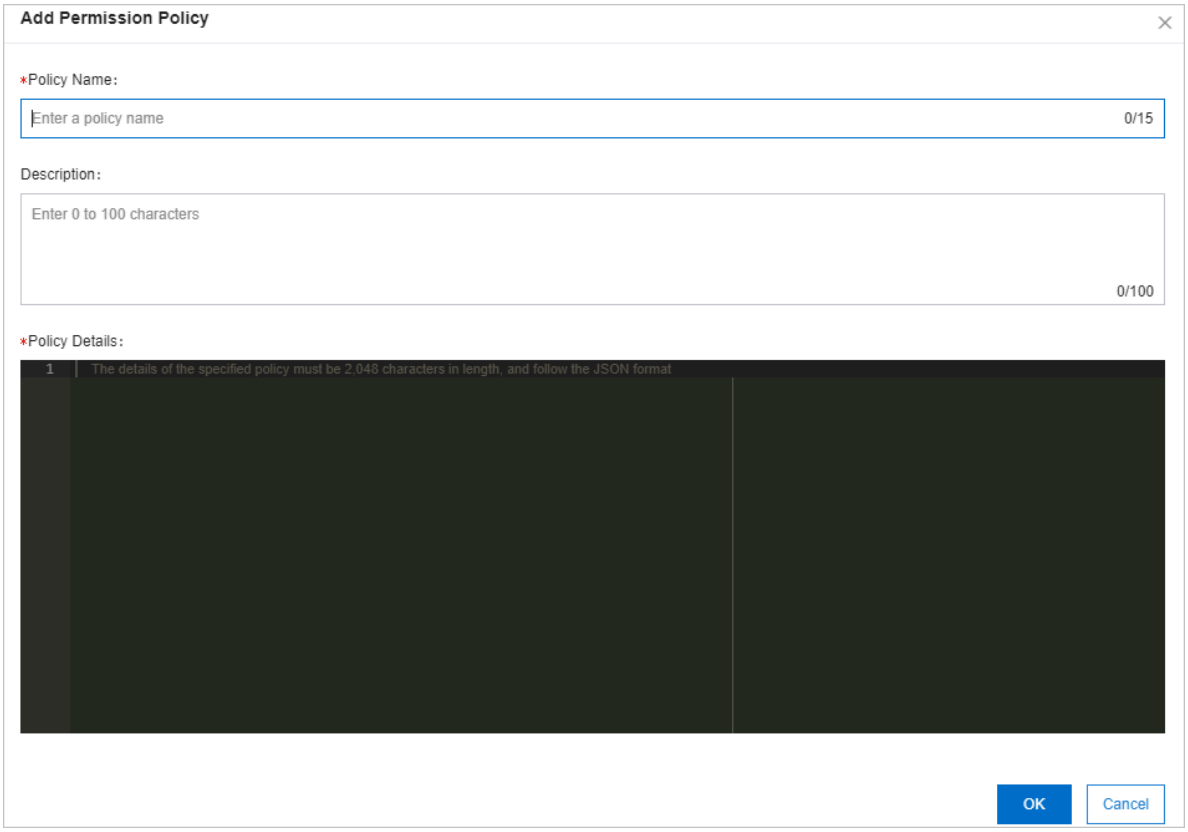
1.11.3.3 Add a permission policy

To use a cloud service to access other cloud resources, you must create a permission policy and attach it to a user group.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Roles.**

4. In the role name list, click More in the Actions column corresponding to a RAM role, and choose Modify from the shortcut menu to go to the Roles page.
5. Click the Permissions tab.
6. Click Add Permission Policy.
7. In the dialog box that appears, enter information about the permission policy.



For more information about how to enter the policy content, see *Permission policy structure and syntax*.

1.11.3.4 Modify the content of a RAM permission policy

You can modify the content of a RAM permission policy as needed.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Roles.
4. In the role name list, click More in the Actions column corresponding to a RAM role, and choose Modify from the shortcut menu to go to the Roles page.
5. Click the Permissions tab.
6. Click the name of a permission policy in the Permission Policy Name column.

7. **In the Modify Permission Policy dialog box that appears, modify the relevant information and click OK.**

   For more information about how to modify the policy content, see *Permission policy structure and syntax*.

1.11.3.5 Modify the name of a RAM permission policy

You can modify the name of a RAM permission policy as needed.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Roles.**
4. **In the role name list, click More in the Actions column corresponding to a RAM role, and choose Modify from the shortcut menu to go to the Roles page.**
5. **Click the Permissions tab. Click the name of a permission policy in the Permission Policy Name column.**
6. **In the Modify Permission Policy dialog box that appears, modify the permission policy name.**

1.11.3.6 Add a RAM role to a user group

You can bind RAM roles to user groups as needed.

Prerequisites

You must create a user group before RAM roles can be added. If no user groups have been created, see *Create a user group*.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Roles.**
4. **In the role name list, click More in the Actions column corresponding to a RAM role, and choose Modify from the shortcut menu to go to the Roles page.**
5. **Click the User Groups tab.**
6. **Click Add User Group. In the dialog box that appears, select a user group.**
7. **Click OK.**

1.11.3.7 Grant permissions to a RAM role

When you grant permissions to a RAM role, all users in the user groups that are assigned this role will share the granted permissions.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Roles.**
4. **In the role name list, click More in the Actions column corresponding to a RAM role, and choose Modify from the shortcut menu to go to the Roles page.**
5. **Click the Permissions tab.**
6. **Click Select Existing Permission Policy.**
7. **In the dialog box that appears, select a RAM permission policy and click OK.**

   If no RAM permission policies are available, see *Add a permission policy*.

1.11.3.8 Remove permissions from a RAM role

You can remove permissions that are no longer needed from RAM roles.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Enterprise.**
3. **In the left-side navigation pane of the Enterprise page, click Roles.**
4. **In the role name list, click More in the Actions column corresponding to a RAM role, and choose Modify from the shortcut menu to go to the Roles page.**
5. **Click the Permissions tab.**
6. **Click Remove in the Actions column corresponding to the permission policy that you want to remove.**

1.11.3.9 Modify a RAM role name

Administrators can modify the names of RAM roles.

Context

> **Note:**
> **The name of a preset role cannot be modified.**

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Roles.
4. In the role name list, click More in the Actions column corresponding to a RAM role, and choose Modify from the shortcut menu to go to the Roles page.
5. Move the pointer over the role name and click  to enter a new role name.

1.11.3.10 Delete a RAM role

This topic describes how to delete a RAM user.

Prerequisites

Before you delete a RAM role, make sure that no policies are attached to the RAM role.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Enterprise.
3. In the left-side navigation pane of the Enterprise page, click Roles.
4. In the role name list, click More in the Actions column corresponding to a RAM role, and choose Delete from the shortcut menu.
5. In the message that appears, click OK.

1.11.4 RAM authorization policies

1.11.4.1 Create a RAM role

You can create authorization policies and grant them to organizations as needed.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the top navigation bar, click Configurations.
3. In the left-side navigation pane of the Configurations page, click RAM Roles.
4. In the upper-right corner of the page, click Create RAM User.
5. On the Create RAM User page, set Organization and Service.
6. Click OK.

1.11.4.2 View the details of a RAM role

You can view the details of a RAM role, including its role name, creation time, description, and Alibaba Cloud Resource Name (ARN).

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Configurations.**
3. **In the left-side navigation pane of the Configurations page, click RAM Roles.**
4. **On the RAM Users page, set Role Name or Service Name, and click Search in the upper-right corner.**

   To perform another search, click Clear.
5. **Find the RAM role that you want to view and click Details in the Actions column.**
6. **Click the Role Details tab to view the details of the RAM role.**

1.11.4.3 View RAM authorization policies

You can view the details of a RAM authorization policy, including its policy name, policy type, default version, description, association time, and policy content.

Prerequisites

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

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the top navigation bar, click Configurations.**
3. **In the left-side navigation pane of the Configurations page, click RAM Roles.**
4. **On the RAM Users page, set Role Name or Service Name, and click Search in the upper-right corner.**

   To perform another search, click Clear.
5. **Find the RAM role that you want to view and click Details in the Actions column.**
6. **Click the Role Policy tab to view information about the role authorization policy. Click Details in the Actions column to view the policy details.**

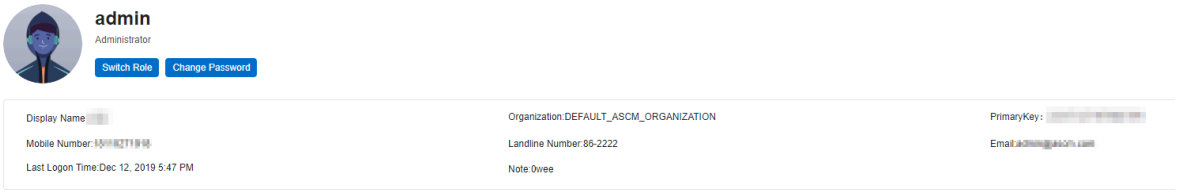
1.12 Personal information management

1.12.1 Modify personal information

You can modify your personal information to keep it up-to-date.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the upper-right corner of the homepage, move the pointer over the user profile picture and choose User Information from the shortcut menu.**



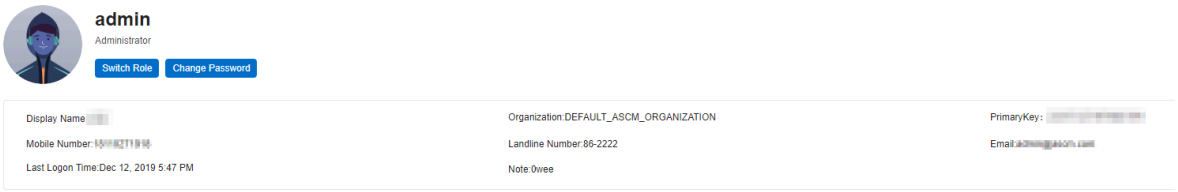
3. **Click**  **next to the item you want to modify.**
4. **In the Modify User Information dialog box that appears, modify the relevant information.**
5. **Click OK.**

1.12.2 Change your logon password

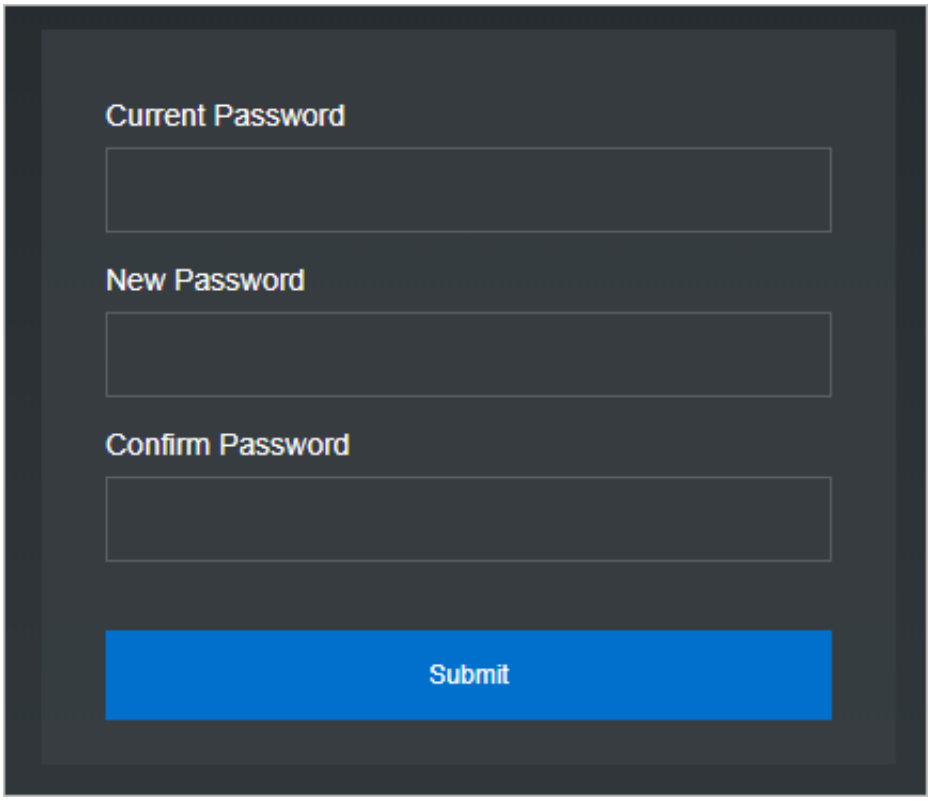
To improve security, you must change your logon password in a timely manner.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the upper-right corner of the homepage, move the pointer over the user profile picture and choose User Information from the shortcut menu.**



3. **Click Change Password. On the page that appears, set Current Password, New Password, and Confirm Password.**



4. **Click Submit.**

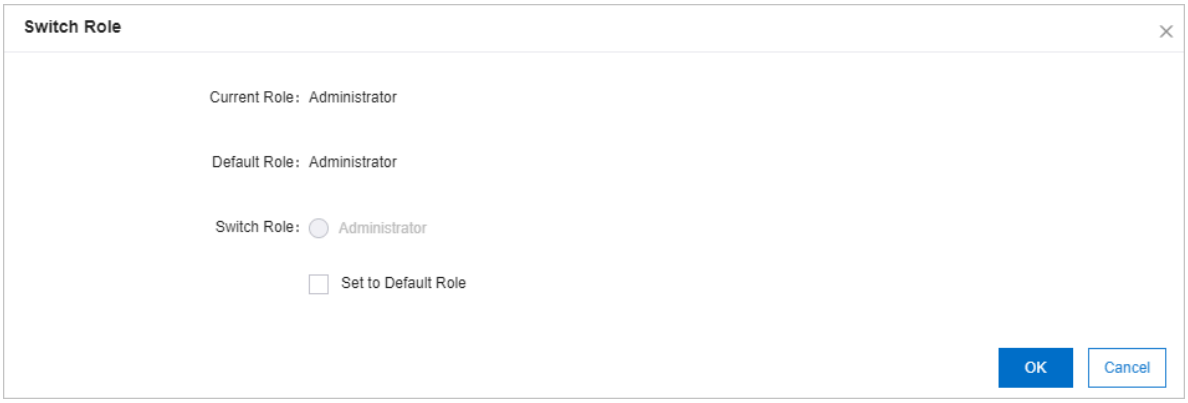
1.12.3 Switch the current role

You can switch the scope of your current role.

Procedure

1. *Log on to the ASCM console* **as an administrator.**
2. **In the upper-right corner of the homepage, move the pointer over the user profile picture and choose User Information from the shortcut menu.**
3. **Click Switch Role.**

4. In the Switch Role dialog box that appears, select the role that you want to switch to.



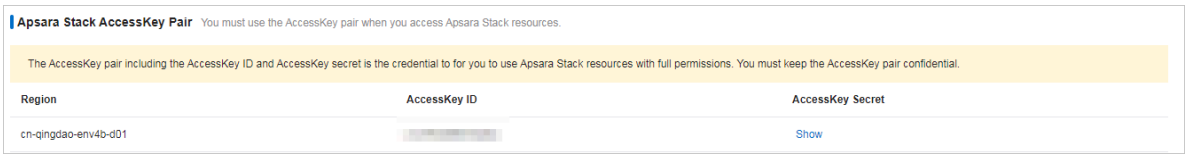
You can also switch back to the default role.

1.12.4 View the AccessKey pair of your Apsara Stack tenant account

To secure cloud resources, the system must verify the identity of visitors and ensure that they have the relevant permissions. You must obtain the AccessKey ID and AccessKey secret of your personal account to access cloud resources.

Procedure

1. *Log on to the ASCM console* as an administrator.
2. In the upper-right corner of the homepage, move the pointer over the user profile picture and choose User Information from the shortcut menu.
3. In the Apsara Stack AccessKey Pair section, view your AccessKey pair.



Note:

The AccessKey pair is made up of the AccessKey ID and AccessKey secret. These credentials provide you full permissions on Apsara Stack resources. You must keep the AccessKey pair confidential.

2 Elastic Compute Service (ECS)

2.1 What is ECS?

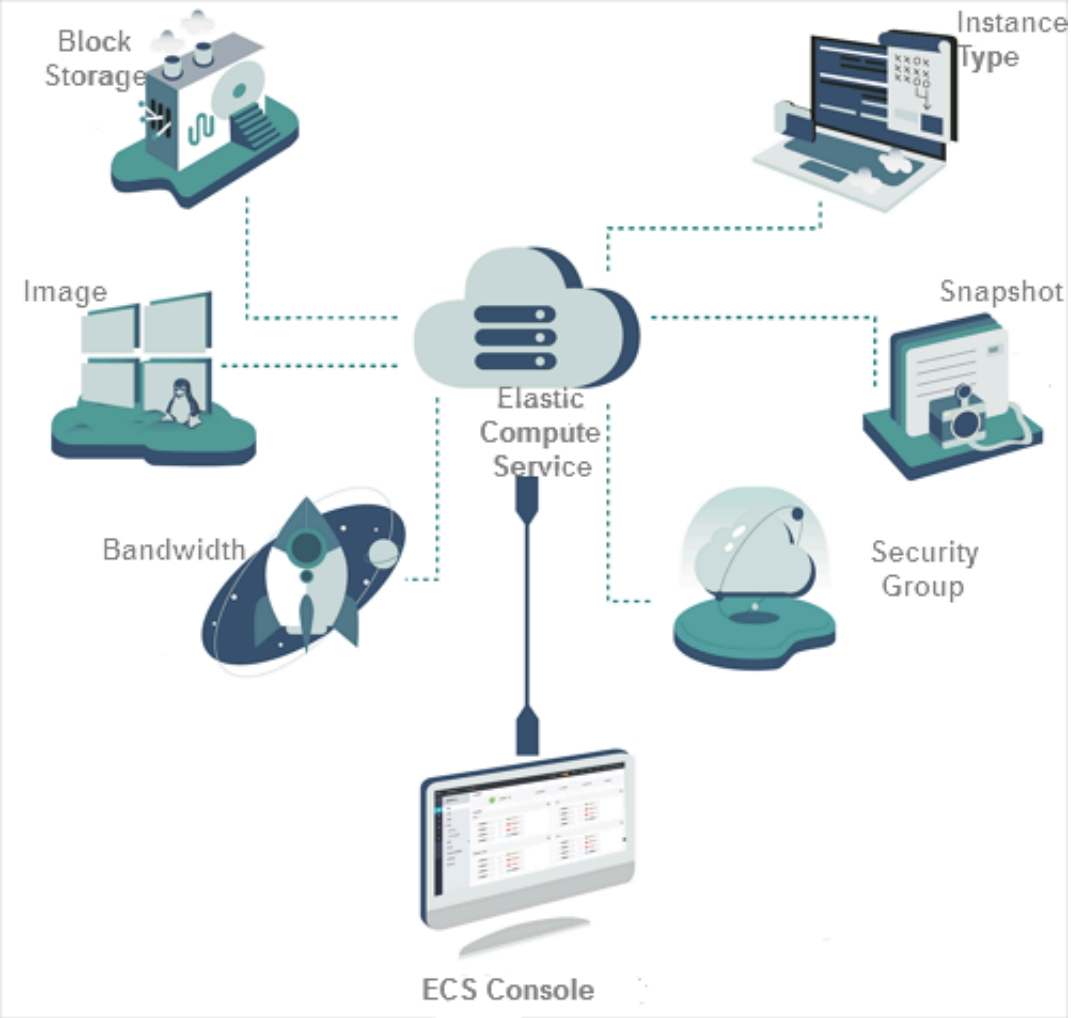
2.1.1 Overview

Elastic Compute Service (ECS) is a type of computing service that features elastic processingcapabilities. Compared with physical servers, ECS can be more efficientlymanaged and is more user-friendly. You can create instances, resize disks, and add or release any number of ECS instances at any time based on your business needs.

An ECS instance is a virtual computing environment that contains the most basic components of computers such as the CPU, memory, and storage. Users perform operations on ECS instances. Instances are core components of ECS, and operations can be performed on instances by using the ECS console. Other resources such as

block storage, images, and snapshots can only be used after they are integrated into ECS instances. For more information, see *Figure 2-1: ECS components*.

Figure 2-1: ECS components



2.1.2 Instance types

An ECS instance is the smallest unit that can provide computing services. The compute capabilities of an ECS instance vary with instance types of the ECS instance.

The ECS instance type defines the basic properties of an ECS instance: CPU (including CPU model and clock speed) and memory. In addition to the instance type, you must also configure the Block Storage, image, and network type when you create an instance.

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (including one primary ENI)
n4	ecs.n4.small	1	2.0	None	0.5	50	1	1
	ecs.n4.large	2	4.0	None	0.5	100	1	1
	ecs.n4.xlarge	4	8.0	None	0.8	150	1	2
	ecs.n4.2xlarge	8	16.0	None	1.2	300	1	2
	ecs.n4.4xlarge	16	32.0	None	2.5	400	1	2
	ecs.n4.8xlarge	32	64.0	None	5.0	500	1	2
mn4	ecs.mn4.small	1	4.0	None	0.5	50	1	1
	ecs.mn4.large	2	8.0	None	0.5	100	1	1
	ecs.mn4.xlarge	4	16.0	None	0.8	150	1	2
	ecs.mn4.2xlarge	8	32.0	None	1.2	300	1	2
	ecs.mn4.4xlarge	16	64.0	None	2.5	400	1	2
	ecs.mn4.8xlarge	32	128.0	None	5.0	500	2	8
xn4	ecs.xn4.small	1	1.0	None	0.5	50	1	1
e4	ecs.e4.small	1	8.0	None	0.5	50	1	1
	ecs.e4.large	2	16.0	None	0.5	100	1	1

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.e4.xlarge	4	32.0	None	0.8	150	1	2
	ecs.e4.2xlarge	8	64.0	None	1.2	300	1	3
	ecs.e4.4xlarge	16	128.0	None	2.5	400	1	8
sn1ne	ecs.sn1ne.large	2	4.0	None	1.0	300	2	2
	ecs.sn1ne.xlarge	4	8.0	None	1.5	500	2	3
	ecs.sn1ne.2xlarge	8	16.0	None	2.0	1,000	4	4
	ecs.sn1ne.3xlarge	12	24.0	None	2.5	1,300	4	6
	ecs.sn1ne.4xlarge	16	32.0	None	3.0	1,600	4	8
	ecs.sn1ne.6xlarge	24	48.0	None	4.5	2,000	6	8
	ecs.sn1ne.8xlarge	32	64.0	None	6.0	2,500	8	8
g5	ecs.g5.large	2	8.0	None	1.0	300	2	2
	ecs.g5.xlarge	4	16.0	None	1.5	500	2	3

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.g5.2xlarge	8	32.0	None	2.5	800	2	4
	ecs.g5.3xlarge	12	48.0	None	4.0	900	4	6
	ecs.g5.4xlarge	16	64.0	None	5.0	1,000	4	8
	ecs.g5.6xlarge	24	96.0	None	7.5	1,500	6	8
	ecs.g5.8xlarge	32	128.0	None	10.0	2,000	8	8
	ecs.g5.16xlarge	64	256.0	None	20.0	4,000	16	8
sn2ne	ecs.sn2ne.large	2	8.0	None	1.0	300	2	2
	ecs.sn2ne.xlarge	4	16.0	None	1.5	500	2	3
	ecs.sn2ne.2xlarge	8	32.0	None	2.0	1,000	4	4
	ecs.sn2ne.3xlarge	12	48.0	None	2.5	1,300	4	6
	ecs.sn2ne.4xlarge	16	64.0	None	3.0	1,600	4	8
	ecs.sn2ne.6xlarge	24	96.0	None	4.5	2,000	6	8

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.sn2ne.8xlarge	32	128.0	None	6.0	2,500	8	8
	ecs.sn2ne.14xlarge	56	224.0	None	10.0	4,500	14	8
se1ne	ecs.se1ne.large	2	16.0	None	1.0	300	2	2
	ecs.se1ne.xlarge	4	32.0	None	1.5	500	2	3
	ecs.se1ne.2xlarge	8	64.0	None	2.0	1,000	4	4
	ecs.se1ne.3xlarge	12	96.0	None	2.5	1,300	4	6
	ecs.se1ne.4xlarge	16	128.0	None	3.0	1,600	4	8
	ecs.se1ne.6xlarge	24	192.0	None	4.5	2,000	6	8
	ecs.se1ne.8xlarge	32	256.0	None	6.0	2,500	8	8
	ecs.se1ne.14xlarge	56	480.0	None	10.0	4,500	14	8
se1	ecs.se1.large	2	16.0	None	0.5	100	1	2

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.se1.xlarge	4	32.0	None	0.8	200	1	3
	ecs.se1.2xlarge	8	64.0	None	1.5	400	1	4
	ecs.se1.4xlarge	16	128.0	None	3.0	500	2	8
	ecs.se1.8xlarge	32	256.0	None	6.0	800	3	8
	ecs.se1.14xlarge	56	480.0	None	10.0	1,200	4	8
ebmg5s	ecs.ebmg5s.24xlarge	96	384.0	None	30.0	4,500	8	32
ebmg5	ecs.ebmg5.24xlarge	96	384.0	None	10.0	4,000	8	32
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
	ecs.i2-pf.20xlarge	80	704.0	10 × 1,788	25.0	4,500	16	8
d1	ecs.d1.2xlarge	8	32.0	4 × 5,500	3.0	300	1	4

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	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
d2	ecs.d2-zyy-d0.4xlarge	16	64.0	None	3.0	300	2	8
	ecs.d2-zyy-d0.6xlarge	24	96.0	None	4.0	400	2	8
	ecs.d2-zyy.4xlarge	16	64.0	6 × 7,500	3.0	300	4	8
	ecs.d2-zyy.6xlarge	24	96.0	12 × 7,500	4.0	400	4	8
	ecs.d2-zyy-m40.8xlarge	32	128.0	12 × 7,500	6.0	600	4	8

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.d2-zyy-m40.12xlarge	48	160.0	12 × 7,500	10.0	1,000	4	8
	ecs.d2-gab.4xlarge	16	64.0	6 × 1,150	3.0	300	4	8
	ecs.d2-gab.8xlarge	32	128.0	12 × 1,150	6.0	600	4	8
sccg5ib	ecs.sccg5ib.24xlarge	96	384.0	None	10.0	4,500	8	32
scch5ib	ecs.scch5ib.16xlarge	64	192.0	None	10.0	4,500	8	32
sn1	ecs.sn1.medium	2	4.0	None	0.5	100	1	2
	ecs.sn1.large	4	8.0	None	0.8	200	1	3
	ecs.sn1.xlarge	8	16.0	None	1.5	400	1	4
	ecs.sn1.3xlarge	16	32.0	None	3.0	500	2	8
	ecs.sn1.7xlarge	32	64.0	None	6.0	800	3	8
sn2	ecs.sn2.medium	2	8.0	None	0.5	100	1	2
	ecs.sn2.large	4	16.0	None	0.8	200	1	3
	ecs.sn2.xlarge	8	32.0	None	1.5	400	1	4

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.sn2.3xlarge	16	64.0	None	3.0	500	2	8
	ecs.sn2.7xlarge	32	128.0	None	6.0	800	3	8
	ecs.sn2.14xlarge	56	224.0	None	10.0	1,200	4	8

The following table describes the FPGA instance family.

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (includi one primary ENI)	FPGAs
f1	ecs.f1-c8f1.2xlarge	8	60.0	None	3.0	400	4	4	Intel Arria 10 GX 1150
	ecs.f1-c8f1.4xlarge	16	120.0	None	5.0	1,000	4	8	2 × Intel Arria 10 GX 1150
	ecs.f1-c28f1.7xlarge	28	112.0	None	5.0	2,000	8	8	Intel Arria 10 GX 1150
	ecs.f1-c28f1.14xlarge	56	224.0	None	10.0	2,000	14	8	2 × Intel Arria 10 GX 1150

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (includi one primary ENI)	FPGAs
f3	ecs.f3-c16f1.4xlarge	16	64.0	None	5.0	1,000	4	8	1 × Xilinx VU9P
	ecs.f3-c16f1.8xlarge	32	128.0	None	10.0	2,000	8	8	2 × Xilinx VU9P
	ecs.f3-c16f1.16xlarge	64	256.0	None	20.0	2,000	16	8	4 × Xilinx VU9P

The following table describes the instance families with GPUs.

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (includi one primary ENI)	GPUs
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

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (includi one primary ENI)	GPUs
	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
	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	None	3.0	300	1	3	1 × NVIDIA M40
	ecs.gn4-c8g1.2xlarge	8	30.0	None	3.0	400	1	4	1 × NVIDIA M40
	ecs.gn4.8xlarge	32	48.0	None	6.0	800	3	8	1 × NVIDIA M40
	ecs.gn4-c4g1.2xlarge	8	60.0	None	5.0	500	1	4	2 × NVIDIA M40

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (includi one primary ENI)	GPUs
	ecs.gn4-c8g1.4xlarge	16	60.0	None	5.0	500	1	8	2 × NVIDIA M40
	ecs.gn4.14xlarge	56	96.0	None	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
	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	None	1.0	100	2	2	1 × NVIDIA P4
	ecs.gn5i-c4g1.xlarge	4	16.0	None	1.5	200	2	3	1 × NVIDIA P4
	ecs.gn5i-c8g1.2xlarge	8	32.0	None	2.0	400	4	4	1 × NVIDIA P4

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (includi one primary ENI)	GPUs
	ecs.gn5i-c16g1.4xlarge	16	64.0	None	3.0	800	4	8	1 × NVIDIA P4
	ecs.gn5i-c16g1.8xlarge	32	128.0	None	6.0	1,200	8	8	2 × NVIDIA P4
	ecs.gn5i-c24g1.12xlarge	48	192.0	None	10.0	2,000	8	8	2 × NVIDIA P4
	ecs.gn5i-c28g1.14xlarge	56	224.0	None	10.0	2,000	14	8	2 × NVIDIA P4
gn5e	ecs.gn5e-c11g1.3xlarge	10	58.0	None	2.0	150	1	6	1 × NVIDIA P4
	ecs.gn5e-c11g1.5xlarge	22	116.0	None	4.0	300	1	8	2 × NVIDIA P4
	ecs.gn5e-c11g1.11xlarge	44	232.0	None	6.0	600	2	8	4 × NVIDIA P4
	ecs.gn5e-c11g1.22xlarge	88	464.0	None	10.0	1,200	4	15	8 × NVIDIA P4

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (includi one primary ENI)	GPUs
gn6i	ecs.gn6i-c10g1.2xlarge	10	42.0	None	5.0	800	2	4	1 × T4
	ecs.gn6i-c10g1.5xlarge	20	84.0	None	8.0	1,000	4	6	2 × T4
	ecs.gn6i-c10g1.10xlarge	40	168.0	None	15.0	2,000	8	8	4 × T4
	ecs.gn6i-c10g1.20xlarge	80	336.0	None	30.0	4,000	16	8	8 × T4
	ecs.gn6i-c14g1.3xlarge	14	56.0	None	5.0	1,000	4	6	1 × T4
	ecs.gn6i-c14g1.7xlarge	28	112.0	None	10.0	2,000	8	8	2 × T4
	ecs.gn6i-c14g1.14xlarge	56	224.0	None	20.0	4,000	12	8	4 × T4
	ecs.gn6i-c20g1.5xlarge	20	80.0	None	10.0	1,500	4	6	1 × T4

Instanc family	Instanc type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwi (Gbit/s)	Packet forward rate (Kpps)	NIC queues	ENIs (includi one primary ENI)	GPUs
	ecs.gn6i-c20g1.10xlarge	40	160.0	None	20.0	3,000	8	8	2 × T4
gn6v	ecs.gn6v-c8g1.2xlarge	8	32.0	None	2.5	800	4	4	1 × NVIDIA V100
	ecs.gn6v-c8g1.8xlarge	32	128.0	None	10.0	2,000	8	8	4 × NVIDIA V100
	ecs.gn6v-c8g1.16xlarge	64	256.0	None	20.0	2,500	16	8	8 × NVIDIA V100
sccgn6p	ecs.sccgn6p.24xlarge	96	768.0	None	30.0	4,500	8	32	8 × NVIDIA V100

The following instance types are only applicable to environments that are upgraded from Apsara Stack V2 to V3.

Instance family	Instance type	vCPUs	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

Instance family	Instance type	vCPUs	Memory (GiB)
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
	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

2.1.3 Instance lifecycle

The lifecycle of an ECS instance begins when it is created and ends when it is released. This topic describes the instance status, status attributes, and corresponding API status.

An instance has several inherent states throughout its lifecycle, as shown in *Table 2-1: Lifecycle description*.

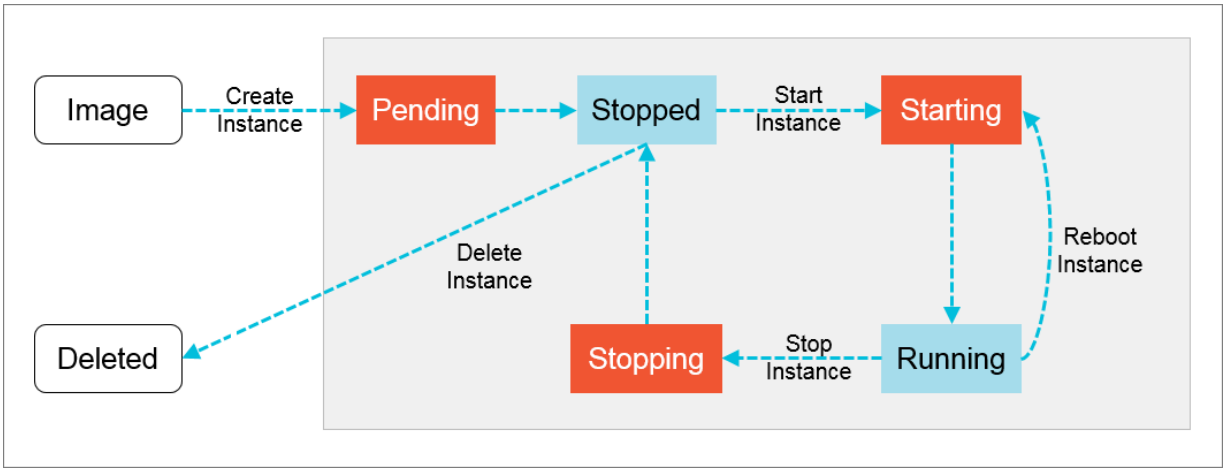
Table 2-1: Lifecycle description

Status	Status attribute	Description	Corresponding API status
`Instance being created`	**Intermediate**	**The instance is being created and is waiting to be enabled. If an instance remains in this status for a long period of time, an exception occurs.**	`Pending`
`Starting`	**Intermediate**	**After an instance is restarted or started from the console or through APIs, the instance enters the starting state before entering the `running` state. If an instance remains in the starting state for a long period of time, an exception occurs.**	`Starting`
`Running`	**Stable**	**Indicates that the instance is running normally and can accommodate your business needs.**	`Running`
`Stopping`	**Intermediate**	**After an instance is stopped from the console or through APIs, the instance enters the stopping state before entering the `stopped` state. If an instance remains in the stopping state for a long period of time, an exception occurs.**	`Stopping`
`Stopped`	**Stable**	**Indicates that an instance has been stopped. An instance in the stopped state cannot provide external services.**	`Stopped`

Status	Status attribute	Description	Corresponding API status
`Reinitializing`	Intermediate	After the system disk or data disk is reinitialized from the console or through APIs the instance enters the reinitializing state before entering the `running` state. If an instance remains in the reinitializing state for a long period of time, an exception occurs.	`Stopped`
`Changing system disk`	Intermediate	After the system disk is changed from the console or through APIs, the instance enters the changing system disk state before entering the `running` state. If an instance remains in the changing system disk state for a long time, an exception occurs.	`Stopped`

Table 2-1: Lifecycle description describes corresponding relationship between instance states in the console and instance states in APIs. *Figure 2-2: Instance status in APIs* shows the instance states in APIs.

Figure 2-2: Instance status in APIs



2.2 Instructions

2.2.1 Restrictions

Learn about restrictions before performing operations on ECS instances.

- **Do not upgrade the kernel or operating system version of an ECS instance.**
- **Do not start SELinux for Linux systems except CentOS and RedHat.**
- **Do not detach PVDriver.**
- **Do not arbitrarily modify the MAC address of the network interface.**

2.2.2 Suggestions

Consider the following suggestions to make more efficient use of ECS:

- **ECS instances with 4 GiB or higher memory must use a 64-bit operating system. 32-bit operating systems have a maximum of 4 GiB of memory addressing.**
- **A 32-bit Windows operating system supports a maximum of 4 CPU cores.**
- **To ensure service continuity and avoid failover-induced service unavailability, we recommend that you configure service applications to boot automatically at system startup.**

2.2.3 Limits

Before using ECS instances, you must be familiar with the limits of instance families.

General limits

- **Windows operating systems support a maximum of 64 vCPUs in instance specifications.**
- **ECS instances do not support the installation of virtualization software and secondary virtualization.**
- **Sound card applications are not supported. Only GPU instances support virtual sound cards. External hardware devices, such as hardware dongles, USB flash drives, external hard disks, and bank U keys, cannot be directly connected to ECS instances.**
- **ECS does not support multicast protocols. If multicasting services are required, we recommend that you use unicast instead.**

Instance family ga1

To create a ga1 instance, you must use one of the following images pre-installed with drivers:

- **Ubuntu 16.04 with an AMD GPU driver pre-installed**
- **Windows Server 2016 English version with an AMD GPU driver pre-installed**
- **Windows Server 2008 R2 English version with an AMD GPU driver pre-installed**

Note:

- **A ga1 instance uses an optimized driver provided by Alibaba Cloud and AMD. The driver is installed in images provided by Alibaba Cloud and is currently unavailable for download.**
- **If the GPU driver malfunctions due to improper removal of related components, you need to replace the system disk to restore GPU related functions.**

   **Note:**
  This operation causes data loss.
- **If the driver malfunctions because an improper image is selected, you need to replace the system disk to reselect an image with an AMD GPU driver pre-installed.**
- **For Windows Server 2008 or earlier, you cannot connect to the VNC after the GPU driver takes effect. The VNC is irresponsive with a black screen or stuck at the splash screen. You can use other methods such as Remote Desktop Protocol (RDP) to access the system.**
- **RDP does not support DirectX, OpenGL, or other related applications. You need to install the VNC and a client, or use other supported protocols such as PCOIP and XenDesktop HDX 3D.**

Instance families gn4, gn5i, and gn5

- **Bandwidth: If you use an image of Windows Server 2008 R2 for a gn4 instance, you cannot use the Connect to VNC function in the ECS console to connect to the instance after the installed GPU driver takes effect. You need to set the bandwidth to a non-zero value or attach an Elastic IP address to the created instance.**

- **Image: If an NVIDIA GPU driver is not required, you can select any image, and then** *Install the CUDA and GPU drivers for a Linux instance* **or** *Install the CUDA and GPU drivers for a Windows instance***.**

2.2.4 Notice for Windows users

Before using Windows-based ECS instances, you must consider the following points:

- **Data loss may occur if a local disk is used as the data disk of an instance. We recommend that you use a cloud disk to create your instance if you are not sure about the reliability of the data architecture.**
- **Do not close the built-in shutdownmon.exe process. Otherwise, the server may require a longer time to restart.**
- **Do not rename, delete, or disable Administrator accounts or it may affect the use of the server.**
- **We do not recommend that you use virtual memory.**
- **When you modify your computer name, you must synchronize the following key values in the registry. Otherwise, the computer name cannot be modified, causing failure when installing 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
```

2.2.5 Notice for Linux users

Before using Linux-based ECS instances, you must consider the following points:

- **Do not modify content of the default /etc/issue files under a Linux instance. Otherwise, the custom image created from the instance cannot be recognized, and instances created based on the image cannot start as expected.**
- **Do not arbitrarily modify the permissions of each directory in the partition where the root directory is located, especially permissions of /etc, /sbin, /bin, /boot, /dev, /usr, and /lib directories. Improper modification of permissions can cause errors.**
- **Do not rename, delete, or disable Linux root accounts.**
- **Do not compile or perform any other operations on the Linux kernel.**

- **We do not recommend the use of Swap for partitioning.**
- **Do not enable the NetWorkManager service. This service conflicts with the internal network service of the system, causing network errors.**

2.2.6 Notice on defense against DDoS attacks

You need to purchase Anti-DDoS Pro to defend against DDoS attacks. For more information, see *Apsara Stack Security Product Introduction*.

2.3 Quick start

2.3.1 Overview

This topic describes how to quickly create and connect to an ECS instance.

Perform the following procedure:

1. *Create a security group*

   A security group is a virtual firewall used to control traffic to and from ECS instances. Each ECS instance must be added to at least one security group. Before creating an instance, you must select a security group to control traffic to and from the instance.

2. *Create an instance*

   An ECS instance is a virtual machine that contains basic computing components such as CPU, memory, operating system, network, and disks. After a security group is created, you can select an instance type based on your business requirements. For more information, see *Instance types*.

3. *Connect to an instance*

   Select a remote connection method based on the network configuration and operating system of the ECS instance and your local operating system. After you log on to the instance, you can perform other operations on it, such as installing applications.

2.3.2 Log on to the ECS console

This topic describes how to log on to the ECS console.

Prerequisites

- Before logging on to the ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel. The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.
- We recommend that you use the Google Chrome browser.

Procedure

1. In the address bar, enter the URL used to access the ASCM console. Press Enter.
2. Enter your username and password.

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

   > **Note:**
   > When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).

3. Click Login to go to the ASCM console homepage.
4. In the top navigation bar, choose Products > Elastic Computing > Elastic Compute Service.

2.3.3 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.

Prerequisites

A Virtual Private Cloud (VPC) has been created. For more information, see *VPC User Guide*.

Context

Instances that belong to the same account and are in the same region and in the same security group can communicate with each other over the internal network. If instances that belong to the same account in the same region are in

different security groups, you can implement internal network communication by authorizing mutual access between two security groups.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Networks and Security > Security Groups.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click New Security Group.
5. Configure the parameters of the security group.

Type	Parameter	Required	Description
Region	Organization	Yes	The organization to which the security group belongs. Make sure that the security group and the VPC belong to the same organization.
	Resource Set	Yes	The resource set to which the security group belongs. Make sure that the security group and the VPC belong to the same resource set.
	Region	Yes	The region to which the security group belongs. Make sure that the security group and the VPC belong to the same region.
	Zone	Yes	The ID of the zone where the security group resides.

Type	Parameter	Required	Description
Basic Settings	VPC	Yes	The VPC to which the security group belongs.
	Security Group Name	No	The name must be 2 to 128 characters in length and start with a letter. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,). It cannot start with http:// or https://.
	Description	No	The description of the security group. We recommend that you provide an informational description to simplify future management operations. The name must be 2 to 256 characters in length and start with a letter. It can contain letters, digits, periods (.), underscores (_), hyphens (-), and commas (,). It cannot start with http:// or https://.

6. Click Submit.

2.3.4 Create an instance

An ECS instance is a virtual machine that contains basic computing components such as CPU, memory, operating system, network, and disks.

Prerequisites

- **A VPC and a VSwitch have been created. For more information, see** *VPC User Guide* **.**
- **A security group is available. If not, create a security group first. For more information, see** *Create a security group***.**

Context

Some limits apply when you create GPU instances. For more information, see *Limits***.**

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, click Instances.**
3. **Click Create Instance.**
4. **Configure the instance parameters.**
   a) **Configure basic settings of the instance.**

Parameter	Required	Description
Organization	Yes	Select the organization to which the instance belongs.
Resource Set	Yes	Select the resource set to which the instance belongs.

   b) **Configure the region of the instance.**

Parameter	Required	Description
Region	Yes	Select the region where the instance is located.

Parameter	Required	Description
Zone	Yes	Select the zone where the instance is located. Zones refer to the physical zones with separate power supplies and networks in the same region. The internal networks of zones are interconnected, and faults in one zone are isolated from the other zone. If you need to increase the availability of your applications, we recommend that you create multiple instances in different zones.

c) Configure the network of the instance.

Parameter	Required	Description
Network Type	Yes	Select the type of the network to which the instance belongs. VPC is available.
VPC	Yes	Select a VPC for the instance.
VSwitch	Yes	Select the VSwitch to which the instance belongs.

Parameter	Required	Description
Private IP	No	Specify a private IP address for the instance. The IP address must be within the CIDR block of the VSwitch. If you do not specify the private IP address, the system automatically allocates a private IP address to the instance.

d) Optional: Select an IPv6 address for the instance.

e) Required: Select the security group to which the instance belongs.

f) Configure the instance type.

Parameter	Required	Description
Instance Family	Yes	Select the instance family to which the instance belongs. After an instance family is selected, you must specify a specific instance type.
Instance Type	Yes	The instance type. Select CPU and memory based on application requirements. Windows images require specific CPU and memory combinations. For more information, see *ECS product introduction* .

g) Configure the image used by the instance.

Parameter	Required	Description
Image Type	Yes	Select the type of the image. Public Image and Custom Image are available.

Parameter	Required	Description
Public Image	Determined by the image type	Select the public image used by the instance. Public images provided by Alibaba Cloud are fully licensed, highly secure, and stable. Public images include Windows Server system images and major Linux system images. This parameter must be specified when the image type is Public Image.
Custom Image	Determined by the image type	Select the custom image used by the instance. Custom images are created from instances or snapshots, or imported from your local device. This parameter must be specified when the image type is Custom Image.

h) Configure the storage settings for the instance.

Parameter	Required	Description
System Disk	Yes	The disk that is used to install the operating system. Ultra Disk and SSD are available.

Parameter	Required	Description
Data Disk	No	Ultra Disk and SSD are available. A maximum of 16 data disks can be added to an instance. The maximum capacity of each data disk is 32 TiB. You can select Release with Instance and Encrypt for the disk as needed. You can also add data disks after the instance is created. For more information, see *Create a disk*.

i) Configure the logon password settings for the instance.

Parameter	Required	Description
Set Password	Yes	Select when to set the password. Now and Later are available. If Later is selected, you can use the password reset feature to set the password at a later time. For more information, see *Change an instance logon password*. **Note:** The password is used to log on to the instance, not to the VNC.

Parameter	Required	Description
Logon Password	No	The password used to log on to the instance. The password must be 8 to 30 characters in length and must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The supported special characters include `( ) ' ~ ! @ # $ % ^ & * - _ + = | { } [ ] : ; ' < > , . ? /`
Confirm Password	No	Confirm the logon password.

j) Optional: Select the deployment set to which the instance belongs.

k) Optional: Enter the instance name.

The name must be 2 to 128 characters in length and start with a letter. It can contain periods (.), underscores (_), colons (:), and hyphens (-).

If the name is not specified, the system will assign an instance name at random.

l) Optional: Enter the custom data that will be automatically run after the instance is started.

Windows supports Batch and PowerShell scripts. Before you perform Base64 encoding, make sure that `[bat]` or `[powershell]` is included in the first line.

Linux supports shell scripts.

m) Required: Enter the number of instances you want to purchase.

5. Click Submit.

Result

The created instance appears in the instance list and is in the Preparing state. After the creation, the instance changes to the Running state.

2.3.5 Connect to an instance

2.3.5.1 Instance connecting overview

After an instance is created, you can connect to the instance to perform operations such as installing applications.

You can use one of the following methods to connect to an instance:

- **Use remote connection tools to connect to instances that have public IP addresses. For more information about the procedure, see the following topics:**
  - *Connect to a Linux-based instance by using SSH commands in Linux or Mac OS X*
  - *Connect to a Linux-based instance by using remote connection tools in Windows*
  - *Connect to a Windows instance by using RDP*
- **Use the VNC feature in the ECS console. For more information, see** *Connect to an ECS instance by using the VNC***.**

The username of a Windows instance is Administrator, and that of a Linux instance is root.

2.3.5.2 Connect to a Linux-based instance by using SSH commands in Linux or Mac OS X

This topic describes how to use SSH commands to connect to a Linux-based instance.

Prerequisites

Create a security group and an instance.

Procedure

1. **Enter the following command:** `ssh root@instance IP`**.**
2. **Enter the password for the** *root* **user to log on to the instance.**

2.3.5.3 Connect to a Linux-based instance by using remote connection tools in Windows

This topic describes how to connect to an instance by using the PuTTY tool.

Prerequisites

Remote connection tools are designed with similar logics. In this example, PuTTY is used to connect to an instance. Download PuTTY at the following URL: *http://www.chiark.greenend.org.uk/~sgtatham/putty/***.**

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 EIP of the instance to be connected.**
   - `Port`**: Select the default port 22.**
   - `Connection Type`**: Select** `SSH`**.**
   - `Saved Session`**: Enter the name of the session. Click Save. After the settings are saved, PuTTY remembers the name and IP address of the instance. This eliminates the need to enter them 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, PuTTY displays security alerts. Click Yes to proceed.
4. **Enter the username** `root` **and press Enter.**
5. **Enter the password for the instance and press Enter.**

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

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

2.3.5.4 Connect to a Windows instance by using RDP

This topic describes how to connect to a Windows instance by using Remote Desktop Protocol (RDP).

Prerequisites

- **The instance and the security group are created.**
- **The instance is in the Running state.**
- **A logon password is set for the instance.**
- **An Elastic IP address (EIP) is bound with the instance.**
- **An inbound security group rule is added to the security group to allow the RDP port.**

Rule direction	Authorization policy	Protocol type	Port range	Priority	Authorization type	Authorization object
Inbound	Accept	TCP	3389/3389	1	IPv4 CIDR block	0.0.0.0/0

- CredSSP-related security updates are installed on the operating system of the instance.

Procedure

1. Activate the Remote Desktop Connection feature by using any of the following methods:
   - Click Start, enter `mstsc` in the search box, and click mstsc in the search result.
   - Press Windows Key + R. In the Run dialog box that appears, enter `mstsc` and click OK.
2. In the Remote Desktop Connection dialog box, enter the EIP of the instance and click Show Options.
3. Enter the username.

   The default username is administrator.
4. Optional: If you do not want to enter the password upon subsequent logons, select Allow me to save credentials.
5. Click Connect.
6. In the Windows Security dialog box that appears, enter the password for the account and click OK.

Result

After you log on to the instance, the Windows desktop appears.

If authentication errors occur or the required function is not supported, install security updates.

1. *Connect to an ECS instance by using the VNC* before proceeding.
2. Choose Start > Control Panel.
3. Click System and Security.
4. Click Check for updates in the Windows Updates pane.
5. If updates are available, click Install updates.
6. Restart the instance.

2.3.5.5 Connect to an ECS instance by using the VNC

You can access your instance by using the VNC in the ECS console when other SSH clients such as PuTTy, Xshell, and SecureCRT do not work properly.

Prerequisites

- The instance is in the Running state.
- The root certificate is imported to your web browser. For more information, see *Install a certificate in Windows*.
- If you log on to an instance for the first time after the instance is created, make sure that you set a new VNC password. For more information, see *Change the VNC password*.

Context

The VNC password is used to log on to the VNC of the ECS console, and the instance password is used to log on to an instance.

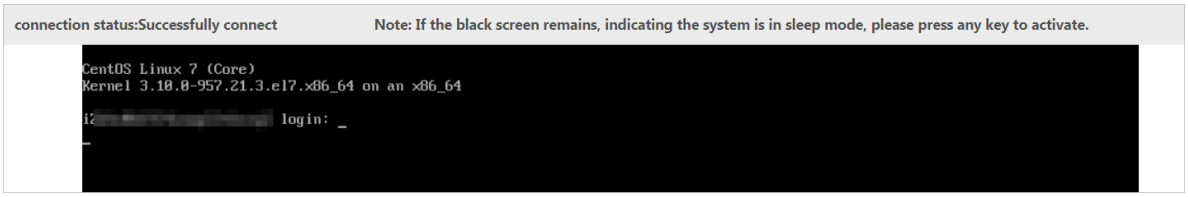
You can use the VNC to connect to an instance to solve issues shown in the following table.

Scenario	Resolution
The instance startup is slowly due to self-check upon startup.	Check the progress of the self check.
The firewall of the operating system is enabled by mistake.	Disable the firewall.
Abnormal processes appear, which consume large amounts of CPU or bandwidth resources.	Troubleshoot and terminate the abnormal processes.

Procedure

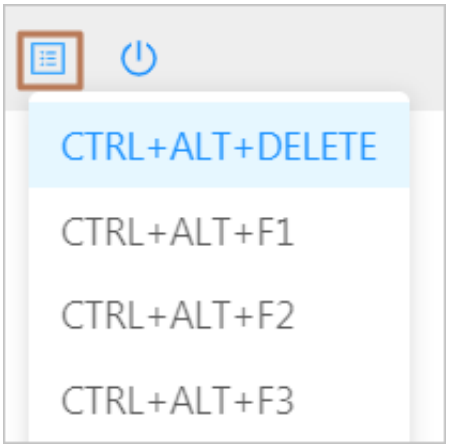
1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Instances.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the ECS instance, and click Remote Connection in the Actions column.
5. Enter the VNC password, and then click OK.

   After the connection is successful, the logon page is displayed, as shown in the following figure.



6. Enter the username and password.
   - For Linux instances: Enter the username `root` and the logon password.

     Note:
     Passwords in Linux are not displayed as you type. Press the Enter key after you enter the password.
   - For Windows instances: To use key combinations such as Ctrl + Alt + Delete, click the corresponding key combination in the upper-right corner of the VNC page.

     Enter the username and password as prompted, and click the Log In icon  .

2.4 Instances

2.4.1 Create an instance

An ECS instance is a virtual machine that contains basic computing components such as CPU, memory, operating system, network, and disks.

Prerequisites

- A VPC and a VSwitch have been created. For more information, see *VPC User Guide* .
- A security group is available. If not, create a security group first. For more information, see *Create a security group*.

Context

Some limits apply when you create GPU instances. For more information, see *Limits*.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Instances.
3. Click Create Instance.
4. Configure the instance parameters.
   a) Configure basic settings of the instance.

Parameter	Required	Description
Organization	Yes	Select the organization to which the instance belongs.
Resource Set	Yes	Select the resource set to which the instance belongs.

   b) Configure the region of the instance.

Parameter	Required	Description
Region	Yes	Select the region where the instance is located.

Parameter	Required	Description
Zone	Yes	Select the zone where the instance is located. Zones refer to the physical zones with separate power supplies and networks in the same region. The internal networks of zones are interconnected, and faults in one zone are isolated from the other zone. If you need to increase the availability of your applications, we recommend that you create multiple instances in different zones.

c) Configure the network of the instance.

Parameter	Required	Description
Network Type	Yes	Select the type of the network to which the instance belongs. VPC is available.
VPC	Yes	Select a VPC for the instance.
VSwitch	Yes	Select the VSwitch to which the instance belongs.

Parameter	Required	Description
Private IP	No	Specify a private IP address for the instance. The IP address must be within the CIDR block of the VSwitch. If you do not specify the private IP address, the system automatically allocates a private IP address to the instance.

d) Optional: Select an IPv6 address for the instance.

e) Required: Select the security group to which the instance belongs.

f) Configure the instance type.

Parameter	Required	Description
Instance Family	Yes	Select the instance family to which the instance belongs. After an instance family is selected, you must specify a specific instance type.
Instance Type	Yes	The instance type. Select CPU and memory based on application requirements. Windows images require specific CPU and memory combinations. For more information, see *ECS product introduction* .

g) Configure the image used by the instance.

Parameter	Required	Description
Image Type	Yes	Select the type of the image. Public Image and Custom Image are available.

Parameter	Required	Description
Public Image	Determined by the image type	Select the public image used by the instance. Public images provided by Alibaba Cloud are fully licensed, highly secure, and stable. Public images include Windows Server system images and major Linux system images. This parameter must be specified when the image type is Public Image.
Custom Image	Determined by the image type	Select the custom image used by the instance. Custom images are created from instances or snapshots, or imported from your local device. This parameter must be specified when the image type is Custom Image.

h) Configure the storage settings for the instance.

Parameter	Required	Description
System Disk	Yes	The disk that is used to install the operating system. Ultra Disk and SSD are available.

Parameter	Required	Description
Data Disk	No	Ultra Disk and SSD are available. A maximum of 16 data disks can be added to an instance. The maximum capacity of each data disk is 32 TiB. You can select Release with Instance and Encrypt for the disk as needed. You can also add data disks after the instance is created. For more information, see *Create a disk*.

i) Configure the logon password settings for the instance.

Parameter	Required	Description
Set Password	Yes	Select when to set the password. Now and Later are available. If Later is selected, you can use the password reset feature to set the password at a later time. For more information, see *Change an instance logon password*. **Note:** The password is used to log on to the instance, not to the VNC.

Parameter	Required	Description
Logon Password	No	The password used to log on to the instance. The password must be 8 to 30 characters in length and must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The supported special characters include `( ) ' ~ ! @ # $ % ^ & * - _ + = | { } [ ] : ; ' < > , . ? /`
Confirm Password	No	Confirm the logon password.

j) Optional: Select the deployment set to which the instance belongs.

k) Optional: Enter the instance name.

The name must be 2 to 128 characters in length and start with a letter. It can contain periods (.), underscores (_), colons (:), and hyphens (-).

If the name is not specified, the system will assign an instance name at random.

l) Optional: Enter the custom data that will be automatically run after the instance is started.

Windows supports Batch and PowerShell scripts. Before you perform Base64 encoding, make sure that `[bat]` or `[powershell]` is included in the first line.

Linux supports shell scripts.

m) Required: Enter the number of instances you want to purchase.

5. Click Submit.

Result

The created instance appears in the instance list and is in the Preparing state. After the creation, the instance changes to the Running state.

2.4.2 Connect to an instance

2.4.2.1 Instance connecting overview

After an instance is created, you can connect to the instance to perform operations such as installing applications.

You can use one of the following methods to connect to an instance:

- **Use remote connection tools to connect to instances that have public IP addresses. For more information about the procedure, see the following topics:**
  - *Connect to a Linux-based instance by using SSH commands in Linux or Mac OS X*
  - *Connect to a Linux-based instance by using remote connection tools in Windows*
  - *Connect to a Windows instance by using RDP*
- **Use the VNC feature in the ECS console. For more information, see** *Connect to an ECS instance by using the VNC*.

The username of a Windows instance is Administrator, and that of a Linux instance is root.

2.4.2.2 Connect to a Linux-based instance by using SSH commands in Linux or Mac OS X

This topic describes how to use SSH commands to connect to a Linux-based instance.

Prerequisites

Create a security group and an instance.

Procedure

1. **Enter the following command:** `ssh root@instance IP`.
2. **Enter the password for the** *root* **user to log on to the instance.**

2.4.2.3 Connect to a Linux-based instance by using remote connection tools in Windows

This topic describes how to connect to an instance by using the PuTTY tool.

Prerequisites

Remote connection tools are designed with similar logics. In this example, PuTTY is used to connect to an instance. Download PuTTY at the following URL: *http://www.chiark.greenend.org.uk/~sgtatham/putty/*.

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 EIP of the instance to be connected.**
   - `Port`**: Select the default port 22.**
   - `Connection Type`**: Select** `SSH`**.**
   - `Saved Session`**: Enter the name of the session. Click Save. After the settings are saved, PuTTY remembers the name and IP address of the instance. This eliminates the need to enter them 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, PuTTY displays security alerts. Click Yes to proceed.
4. **Enter the username** `root` **and press Enter.**
5. **Enter the password for the instance and press Enter.**

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

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

2.4.2.4 Connect to a Windows instance by using RDP

This topic describes how to connect to a Windows instance by using Remote Desktop Protocol (RDP).

Prerequisites

- **The instance and the security group are created.**
- **The instance is in the Running state.**
- **A logon password is set for the instance.**
- **An Elastic IP address (EIP) is bound with the instance.**
- **An inbound security group rule is added to the security group to allow the RDP port.**

Rule direction	Authorization policy	Protocol type	Port range	Priority	Authorization type	Authorization object
Inbound	Accept	TCP	3389/3389	1	IPv4 CIDR block	0.0.0.0/0

- CredSSP-related security updates are installed on the operating system of the instance.

Procedure

1. Activate the Remote Desktop Connection feature by using any of the following methods:
   - Click Start, enter `mstsc` in the search box, and click mstsc in the search result.
   - Press Windows Key + R. In the Run dialog box that appears, enter `mstsc` and click OK.
2. In the Remote Desktop Connection dialog box, enter the EIP of the instance and click Show Options.
3. Enter the username.

   The default username is administrator.
4. Optional: If you do not want to enter the password upon subsequent logons, select Allow me to save credentials.
5. Click Connect.
6. In the Windows Security dialog box that appears, enter the password for the account and click OK.

Result

After you log on to the instance, the Windows desktop appears.

If authentication errors occur or the required function is not supported, install security updates.

1. *Connect to an ECS instance by using the VNC* before proceeding.
2. Choose Start > Control Panel.
3. Click System and Security.
4. Click Check for updates in the Windows Updates pane.
5. If updates are available, click Install updates.
6. Restart the instance.

2.4.2.5 Install a certificate in Windows

Before you log on to the VNC, you must export the certificate from sites such as the Apsara Stack Cloud Management (ASCM) console and install it in your web browser.

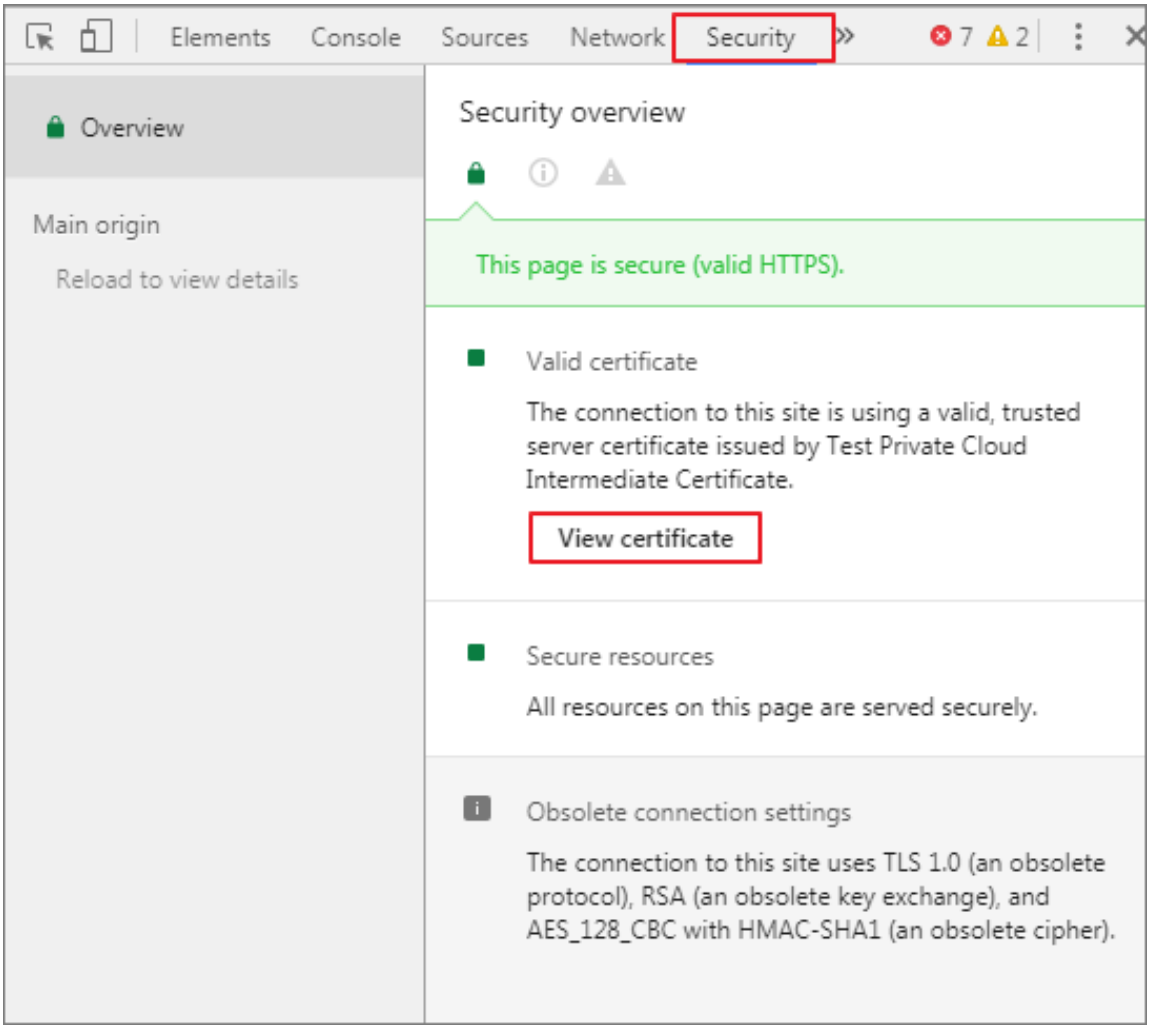
Context

The VNC feature is provided by the VNC proxy service. The VNC proxy service uses a different certificate from that of Apsara Infrastructure Management Framework. The certificate must be manually imported.

Procedure

1. **Export the certificate from the ASCM console.**
   a) **Log on to the ASCM console. Press the F12 key or Fn + F12 to view and select a certificate.**

      For example, in the Chrome browser, press the F12 key to open Chrome DevTools.



   b) **In the Certificate dialog box, click the Certificate Path tab, select the root certificate, and then click View Certificate.**
   c) **In the Certificate dialog box, click the Details tab, and click Copy to File.**
   d) **In the Certificate Export Wizard dialog box, click Next.**
   e) **Select DER encoded binary X.509 (.CER) as the format, and then click Next.**
   f) **Click Browse, select where to save the certificate, and then click Next.**
   g) **Enter the file name, and then click Save.**
   h) **Click Finish.**
   i) **Click OK.**

2. Install the certificate in your web browser.
   a) Double-click the certificate.
   b) In the Certificate dialog box, click Install Certificate.
   c) In the Certificate Import Wizard dialog box, click Next.
   d) In the dialog box that appears, select Place all certificates in the following store, and click Browse.
   e) In the Select Certificate Store dialog box, select Trusted Root Certificate Authority, and then click OK.
   f) In the Certificate Import Wizard dialog box, click Next.
   g) Click Finish.
   h) When a security warning message appears, click Yes.
3. Restart your web browser and log on to the ASCM console.

   If no security warning message is displayed in the left part of the address bar, the certificate is installed successfully.

2.4.2.6 Connect to an ECS instance by using the VNC

You can access your instance by using the VNC in the ECS console when other SSH clients such as PuTTy, Xshell, and SecureCRT do not work properly.

Prerequisites

- The instance is in the Running state.
- The root certificate is imported to your web browser. For more information, see *Install a certificate in Windows*.
- If you log on to an instance for the first time after the instance is created, make sure that you set a new VNC password. For more information, see *Change the VNC password*.

Context

The VNC password is used to log on to the VNC of the ECS console, and the instance password is used to log on to an instance.

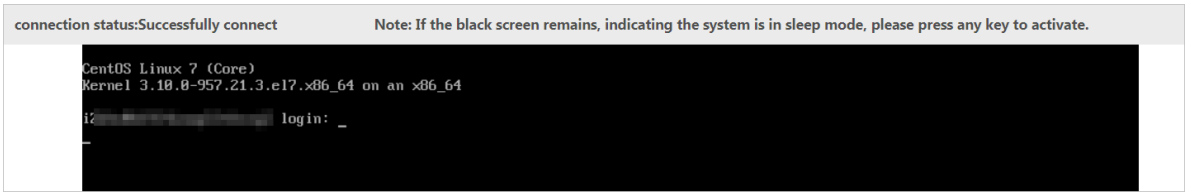
You can use the VNC to connect to an instance to solve issues shown in the following table.

Scenario	Resolution
The instance startup is slowly due to self-check upon startup.	Check the progress of the self check.
The firewall of the operating system is enabled by mistake.	Disable the firewall.
Abnormal processes appear, which consume large amounts of CPU or bandwidth resources.	Troubleshoot and terminate the abnormal processes.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Instances.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the ECS instance, and click Remote Connection in the Actions column.
5. Enter the VNC password, and then click OK.

   After the connection is successful, the logon page is displayed, as shown in the following figure.

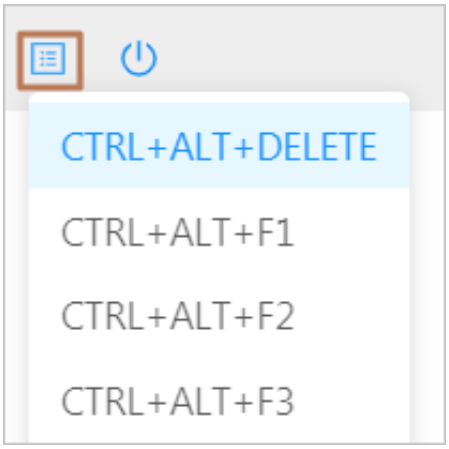


6. Enter the username and password.
   - For Linux instances: Enter the username `root` and the logon password.

     Note:

Passwords in Linux are not displayed as you type. Press the Enter key after you enter the password.

- **For Windows instances: To use key combinations such as Ctrl + Alt + Delete, click the corresponding key combination in the upper-right corner of the VNC page.**



Enter the username and password as prompted, and click the Log In icon  **.**

2.4.3 View instances

You can view the list of created instances and the details of a single instance. The details include basic configurations, disks, snapshots, security groups, and Elastic Network Interfaces (ENIs) of the instance.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, click Instances.**
3. **In the top navigation bar, select an organization, a resource set, and a region. The created instances that match the specified criteria are displayed.**
4. **Select a filtering option from the drop-down list, enter the relevant information in the search bar, and click Search.**

   You can select multiple filtering options to narrow down the search results.

Filtering option	Description
Instance Name	Enter an instance name to search for the instance.

Filtering option	Description
Instance ID	Enter an instance ID to search for the instance.
IP Address	Enter the IP address of an instance to search for the instance.
VPC ID	Enter a VPC ID to search for the instances that belong to the VPC.
Image ID	Enter an image ID to search for the instances that use the image.
Status	Select an instance status to search for instances in that status. Options are: · Running · Stopped · Starting · Stopping
Security Group ID	Enter a security group ID to search for instances in the security group.
Operating System	Enter an operating system to search for instances that use this operating system.
Region	Enter a region ID to search for instances in this region.
Organization ID	Enter an organization ID to search for instances that belong to this organizati on.
Resource Set ID	Enter a resource set ID to search for instances in this resource set.
Deployment Set ID	Enter a deployment set ID to search for instances that belong to this deployment set.

5. Select one of the following methods to access the details page of an instance.
   - In the Instance ID/Name column, click the instance ID.
   - Click Manage in the Actions column.
   - Choose More > Show Details from the Actions column.

2.4.4 Modify an instance

You can modify the name, description, and custom data of an existing instance.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, click Instances.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the instance and choose More > Modify from the Actions column.**
5. **Modify the name, description, and custom data for the instance.**

   The name must be 2 to 128 characters in length. The description must be 2 to 256 characters in length. The custom data must be 2 to 999 characters in length.
6. **Click OK.**

2.4.5 Stop an instance

You can stop an instance that is temporarily not in use. Stopping an instance will cause service interruption. Exercise caution when you stop an instance.

Prerequisites

The instance is in the Running state.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, click Instances.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Select one of the following methods to stop instances.**
   - **To stop one instance, find the instance and choose More > Instance Status > Stop from the Actions column.**
   - **To stop multiple instances, select the instances and click Stop at the bottom of the instance list.**
5. **Click OK.**

Result

In the Status column, the instance status changes from Running to Stopping. After the instance is stopped, the status changes to Stopped.

2.4.6 Start an instance

You can start a stopped instance.

Prerequisites

The instance is in the Stopped state.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, click Instances.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Select one of the following methods to start instances.**
   - **To start one instance, find the instance and choose More > Instance Status > Start from the Actions column.**
   - **To start multiple instances, select the instances and click Start at the bottom of the instance list.**
5. **Click OK.**

Result

In the Status column, the instance status changes from Stopped to Starting. After the instance is started, the status changes to Running.

2.4.7 Restart an instance

You must restart an instance after you change its logon password or install system updates. Restarting an instance will stop the instance for a period of time. As a result, services provided by the instance are interrupted. Exercise caution when you restart an instance.

Prerequisites

The instance is in the Running state.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, click Instances.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**

4. Select one of the following methods to restart instances.
   - To restart one instance, find the instance and choose More > Instance Status > Restart from the Actions column.
   - To restart multiple instances, select the instances and click Restart at the bottom of the instance list.
5. Select a restart mode.
   - Restart: Restart the instance normally.
   - Force Restart: Forcibly restart the instance. This may cause data loss if data in the instance is not yet written to the disk.
6. Click OK.

2.4.8 Delete an instance

You can delete an instance that is no longer needed to release its resources. After an instance is deleted, it cannot be recovered. We recommend that you back up data before you delete an instance. If a data disk is released with the instance, the data on the disk cannot be recovered.

Prerequisites

The instance is in the Stopped state.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Instances.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Select the instance and click Delete at the bottom of the instance list.
5. Click OK.

2.4.9 Change the instance type

You can change the instance type to suit your business changes, saving you from creating a new instance.

Prerequisites

The instance is in the Stopped state.

Procedure

1. *Log on to the ECS console*.

2. In the left-side navigation pane, click Instances.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance and click Upgrade/Downgrade in the Actions column.
5. Select a new instance type and click Submit.

   The instance types available for selection are displayed on the Change Configuration page.
6. Restart the instance in the console or by calling an API operation to make the new instance type take effect.

   For more information, see *Restart an instance* or *the RebootInstance section* in ECS Developer Guide.

2.4.10 Change an instance logon password

If you did not set a logon password when creating an instance or forgot the password, you can reset the password.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Instances.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance and select one of the following methods to access the instance details page.
   - In the Instance ID/Name column, click the instance ID.
   - Click Manage in the Actions column.
   - In the Actions column, choose More > Show Details.
5. Click Change Password.
6. Enter and confirm the new password, and then click OK.

   The password must be 8 to 30 characters in length and must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The supported special characters include `( ) ' ~ ! @ # $ % ^ & * - _ + = | { } [ ] : ; ' < > , . ? /`

7. **Restart the instance in the console or by calling an API operation to make the new password take effect.**

   For more information, see *Restart an instance* **or** *the RebootInstance section* **in ECS Developer Guide.**

2.4.11 Change the VNC password

If you log on to the VNC for the first time or forget the VNC password, you can reset the password.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, click Instances.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the instance and select one of the following methods to access the instance details page.**
   - **In the Instance ID/Name column, click the instance ID.**
   - **Click Manage in the Actions column.**
   - **In the Actions column, choose More > Show Details.**
5. **Click Change VNC Password.**
6. **Enter and confirm the new password, and click OK.**

   The password must be 6 characters in length and can contain digits and uppercase and lowercase letters. It does not support special characters.
7. **Restart the instance in the console or by calling an API operation to make the new password take effect.**

   For more information, see *Restart an instance* **or** *the RebootInstance section* **in ECS Developer Guide.**

2.4.12 Add an ECS instance to a security group

You can add a created instance to a security group and use security group rules to control network access to the instance.

Context

A security group acts as a virtual firewall and is used to provide security isolation. A security group controls access to ECS instances.

Instances that belong to the same account and are in the same region and in the same security group can communicate with each other over the internal network. If instances that belong to the same account in the same region are in different security groups, you can implement internal network communication by authorizing mutual access between two security groups.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Networks and Security > Security Groups.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the security group and click Manage Instances in the Actions column.**
5. **Click Add Instance.**
6. **Select the instance and click OK.**

   An instance can be added to five security groups. After an instance is added, the security group rules automatically apply to the instance.

2.4.13 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 instructions

- **Limits**

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

  - **Network type: VPC**
  - **Image: a system image or a custom image that is inherited from the system image**
  - **Operating system: one type included in** *Table 2-2: Supported operating systems*

    Table 2-2: Supported operating systems

Windows	Linux
■ Windows Server 2016 64-bit ■ Windows Server 2012 64-bit ■ Windows Server 2008 64-bit	■ CentOS ■ Ubuntu ■ SUSE Linux Enterprise ■ OpenSUSE ■ Debian ■ Aliyun Linux

  - **When you configure instance data 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` script in `Cloud Config` is modified for Linux instances, the script is automatically executed each time instances are restarted.**
- **OpenAPI: You can also use OpenAPI to customize instance data. For more information, see CreateInstance and ModifyInstanceAttribute in** *ECS Developer Guide* **.**

Linux instance data customization scripts

Linux instance data 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 data customization scripts

- **Linux instance customization scripts are executed after instances are started and before `/etc/init` is executed.**
- **Linux instance customization scripts can only be 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 in `#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/
```

2.4.14 Modify a private IP address

Each instance is assigned a private NIC and bound with a private IP address. You can modify the private IP address of the instance. The private IP address you use

must be within the CIDR block of the VSwitch to which the instance belongs and cannot be used by another instance.

Prerequisites

The instance is in the Stopped state.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Instances.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance and choose More > Change Private IP Address from the Actions column.
5. Enter a new private IP address and click OK.

   The private IP address you can use must be within the CIDR block of the VSwitch to which the instance belongs and cannot be used by another instance or for a specific purpose.

   For example, if the CIDR block of the VSwitch is 192.168.1.0/24, you can use an IP address in the range of 192.168.1.1 to 192.168.1.254. The first address 192.168.1.0 identifies the subnet itself, and the last address 192.168.1.255 identifies the broadcast address. Both addresses are reserved and cannot be used.

2.4.15 Install the CUDA and GPU drivers for a Linux instance

You must install a GPU driver on GPU instances to use the GPU. If the image you use does not contain a pre-installed GPU driver, you must manually install the CUDA and GPU drivers for the instance.

Prerequisites

If your instance cannot connect to the Internet, the installation file cannot be downloaded. You can install an FTP client on the instance to transfer the installation file to 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 on the instance.

Procedure

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

      A similar output is displayed:

      - **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.**

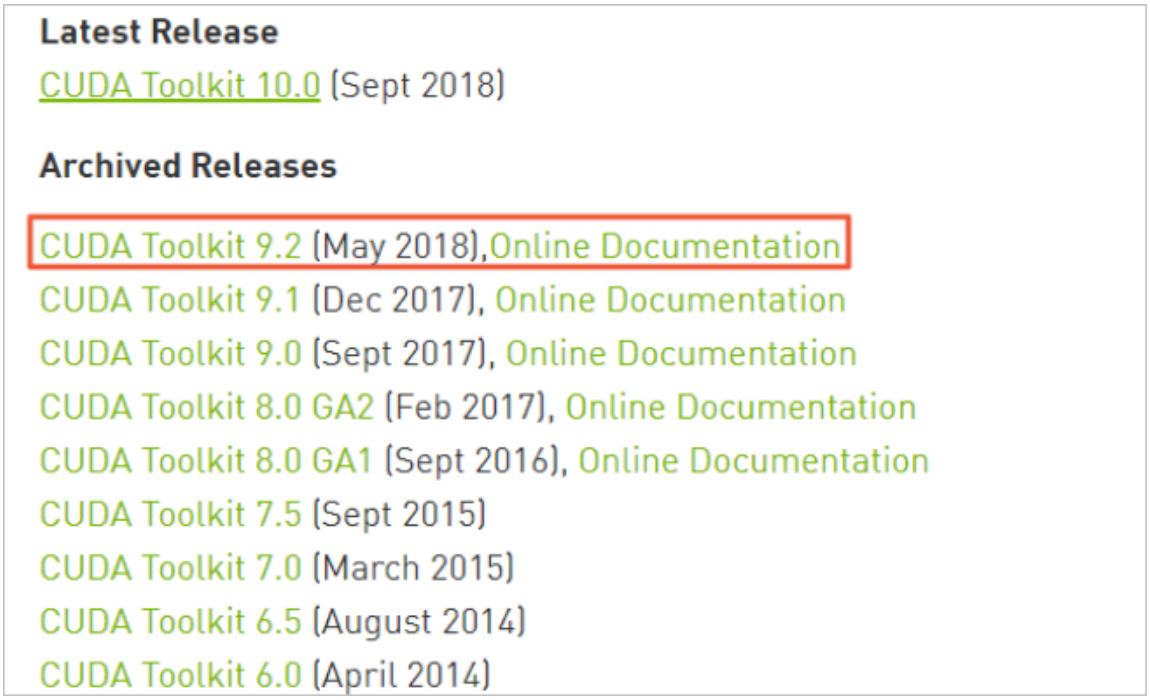
      - **CentOS: Copy the RPM package of the `kernel-devel` component and run the `rpm -ivh 3.10.0-862.14.4.el7.x86_64.rpm` command to install the package. *`3.10.0-862.14.4.el7.x86_64.rpm`* is used as an example. Replace it with the actual package name.**
      - **Ubuntu: Copy the DEB package of the `linux-headers`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

2. **Download the CUDA Toolkit.**

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

   This example uses *CUDA Toolkit 9.2*.

   Figure 2-3: Download the CUDA Toolkit



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

NVIDIA drivers are already included in the CUDA Toolkit.

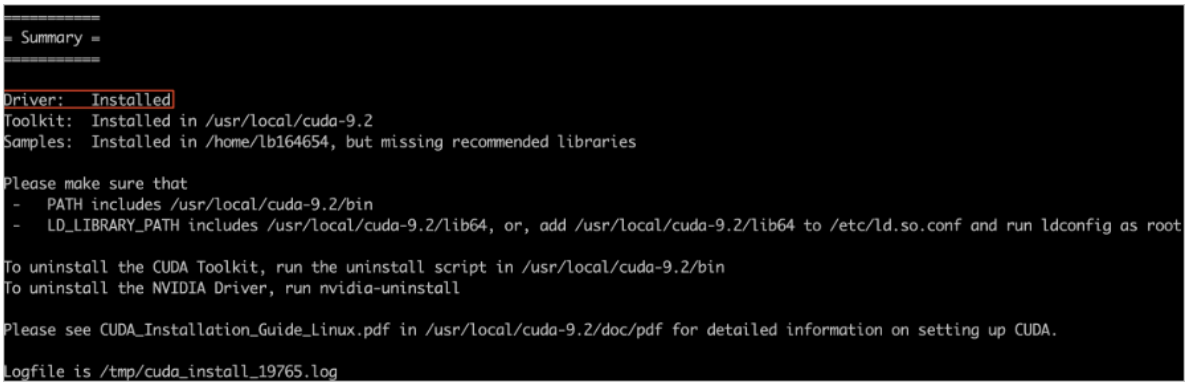
Figure 2-4: Download the drivers

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. Replace it with the actual file name.**
4. **Run the** `sudo sh ./cuda_9.2.148_396.37_linux.run --silent --verbose --driver --toolkit --samples` **command to install the CUDA driver.** *cuda_9.2*

.148_396.37_linux.run **is used as an example. Replace it with the actual file name.**

The installation takes about 10 to 20 minutes. When `Driver: Installed` **is displayed, the installation is successful.**

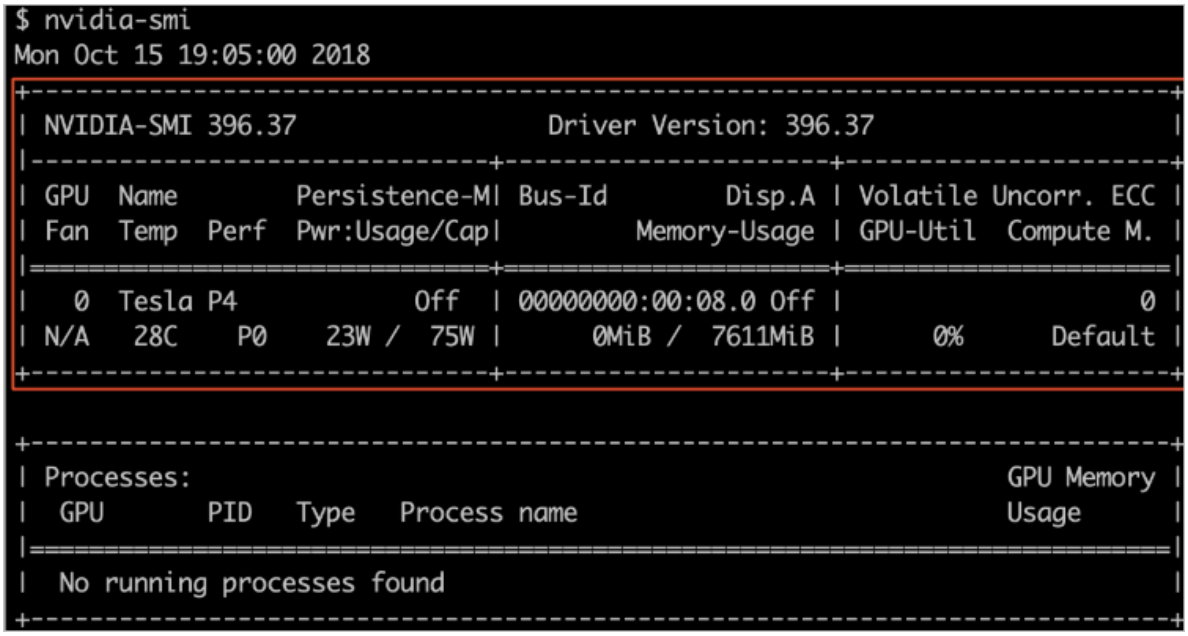
Figure 2-5: View the CUDA installation result



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

   If the output displays the details of the GPU driver, the driver is running properly.

   Figure 2-6: View the GPU driver status



What's next

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

2.4.16 Install the CUDA and GPU drivers for a Windows instance

You must install a GPU driver on GPU instances to use the GPU. If the image you use does not contain a pre-installed GPU driver, you must manually install the CUDA and GPU drivers for the instance.

Prerequisites

- **If your instance cannot connect to the Internet, the installation file cannot be downloaded. You can install an FTP client on the instance to transfer the installation file to the instance.**
- **To compile CUDA programs, first install a Windows compiling environment, such as Visual Studio 2015. If you do not need to compile CUDA programs, ignore it.**

Procedure

1. **Download the CUDA Toolkit.**
   a) **Access the** *official CUDA download page*. **Choose the version based on the GPU application requirements for CUDA.**
      This example uses *CUDA Toolkit 9.2*.
   b) **Choose a platform based on your operating system. Set Installer Type to exe (local) and click Download.**
      NVIDIA drivers are already included in the CUDA Toolkit.
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. Replace it with the actual file name.**
3. **Double-click** *cuda_9.2.148_windows.exe* **and follow the installation wizard to install the CUDA driver.** *cuda_9.2.148_windows.exe* **is used as an example. Replace it with the actual file name.**
   The installation takes about 10 to 20 minutes. When `Installed: - Nsight Monitor and HUD Launcher` **is displayed, the driver is installed.**
4. **Press** `Win + R` **and enter** `devmgmt.msc`.
   The NVIDIA device is displayed in Display Adapter.
5. **Press** `Win + R`, **enter** `cmd`, **and run the** `"C:\Program Files\NVIDIA Corporation\NVSMI\nvidia-smi"` **command.**
   If the output displays the details of the GPU driver, the driver is running properly.

What's next

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

2.5 Disks

2.5.1 Create a disk

You can create a data disk separately for an ECS instance to increase the storage space of the instance. This topic describes how to create an empty data disk. You cannot create a system disk separately.

Context

We recommend that you plan the number and size of data disks before you create the disks. The following limits apply to data disks:

- A maximum of 16 data disks can be attached to an instance. Disks and Shared Block Storage devices share the quota for data disks.
- A Shared Block Storage device can be attached to a maximum of four ECS instances.
- Each ultra disk, ultra Shared Block Storage device, standard SSD, and SSD Shared Block Storage device have a maximum capacity of 32 TiB.
- Created disks cannot be combined in ECS. They are independent of each other and cannot be formatted and combined into one disk.

We recommend that you do not use Logical Volume Manager (LVM) to create logical volumes on multiple disks, because a snapshot can only back up data of a single disk. If you create a logical volume on several disks by using LVM, data discrepancies will occur when you roll back these disks.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Storage > Disks.
3. Click Create Disk.

4. Configure the disk parameters.

Type	Parameter	Required	Description
Region	Organization	Yes	Select the organization to which the disk belongs.
	Resource Set	Yes	Select the resource set to which the disk belongs.
	Region	Yes	Select the region where the disk is located.
	Zone	Yes	Select the zone where the disk is located.
Basic settings	Name	Yes	Enter a disk name. The name must be 2 to 128 characters in length. It must start with a letter but cannot start with http:// or https://. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,).

Type	Parameter	Required	Description
	Specifications	Yes	Select a category of the disk and specify the disk size. Options are · SSD Disk · Ultra Disk · Shared SSD Disk: SSD Shared Block Storage device · Shared Ultra Disk: Ultra Block Storage device The valid values of the disk size are 20 to 32768 GiB.
	Encrypt	No	Specify whether to encrypt the disk. If Yes is selected, the Use Snapshot parameter is automatically set to No. You cannot use a snapshot to create a disk.

Type	Parameter	Required	Description
	Use Snapshot	No	Specify whether to create a disk from a snapshot. If you select Yes, you must specify a snapshot. The size of the created disk may be affected by the size of the specified snapshot: · If the specified disk size is greater than the size of the specified snapshot, the disk still uses the specified disk size. · If the specified disk size is smaller than the size of the specified snapshot, the disk uses the size of the specified snapshot.

5. Click Submit.

Result

The created disk is displayed in the disk list and in the Pending state.

What's next

After the disk is created, you must attach the disk to an instance and partition and format the disk. For more information, see:

- *Attach a disk*
- *Format a data disk for a Linux instance*

- *Format a data disk of a Windows instance*

2.5.2 View disks

You can view the list of created disks or the details of a single disk.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Storage > Disks.**
3. **In the top navigation bar, select an organization, a resource set, and a region. The created disks that match the specified criteria are displayed.**
4. **Select a filtering option from the drop-down list, enter the relevant information in the search bar, and click Search.**

   You can select multiple filtering options to narrow down the search results.

Filtering option	Description
Disk Name	Enter a disk name to search for the disk.
Disk ID	Enter a disk ID to search for the disk.
Instance ID	Enter an instance ID to search for the disks that are attached to the instance.
Disk Status	Select a disk state to search for disks. Options are: · Running · Pending · Attaching · Detaching · Creating · Deleting · Deleted Note: The deleted disks are no longer displayed in the disk list. · Initializing · All Status

Filtering option	Description
Disk Properties	Select a disk type to search for disks. Options are: · All · System Disk · Data Disk
Policy ID	Enter the ID of an automatic snapshot policy to search for the disks that use this policy.
Encryption Key ID	Enter the ID of an encryption key to search for the disks that are encrypted with this key.

5. In the Disk ID/Name column, click a disk ID to view the disk details.

   The details page of the disk displays its properties and attaching information.

2.5.3 Roll back a disk

If you have created snapshots for a disk, you can use a snapshot to roll back the disk to the state it was when the snapshot was taken. Rolling back a disk is irreversible. Once the disk is rolled back, the disk data before the rollback time cannot be restored. Exercise caution when you perform this operation.

Prerequisites

- Snapshots have been created for the disk.
- The disk is not released.
- The instance where the target disk resides must be in the Stopped state.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Snapshots and Images > Snapshots.
3. In the top navigation bar, select an organization, a resource set, and a region.

4. Select a filtering option from the drop-down list, enter the relevant information in the search bar, and click Search.

   You can select multiple filtering options to narrow down the search results.

Filtering option	Description
Snapshot Name	Enter a snapshot name to search for the snapshot.
Snapshot ID	Enter a snapshot ID to search for the snapshot.
Instance ID	Enter an instance ID to search for the snapshots related to the instance.
Disk ID	Enter a disk ID to search for the snapshots related to the disk.
Snapshot Type	Select a snapshot type to search for the snapshots of that type. Options include: · All · User Snapshots: manual snapshots. · Automatic snapshots: automatic snapshots.
Creation Time	Enter a creation time to search for the snapshots that were created at that time.

5. Find the snapshot and click Restore in the Actions column.
6. Click OK.

2.5.4 Modify the disk properties

You can modify the properties of a created disk, such as changing the settings of the Release Disk with Instance and Release Automatic Snapshots with Disk options.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Storage > Disks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the disk and choose More > Modify Disk Properties from the Actions column.

5. Modify the release mode.
   - Release Disk with Instance: When this option is selected, the disk is released together when the instance it is attached to is deleted. When this option is not selected, the disk changes to the Pending state when the instance it is attached to is deleted.
   - Release Automatic Snapshots with Disk: When this option is selected, the automatic snapshots created for the disk is released together when the disk is deleted. When this option is not selected, the automatic snapshots are retained.
6. Click OK.

2.5.5 Modify the disk description

You can modify the name and description of a created disk.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Storage > Disks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the disk and choose More > Modify Disk Description from the Actions column.
5. Modifies the name and description of the disk.

   The name must be 2 to 128 characters in length and start with a letter. It can contain periods (.), underscores (_), colons (:), and hyphens (-).

   The description must be 2 to 256 characters in length and cannot start with http :// and https://.
6. Click OK.

2.5.6 Attach a disk

You can attach a disk that is created separately to an ECS instance as a data disk. The disk and the instance must be in the same region and the same zone.

Prerequisites

The disk is in the Pending state.

Context

- You do not need to attach data disks that are created at the same time as an instance.
- A disk can only be attached to an instance that is in the same zone and region as the disk.
- You can attach a disk to a single ECS instance at a time.
- A Shared Block Storage device can be attached to a maximum of four ECS instances at the same time.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Storage > Disks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the disk and choose More > Attach from the Actions column.
5. Specify the destination instance and configure the release mode as needed.
   - If you select Release Disk with Instance, the disk is released together when the instance it is attached to is deleted.
   - If you do not select Release Disk with Instance, the disk changes to the Pending state when the instance it is attached to is deleted.
6. Click OK.

2.5.7 Partition and format disks

2.5.7.1 Format a data disk for a Linux instance

Data disks created separately are not partitioned or formatted. This topic describes how to partition and format a data disk of a Linux instance.

Prerequisites

The disk has been attached to the instance.

Procedure

1. *Connect to the instance*.
2. Run the `fdisk -l` command to view all data disks attached to the ECS instance.

   If */dev/vdb* is not displayed in the command output, the ECS instance does not have a data disk. Check whether the data disk is attached to the instance.

```
[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
```

3. **Create partitions for the data disk.**
   a) **Run the `fdisk /dev/sdb` command.**
   b) **Enter `n` to create a new partition.**
   c) **Enter `p` to set the partition as the primary partition.**
   d) **Enter a partition number and press the Enter key. In this example, `1` is entered to create Partition 1.**
   e) **Enter the number of the first available sector. This example uses the default value. This is done by pressing the Enter key. You can also enter a value from 1 to 41610 and press the Enter key.**
   f) **Enter the number of the last sector. This example uses the default value. This is done by pressing the Enter key. You can also enter a value from 1 to 11748 and press the Enter key.**
   g) **Optional: Optional. To create multiple partitions, repeat steps b through f until all four primary partitions are created.**
   h) **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!
```

4. **Run the `fdisk -l` command to view the partitions.**

   If *`/dev/vdb1`* is displayed in the command output, new partition *`vdb1`* is 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
```

5. **Format the new partition. In this example, the new partition is formatted as ext3 after you run the `mkfs.ext3 /dev/vdb1` command.**

   The time required for formatting depends on the disk size. You can also format the new partition to another file system. For example, you can run the `mkfs.ext4 /dev/vdb1` command to format the partition as ext4.

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

6. **Run the `echo '/dev/vdb1 /mnt ext3 defaults 0 0' >> /etc/fstab` command to write the information of the new partition to the /etc/fstab file. You can run the `cat /etc/fstab` command to view the new partition information.**

   Ubuntu 12.04 does not support barriers. To write the information of the new partition into the /etc/fstab file, you must run the `echo '/dev/vdb1 /mnt ext3 barrier=0 0 0' >> /etc/fstab` command.

   In this example, the partition information is added to the ext3 file system. You can also modify the ext3 parameter to add the partition information to another type of file system.

   To attach the data disk to a specific folder, for example, to store web pages, modify the /mnt part of the preceding command.

```
[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-bc96-057dd64f**** / 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
```

7. **Mount the new partitions. Run the `mount -a` command to mount all the partitions listed in *`/etc/fstab`* and run the `df -h` command to view the result.**

   If the following information is displayed, the new partitions are mounted and 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
```

2.5.7.2 Format a data disk of a Windows instance

Data disks created separately are not partitioned or formatted. This topic describes how to partition and format a data disk of a Windows instance. This example uses Windows Server 2008.

Prerequisites

The disk has been attached to an instance.

Procedure

1. **In the lower-left corner of the screen, click the Server Manager icon.**
2. **In the left-side navigation pane of the Server Manager window, choose Storage > Disk Management.**
3. **Right-click an empty partition and select New Simple Volume from the shortcut menu.**

   If the disk status is Offline, change it to Online.
4. **Click Next.**
5. **Set the size of the simple volume, which is the partition size. Then click Next.**

   The default value is the maximum value of the disk space. You can specify the partition size as needed.
6. **Specify the drive letter and then click Next.**
7. **Specify the formatting options and then click Next.**

   We recommend that you format the partition with the default settings provided by the wizard.
8. **When the wizard prompts that the partition has been completed, click Finish to close the wizard.**

2.5.8 Resize a system disk

You can increase the size of a system disk by replacing the system disk as your business needs evolve.

Context

Note the following items when you resize a system disk:

- **You cannot reduce the disk capacity by resizing a system disk.**
- **You cannot convert system disks to data disks.**
- **You cannot resize a system disk on Windows Server 2003.**
- **Resizing a system disk does not change the IP address and MAC address of the instance.**
- **Make sure that you have sufficient snapshot quota to fulfill the automatic snapshot policy of the new system disk.**

The following risks may occur when you resize a system disk:

- **Your workloads are interrupted because you must stop the instance to resize the system disk.**
- **After replacing a system disk, you must redeploy the business runtime environment in the new system disk. This may result in a long business interruption. Exercise caution when you perform this operation.**
- **After you replace a system disk, manual snapshots of the disk are retained. These snapshots can still be used to create custom images. Manual snapshots of the original system disk cannot be used to roll back the new system disk because the disk ID is changed after the system disk is replaced.**
- **Resizing a system disk will release the original system disk.**

Procedure

If you want to resize a system disk, perform the following steps:

1. *(Optional) Step 1: Create a snapshot of a system disk*
2. *(Optional) Step 2: Create an image from a snapshot*
3. *Step 3: Replace a system disk*

If you do not plan to save the data of the system disk, skip to step 3 to replace the system disk. Doing so will remove the original system disk directly.

Step 1: Create a snapshot of a system disk

Before you create a snapshot, make sure that the instance is in the Running or Stopped state and the system disk is in the Running state.

To avoid impact on your business, try to create snapshots during off-peak hours. The initial snapshot of a 40 GiB disk takes about 40 minutes to create. Therefore, plan for enough time to create a snapshot. When you create a snapshot, make sure that the system disk has sufficient free space. We recommend that you reserve an additional 1 GiB free space. Otherwise, the system may not start properly after the system disk is resized.

1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Instances.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance, select filtering options, enter the appropriate information in the search bar, and then click Search.
5. Use one of the following methods to open the details page of an instance:
   - In the Instance ID/Name column, click the instance ID.
   - Click Manage in the Actions column corresponding to the instance.
   - Choose More > Show Details in the Actions column corresponding to the instance.
6. Click Disks.
7. Find the system disk and click Create Snapshot in the Actions column.
8. Enter the snapshot name and description, and then click OK.

   Note:
   The snapshot name cannot start with auto. auto is a reserved prefix for automatic snapshots.

   You can view the progress of snapshot creation and the snapshot status on the snapshot list page. For more information, see *View snapshots*.

Step 2: Create an image from a snapshot

When you create an image, make sure that the system disk has sufficient free space . We recommend that you reserve an additional 1 GiB free space. Otherwise, the system may not start properly after the system disk is resized.

1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Instances.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance, select filtering options, enter the appropriate information in the search bar, and then click Search.
5. Use one of the following methods to open the details page of an instance:
   - In the Instance ID/Name column, click the instance ID.
   - Click Manage in the Actions column corresponding to the instance.
   - Choose More > Show Details in the Actions column corresponding to the instance.
6. Click the Snapshots tab.
7. Find the snapshot and click Create Custom Image in the Actions column.
8. Enter the name and description of the custom image, and then click OK.

   The name must be 2 to 128 characters in length and can contain periods (.), underscores (_), hyphens (-), and colons (:). It cannot start with a special character or digit.

   The description must be 2 to 256 characters in length and cannot start with http :// or https://.

   > **Notice:**
   > - Keep the image name in mind. The custom image will be used when you replace the system disk.
   > - You cannot select a data disk snapshot to create a custom image.

   You can view the created image on the image list page. For more information, see *View images*.

Step 3: Replace a system disk

Before replacing a system disk, make sure that:

- You have backed up all the data of the original system disk to avoid data loss.
- The instance whose system disk you want to replace is in the Stopped state.

After you replace a system disk,

- Manual snapshots of the original system disk are retained. The original automatic snapshot policy becomes invalid and must be reconfigured.
- The system disk ID changes and the original system disk is released.

1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Instances.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance, select filtering options, enter the appropriate information in the search bar, and then click Search.
5. Use one of the following methods to open the details page of an instance:
   - In the Instance ID/Name column, click the instance ID.
   - Click Manage in the Actions column corresponding to the instance.
   - Choose More > Show Details in the Actions column corresponding to the instance.
6. Click Replace System Disk.
7. Configure the properties of the new system disk.

   Note:
   Before you replace the system disk, we recommend that you read the prerequisites and background information.

   - Image Type: If you want to save the data of the original system disk, select the custom image created in *Step 2: Create an image from a snapshot*. Otherwise, select a public image.
   - System Disk: You cannot change the disk type, but can specify the size of the new disk. The new disk size cannot be smaller than the original disk size. The maximum value is 500 GiB.
8. Click OK.

   It takes about 10 minutes to replace the system disk. After that, the instance will start automatically.

What's next

If an automatic snapshot policy is configured for the original system disk, you must reconfigure the snapshot policy for the new system disk. For more information, see *Configure an automatic snapshot policy for multiple disks*.

2.5.9 Reinitialize a disk

You can reinitialize a disk to restore it to its initial state.

Prerequisites

- The disk is in the Running state.
- The instance is in the Stopped state.
- After a disk is reinitialized, its data is lost and cannot be recovered. Exercise caution when you perform this operation. We recommend that you back up the data of the disk or create snapshots before you reinitialize the disk. For more information, see *Create a snapshot*.

Context

The result of disk reinitialization depends on the disk type and how the disk is created.

- System disk:
  - The disk is restored to the initial state of the image used by the disk.
  - If the original image is deleted, the disk cannot be reinitialized.
- Data disk:
  - If the disk is empty when created, the disk is restored to an empty disk.
  - If the disk is created from a snapshot, the disk is restored to a disk with only the data of the source snapshot.
  - If the disk is created from a snapshot and the snapshot is deleted, the disk cannot be reinitialized.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Storage > Disks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the data disk and click Reinitialize in the Actions column.
5. Perform the following operations based on the disk type.
   - For a system disk, enter and confirm a new logon password, select the Start Instance After Reinitializing Disk option as needed, and then click OK.

     The password must be 8 to 30 characters in length and must contain at least three of the following character types: uppercase letters, lowercase letters,

digits, and special characters. The supported special characters include `( ) ' ~ ! @ # $ % ^ & * - _ + = | { } [ ] : ; ' < > , . ? /`

- For a data disk, click OK.

Result

When the disk is being reinitialized, the disk enters the Initializing state. After the reinitialization, it changes to the Running state.

2.5.10 Detach a data disk

You can detach a data disk, not a system disk.

Prerequisites

- For a Windows instance, you must bring the data disk offline in Disk Management.

  Note:
  To guarantee data integrity, we recommend that you stop read/write operations on the data disk when you detach the disk. Otherwise, data may be lost.

- For a Linux instance, you must connect to the instance and unmount the partitions on the disk.

  Note:
  If you have configured the `/etc/fstab` file to automatically mount the disk partitions upon instance startup, you must delete the mounting information from the `/etc/fstab` file before you detach the disk. Otherwise, you cannot connect to the instance after the instance is restarted.

- The data disk to be detached is in the Running state.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Storage > Disks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the data disk and choose More > Detach from the Actions column.
5. Click OK.

2.5.11 Release a data disk

You can release a data disk that is no longer needed. The released data disk cannot be recovered. Exercise caution when you release a data disk.

Prerequisites

The data disk is in the Pending state. If the data disk is attached to an instance, detach the disk from the instance first.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Storage > Disks.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the data disk and choose More > Release from the Actions column.**
5. **Click OK.**

2.6 Images

2.6.1 Create a custom image

You can create a custom image and use it to create identical instances or replace the system disks of existing instances. This allows you to have multiple instances with the same operating system and data environment.

Create a custom image from a snapshot

You can create a custom image from a system disk snapshot to fully load the operating system and data environment of the snapshot to the image. Before you perform this operation, make sure that a snapshot of system disks is used. You cannot create a custom image from snapshots of data disks.

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Snapshots and Images > Snapshots.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the snapshot and click Create Custom Image in the Actions column.**

5. **Enter the name and description of the image, and then click OK.**

   The name must be 2 to 128 characters in length and can contain periods (.), underscores (_), hyphens (-), and colons (:). It cannot start with a special character or digit.

   The description must be 2 to 256 characters in length and cannot start with http :// and https://.

Create a custom image from an instance

You can create a custom image from an instance to completely replicate the data of all disks of the instance, including the system disk and data disks.

Note:
To avoid data security risks, delete sensitive data before you create a custom image.

When you create a custom image from an instance, a snapshot is generated for each disk in the instance, and all the snapshots constitute a complete custom image.

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, click Instances.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the instance and choose More > Create Custom Image from the Actions column.**
5. **Enter the name and description of the custom image, and then click OK.**

   The name must be 2 to 128 characters in length and can contain periods (.), underscores (_), hyphens (-), and colons (:). It cannot start with a special character or digit.

   The description must be 2 to 256 characters in length and cannot start with http :// and https://.

2.6.2 View images

You can view the list of created images.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Snapshots and Images > Images.**

3. In the top navigation bar, select an organization, a resource set, and a region.
   The created images that match the specified criteria are displayed.
4. Select the tab based on the type of images you want to view.
   You can select the Custom Images or Public Images tab.
5. Select a filtering option from the drop-down list, enter the relevant information in the search bar, and click Search.
   You can select multiple filtering options to narrow down the search results.

Filtering option	Description
Image Name	Enter an image name to search for the image.
Image ID	Enter an image ID to search for the image.
Snapshot ID	Enter a snapshot ID to search for the images associated with the snapshot. This option is not available for public images.

2.6.3 Share custom images

You can share a custom image that you create to organizations that you manage to create multiple identical ECS instances in a short time.

Context

Only custom images can be shared. Shared images are not counted towards the image quota assigned to the organization.

The organization can use the shared image to create instances or replace system disks of existing instances.

You can delete shared images. After a shared image is deleted, the image is no longer visible to the organization to which the image was shared. The system disk of instances created from the shared image can no longer be reinitialized.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Snapshots and Images > Images.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the image and click Share Image in the Actions column.

5. **Select the organization to which you want to share the image and click OK.**

   An image can be shared only to the organizations that the image owner manages.

2.6.4 Import custom images

2.6.4.1 Limits on importing custom images

This topic describes the limits on importing images. You must understand the limits to ensure image availability and improve import efficiency.

The following limits apply when you import custom images:

- *Limits on importing custom images in Linux*
- *Limits on importing custom images in Windows*

Limits on importing custom images in Linux

When you import custom images in Linux, note 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 at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters.**
- **The firewall is disabled, and port 22 is opened.**
- **The Linux system disk size ranges from 40 GiB to 500 GiB.**
- **DHCP must be enabled in the image.**
- **SELinux is disabled.**
- **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.**

Table 2-3: Limits

Item	Standard operating system image	Non-standard operating system image
Description	The supported standard 32-bit and 64-bit operating systems include: · CentOS · Ubuntu · SUSE · openSUSE · Red Hat · Debian · CoreOS · Aliyun Linux Note: Support for standard operating systems may be subject to version changes. You can access the ECS console to check the latest supported operating systems.	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 perform any processing on the instances created from these images. After you create an instance, you must connect to the instance by clicking Connect to VPN in the ECS console. You can then configure the IP address, route, and password for the instance.

Item	Standard operating system image	Non-standard operating system image
Critical system configuration files	• Do not modify `/etc/issue*`. Otherwise, the version of the operating system cannot be identified, which leads to system creation failure. • Do not modify `/boot/grub/menu.lst`. Otherwise, the system may fail to start. • Do not modify `/etc/fstab`. Otherwise, partitions cannot be loaded, which leads to system startup failure. • Do not modify `/etc/shadow` to read-only. Otherwise, the password file cannot be modified, which leads to system creation failure. • Do not modify `/etc/selinux/config` to enable SELinux. Otherwise, the system may fail to start.	Requirements for standard operating system images are not met.

Item	Standard operating system image	Non-standard operating system image
Requirements for the basic system environment	· Do not adjust the system disk partitions. Only disks with a single root partition are supported. · Make sure that the system disk has sufficient storage 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` on a custom image. Otherwise, some of the services that Alibaba Cloud needs may be unavailable.	
Image file formats	Only images in the RAW, VHD, or qcow2 format 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 the VHD format, which has a smaller transmission footprint.	

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 at least 40 GiB.	

Limits on importing custom images in Windows

When you import custom images in Windows, note the following limits:

- **The password must be 8 to 30 characters in length. It must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters.**
- **Imported Windows images do not provide the Windows activation service.**
- **The firewall must be disabled. Otherwise, remote logon is unavailable. Port 3389 must be opened.**
- **The Windows system disk size ranges from 40 GiB to 500 GiB.**

Table 2-4: Limits

Item	Description
Operating system versions	Alibaba Cloud supports importing the following versions of 32-bit and 64-bit operating system images: · Microsoft Windows Server 2016 · Microsoft Windows Server 2012, including: - Microsoft Windows Server 2012 R2 (Standard Edition) - Microsoft Windows Server 2012 (Standard Edition and Datacenter Edition) · Microsoft Windows Server 2008, including: - 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, including: - 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, including Service Pack 1 (SP1) · Microsoft Windows 7, including: - Microsoft Windows 7 (Professional Edition) - Microsoft Windows 7 (Enterprise Edition) Note: Support for standard operating systems may be subject to version changes. You can access the ECS console to check the latest supported operating systems.

Item	Description
Requirements for the basic system environment	· Multi-partition system disks are supported. · Make sure that the system disk has sufficient storage 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 on an imported image. If it is installed, some of the services that Alibaba Cloud needs may be unavailable.
Image file formats	· RAW · VHD · qcow2 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 range from 40 GiB to 500 GiB. Note: We recommend that you import images in the VHD format, which has a smaller transmission footprint.

2.6.4.2 Convert the image file format

You can only import image files in the RAW, VHD, and qcow2 formats to ECS. If you want to import images in other formats, you must convert the image into a supported format. This topic describes how to convert the image format in Windows and Linux.

Context

You can use the qemu-img tool to convert an image from VMDK, VDI, VHDX, qcow1, or QED to RAW, VHD, or qcow2, or implement conversion between RAW, VHD, and qcow2.

Note:
We recommend that you use the qcow2 format if your application environment supports this format.

Windows

1. **Download qemu.**

   Visit *QEMU Binaries for Windows (64 bit)* **to download the qemu tool. Select a qemu version based on your operating system.**

2. **Install qemu.**

   The installation path in this example is `C:\Program Files\qemu`**.**

3. **Configure the environment variables for qemu.**

   a) **Choose Start > Computer, right-click Computer, and choose Properties from the shortcut menu.**

   b) **In the left-side navigation pane, click Advanced System Settings.**

   c) **In the System Properties dialog box that appears, click the Advanced tab and then click Environment Variables.**

   d) **In the Environment Variables dialog box that appears, find the Path variable from the System variables section.**

      - **If the Path variable exists, click Edit.**
      - **If the Path variable does not exist, click New.**

   e) **Add a system variable value.**

      - **In the Edit System Variable dialog box that appears, add** `C:\Program Files\qemu` **to the Variable value field, separate different variable values with semicolons (;), and then click OK.**
      - **In the New System Variable dialog box that appears, enter** `Path` **in the Variable name field, enter** `C:\Program Files\qemu` **in the Variable value field, and then click OK.**

4. **Open Command Prompt in Windows and run the** `qemu-img --help` **command. If a successful response is displayed, the tool is installed.**

5. **In the Command Prompt window, run the** `cd [Directory of the source image file]` **command to switch to a new file directory,**

   for example, `cd D:\ConvertImage`**.**

6. **In the Command Prompt window, run the `qemu-img convert -f raw -O qcow2 centos.raw centos.qcow2` command to convert the image file format.**

   The parameters are described as follows:

   - **The `-f` parameter is followed by the source image format.**
   - **The `-O` parameter (case-sensitive) is followed by the destination image format, source file name, and destination file name.**

   After the conversion is complete, the destination file appears in the directory of the source image file.

Linux

1. **Install the qemu-img tool.**
   - **For Ubuntu, run the `apt install qemu-img` command.**
   - **For CentOS, run the `yum install qemu-img` command.**
2. **Run the `qemu-img convert -f raw -O qcow2 centos.raw centos.qcow2` command to convert the image file format.**

   The parameters are described as follows:

   - **The `-f` parameter is followed by the source image format.**
   - **The `-O` parameter (case-sensitive) is followed by the destination image format, source file name, and destination file name.**

2.6.4.3 Import a custom image

After you upload a local image to an OSS bucket, you can import the image to the ECS environment as a custom image.

Prerequisites

- **An image that meets the limits and requirements for image import has been made. The image must be in the RAW, VHD, or qcow2 format. For more information, see** *Limits on importing custom images* **and** *Convert the image file format*.
- **You have been authorized to import images. For more information, see** *the RAM authorization section* **in ASCM Console User Guide.**

- **A local image has been uploaded to a bucket by using the OSS console or calling an OSS API operation. For more information, see** *the Upload objects section in OSS User Guide* **or** *the PutObject section* **in OSS Developer Guide.**

> **Note:**
> **Make sure that the bucket is in the same region as the custom image you want to create.**

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Snapshots and Images > Images.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click Import Image.
5. Configure the parameters of the image.

Parameter	Required	Description
Region	Yes	The region of the custom image to be imported.
Organization	Yes	The organization to which the custom image belongs.
Resource Set	Yes	The resource set of the custom image.
OSS Bucket Name	Yes	The name of the OSS bucket where the image to be imported resides.
OSS Object Name	Yes	The endpoint of the OSS object where the image is stored. For information about how to obtain the endpoint of an OSS object, see *the Obtain object URLs section* in OSS User Guide.
Image Name	Yes	The name of the custom image. The name must be 2 to 128 characters in length. It must start with a letter and can contain letters, periods (.), underscores (_), and hyphens (-).
Operating System	Yes	Linux and Windows are available.

Parameter	Required	Description
System Disk	Yes	The size of the system disk of the ECS instance. Unit: GiB.
System Architecture	Yes	x86_64 and i386 are available.
Platform	Yes	Linux: · CentOS · Ubuntu · SUSE · OpenSUSE · Debian · CoreOS · Aliyun · Others Linux · Customized Linux Windows: · Windows Server 2003 · Windows Server 2008 · Windows Server 2012
Image Format	Yes	The format of the custom image. RAW, VHD, and qcow2 are available.
Description	No	The description of the custom image.

6. Click OK.

Result

You can go to the Images page to view the progress of custom image creation. For more information, see *View images*. When 100% is displayed in the Progress column of the Images page, the image is created.

2.6.5 Export a custom image

You can export a created custom image to an OSS bucket and then download it to your local device.

Prerequisites

- You have activated OSS and created an OSS bucket. For more information, see *the Create buckets section* in OSS User Guide.

- You have been authorized to export images. For more information, see *the RAM authorization section* in ASCM Console User Guide.

Context

After a custom image is exported to an OSS bucket, you can download the image to your local device. For more information, see *the Obtain object URLs section* in OSS User Guide.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Snapshots and Images > Images.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the custom image and click Export Image in the Actions column.
5. For OssBucket, select a bucket. For OSS Prefix, enter a prefix as needed. Then click OK.

   The OSS Prefix field is optional. The OSS prefix must be 1 to 30 characters in length and can contain digits and letters.

2.6.6 Delete a custom image

You can delete a custom image that is no longer needed. However, public images cannot be deleted.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Snapshots and Images > Images.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Select one of the following methods to delete custom images.
   - To delete one image, find the image and click Delete Image in the Actions column.
   - To delete multiple images, select images and click Delete at the bottom of the image list.
5. Click OK.

2.7 Snapshots

2.7.1 Create a snapshot

You can manually create a snapshot for a disk to back up disk data.

Prerequisites

- **The instance is in the Running or Stopped state.**
- **The disk is in the Running state.**

Context

You can create up to 64 snapshots for each disk.

Snapshots can be used in the following scenarios:

- **Restore the source disk from its snapshot**
- **Create a custom image**

  For more information, see *Create a custom image from a snapshot*. **Data disk snapshots cannot be used to create custom images.**
- **Create a new data disk by using a data disk snapshot as a baseline**

  To create a data disk from a snapshot, select Yes for Use Snapshot and specify a snapshot. For more information, see *Create a disk*. **The disk size is determined by the size of the specified snapshot and cannot be changed. When you reset the data disk, the disk is restored to the status of the snapshot that is used to create the disk.**

The following considerations apply to snapshot creation:

- **When you create a complete snapshot of a disk for the first time, it requires an extended period of time. Then you can create an incremental snapshot of the disk and it only requires a short time. The duration depends on the volume of data that is changed since the last snapshot. The more data that is changed, the longer the duration.**
- **Avoid creating snapshots during peak hours.**

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Storage > Disks.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**

4. Find the disk and click Create Snapshot in the Actions column.
5. Enter a snapshot name and description, and then click OK.

> **Note:**
> The manual snapshot name cannot start with auto. auto is a reserved prefix for automatic snapshots.

You can go to the Snapshots page to view the progress of snapshot creation. For more information, see *View snapshots*. When 100% is displayed in the Progress column, the snapshot is created.

2.7.2 View snapshots

You can view the list of created snapshots.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Snapshots and Images > Snapshots.
3. In the top navigation bar, select an organization, a resource set, and a region. The created snapshots that match the specified criteria are displayed.
4. Select a filtering option from the drop-down list, enter the relevant information in the search bar, and click Search.

   You can select multiple filtering options to narrow down the search results.

Filtering option	Description
Snapshot Name	Enter a snapshot name to search for the snapshot.
Snapshot ID	Enter a snapshot ID to search for the snapshot.
Instance ID	Enter an instance ID to search for the snapshots related to the instance.
Disk ID	Enter a disk ID to search for the snapshots related to the disk.

Filtering option	Description
Snapshot Type	Select a snapshot type to search for the snapshots of that type. Options include: · All · User Snapshots: manual snapshots. · Automatic snapshots: automatic snapshots.
Creation Time	Enter a creation time to search for the snapshots that were created at that time.

2.7.3 Delete a snapshot

You can delete a snapshot that is no longer needed. After the snapshot is deleted, it cannot be recovered. You cannot delete system disk snapshots that have been used to create custom images.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Snapshots and Images > Snapshots.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Select one of the following methods to delete snapshots.
   - To delete one snapshot, find the snapshot and click Delete in the Actions column.
   - To delete multiple snapshots, select snapshots and click Delete at the bottom of the snapshot list.
5. Click OK.

2.8 Automatic snapshot policies

2.8.1 Create an automatic snapshot policy

Automatic snapshot policies can apply to both system disks and data disks and can be used to create periodical snapshots for the disks. Using automatic snapshot policies can improve data availability and operation error tolerance.

Context

Automatic snapshot policies can effectively eliminate the following risks associated with manual snapshot creation:

- When applications such as personal websites or databases deployed on an ECS instance encounter attacks or system vulnerabilities, you may not be able to manually create snapshots. In this case, you can use the latest automatic snapshot to roll back the affected disks to restore your data and reduce losses.
- You can also specify an automatic snapshot policy to create snapshots before regular system maintenance tasks. This eliminates the need to manually create snapshots and ensures that snapshots are always created before maintenance.

You can create up to 64 snapshots for each disk. If the maximum number of snapshots for a disk is reached when a new snapshot is being created, the system deletes the oldest automatic snapshot.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Snapshots and Images > Automatic Snapshot Policies.
3. Click Create Policy.
4. Configure the properties of the automatic snapshot policy.

Parameter	Required	Description
Region	Yes	The ID of the region to which the automatic snapshot policy applies.
Organization	Yes	The organization to which the automatic snapshot policy applies.
Policy Name	Yes	The name of the automatic snapshot policy. The name must be 2 to 128 characters in length and cannot start with a special character or digit. It can contain periods (.), underscores (_), hyphens (-), and colons (:).

Parameter	Required	Description
Creation Time	Yes	The time when a snapshot is automatically created. You can select any hour from 00:00 to 23:00. Note: The default time zone for the snapshot policy is UTC+8. You can change the time zone based on your business requirements. The creation of an automatic snapshot is canceled if the scheduled time for creating the snapshot is reached but the previous automatic snapshot is still being created. This may occur if the disk stores a large volume of data. For example, you can specify a policy for the system to create automatic snapshots at the following points in time: 00:00, 01:00, and 02:00. When the system starts creating a snapshot at 00:00, it takes 70 minutes for the system to complete the snapshot task. Therefore, the system does not create another snapshot at 01:00. Instead, after the system completes the snapshot task at 01:10, the system creates the next snapshot at 02:00.
Frequency	Yes	The day when a snapshot is created. You can select multiple values. The day ranges from Monday to Sunday.

Parameter	Required	Description
Retention Policy	No	The retention policy of the automatic snapshot. The default value of the retention time is 30 days. You can configure the following parameters: · Keep for: Specify the number of days during which the snapshots can be retained. Valid values: 1 to 65536. · Always keep the snapshots until the number of snapshots reaches the upper limit: Select this option to retain the snapshots until the maximum number of snapshots is reached.

5. Click OK.

What's next

After the automatic snapshot policy is created, you need to apply it to a disk to automatically create snapshots for the disk. For more information, see *Configure an automatic snapshot policy for multiple disks*.

2.8.2 View automatic snapshot policies

You can view the list of automatic snapshot policies.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Snapshots and Images > Automatic Snapshot Policies.
3. In the top navigation bar, select an organization, a resource set, and a region. The created automatic snapshot policies that match the specified criteria are displayed.
4. View the list of automatic snapshot policies.

2.8.3 Modify an automatic snapshot policy

You can modify the properties of an automatic snapshot policy, such as the name, creation time, frequency, and retention policy.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Snapshots and Images > Automatic Snapshot Policies.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the automatic snapshot policy and click Modify Policy in the Actions column.**
5. **Modify the properties of the policy.**

   Changes made to the retention duration do not take effect on existing snapshots, but take effect only on newly created snapshots.
6. **Click OK.**

2.8.4 Configure an automatic snapshot policy

After you apply an automatic snapshot policy to a disk, snapshots will be created automatically for the disk based on the policy settings. You can cancel an applied automatic snapshot policy at any time.

Context

We recommend that you configure the automatic snapshot policy to create automatic snapshots during off-peak hours. You can also manually create a snapshot for the disk that already has an automatic snapshot policy applied. When an automatic snapshot is being created, you must wait until the snapshot is complete before you can create a manual snapshot. The automatic snapshot is named in the following format: auto_yyyyMMdd_1, such as auto_20140418_1.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Storage > Disks.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the disk and click Configure Automatic Snapshot Policy in the Actions column.**

5. Select a procedure based on the operation you want to perform on the policy.
   - To apply an automatic snapshot policy, turn on Automatic Snapshot Policy, select a policy, and then click OK.
   - To cancel an automatic snapshot policy, turn off Automatic Snapshot Policy and click OK.

2.8.5 Configure an automatic snapshot policy for multiple disks

After you apply an automatic snapshot policy to a disk, snapshots will be created automatically for the disk based on the policy settings. You can cancel an applied automatic snapshot policy at any time.

Context

We recommend that you configure the automatic snapshot policy to create automatic snapshots during off-peak hours. You can also manually create a snapshot for the disk that already has an automatic snapshot policy applied. When an automatic snapshot is being created, you must wait until the snapshot is complete before you can create a manual snapshot. The automatic snapshot is named in the following format: auto_yyyyMMdd_1, such as auto_20140418_1.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Snapshots and Images > Automatic Snapshot Policies.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the automatic snapshot policy and click Apply Policy in the Actions column.
5. Select a tab based on the operation you want to perform on the disks.
   - To apply the automatic snapshot policy, click the Disks Without Policy Applied tab, select one or more disks, and click Apply Policy at the bottom of the disk list.
   - To cancel the automatic snapshot policy, click the Disks With Policy Applied tab, select one or more disks, and click Disable Automatic Snapshot Policy at the bottom of the disk list.

2.8.6 Delete an automatic snapshot policy

You can delete an automatic snapshot policy that is no longer needed. After you delete the automatic snapshot policy, the policy is automatically canceled for disks that have it applied.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Snapshots and Images > Automatic Snapshot Policies.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the automatic snapshot policy and click Delete Policy in the Actions column.**
5. **In the message that appears, click OK.**

2.9 Security groups

2.9.1 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.

Prerequisites

A Virtual Private Cloud (VPC) has been created. For more information, see *VPC User Guide* **.**

Context

Instances that belong to the same account and are in the same region and in the same security group can communicate with each other over the internal network. If instances that belong to the same account in the same region are in different security groups, you can implement internal network communication by authorizing mutual access between two security groups.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Networks and Security > Security Groups.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Click New Security Group.**

5. Configure the parameters of the security group.

Type	Parameter	Required	Description
Region	Organization	Yes	The organization to which the security group belongs. Make sure that the security group and the VPC belong to the same organization.
	Resource Set	Yes	The resource set to which the security group belongs. Make sure that the security group and the VPC belong to the same resource set.
	Region	Yes	The region to which the security group belongs. Make sure that the security group and the VPC belong to the same region.
	Zone	Yes	The ID of the zone where the security group resides.
Basic Settings	VPC	Yes	The VPC to which the security group belongs.

Type	Parameter	Required	Description
	Security Group Name	No	The name must be 2 to 128 characters in length and start with a letter. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,). It cannot start with http:// or https://.
	Description	No	The description of the security group. We recommend that you provide an informational description to simplify future management operations. The name must be 2 to 256 characters in length and start with a letter. It can contain letters, digits, periods (.), underscores (_), hyphens (-), and commas (,). It cannot start with http:// or https://.

6. Click Submit.

2.9.2 View security groups

You can view the list of security groups that you create.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Networks and Security > Security Groups.

3. In the top navigation bar, select an organization, a resource set, and a region. The created security groups that match the specified criteria are displayed.
4. Select a filtering option from the drop-down list, enter the relevant information in the search bar, and click Search.

   You can select multiple filtering options to narrow down the search results.

Filtering option	Description
Security Group ID	Enter a security group ID to search for the security group.
Security Group Name	Enter a security group name to search for the security group.
VPC ID	Enter a VPC ID to search for the security groups that belong to the VPC.

2.9.3 Modify a security group

You can modify the name and description of a created security group.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Networks and Security > Security Groups.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the security group and click Modify in the Actions column.
5. Modify the name and description of the security group.
6. Click OK.

2.9.4 Add a security group rule

You can use security group rules to control access to the ECS instances in a security group over the Internet and the internal network.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Networks and Security > Security Groups.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the security group and click Rules in the Actions column.

5. Click Create Rule.
6. Configure the parameters of the security group rule.

Parameter	Required	Description
ENI Type	Yes	Internal Network ENI is available. Public NICs are not available for ECS instances in a VPC. You can add security group rules for an internal network ENI. However, the security group rules apply to both the internal network and the Internet.
Direction	Yes	· Outbound: Your ECS instances access other ECS instances in the internal network or resources in the Internet. · Inbound: Other ECS instances in the internal network or resources in the Internet access your ECS instances.
Action	Yes	· Allow: Access requests on the specified port or ports are allowed. · Deny: Data packets are discarded with no messages returned. If two security group rules are only different in the Action parameter, the Deny policy takes precedence over the Allow policy.
Protocol	Yes	· all: It is used in total trust scenarios. · tcp: It can be used to allow or deny one or several successive ports. · udp: It can be used to allow or deny one or several successive ports. · icmp: It is used in scenarios where the `ping` command is used to test the network connection status between instances. · icmpv6: It is used in scenarios where the `ping6` command is used to test the network connection status between instances. · gre: It is used for VPN.

Parameter	Required	Description
Port Range	Yes	The port range depends on the protocol type. · For the all protocol type: -1/-1 is displayed, indicating all ports. You cannot set a port range for this protocol type. · For the tcp protocol type: Set a port range. Valid values: 1 to 65535. You must use the <start port>/<end port> format to specify the range of ports. Set the start port and end port to the same number to specify a single port. For example, use 22/22 to indicate port 22. · For the udp protocol type: Set a port range. Valid values: 1 to 65535. You must use the <start port>/<end port> format to specify the range of ports. Set the start port and end port to the same number to specify a single port. For example, use 3389/3389 to indicate port 3389. · For the icmp protocol type: -1/-1 is displayed, indicating all ports. You cannot set a port range for this protocol type. · For the icmpv6 protocol type: -1/-1 is displayed, indicating all ports. You cannot set a port range for this protocol type. · For the gre protocol type: -1/-1 is displayed, indicating all ports. You cannot set a port range for this protocol type.
Priority	Yes	The priority of the rule. Valid values: 1 to 100. The default value is 1, indicating the highest priority.
Authorization Type	Yes	· IPv4 Addresses: Authorizes IPv4 addresses or an IPv4 CIDR block to access the ECS instances of the security group. · IPv6 Addresses: Authorizes IPv6 addresses or an IPv6 CIDR block to access the ECS instances of the security group. · Security Groups: Authorizes instances in other security groups to access the ECS instances of the security group. This authorization type only takes effect for the internal network.

Parameter	Required	Description
Authorization Object	Yes	Authorization objects depend on the authorization type. For the IPv4 Addresses authorization type: · Enter an IPv4 address or IPv4 CIDR block, in the format of `12.1.1.1` or `13.1.1.1/25`. · You can enter up to 10 authorization objects at a time. Separate multiple objects with commas (,). · Specifying `0.0.0.0/0` will allow or deny all IP addresses based on the Action parameter. Exercise caution when you specify 0.0.0.0/0. For the IPv6 Addresses authorization type: · Enter an IPv6 address or IPv6 CIDR block, in the format of `2001:0db8::1428:****` or `2001:0db8::1428:****/128`. · You can enter up to 10 authorization objects at a time. Separate multiple objects with commas (,). · Specifying `::/0` will allow or deny all IP addresses based on the Action parameter. Exercise caution when you specify ::/0. For the Security Groups authorization type: Select a security group ID. For a security group of the VPC type, the security group to be authorized must be in the same VPC.
Expiration Time	Yes	The time when the security group rule expires. You can select a date and time, or set the time to Never.
Description	No	The description of the security group rule. We recommend that you provide an informational description to simplify future management operations. The description must be 2 to 256 characters in length and cannot start with http:// or https://.

7. Click OK.

2.9.5 Add an instance

You can add an existing instance to a security group in the same region. After the instance is added, the security group rules of the security group automatically apply to the instance.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Networks and Security > Security Groups.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the security group and click Manage Instances in the Actions column.**
5. **Click Add Instance.**
6. **Select an instance and click OK.**

2.9.6 Remove instances from a security group

You can remove instances from their security groups to which they belong.

Prerequisites

An instance has been added to two or more security groups.

Context

After an ECS instance is removed, the instance will be isolated from other ECS instances in the security group. We recommend that you perform a full test in advance to ensure that the business can run properly after you remove the instance.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Networks and Security > Security Groups.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the security group and click Manage Instances in the Actions column.**
5. **Select one or more instances and click Remove at the bottom of the instance list.**
6. **Click OK.**

2.9.7 Delete a security group

You can delete a security group that you no longer use.

Prerequisites

All instances have been removed from the security group.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Networks and Security > Security Groups.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Select one of the following methods to delete security groups.**
   - **To delete one security group, find the security group and click Delete in the Actions column.**
   - **To delete multiple security groups, select security groups and click Delete at the bottom of the security group list.**
5. **Click OK.**

2.10 Elastic Network Interfaces

2.10.1 Create an ENI

You can bind an Elastic Network Interface (ENI) to multiple instances to create high-availability clusters and implement fine-grained network management. You can also unbind an ENI from an instance and bind it to another instance to implement a low-cost failover solution.

Prerequisites

- **A VPC and a VSwitch have been created. For more information, see** *the Create a VPC and Create a VSwitch sections in* **VPC User Guide.**
- **A security group is available in the VPC. Otherwise, you must create a security group. For more information, see** *Create a security group*.

Context

ENIs are divided into primary ENIs and secondary ENIs.

A primary ENI is created by default when an instance in a VPC is created. The primary ENI has the same lifecycle as the instance. It cannot be unbound from the instance.

ENIs that are created separately are secondary ENIs. You can bind and unbind a secondary ENI to and from an instance. This topic describes how to create a secondary ENI.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Networks and Security > ENIs.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click Create ENI.
5. Configure the parameters of the ENI.

Category	Parameter	Required	Description
Region	Organization	Yes	The organization to which the ENI belongs.
	Resource Set	Yes	The resource set to which the ENI belongs.
	Region	Yes	The region to which the ENI belongs.
	Zone	Yes	The zone where the ENI resides.

Category	Parameter	Required	Description
Basic Settings	VPC	Yes	The VPC where the ENI resides. The secondary ENI can only be bound to an instance in the same VPC as the ENI. Note: After the ENI is created, you cannot change its VPC.
	VSwitch	Yes	The VSwitch to which the ENI belongs. The secondary ENI can only be bound to an instance in the same VPC. Select a VSwitch that is in the same zone as the instance to which the ENI is bound. However, the ENI and the instance can belong to different VSwitches. Note: After the ENI is created, you cannot change its VSwitch.

Category	Parameter	Required	Description
	Security Group	Yes	The security group of the ENI in the VPC. The rules of the security group automatically apply to the ENI.
	ENI Name	Yes	The name must be 2 to 128 characters in length. It must start with a letter but cannot start with http:// or https://. It can contain digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,).

Category	Parameter	Required	Description
	Description	No	The description of the ENI. We recommend that you provide an informational description to simplify future management operations. The description must be 2 to 256 characters in length. It must start with a letter but cannot start with http:// or https://. It can contain digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,).
	Primary Private IP	No	The primary private IPv4 address of the ENI. The IPv4 address must be within the CIDR block of the specified VSwitch. If the address is not specified, the system will assign a private IP address to the ENI after it is created.

6. Click Submit.

Result

The created ENI is displayed in the ENI list and is in the Available state.

2.10.2 View ENIs

You can view the list of created ENIs.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Networks and Security > ENIs.
3. In the top navigation bar, select an organization, a resource set, and a region. The created ENIs that match the specified criteria are displayed.
4. Select a filtering option from the drop-down list, enter the relevant information in the search bar, and click Search.

   You can select multiple filtering options to narrow down the search results.

Filtering option	Description
ENI Name	Enter an ENI name to search for the ENI.
ENI ID	Enter an ENI ID to search for the ENI.
VSwitch ID	Enter a VSwitch ID to search for the ENIs that are associated with the VSwitch.
Security Group ID	Enter a security group ID to search for the ENIs that belong to the security group.
Instance ID	Enter an instance ID to search for the ENIs that are bound to the instance.

2.10.3 Modify the properties of a secondary ENI

You can modify the properties of a secondary ENI, including the name, security group, and description.

Prerequisites

The secondary ENI is in the Available state.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Networks and Security > ENIs.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the secondary ENI and click Modify in the Actions column.

5. Modify the name, security group, and description of the ENI.
6. Click OK.

2.10.4 Bind a secondary ENI to an instance

You can bind a secondary ENI to an instance so that the instance can process the traffic on this ENI.

Prerequisites

- The secondary ENI is in the Available state.
- The instance is in the Running or Stopped state.
- The instance and the secondary ENI belong to the same VPC.
- The VSwitch to which the secondary ENI belongs must be in the same zone as the VSwitch to which the instance belongs. The VSwitches of the ENI and the instance can be different, but they must be in the same zone.

Context

The following limits apply when you bind an ENI to an instance:

- You can manually bind only secondary ENIs. The primary ENI shares the same lifecycle as the instance and cannot be bound.
- An ENI can only be bound to one ECS instance at a time. However, an ECS instance can be bound with multiple ENIs. The maximum number of ENIs that can be bound to an instance depends on the instance type. For more information about the number of ENIs that can be bound to each instance type, see *the Instance families section* in ECS Product Introduction.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, choose Networks and Security > ENIs.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the secondary ENI and click Bind in the Actions column.
5. Select the instance and click OK.

Result

In the Status column, the status of the secondary ENI changes to Bound.

2.10.5 Unbind a secondary ENI from an instance

You can unbind a secondary ENI from an instance. After the ENI is unbound, the instance will not process the traffic on this ENI.

Prerequisites

- **The secondary ENI must be in the Bound state.**
- **The instance is in the Running or Stopped state.**

Context

Only secondary ENIs can be unbound. The primary ENI shares the same lifecycle as the instance and cannot be unbound.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Networks and Security > ENIs.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the secondary ENI and click Unbind in the Actions column.**
5. **Click OK.**

Result

In the Status column, the status of the secondary ENI changes to Available.

2.10.6 Delete a secondary ENI

You can delete a secondary ENI that is no longer needed.

Prerequisites

The ENI must be in the Available state.

Context

You can delete only secondary ENIs. The primary ENI shares the same lifecycle as the instance and cannot be deleted.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, choose Networks and Security > ENIs.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the secondary ENI and click Delete in the Actions column.**
5. **Click OK.**

2.11 Deployment sets

2.11.1 Create a deployment set

You can use a deployment set to distribute or aggregate instances involved in your business. You can select hosts, racks, or network switches to improve service availability or network performance based on your needs.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Deployment Sets.
3. Click Create Deployment Set.
4. Configure the parameters of the deployment set.

Category	Parameter	Required	Description
Region	Organization	Yes	The organization to which the deployment set belongs.
	Resource Set	Yes	The resource set to which the deployment set belongs.
	Region	Yes	The region where the deployment set is located.
	Zone	Yes	The zone where the deployment set is located.
Basic Settings	Deployment Domain	Yes	This parameter setting determines the Deployment Target options. Valid values: · Default: When Default is selected, the deployment target options are Host, Rack, and Network Switch. · Switch: When Switch is selected, the deployment target options are Host and Rack.

Category	Parameter	Required	Description
	Deployment Target	Yes	The basic unit that can be scheduled when you deploy instances. · Host: Instances are distributed or aggregated at the host level. · Rack: Instances are distributed or aggregated at the rack level. · VSwitch: Instances are distributed or aggregated at the VSwitch level.
	Deployment Policy	Yes	The dispersion policies are used to improve service availability to avoid business impact when a host, rack, or switch fails. The aggregation policies are used to improve network performance to minimize the access latency between instances. Options are: · Loose Dispersion · Strict Dispersion · Loose Aggregation · Strict Aggregation
	Deployment Set Name	No	The name of the deployment set. The name must be 2 to 128 characters in length. It must start with a letter but cannot start with http:// or https ://. It can contain digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,).
	Description	No	The description of the deployment set. We recommend that you provide an informational description to simplify future management operations. The description must be 2 to 256 characters in length. It must start with a letter but cannot start with http:// or https://. It can contain digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,).

5. Click Submit.

2.11.2 View deployment sets

You can view the list of deployment sets that you create.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, click Deployment Sets.**
3. **In the top navigation bar, select an organization, a resource set, and a region. The created deployment sets that match the specified criteria are displayed.**
4. **Select a filtering option from the drop-down list, enter the relevant information in the search bar, and then click Search.**

   You can select multiple filtering options to narrow down the search results.

Filtering option	Description
Deployment Set Name	Enter a deployment set name to search for the deployment set.
Deployment Set ID	Enter a deployment set ID to search for the deployment set.
Resource Set	Enter a resource set name to search for the deployment sets that belong to the resource set.

2.11.3 Modify a deployment set

You can modify the name and description of a deployment set.

Procedure

1. *Log on to the ECS console*.
2. **In the left-side navigation pane, click Deployment Sets.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the deployment set and click Modify in the Actions column.**
5. **Modify the name and description of the deployment set.**
6. **Click OK.**

2.11.4 Delete a deployment set

You can delete a deployment set that is no longer needed.

Prerequisites

ECS instances must have been removed from the deployment set.

Procedure

1. *Log on to the ECS console*.
2. In the left-side navigation pane, click Deployment Sets.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the deployment set and click Delete in the Actions column.
5. Click OK.

2.12 Install FTP software

2.12.1 Overview

File Transfer Protocol (FTP) transfers files between a client and a server by establishing two TCP connections. One is the command link for transferring commands between a client and a server. The other is the data link used to upload or download data. Before uploading files to an instance, you must build an FTP site for the instance.

2.12.2 Install and configure vsftp in CentOS

This topic describes how to install and configure vsftp in CentOS to transfer files.

Procedure

1. Install vsftp.

```
yum install vsftpd -y
```

2. Add an FTP account and a directory.

   a) Check the location of the *nologin* file,

   which is usually under the */usr/sbin* or */sbin* directory.

   b) Create an FTP account.

   Run the following commands to create the */alidata/www/wwwroot* directory and specify this directory as the home directory of the account pwftp. You can also customize the account name and directory.

```
mkdir -p /alidata/www/wwwroot
```

```
useradd -d /alidata/www/wwwroot -s /sbin/nologin pwftp
```

c) **Modify the account password.**

```
passwd pwftp
```

d) **Modify the permissions on the specified directory.**

```
chown -R pwftp.pwftp /alidata/www/wwwroot
```

3. **Configure vsftp.**

a) **Open the vsftp configuration file.**

```
vi /etc/vsftpd/vsftpd.conf
```

b) **Change the value of anonymous_enable from `YES` to `NO`.**

c) **Delete the comment delimiter (`#`) from the following configuration lines:**

```
local_enable=YES
        write_enable=YES
        chroot_local_user=YES
```

d) **Press the Esc key to exit the edit mode, and enter `:wq` to save the modifications and exit.**

4. **Modify the shell configuration.**

a) **Open the shell configuration file.**

```
vi /etc/shells
```

b) **If the file does not contain *`/usr/sbin/nologin`* or *`/sbin/nologin`*, add it to the file.**

5. **Start vsftp and perform a logon test.**

a) **Start vsftp.**

```
service vsftpd start
```

b) **Use the account pwftp to perform an FTP logon test.**

This example uses the directory *`/alidata/www/wwwroot`*.

2.12.3 Install vsftp in Ubuntu or Debian

This topic describes how to install and configure vsftp in an instance running Ubuntu or Debian to transfer files.

Procedure

1. **Update the software source.**

```
apt-get update
```

2. **Install vsftp.**

```
apt-get install vsftpd -y
```

3. **Add an FTP account and a directory.**

   a) **Check the location of the** *nologin* **file,**

   which is typically under the */usr/sbin***or** */sbin* **directory.**

   b) **Create an FTP account.**

   Run the following commands to create the */alidata/www/wwwroot* **directory and specify this directory as the home directory of the account pwftp. You can also customize the account name and directory.**

```
mkdir -p /alidata/www/wwwroot
useradd -d /alidata/www/wwwroot -s /sbin/nologin pwftp
```

   c) **Modify the account password.**

```
passwd pwftp
```

   d) **Modify the permissions on the specified directory.**

```
chown -R pwftp.pwftp /alidata/www/wwwroot
```

4. **Configure vsftp.**

   a) **Open the vsftp configuration file.**

```
vi /etc/vsftpd.conf
```

   b) **Change the value of anonymous_enable from** `YES` **to** `NO`**.**

   c) **Delete the comment delimiter (**`#`**) from the following configuration lines:**

```
local_enable=YES
        write_enable=YES
        chroot_local_user=YES
        chroot_list_enable=YES
        chroot_list_file=/etc/vsftpd.chroot_list
```

   d) **Press the Esc key to exit the edit mode, and enter** `:wq` **to save the modifications and exit.**

   e) **Open the** */etc/vsftpd.chroot_list* **file and add the FTP account name to the file. Save the modifications and exit.**

   You can follow steps a to d to open and save the file.

5. **Modify shell configurations.**
   a) **Open the shell configuration file.**

   ```
   vi /etc/shells
   ```

   b) **If the file does not contain** *`/usr/sbin/nologin`* **or** *`/sbin/nologin`*, **add it to the file.**
6. **Start vsftp and perform a logon test.**
   a) **Start vsftp.**

   ```
   service vsftpd restart
   ```

   b) **Use the account pwftp to perform an FTP logon test.**

   This example uses the directory *`/alidata/www/wwwroot`*.

2.12.4 Build an FTP site in Windows Server 2008

This topic describes how to build an FTP site on an instance running Windows Server 2008.

Prerequisites

You have added the Web Server (IIS) role and installed FTP on an instance.

Procedure

1. *Connect to an instance*.
2. **Choose Start > Adminstrative Tools > Internet Information Services (IIS) Manager.**
3. **Right-click the server name and select Add FTP Site from the shortcut menu.**
4. **Enter an FTP site name and a physical path, and then click Next.**
5. **Set IP Address to All Unassigned and SSL to No SSL, and then click Next.**
6. **Set Authentication to Basic, Authorization to All Users, and Permissions to Read and Write, and click Finish.**

Result

Then you can use the administrator account and password to upload and download files through FTP. Make sure that the following conditions are met:

- **The port for the FTP site is not in use by other applications, and Windows firewall is not blocking the port.**
- **The security group of the instance contains a security group rule that allows inbound access to the FTP port.**

2.12.5 Build an FTP site in Windows Server 2012

This topic describes how to build an FTP site on an instance running Windows Server 2012.

Prerequisites

You have added the Web Server (IIS) role and installed FTP on an instance.

Procedure

1. *Connect to an instance*.
2. **Click the Server Manager icon.**
3. **In the left-side navigation pane, click IIS.**
4. **In the Server area, right-click the server name and select Internet Information Services (IIS) Manager from the shortcut menu.**
5. **Right-click the server name and select Add FTP Site from the shortcut menu.**
6. **Enter an FTP site name and a physical path, and then click Next.**
7. **Set IP Address to All Unassigned and SSL to No SSL, and then click Next.**
8. **Set Authentication to Basic, Authorization to All Users, and Permissions to Read and Write, and click Finish.**

Result

Then you can use the administrator account and password to upload and download files through FTP. Make sure that the following conditions are met:

- **The port for the FTP site is not in use by other applications, and Windows firewall is not blocking the port.**
- **The security group of the instance contains a security group rule that allows inbound access to the FTP port.**

3 Auto Scaling (ESS)

3.1 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.

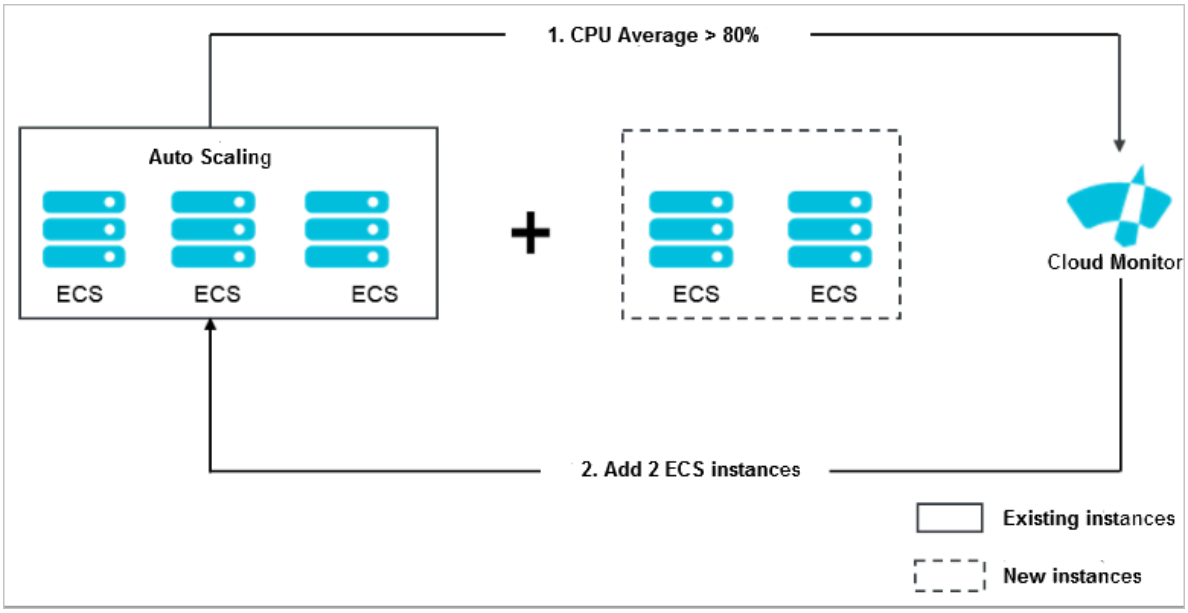
Based on user-defined scaling rules, ESS automatically adds ECS instances as business loads increase to ensure sufficient computing capabilities. When your business loads decrease, ESS automatically removes ECS instances to reduce running costs.

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. *Figure 3-1: Elastic scale-out* **shows the process.**

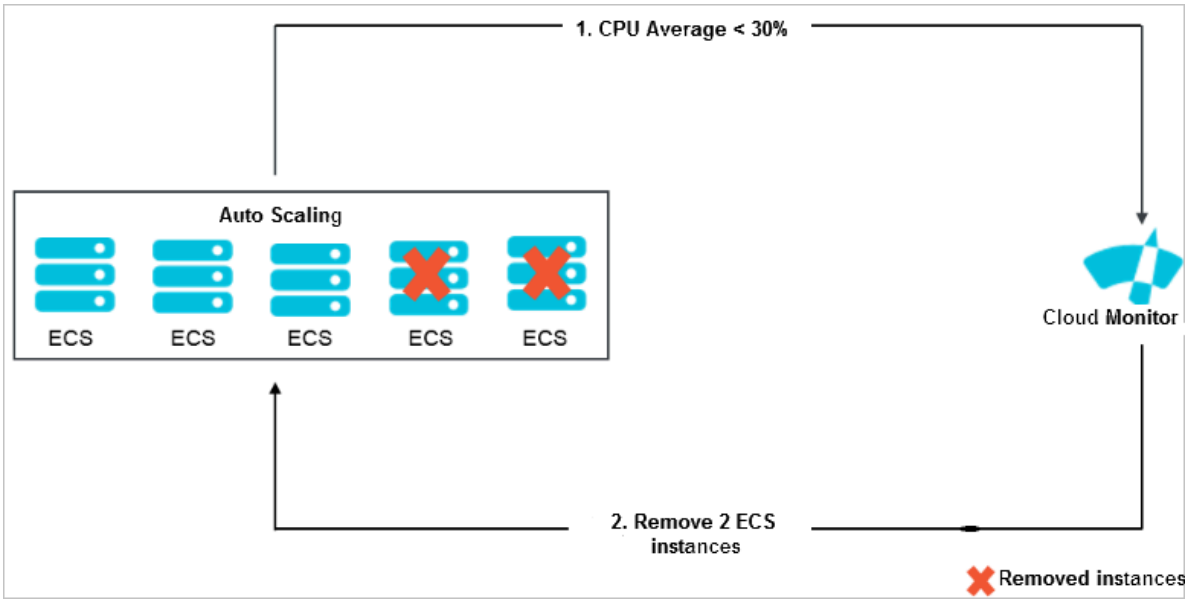
  Figure 3-1: Elastic scale-out



- **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. *Figure 3-2: Elastic scale-in* **shows the process.**

  Figure 3-2: Elastic scale-in



- **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.

3.2 Notes

3.2.1 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.**

3.2.2 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.

Resource	Manual intervention	Processing method
SLB	An SLB instance is unavailable because of system-related reasons.	All scaling activities fail except for instance removal tasks that are manually executed.
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.

3.2.3 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 3-1: Scaling group statuses* **describes the details.**

Table 3-1: Scaling group statuses

Status	Status in open API
Creating	**Inactive**
Created	**Inactive**
Enabling	**Inactive**
Running	**Active**
Disabling	**Inactive**
Stopped	**Inactive**
Deleting	**Deleting**

3.2.4 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.

3.2.5 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.

3.2.6 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.

3.2.7 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.**

3.3 Quick start

3.3.1 Overview

This topic describes how to create a scaling group and how to add or remove ECS instances.

You can perform the following steps to create a scaling group, and add or remove ECS instances.

1. *Create a scaling group*

   Set the parameters for the scaling group, such as the Maximum Capacity and Minimum Capacity of ECS instances.

2. *Create a scaling configuration*

   Set the parameters for the scaling configuration, such as Instance Type and Image.

3. *Enable a scaling group*

   Enable the scaling group after creating the scaling configuration.

4. *Create a scaling rule*

   Specify how to add or remove ECS instances. For example, add an ECS instance to a scaling group.

5. *Create a scheduled task*

   Create scheduled tasks to add or remove instances at a specified time point. Auto Scaling executes the scheduled tasks and scaling rules at the specified time. For example, Auto Scaling can trigger a task to execute a specific scaling rule at 08:00 everyday.

3.3.2 Log on to the Auto Scaling console

This topic describes how to log on to the Auto Scaling console.

Prerequisites

- **Before logging on to the ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel . The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.**
- **We recommend that you use the Google Chrome browser.**

Procedure

1. **In the address bar, enter the URL used to access the ASCM console. Press Enter.**
2. **Enter your username and password.**

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

   > **Note:**
   > **When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).**

3. **Click Login to go to the ASCM console homepage.**
4. **In the top navigation bar, choose Products > Elastic Computing > Auto Scaling.**

3.3.3 Create a scaling group

This topic describes how to create a scaling group. A scaling group is a group of ECS instances that is dynamically scaled based on the configured scenario. You can specify the minimum and maximum numbers of ECS instances in a scaling group.

Prerequisites

You must create a VPC and a VSwitch before you create a scaling group. For more information, see the Create a VPC and a VSwitch topic in *VPC User Guide* **.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Click Create Scaling Group.**

4. Set the parameters for the scaling group.

Parameter	Required	Description
Scaling Group	Yes	The name of the scaling group. It must be 2 to 40 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or digit.
Organization/ Resource Group	Yes	The organization and resource set to which the scaling group belongs.
Maximum Capacity	Yes	The maximum number of instances that a scaling group must contain. This helps control costs within an expected amount. Valid values: 0 to 1000.
Minimum Capacity	Yes	The minimum number of instances that a scaling group must contain. This helps guarantee the availability of your services. When the scaling group is enabled, Auto Scaling automatically creates this number of ECS instances. Valid values: 0 to 1000.
Cooldown	Yes	The cooldown time refers to the period during which Auto Scaling cannot execute any new scaling activities. This occurs after the scaling group executes a successful scaling activity. During the cooldown time, the scaling group rejects all scaling activity requests triggered by alarm tasks from CloudMonitor. However, scaling activities triggered by other types of tasks (manually triggered tasks and scheduled tasks) are not limited by the cooldown time. These tasks are executed immediately. The value must be an integer that is greater than or equal to zero. Unit: seconds.

Parameter	Required	Description
Scale-In Policy	Yes	The policy to remove ECS instances from a scaling group. This parameter contains two fields. Valid values for the first field: · Oldest Instance · Newest Instances · Instance with Oldest Scaling Configuration Valid values for the second field: · None · Oldest Instance · Newest Instances For example, you can select Instance with Oldest Scaling Configuration for the first field and select Oldest Instance for the second field. This means to filter ECS instances that are created based on the oldest scaling configuration and then select the oldest instance from the found ECS instances.
VPC	Yes	The ID of the VPC to which the scaling group belongs.
VSwitch	Yes	The ID of the VSwitch to which the scaling group belongs.

5. Click OK.

Result

The specified scaling group is displayed on your scaling group list and is in the Disabled state. To enable the scaling group, you must create a scaling configuration. For more information, see *Create a scaling configuration*.

3.3.4 Create a scaling configuration

This topic describes how to create a scaling configuration for a scaling group.

Prerequisites

Make sure that at least one security group is available. If you do not have any security groups, you must create a security group. For more information, see the Create a security group topic in *ECS User Guide* **.**

Context

You can create only a limited number of scaling configurations for a scaling group. For more information, see the Limits topic in *Auto Scaling Product Introduction* **.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.**
4. **In the left-side navigation pane, click Scaling Configuration.**
5. **Choose Create > Create Scaling Configuration.**
6. **Set the parameters for the scaling configuration.**

Category	Parameter	Required	Description
Region	Region	Yes	The region where the ECS instance is located.
	Zone	Yes	The zone where the ECS instance is located.
Security Group	Security Group	Yes	The security group to which the ECS instance belongs.
Instance	Instance Family	Yes	The instance family to which the ECS instance belongs.
	Instance Type	Yes	The instance type of the ECS instance.

Category	Parameter	Required	Description
Image	Image Type	Yes	· Public Image: Public images provided by Alibaba Cloud are fully licensed to offer a secure and stable operating environment for applications on ECS instances. · Custom Image: You can create custom images to install software or deploy projects that have special requirements.
Storage	System Disk	Yes	Specify the category and size of the system disk. The operating system is installed on the system disk. You can select Ultra Disk or SSD Disk.
	Data Disk	No	Specify the category and size of the system disk. You can select Ultra Disk or SSD Disk. You can add a maximum of 16 data disks. The maximum capacity of each data disk is 32 TiB. You can set Release with Instance and Encrypt for each data disk.

Category	Parameter	Required	Description
Password	Set Password	Yes	Select when to set password. You can select Now or Later. If you select Later, you can use the Change Password feature in the console to set the password. For more information, see the Change Password topic in *ECS User Guide* .
	Logon Password	No	The password used to log on to the ECS instance. The password must be 8 to 30 characters in length and must contain at least three of the following character types: digits, uppercase letter, lowercase letters, and special characters. **Note:** The password is used to log on to the operating system and is not the VNC password.
	Confirm Password	No	Enter the password again.
Deployment Set	Deployment Set	No	The deployment set to which the instance belongs.
Instance Name	Configuration Name	No	The name of the scaling configuration.
	Instance Name	No	The name of the ECS instance.

Category	Parameter	Required	Description
User Data	User Data	No	Windows supports two formats: Bat and Powershell. When you perform Base64 encoding, make sure to include `[bat]` or `[powershell]` as the first line. You can run Shell scripts for Linux-based ECS instances.
Quantity	Quantity	No	The number of instances to purchase.

7. Click Submit.

Result

After the scaling configuration is created, it is in the Disabled state and is displayed on your scaling configuration list. To automatically create ECS instances, you must apply a scaling configuration. For more information, see *Apply a scaling configuration*.

3.3.5 Enable a scaling group

This topic describes how to enable a scaling group. You can enable a scaling group to trigger scaling activities.

Prerequisites

- The scaling group is in the Disabled state.
- The scaling group has scaling configurations that are in the Enable state.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Find the target scaling group and click Enable in the Actions column.
4. Click OK.

Result

In the Status column, the state of the scaling group is changed from Disabled to Enable.

3.3.6 Create a scaling rule

This topic describes how to create a scaling rule. You can scale in or out ECS instances by creating scaling rules. For example, you can add an ECS instance to a scaling group.

Context

- **You can create only a limited number of scaling rules for a scaling group. For more information, see the Limits topic in** *Auto Scaling Product Introduction* **.**
- **If a scaling rule is executed and the resulting number of ECS instances in the scaling group is less than the minimum number or greater than the maximum number, Auto Scaling will automatically adjust the number of ECS instances to within the valid range.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.**
4. **In the left-side navigation pane, click Scaling Rules.**
5. **Click Create Scaling Rule.**
6. **Set the parameters for the scaling rule.**

Parameter	Required	Description
Rule Name	Yes	The name of the scaling rule. It must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or digit.
Rule Type	Yes	The type of the scaling rule. You can select Simple Scaling Rule for this parameter. Note that this parameter cannot be modified after the scaling rule is created.

Parameter	Required	Description
Scaling Activity	Yes	The operation that is executed when the scaling rule is triggered. The operations include: · Change to N instances: When the scaling rule is executed, the number of instances in the scaling group changes to N. · Add N instances: When the scaling rule is executed, N instances are added to the scaling group. · Remove N instances: When the scaling rule is executed, N instances are removed from the scaling group.
Default Cooldown (Seconds)	No	The cooldown period. If this parameter is not specified, the default value is used.

7. Click OK.

3.3.7 Create a scheduled task

This topic describes how to create a scheduled task to scale computing resources in response to predictable business changes in the future. Scheduled tasks enable the system to automatically obtain sufficient computing resources before business peaks and release idle computing resources after the peaks.

Context

A scheduled task is preconfigured to execute a specified scaling rule at a specified time in the future. When the specified time arrives, the scheduled task scales computing resources to meet the business requirements. This helps you optimize cost by paying only for additional resources when you need them. You can also specify the recurrence for scheduled tasks if business changes are regular.

If multiple scheduled tasks need to be executed in one minute, Auto Scaling executes the newest scheduled task.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the left-side navigation pane, choose Scaling Tasks > Scheduled Tasks.

3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click Create Scheduled Task.
5. In the dialog box that appears, set the parameters for the scheduled task.

Parameter	Required	Description
Task Name	Yes	The name of the scheduled task. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or digit.
Description	Yes	The description of the scheduled task.
Organization/ Resource Group	Yes	The organization and resource set to which the scheduled task belongs.
Start Time	Yes	The time when the scheduled task is executed.
Scaling Rules	Yes	The scaling group that is monitored and the scaling rule that is executed.
Retry Expiry Time	No	The expiration time for the scheduled task retries. Unit: seconds. If a scaling activity fails to execute at a specified time, Auto Scaling will execute the scheduled task again within the retry expiry time.
Recurrence Settings	No	Specifies whether to execute the scheduled task repeatedly. Select Recurrence Settings and set the Recurrence and Expire parameters. The recurrence options include Daily, Weekly, and Monthly.

6. Click OK.

Result

The scheduled task that you created is displayed on your scheduled task list.

3.3.8 Create an event-triggered task

This topic describes how to create an event-triggered task in response to emergent or unpredictable business changes. After you create and enable an event-triggered task, Auto Scaling collects data for the specified metric in real time. When an alert is generated, it triggers a scaling rule. Auto Scaling triggers the scaling rule to dynamically scale ECS instances in the scaling group.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the left-side navigation pane, choose Scaling Tasks > Event-triggered Tasks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click Create Alert.
5. In the dialog box that appears, set the parameters for the event-triggered task.

Parameter	Required	Description
Task Name	Yes	The name of the event-triggered task. It must be 2 to 40 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or digit.
Description	No	The description of the event-triggered task.
Organization /Resource Group	Yes	The organization and resource set to which the event-triggered task belongs.
Monitoring Metrics/ Scaling Rules	Yes	Select the scaling group that is monitored and the scaling rule that is executed.
Monitoring Type	Yes	It supports System-Level Monitoring.
Monitoring Metrics	Yes	Select the metric that you want to monitor. Valid values: · Average CPU Utilization · Memory Usage · System Average Load · Outbound Traffic · Inbound Traffic
Monitoring Period	Yes	Specifies the time during which data is aggregated and analyzed. The shorter the monitoring period, the higher the frequency that the alert is triggered. Unit: minutes. Valid values: · 1 · 2 · 5 · 15

Parameter	Required	Description
Statistic	Yes	The rule that triggers the alert. Select Average, Max Capacity, or Min Capacity, and specify a threshold value. For example, alerts are triggered when the CPU utilization exceed 80%: · Average: Alerts are triggered when the average CPU utilization of all the ECS instances in the scaling group exceeds 80%. · Max Capacity: Alerts are triggered when the highest CPU utilization among the ECS instances in the scaling group exceed 80%. · Min Capacity: Alerts will be triggered when the lowest CPU utilization among the ECS instances in the scaling group exceeds 80%.
Trigger after occurrences	Yes	The consecutive number of times that the threshold must be exceeded before the alert is triggered. Valid values: · 1 · 2 · 3 · 5

6. Click OK.

3.4 Scaling group

3.4.1 Create a scaling group

This topic describes how to create a scaling group. A scaling group is a group of ECS instances that is dynamically scaled based on the configured scenario. You can specify the minimum and maximum numbers of ECS instances in a scaling group.

Prerequisites

You must create a VPC and a VSwitch before you create a scaling group. For more information, see the Create a VPC and a VSwitch topic in *VPC User Guide* .

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Click Create Scaling Group.

4. Set the parameters for the scaling group.

Parameter	Required	Description
Scaling Group	Yes	The name of the scaling group. It must be 2 to 40 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or digit.
Organization/ Resource Group	Yes	The organization and resource set to which the scaling group belongs.
Maximum Capacity	Yes	The maximum number of instances that a scaling group must contain. This helps control costs within an expected amount. Valid values: 0 to 1000.
Minimum Capacity	Yes	The minimum number of instances that a scaling group must contain. This helps guarantee the availability of your services. When the scaling group is enabled, Auto Scaling automatically creates this number of ECS instances. Valid values: 0 to 1000.
Cooldown	Yes	The cooldown time refers to the period during which Auto Scaling cannot execute any new scaling activities. This occurs after the scaling group executes a successful scaling activity. During the cooldown time, the scaling group rejects all scaling activity requests triggered by alarm tasks from CloudMonitor. However, scaling activities triggered by other types of tasks (manually triggered tasks and scheduled tasks) are not limited by the cooldown time. These tasks are executed immediately. The value must be an integer that is greater than or equal to zero. Unit: seconds.

Parameter	Required	Description
Scale-In Policy	Yes	The policy to remove ECS instances from a scaling group. This parameter contains two fields. Valid values for the first field: · Oldest Instance · Newest Instances · Instance with Oldest Scaling Configuration Valid values for the second field: · None · Oldest Instance · Newest Instances For example, you can select Instance with Oldest Scaling Configuration for the first field and select Oldest Instance for the second field. This means to filter ECS instances that are created based on the oldest scaling configuration and then select the oldest instance from the found ECS instances.
VPC	Yes	The ID of the VPC to which the scaling group belongs.
VSwitch	Yes	The ID of the VSwitch to which the scaling group belongs.

5. Click OK.

Result

The specified scaling group is displayed on your scaling group list and is in the Disabled state. To enable the scaling group, you must create a scaling configuration. For more information, see *Create a scaling configuration*.

3.4.2 Enable a scaling group

This topic describes how to enable a scaling group. You can enable a scaling group to trigger scaling activities.

Prerequisites

- The scaling group is in the Disabled state.
- The scaling group has scaling configurations that are in the Enable state.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Find the target scaling group and click Enable in the Actions column.
4. Click OK.

Result

In the Status column, the state of the scaling group is changed from Disabled to Enable.

3.4.3 Query scaling groups

This topic describes how to query the scaling group list and the details of a specific scaling group.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set and a region. The scaling groups that correspond to the specified organization, resource set, and region are displayed.
3. Select the filtering option, enter the corresponding information, and click Search.

   You can select multiple filtering options step by step to narrow down the search results.

Option	Description
Scaling Group	Enter the name of the scaling group that you want to query.
Scaling Group ID	Enter the ID of the scaling group that you want to query.

4. Click the name of the scaling group in the Scaling Group Name/ID column.

5. View the details of the specified scaling group.

Parameter	Description
Basic Information	The configurations of the scaling group, such as the scaling group ID, scaling group name, total instances , minimum number of instances, maximum number of instances, and scale-in policy.
ECS Instances	The details of ECS instances, such as the list of automatically created ECS instances, the list of manually added ECS instances, and the number of ECS instances that are in service.
Scaling Activities	All the scaling activities that have been executed in the scaling group.
Scaling Configuration	The information of scaling configurat ions in the scaling group.
Scaling Rules	The information of scaling rules.

3.4.4 Modify a scaling group

This topic describes how to modify a scaling group. You can modify the parameters of a specified scaling group, such as the minimum and maximum numbers of ECS instances.

Context

After you modify the minimum or maximum number of ECS instances that a scaling group can have, if the number of instances in the scaling group is outside this range, Auto Scaling automatically creates or removes ECS instances until the number of instances are within the range.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Find the target scaling group and click Edit in the Actions column.

4. **Modify the parameters of the scaling group.**

   You can modify the scaling configuration and other parameters, but not the organization and resource set. For more information about other parameters, see *Create a scaling group*.

5. **Click OK.**

3.4.5 Disable a scaling group

This topic describes how to disable a scaling group.

Prerequisites

- **Make sure that the scaling group that you want to disable does not have any scaling activities in progress.**
- **The specified scaling group is in the Enable state.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Find the target scaling group and click Disable in the Actions column.**
4. **Click OK.**

Result

The status of the scaling group is changed from Enable to Disabled in the Status column.

3.4.6 Delete a scaling group

This topic describes how to delete a scaling group. When you delete a scaling group, Auto Scaling removes and releases ECS instances that were automatically created, and removes ECS instances that were manually added. However, the scheduled tasks and event-triggered tasks that are associated with the scaling group are not deleted.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Find the target scaling group and click Delete in the Actions column.**
4. **Click OK.**

3.4.7 Query ECS instances

You can query all ECS instances in a scaling group and their states.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.**
4. **In the left-side navigation pane, click ECS Instances.**
5. **View the details of ECS instances.**

Category	Description
Automatically created ECS instances	The ECS instances that are automatically created based on the enabled scaling configuration when the scaling rule is triggered.
Manually added ECS instances	The ECS instances that are manually added to the specified scaling group.
States of the ECS instances in a scaling group	The number of ECS instances in various states. The states are as follows: · Total: All the ECS instances in a scaling group. · In Service: The ECS instances that are in normal use. · On Standby: The ECS instances that are on standby. · Protected: The ECS instances that are protected. · Disabled: The ECS instances that are stopped. · Adding: The ECS instances that are being added to the scaling group. · Removing: The ECS instances that are being removed from the scaling group.

3.4.8 Switch an ECS instance to the Standby state

This topic describes how to switch an ECS instance to the Standby state. Auto Scaling does not perform health check on or release ECS instances in the Standby state.

Context

After an ECS instance is switched to the Standby state:

- **The ECS instance stays in the Standby state until you change its state.**
- **Auto Scaling stops managing the lifecycle of the ECS instance. You must manually manage the lifecycle of the ECS instance.**
- **If a scale-in activity is triggered, Auto Scaling skips this ECS instance when selecting ECS instances to remove.**
- **When the ECS instance is stopped or restarted, its health check status is not affected.**
- **To release the ECS instance, you must first remove it from the scaling group.**
- **If you delete the scaling group, the ECS instance is automatically removed from the Standby state and is released.**
- **You can also perform other operations on the ECS instance, such as stopping, restarting, changing the specifications of, or changing the operating system of the ECS instance.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.**
4. **In the left-side navigation pane, click ECS Instances.**
5. **Select the source of the ECS instance.**
   - **To select an automatically created ECS instance, click the Auto Created tab.**
   - **To select a manually added ECS instance, click the Manually Added tab.**
6. **Find the target ECS instance and choose Actions > Switch to Standby in the Actions column.**
7. **Click OK.**

3.4.9 Remove an ECS instance from the Standby state

This topic describes how to remove an ECS instance from the Standby state. You can remove an instance from the Standby state to reuse it.

Context

After an ECS instance is removed from the Standby state:

- **The ECS instance enters the InService state.**
- **When the ECS instance is stopped or restarted, its health status is updated.**
- **Auto Scaling continues to manage the lifecycle of the ECS instance and can remove the ECS instance from the scaling group during a scaling activity.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.**
4. **In the left-side navigation pane, click ECS Instances.**
5. **Select the source of the ECS instance.**
   - **To select an automatically created ECS instance, click the Auto Created tab.**
   - **To select a manually added ECS instance, click the Manually Added tab.**
6. **Find the target ECS instance and choose Actions > Move Out Of Standby in the Actions column.**
7. **In the message that appears, click OK.**

3.4.10 Switch an ECS instance to the Protected state

This topic describes how to switch an ECS instance to the Protected state. Auto Scaling does not perform health checks on or release ECS instances that are in the Protected state.

Context

After the state of an ECS instance is switched to Protected:

- **The ECS instance stays in the Protected state until you change its state.**
- **If a scale-in activity is triggered, Auto Scaling skips this ECS instance when selecting the ECS instances to remove. To release the ECS instance, you must manually remove it from the scaling group.**

- When the ECS instance is stopped or restarted, its health check status is not affected.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.
4. In the left-side navigation pane, click ECS Instances.
5. Select the source of the ECS instance.
   - To select an automatically created ECS instance, click the Auto Created tab.
   - To select a manually added ECS instance, click the Manually Added tab.
6. Find the target ECS instance and choose Actions > Switch to Protection in the Actions column.
7. In the message that appears, click OK.

3.4.11 Remove an ECS instance from the Protected state

This topic describes how to remove an ECS instance from the Protected state. After an ECS instance is removed from the Protected state, Auto Scaling continues to manage the lifecycle of the ECS instance.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.
4. In the left-side navigation pane, click ECS Instances.
5. Select the source of the ECS instance.
   - To select an automatically created ECS instance, click the Auto Created tab.
   - To select a manually added ECS instance, click the Manually Added tab.
6. Find the ECS instance that you want to remove from protection and choose Actions > Move Out Of Protection in the Actions column.
7. In the message that appears, click OK.

3.5 Scaling configuration

3.5.1 Create a scaling configuration

This topic describes how to create a scaling configuration for a scaling group.

Prerequisites

Make sure that at least one security group is available. If you do not have any security groups, you must create a security group. For more information, see the Create a security group topic in *ECS User Guide* **.**

Context

You can create only a limited number of scaling configurations for a scaling group. For more information, see the Limits topic in *Auto Scaling Product Introduction* **.**

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.**
4. **In the left-side navigation pane, click Scaling Configuration.**
5. **Choose Create > Create Scaling Configuration.**
6. **Set the parameters for the scaling configuration.**

Category	Parameter	Required	Description
Region	Region	Yes	The region where the ECS instance is located.
	Zone	Yes	The zone where the ECS instance is located.
Security Group	Security Group	Yes	The security group to which the ECS instance belongs.
Instance	Instance Family	Yes	The instance family to which the ECS instance belongs.
	Instance Type	Yes	The instance type of the ECS instance.

Category	Parameter	Required	Description
Image	Image Type	Yes	· Public Image: Public images provided by Alibaba Cloud are fully licensed to offer a secure and stable operating environment for applications on ECS instances. · Custom Image: You can create custom images to install software or deploy projects that have special requirements.
Storage	System Disk	Yes	Specify the category and size of the system disk. The operating system is installed on the system disk. You can select Ultra Disk or SSD Disk.
	Data Disk	No	Specify the category and size of the system disk. You can select Ultra Disk or SSD Disk. You can add a maximum of 16 data disks. The maximum capacity of each data disk is 32 TiB. You can set Release with Instance and Encrypt for each data disk.

Category	Parameter	Required	Description
Password	Set Password	Yes	Select when to set password. You can select Now or Later. If you select Later, you can use the Change Password feature in the console to set the password. For more information, see the Change Password topic in *ECS User Guide* .
	Logon Password	No	The password used to log on to the ECS instance. The password must be 8 to 30 characters in length and must contain at least three of the following character types: digits, uppercase letter, lowercase letters, and special characters. **Note:** The password is used to log on to the operating system and is not the VNC password.
	Confirm Password	No	Enter the password again.
Deployment Set	Deployment Set	No	The deployment set to which the instance belongs.
Instance Name	Configuration Name	No	The name of the scaling configuration.
	Instance Name	No	The name of the ECS instance.

Category	Parameter	Required	Description
User Data	User Data	No	Windows supports two formats: Bat and Powershell. When you perform Base64 encoding, make sure to include `[bat]` or `[powershell]` as the first line. You can run Shell scripts for Linux-based ECS instances.
Quantity	Quantity	No	The number of instances to purchase.

7. Click Submit.

Result

After the scaling configuration is created, it is in the Disabled state and is displayed on your scaling configuration list. To automatically create ECS instances, you must apply a scaling configuration. For more information, see *Apply a scaling configuration*.

3.5.2 Query scaling configurations

This topic describes how to query scaling configurations.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region. The scaling groups that correspond to the specific organization, resource set, and region are displayed.
3. Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.
4. In the left-side navigation pane, click Scaling Configuration.
5. View the list of scaling configurations.

3.5.3 Modify a scaling configuration

This topic describes how to modify a scaling configuration. You can modify the parameters of a scaling configuration based on your actual needs.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.
4. In the left-side navigation pane, click Scaling Configuration.
5. Find the scaling configuration that you want to modify and click its ID in the Scaling Configuration Name/ID column.
6. Modify the parameters of the scaling configuration.

   For more information about parameters of the scaling configuration, see *Create a scaling configuration*.
7. Click OK.

3.5.4 Apply a scaling configuration

This topic describes how to apply a scaling configuration. You can create multiple scaling configurations for a scaling group and apply one as required.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.
4. In the left-side navigation pane, click Scaling Configuration.
5. Find the scaling configuration that you want to apply and click Select in the Actions column.

   Only one scaling configuration can be in the Enable state in a scaling group. After a scaling configuration is applied, other scaling configurations are switched to the Disabled state.
6. In the message that appears, click OK.

Result

The status of the scaling configuration is switched from Disabled to Enable in the Status column.

3.5.5 Delete a scaling configuration

This topic describes how to delete a scaling configuration. If you no longer need a scaling configuration, you can delete it. After you delete a scaling configuration, existing ECS instances that were created from the scaling configuration are not removed.

Prerequisites

The scaling configuration is in the Disabled state.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.**
4. **In the left-side navigation pane, click Scaling Configuration.**
5. **Find the scaling configuration that you want to delete and click Delete in the Actions column.**
6. **In the message that appears, click OK.**

3.6 Scaling rule

3.6.1 Create a scaling rule

This topic describes how to create a scaling rule. You can scale in or out ECS instances by creating scaling rules. For example, you can add an ECS instance to a scaling group.

Context

- **You can create only a limited number of scaling rules for a scaling group. For more information, see the Limits topic in** *Auto Scaling Product Introduction* **.**
- **If a scaling rule is executed and the resulting number of ECS instances in the scaling group is less than the minimum number or greater than the maximum number, Auto Scaling will automatically adjust the number of ECS instances to within the valid range.**

Procedure

1. *Log on to the Auto Scaling console*.

2. In the top navigation bar, select an organization, a resource set, and a region.
3. Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.
4. In the left-side navigation pane, click Scaling Rules.
5. Click Create Scaling Rule.
6. Set the parameters for the scaling rule.

Parameter	Required	Description
Rule Name	Yes	The name of the scaling rule. It must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or digit.
Rule Type	Yes	The type of the scaling rule. You can select Simple Scaling Rule for this parameter. Note that this parameter cannot be modified after the scaling rule is created.
Scaling Activity	Yes	The operation that is executed when the scaling rule is triggered. The operations include: · Change to N instances: When the scaling rule is executed, the number of instances in the scaling group changes to N. · Add N instances: When the scaling rule is executed, N instances are added to the scaling group. · Remove N instances: When the scaling rule is executed, N instances are removed from the scaling group.
Default Cooldown (Seconds)	No	The cooldown period. If this parameter is not specified, the default value is used.

7. Click OK.

3.6.2 Query scaling rules

This topic describes how to query scaling rules.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region. The scaling groups that correspond to the specific organization, resource set, and region are displayed.**
3. **Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.**
4. **In the left-side navigation pane, click Scaling Rules.**
5. **View the list of scaling rules.**

3.6.3 Modify a scaling rule

This topic describes how to modify a scaling rule. You can modify the following parameters for a scaling rule: Rule Name, Scaling Activity, and Default Cooldown.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**
3. **Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.**
4. **In the left-side navigation pane, click Scaling Rules.**
5. **Find the scaling rule that you want to modify and click Edit in the Actions column.**
6. **Modify the Rule Name, Scaling Activity, and Default Cooldown parameters.**
7. **Click OK.**

3.6.4 Delete a scaling rule

This topic describes how to delete a scaling rule. If you no longer need a scaling rule, you can delete it.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the top navigation bar, select an organization, a resource set, and a region.**

3. Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.
4. In the left-side navigation pane, click Scaling Rules.
5. Find the scaling rule that you want to delete and click Delete in the Actions column.
6. In the message that appears, click OK.

3.7 Trigger tasks

3.7.1 Manually execute a scaling rule

This topic describes how to manually execute a scaling rule to add or remove ECS instances.

Prerequisites

- The scaling group to which the scaling rule belongs is in the Enable state.
- No scaling activity is in progress in the scaling group to which the scaling rule belongs.

Context

After the scaling rule is executed, if the number of ECS instances is greater than the maximum number or less than the minimum number, Auto Scaling automatically adjusts the number of ECS instances to within the valid range.

Auto Scaling enables you to execute scaling rules manually. You can also associate an event-triggered task or scheduled task with the scaling rule to automatically adjust the instances. For more information, see *Create a scheduled task* and *Create an event-triggered task*.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.
4. In the left-side navigation pane, click Scaling Rules.
5. Find the scaling rule that you want to execute and click Run in the Actions column.

6. In the message that appears, click OK.

Result

The Scaling Activities page appears. You can view the details of your scaling activity.

3.7.2 Manually add an ECS instance

This topic describes how to manually add an ECS instance to a scaling group. You can add existing ECS instances to a scaling group to take full advantage of the computing resources.

Prerequisites

The ECS instance to be added must meet the following conditions:

- The ECS instance that you want to add and the scaling group share the same region, organization, and resources set.
- The ECS instance to be added is in the Running state.
- The ECS instance is not added to any other scaling groups.
- The ECS instance and the scaling group are in the same VPC.

The scaling group must meet the following conditions:

- The scaling group is in the Enable state.
- No scaling activity is in progress in the scaling group.

Context

- When no scaling activity is being executed in the scaling group, you can immediately remove an ECS instance from the scaling group without the need to wait for the cooldown time to expire.
- After ECS instances are added, the number of instances in the scaling group cannot be greater than the maximum number of instances. Otherwise, the ECS instances cannot be added.
- The manually added ECS instance is not limited by the scaling configuration. The specifications of the manually added instance can be different from those of the scaling configuration in the Enable state.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.

3. Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.
4. In the left-side navigation pane, click ECS Instances.
5. Click Add Instance.
6. Select the ECS instance to be added and click OK.

Result

The manually added instance is displayed on the Manually Added tab.

3.7.3 Manually remove an ECS instance

This topic describes how to manually remove an ECS instance from a scaling group. If you no longer need an ECS instance in a scaling group, you can remove it.

Prerequisites

The scaling group must meet the following conditions:

- The scaling group is in the Enable state.
- No scaling activity is in progress in the scaling group.

Context

- When no scaling activity is being executed in the scaling group, you can immediately remove an ECS instance from the scaling group without the need to wait for the cooldown time to expire.
- After ECS instances are removed, the number of instances in the scaling group must be equal to or greater than the minimum number of instances. Otherwise, the ECS instances cannot be removed.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Find the target scaling group and click the name of the scaling group in the Scaling Group Name/ID column.
4. In the left-side navigation pane, click ECS Instances.
5. Select the source of the ECS instance.
   - To select an automatically created ECS instance, click the Auto Created tab.
   - To select a manually added ECS instance, click the Manually Added tab.

6. Select a method to remove the ECS instances.

   The manually added ECS instances can only be removed, but cannot be released.

   - Find the ECS instance that you want to remove and choose Actions > Remove from Scaling Group in the Actions column.
   - Find the ECS instance that you want to remove and release, and choose Actions > Remove from Scaling Group and Release in the Actions column.

7. In the message that appears, click OK.

3.8 Scheduled tasks

3.8.1 Create a scheduled task

This topic describes how to create a scheduled task to scale computing resources in response to predictable business changes in the future. Scheduled tasks enable the system to automatically obtain sufficient computing resources before business peaks and release idle computing resources after the peaks.

Context

A scheduled task is preconfigured to execute a specified scaling rule at a specified time in the future. When the specified time arrives, the scheduled task scales computing resources to meet the business requirements. This helps you optimize cost by paying only for additional resources when you need them. You can also specify the recurrence for scheduled tasks if business changes are regular.

If multiple scheduled tasks need to be executed in one minute, Auto Scaling executes the newest scheduled task.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the left-side navigation pane, choose Scaling Tasks > Scheduled Tasks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click Create Scheduled Task.

5. In the dialog box that appears, set the parameters for the scheduled task.

Parameter	Required	Description
Task Name	Yes	The name of the scheduled task. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or digit.
Description	Yes	The description of the scheduled task.
Organization/ Resource Group	Yes	The organization and resource set to which the scheduled task belongs.
Start Time	Yes	The time when the scheduled task is executed.
Scaling Rules	Yes	The scaling group that is monitored and the scaling rule that is executed.
Retry Expiry Time	No	The expiration time for the scheduled task retries. Unit: seconds. If a scaling activity fails to execute at a specified time, Auto Scaling will execute the scheduled task again within the retry expiry time.
Recurrence Settings	No	Specifies whether to execute the scheduled task repeatedly. Select Recurrence Settings and set the Recurrence and Expire parameters. The recurrence options include Daily, Weekly, and Monthly.

6. Click OK.

Result

The scheduled task that you created is displayed on your scheduled task list.

3.8.2 Query scheduled tasks

This topic describes how to query scheduled tasks.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the left-side navigation pane, choose Scaling Tasks > Scheduled Tasks.
3. In the top navigation bar, select an organization, a resource set, and a region. The scheduled tasks that correspond to the specific organization, resource set, and region are displayed.

4. Select the filtering option, enter the corresponding information, and click Search.

   You can select multiple filtering options step by step to query scheduled tasks.

Option	Description
Task Name	Enter a task name to query the scheduled task.
Task ID	Enter a scheduled task ID to query the scheduled task.

5. View the list of scheduled tasks.

3.8.3 Modify a scheduled task

This topic describes how to modify a scheduled task. You can modify the following parameters for a scheduled task: Start Time, Scaling Rules, and Retry Expiry Time.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the left-side navigation pane, choose Scaling Tasks > Scheduled Tasks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the scheduled task that you want to modify and click Edit in the Actions column.
5. Modify the parameters of the scheduled task.

   You can modify the Recurrence and Expire parameters if you enable the Recurrence Settings feature when creating the scheduled task, but the Recurrence Settings feature cannot be disabled. For more information about other parameters of the scheduled task, see *Create a scheduled task*.
6. Click OK.

3.8.4 Disable a scheduled task

This topic describes how to disable a scheduled task. You can disable a scheduled task if you do not want to use it to trigger a scaling activity.

Prerequisites

The scheduled task is in the Running state.

Procedure

1. *Log on to the Auto Scaling console*.

2. In the left-side navigation pane, choose Scaling Tasks > Scheduled Tasks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the scheduled task that you want to disable and click Disabled in the Actions column.
5. In the message that appears, click OK.

Result

The status of the scheduled task is switched from Running to Stop in the Status column.

3.8.5 Enable a scheduled task

This topic describes how to enable a scheduled task. You can enable a scheduled task that has been disabled and use it to trigger a scaling activity at a specified time point.

Prerequisites

The scheduled task is in the Stop state.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the left-side navigation pane, choose Scaling Tasks > Scheduled Tasks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the scheduled task that you want to enable and click Enable in the Actions column.
5. In the message that appears, click OK.

Result

The status of the scheduled task is switched from Stop to Running in the Status column.

3.8.6 Delete a scheduled task

This topic describes how to delete a scheduled task. If you no longer need a scheduled task, you can delete it.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the left-side navigation pane, choose Scaling Tasks > Scheduled Tasks.
3. In the top navigation bar, select an organization, a resource set, and a region.

4. Find the scheduled task that you want to delete and click Delete in the Actions column.
5. In the message that appears, click OK.

3.9 Monitoring tasks

3.9.1 Create an event-triggered task

This topic describes how to create an event-triggered task in response to emergent or unpredictable business changes. After you create and enable an event-triggered task, Auto Scaling collects data for the specified metric in real time. When an alert is generated, it triggers a scaling rule. Auto Scaling triggers the scaling rule to dynamically scale ECS instances in the scaling group.

Procedure

1. *Log on to the Auto Scaling console*.
2. In the left-side navigation pane, choose Scaling Tasks > Event-triggered Tasks.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click Create Alert.
5. In the dialog box that appears, set the parameters for the event-triggered task.

Parameter	Required	Description
Task Name	Yes	The name of the event-triggered task. It must be 2 to 40 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or digit.
Description	No	The description of the event-triggered task.
Organization /Resource Group	Yes	The organization and resource set to which the event-triggered task belongs.
Monitoring Metrics/ Scaling Rules	Yes	Select the scaling group that is monitored and the scaling rule that is executed.
Monitoring Type	Yes	It supports System-Level Monitoring.

Parameter	Required	Description
Monitoring Metrics	Yes	Select the metric that you want to monitor. Valid values: · Average CPU Utilization · Memory Usage · System Average Load · Outbound Traffic · Inbound Traffic
Monitoring Period	Yes	Specifies the time during which data is aggregated and analyzed. The shorter the monitoring period, the higher the frequency that the alert is triggered. Unit: minutes. Valid values: · 1 · 2 · 5 · 15
Statistic	Yes	The rule that triggers the alert. Select Average, Max Capacity, or Min Capacity, and specify a threshold value. For example, alerts are triggered when the CPU utilization exceed 80%: · Average: Alerts are triggered when the average CPU utilization of all the ECS instances in the scaling group exceeds 80%. · Max Capacity: Alerts are triggered when the highest CPU utilization among the ECS instances in the scaling group exceed 80%. · Min Capacity: Alerts will be triggered when the lowest CPU utilization among the ECS instances in the scaling group exceeds 80%.
Trigger after occurrences	Yes	The consecutive number of times that the threshold must be exceeded before the alert is triggered. Valid values: · 1 · 2 · 3 · 5

6. Click OK.

3.9.2 Query event-triggered tasks

This topic describes how to query event-triggered tasks.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the left-side navigation pane, choose Scaling Tasks > Event-triggered Tasks.**
3. **In the top navigation bar, select an organization, a resource set, and a region. The event-triggered tasks that correspond to the specific organization, resource set, and region are displayed.**
4. **Select the filtering option, enter the corresponding information, and click Search.**

   You can select multiple filtering options to narrow down the search results.

Option	Description
Alert Name	Enter an event-triggered task name to query the event-triggered task.
Scaling Group ID	Enter a scaling group ID to query the event-triggered task.

3.9.3 Modify an event-triggered task

This topic describes how to modify an event-triggered task. You can modify one or more of the following parameters: Scaling Rules, Monitoring Type, and Statistic.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the left-side navigation pane, choose Scaling Tasks > Event-triggered Tasks.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the event-triggered task that you want to modify and click Edit in the Actions column.**
5. **Modify the parameters of the event-triggered task.**

   You cannot modify the following parameters: Organization and Resource Group, Monitoring Metrics, and Monitoring Period. For more information about other parameters of the event-triggered task, see *Create an event-triggered task*.
6. **Click OK.**

3.9.4 Disable an event-triggered task

This topic describes how to disable an event-triggered task. You can disable an event-triggered task if you no longer want to use it to trigger a scaling activity.

Prerequisites

The event-triggered task is in the Normal, Alerts, or Insufficient Data state.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the left-side navigation pane, choose Scaling Tasks > Event-triggered Tasks.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the event-triggered task that you want to disable and click Disable in the Actions column.**
5. **In the message that appears, click OK.**

Result

The status of the event-triggered task is switched to Stopped in the Status column.

3.9.5 Enable an event-triggered task

This topic describes how to enable an event-triggered task. You can enable an event-triggered task that has been disabled to continue monitoring metrics and trigger scaling activities of a scaling group.

Prerequisites

The event-triggered task is in the Stopped state.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the left-side navigation pane, choose Scaling Tasks > Event-triggered Tasks.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the event-triggered task that you want to enable and click Enable in the Actions column.**
5. **In the message that appears, click OK.**

Result

The status of the event-triggered task is switched from Stopped to Normal in the Status column.

3.9.6 Delete an event-triggered task

This topic describes how to delete an event-triggered task. If you no longer need an event-triggered task, you can delete it.

Procedure

1. *Log on to the Auto Scaling console*.
2. **In the left-side navigation pane, choose Scaling Tasks > Event-triggered Tasks.**
3. **In the top navigation bar, select an organization, a resource set, and a region.**
4. **Find the event-triggered task that you want to delete and click Delete in the Actions column.**
5. **In the message that appears, click OK.**

4 Object Storage Service (OSS)

4.1 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 outstandin g 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.**

4.2 Instructions

Before you use OSS, you must understand the following content:

- **To allow other users to use all or part of OSS functions, you must create RAM users and grant permissions to the users by configuring their RAM policies.**

- **Before you use OSS, you must also understand the following limits.**

Item	Description
Bucket	- **You can create up to 100 buckets.** - **After a bucket is created, its name and region cannot be modified.**
Object upload	- **Objects larger than 5 GB cannot be uploaded 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 multipart upload. The size of an object uploaded using multipart upload cannot exceed 48.8 TB.** - **You can upload an object with the same name as that of an existing object, but the existing object is overwritten.**
Object deletion	- **Deleted objects cannot be recovered.** - **You can delete up to 50 objects at a time in the console. To delete more objects at a time, you must use APIs or SDKs.**
Lifecycle	**You can configure up to 1,000 lifecycle rules for each bucket.**

4.3 Quick start

4.3.1 Log on to the OSS console

This topic describes how to log on to the OSS console.

Prerequisites

- **Before you log on to the Apsara Stack Cloud Management (ASCM) console, you must obtain the IP address or domain name of the ASCM console from the deployment personnel. The URL to access the ASCM console is in the format of http:// the IP address or domain name of the ASCM console/manage.**
- **We recommend that you use the Google Chrome browser.**

Procedure

1. **In the address bar, enter the URL to access the ASCM console. Press Enter.**

2. Enter your username and password.

   The system has a default super administrator with username super. The super administrator can create system administrators. The system administrator can create other users and notify them of their default passwords by SMS or email.

   > **Note:**
   > You must modify the password of your username as instructed when you log on to the ASCM console for the first time. To improve security, the password must meet the minimum complexity requirements. The password must be 8 to 20 characters in length and contain at least two types of the following characters: letters, numbers, or special characters such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).

3. Click Login to go the ASCM console homepage.
4. In the top navigation bar, choose Products > Object Storage Service.

4.3.2 Create buckets

Objects uploaded to buckets are stored in OSS. Before you upload an object to a bucket, you must have a bucket to store objects.

Context

The attributes of a bucket include the region, ACL, and storage class.

Procedure

1. *Log on to the OSS console*.
2. Click Create Bucket if no buckets are available. In the Create OSS Bucket dialog box that appears, configure parameters.

   > **Note:**

In the left-side navigation pane, click the + icon next to Buckets if there are buckets available. The Create OSS Bucket dialog box appears.

The following table describes the parameters for creating a bucket.

Table 4-1: Parameter descriptions

Parameter	Configuration method
Organization	**Select an organization from the drop-down list for the bucket.**
Resource Set	**Select a resource set from the drop-down list for the bucket.**
Region	**Select a region from the drop-down list for the bucket.**  **Note:** · **After a bucket is created, the region cannot be changed.** · **If you want to access OSS from your ECS instance over the internal network, select the region where your ECS instance is deployed.**
Cluster	**Select a cluster for the bucket. Two OSS clusters can be deployed in Apsara Stack.**
Bucket Name	**Enter the name of the bucket.**  **Note:** · **The bucket name must comply with the naming conventions.** · **The bucket name must be globally unique in Apsara Stack OSS.** · **The bucket name cannot be changed after the bucket is created .**
Storage Type	**Set the value to Standard.**

Parameter	Configuration method
Bucket Access	Set the ACL for the bucket. The following options are available: · Private: Only the owner or authorized users of this bucket can read and write objects in the bucket. Other users, including anonymous users cannot access the objects in the bucket without authorization. · Public Read: Only the owner or authorized users of this bucket can write objects in the bucket. Other users, including anonymous users can only read the objects in the bucket. · Public Read/Write: Any users, including anonymous users can read and write objects in the bucket. Fees incurred by such operations are paid by the owner of the bucket. Exercise caution when you configure this option. Note: After a bucket is created, you can modify its ACL. For more information, see *Modify bucket ACLs*.

3. Click Submit.

4.3.3 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 5 GB to a bucket. To upload an object larger than 5 GB, use an SDK or call an API operation.

Procedure

1. *Log on to the OSS console*.

2. In the left-side navigation pane, click the name of a bucket to go to the bucket details page.
3. Click the Files tab.
4. Click Upload. The Upload dialog box appears.
5. In the Upload To section, set the directory to which the object will be uploaded.
   - Current: Objects are uploaded to the current folder.
   - Specified: Objects are uploaded to the specified folder. OSS creates the specified folder automatically and uploads the object to it.

     Note:
     For more information about folders, see *Create folders*.
6. In the File ACL section, select the ACL of the object to be uploaded. By default, an object inherits the ACL of the bucket to which it belongs.
7. Drag and drop one or more objects to be uploaded to the Upload section, or click Upload to select one or more objects to upload.

   Note:
   - If the uploaded object has the same name as an existing object in the bucket, the existing object will be overwritten.
   - Do not refresh or close the upload page when objects are being uploaded . Otherwise, the upload tasks are interrupted and the upload object list is cleared.
   - The name of the uploaded object must comply with the following conventions:
     - The name can contain only UTF-8 characters.
     - The name is case-sensitive.
     - The name must be 1 to 1,023 bytes in length.
     - The name cannot start with a forward slash (/) or backslash (\).
8. After the object is uploaded, refresh the Files tab to view the uploaded object.

4.3.4 Obtain object URLs

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 must have a bucket and upload an object to it.

Procedure

1. *Log on to the OSS console*.
2. **In the left-side navigation pane, click the name of a bucket to go to the bucket details page.**
3. **Click the Files tab. The list of objects appears.**
4. **Click the name of the target object. In the Preview dialog box that appears, click Copy File URL next to the URL field. You can also choose More > Copy File URL in the Actions column corresponding to the object. In the dialog box that appears, click Copy.**

What's next

You can send the URL to other users so that they can view or download the object.

4.4 Buckets

4.4.1 View bucket details

You can view the details of created buckets in the OSS console.

Prerequisites

Before you view a bucket, you must complete the procedure instructed in *Create buckets*, **or at least one bucket exists in the current region.**

Procedure

1. *Log on to the OSS console*.
2. **In the left-side navigation pane, click the name of a bucket to go to the bucket details page.**
3. **On the bucket details page that appears, click the Overview tab. View the bucket domain names and basic settings.**

4.4.2 Delete buckets

You can delete buckets in the OSS console.

Prerequisites

Before you delete a bucket, you must complete the procedure instructed in *Create buckets*, or at least one bucket exists in the current region.

> **Note:**
> **To delete a bucket, ensure that all objects in the bucket are deleted, including parts generated from incomplete multipart upload operations. Otherwise, the bucket cannot be deleted.**

Procedure

1. ***Log on to the OSS console*.**
2. **In the left-side navigation pane, click the name of a bucket to go to the bucket details page.**
3. **On the bucket details page that appears, click Delete Bucket in the upper-right corner. In the message that appears, click OK.**

> **Notice:**
> **Deleted buckets cannot be recovered. Exercise caution when you perform this operation.**

4.4.3 Modify bucket ACLs

You can modify the Access Control List (ACL) of a bucket in the OSS console to control access to the bucket.

Prerequisites

Before you modify the ACL of a bucket, you must complete the procedure instructed in *Create buckets*, or at least one bucket exists in the current region.

Context

OSS provides ACL to control access to buckets. By default, the ACL of a bucket is private when you create the bucket. You can modify the ACL of the bucket after the bucket is created.

OSS provides ACL for buckets. The following ACLs are available for a bucket.

- **Private: Only the owner or authorized users of this bucket can read and write objects in the bucket. Other users, including anonymous users cannot access the objects in the bucket without authorization.**
- **Public Read: Only the owner or authorized users of this bucket can write objects in the bucket. Other users, including anonymous users can only read the objects in the bucket.**
- **Public Read/Write: Any users, including anonymous users can read and write objects in the bucket. Fees incurred by such operations are paid by the owner of the bucket. Exercise caution when you configure this option.**

 **Notice:**
If you set ACL to public read or public read/write, other users can directly read the data in the bucket without authentication, resulting in security risks. For data security reasons, we recommend that you set the bucket ACL to private.

Procedure

1. *Log on to the OSS console*.
2. **In the left-side navigation pane, click the name of a bucket to go to the bucket details page.**
3. **On the bucket details page that appears, click the Basic Settings tab. Find the Access Control List (ACL) section.**
4. **Click Configure. Modify the bucket ACL.**
5. **Click Save.**

4.4.4 Configure static website hosting

You can configure static website hosting in the OSS console so that users can access the static website through the bucket domain name.

Prerequisites

Before you configure static website hosting for a bucket, you must complete the procedure instructed in *Create buckets*, or at least one bucket exists in the current region.

Context

Static website hosting is not enabled if the default pages are not specified.

After the default homepage is configured, the default homepage is displayed if you 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. In the left-side navigation pane, click the name of a bucket to go to the bucket details page.
3. On the bucket details page that appears, click the Basic Settings tab. Find the Static Pages section.
4. Click Configure. Configure the following parameters:
   - Default Homepage: Specify the name of the index document that links to the index page. The index page functions similar to index.html. Only the HTML object in the root folder can be used. The default homepage is not enabled if you do not specify this parameter
   - Default 404 Page: Specify the name of the error document that links to the error page displayed when the requested resource does not exist. Only the HTML, JPG, PNG, BMP, or WebP object in the root folder can be used. The default 404 page is not enabled if you do not specify this parameter.
5. Click Save.

4.4.5 Configure hotlink protection

You can configure hotlink protection for a bucket in the OSS console to prevent unauthorized domain names from accessing the data in your bucket.

Prerequisites

Before you configure hotlink protection for a bucket, you must complete the procedure instructed in *Create buckets*, or at least one bucket exists in the current region.

Context

OSS provides hotlink protection to prevent other domain names from accessing your data in OSS. You can configure the Referer field in the HTTP header to implement hotlink protection. You can configure a Referer whitelist in the OSS console for a bucket or configure whether to accept requests whose Referer field is empty. For example, you can add `http://www.aliyun.com` to the Referer whitelist

for a bucket named oss-example. Then, requests whose Referer field is set to `http://www.aliyun.com` can access the objects in the *oss-example* bucket.

Procedure

1. *Log on to the OSS console*.
2. In the left-side navigation pane, click the name of a bucket to go to the bucket details page.
3. On the bucket details page that appears, click the Basic Settings tab. Find the Hotlinking Protection section.
4. Click Configure. Configure the following parameters:
   - Referer Whitelist: Add URLs to the whitelist. Referers are typically in URL format. Separate multiple Referers with new lines. You can use question marks (?) and asterisks (*) as wildcard characters.
   - Allow Empty Referer: Specify whether to allow requests whose Referer field is empty. If you do not allow empty Referers fields, only HTTP or HTTPS requests which include an allowed Referer field value can access the objects in the bucket.
5. Click Save.

4.4.6 Configure logging

You can enable or disable bucket logging in the OSS console.

Prerequisites

Before you enable or disable logging for a bucket, you must complete the procedure instructed in *Create buckets*, or at least one bucket exists in the current region.

Context

You can store access logs in the current bucket or in a new bucket.

Procedure

1. *Log on to the OSS console*.
2. In the left-side navigation pane, click the name of a bucket to go to the bucket details page.
3. On the bucket details page that appears, click the Basic Settings tab. Find the Logging section.

4. Click Configure. Turn on Logging. Set Destination Bucket and Log Prefix.
   - Destination Bucket: Select the name of the bucket in which access logs are to be stored from the drop-down list. You must be the owner of the selected bucket and the bucket must be in the same region as the bucket for which logging is enabled.
   - Log Prefix: Enter the directory where the access logs are generated and the prefix of the access logs. The access logs are stored in the specified directory. Example: `log/<TargetPrefix>`. The access logs are stored in the *log* / directory.
5. Click Save.

4.4.7 Configure CORS

You can configure cross-origin resource sharing (CORS) in the OSS console to enable cross-origin access.

Prerequisites

Before you configure CORS, you must complete the procedure instructed in *Create buckets*, or at least one bucket exists in the current region.

Context

OSS provides CORS in HTML5 to enable cross-origin access. When OSS receives a cross-origin request (or an OPTIONS request), it reads the CORS rules of the bucket and then checks the relevant permissions. OSS matches the rules one by one. When OSS finds the first match, it returns a corresponding header. If none of the rules match, OSS does not attach any CORS header.

Procedure

1. *Log on to the OSS console*.
2. In the left-side navigation pane, click the name of a bucket to go to the bucket details page.
3. On the bucket details page, click the Basic Settings tab. Find the Cross-Origin Resource Sharing (CORS) section. Click Configure.

4. Click Create Rule. In the Create Rule dialog box that appears, configure the following parameters.

Parameter	Required	Description
Sources	Yes	Specifies the sources from which you want to allow cross-origin requests. You can configure multiple origins and separate them with new lines. Each origin can contain only one asterisk (*) wildcard. If Sources is set to asterisk (*), all cross-origin requests are allowed.
Allowed Methods	Yes	Specifies the cross-origin request methods that are allowed.
Allowed Headers	No	Specifies the allowed headers in a cross-origin request. Allowed headers are case-insensitive. You can configure multiple headers and separate them with new lines. Each rule can contain only one asterisk (*) wildcard. If there are no special header requirements, we recommend that you set Allowed Headers to asterisk (*) to allow all requests.
Exposed Headers	No	Specifies the list of headers that can be exposed to the browser. The headers are the response headers that allow access from an application such as XMLHttpRequest in JavaScript. No asterisk (*) wildcards are allowed.
Cache Timeout (Seconds)	No	Specifies the time the browser can cache the response to a preflight (OPTIONS) request to a specific resource.

 **Note:**

You can configure up to 10 rules for each bucket.

5. Click OK.

4.4.8 Manage lifecycle rules

You can define and manage lifecycle rules for a bucket in the OSS console.

Prerequisites

Before you manage lifecycle rules, you must complete the procedure instructed in *Create buckets*, **or at least one bucket exists 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 manage lifecycle rules to perform operations, such as 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 based on a lifecycle rule cannot be recovered. Exercise caution when you configure a rule.**

Procedure

1. *Log on to the OSS console*.
2. **In the left-side navigation pane, click the name of a bucket to go to the bucket details page.**
3. **Click the Basic Settings tab. Find the Lifecycle section. Click Configure.**
4. **Click Create Rule. In the Create Rule dialog box that appears, configure the following parameters:**
   - **Status: Configure the status of the rule: Enabled or Disabled.**
   - **Applied To: You can select Files with Specified Prefix or Whole Bucket. Files with Specified Prefix indicates that this rule applies to objects that have a specified prefix. Whole Bucket indicates that this rule applies to all objects in the bucket.**

      **Note:**
     If you select Files with Specified Prefix, you can configure multiple lifecycle rules that have different prefixing configurations for objects. If you select `Whole Bucket`**, only one lifecycle rule can be configured. In addition, if you**

have created a rule that has Files with Specified Prefix configured, you cannot create another rule that has Whole Bucket configured for the same bucket.

- **Prefix: If you set Applied To to Files with Specified Prefix, you must enter the prefix of the objects to which to apply the rule. If you want to match objects whose names start with img, enter `img`.**
- **File Lifecycle: Configure the operation to perform on expired objects. You can select Validity Period (Days), Expiration Date, or Disabled.**
  - **Validity Period (Days): Specify the number of days objects within which objects can be retained after they are last modified, and the operation to perform on these objects after they expire. Objects that meet the specified conditions are retained within the specified validity period after the objects are last modified. The specified operation is performed on these objects after they expire. If you select Delete and set Validity Period (Days) to 30, objects that are last modified before January 1, 2016 are scanned for by the backend application and deleted on January 31, 2016.**
  - **Expiration Date: Specify a date before which objects that are last modified expire and the operation to perform on these objects after they expire. All objects that are last modified before this date expire and the specified operation is performed on these objects. If you select Transit to Archive Storage Class and set Expiration Date to 2012-12-21, the backend application scans for objects that are last modified before December 21, 2012 and converts their storage class to Archive.**
  - **Disabled: The automatic object deletion or storage class conversion function is not enabled.**
- **Delete: If you select Validity Period (Days) or Expiration Date for File Lifecycle, you can select Delete to delete objects based on the validity period or expiration time. If you select Disabled, the rule becomes invalid.**
- **Fragment Lifecycle: Configure the delete operation to perform on expired parts. You can select Validity Period (Days), Expiration Date, or Disabled.**
  - **Validity Period (Days): Specify the number of days parts within which parts can be retained after they are last modified. After the validity period, expired parts are deleted. If you set Validity Period (Days) to 30, the**

backend application scans for parts that are last modified before January 1, 2016 and deletes them on January 31, 2016.

- Expiration Date: Specify the date before which parts that are last modified expire and the delete operation to perform on these parts after they expire. If you set Expiration Date to 2012-12-21, parts that are last modified before this date are scanned for and deleted by the backend application.
- Disabled: The automatic part deletion function is not enabled.

· Delete: If you select Validity Period (Days) or Expiration Date for Fragment Lifecycle, you can select Delete to delete parts based on the validity period or expiration time. If you select Disabled, the rule becomes invalid.

Notice:
In each lifecycle rule, you must configure at least object lifecycle or part lifecycle. In other words, you must select Delete or configure conversion actions for object lifecycle or select Delete for part lifecycle.

5. Click OK.

Notice:

· Lifecycle rules are run after they are configured and saved. Check the configurations before you save them.
· Object deletion is irreversible. Configure lifecycle rules as needed.

4.5 Objects

4.5.1 Search for objects

You can search buckets or folders for objects whose names contain a specified prefix in the OSS console.

Prerequisites

Before you search for objects, you must complete the procedure instructed in *Create buckets* and *Upload objects*, or at least one bucket exists in the current region and at least one object exists in the bucket.

Context

When you search for objects 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. In the left-side navigation pane, click the name of a bucket to go to the bucket details page.
3. On the bucket details page that appears, click the Files tab. The Files tab appears.
4. On the right side of the Files tab, enter a prefix in the search box and press Enter or click the search icon to search for the related objects.

   The system lists the names of objects and folders that match the prefix in the root directory of the bucket.

   Note:
   To search a specific folder for objects, open the folder and enter a prefix in the search box. The system lists the names of objects and subfolders in the folder that match the prefix.

4.5.2 Delete objects

You can delete uploaded objects in the OSS console.

Prerequisites

Before you delete objects, you must complete the procedure instructed in *Create buckets* and *Upload objects*, or at least one bucket exists in the current region and at least one object exists in the bucket.

Context

You can delete one or more objects at a time. A maximum of 1,000 objects can be deleted at a time. You can use an SDK or call an API operation to delete a specific object or more than 1,000 objects.

Notice:
Deleted objects cannot be recovered. Exercise caution when you delete objects.

Procedure

1. *Log on to the OSS console*.
2. In the left-side navigation pane, click the name of a bucket to go to the bucket details page.
3. On the bucket details page that appears, click the Files tab.
4. Select one or more objects. Choose Batch Operation > Delete. You can also choose More > Delete in the Actions column corresponding to the target object.
5. In the Delete File message that appears, click OK.

4.5.3 Configure ACL for objects

You can configure ACL for an object in the OSS console to control access to the object.

Prerequisites

Before you configure ACL for an object, you must complete the procedure instructed in *Create buckets* and *Upload objects*, or at least one bucket exists in the current region and at least one object exists in the bucket.

Procedure

1. *Log on to the OSS console*.
2. In the left-side navigation pane, click the name of a bucket to go to the bucket details page.
3. On the bucket details page that appears, click the Files tab.
4. On the Files tab, click the name of the target object. The Preview dialog box appears.

5. Click Set ACL on the right side of File ACL. The Set ACL dialog box appears.

   ACLs are described as follows:

   - Inherited from Bucket: The ACL of each object is the same as that of the bucket.
   - Private: Only the owner or authorized users of this bucket can read and write objects in the bucket. Other users, including anonymous users cannot access the objects in the bucket without authorization.
   - Public Read: Only the owner or authorized users of this bucket can write objects in the bucket. Other users, including anonymous users can only read the objects in the bucket.
   - Public Read/Write: Any users, including anonymous users can read and write objects in the bucket. Fees incurred by such operations are paid by the owner of the bucket. Exercise caution when you configure this option.

    **Note:**
   You can also choose More > Set ACL in the Actions column corresponding to the target object to open the Set ACL dialog box.

6. Click OK.

4.5.4 Create folders

You can create a folder in a bucket in the OSS console.

Prerequisites

Before you create a folder, you must complete the procedure instructed in *Create buckets*, or at least one bucket exists in the current region.

Context

OSS does not use a hierarchical structure for objects, but instead uses a flat structure. All elements are stored as objects in buckets. However, OSS supports folders as a concept to group objects and simplify management. In the OSS console, a folder is an object whose size is 0 and whose name ends with a forward slash (/). A folder is used to sort objects of the same type and process them at a time. The OSS console displays objects whose names end with a forward slash (/) as folders. These objects can be uploaded and downloaded. You can use OSS folders in the OSS console the same way you use folders in Windows.

 **Note:**

The OSS console displays any objects whose names end with a forward slash (/) as folders, regardless of whether these objects contain data. You can download these objects only by calling an API operation or using an SDK.

Procedure

1. *Log on to the OSS console*.
2. **In the left-side navigation pane, click the name of a bucket to go to the bucket details page.**
3. **Click the Files tab. On the Files tab that appears, click Create Folder.**
4. **In the Create Folder dialog box that appears, set Folder Name.**

   The folder name must comply with the following conventions:

   - **The name can contain only UTF-8 characters. The name cannot contain emojis .**
   - **The name cannot start with a forward slash (/) or backslash (\). The name cannot contain consecutive forward slashes (/). Folders are separated with forward slashes (/). Subfolders in this path are automatically created.**
   - **The subfolder name cannot contain two consecutive periods (. .) .**
   - **The name must be 1 to 254 characters in length.**
5. **Click OK.**

5 Table Store

5.1 What is Table Store?

Table Store is a NoSQL database service independently developed by Alibaba Cloud. Table Store is a proprietary software program that is certified by the relevant authorities in China. Table Store is built on the Apsara system of Alibaba Cloud, and can store large amounts of structured data and allow real-time access to these data.

Table Store provides the following features:

- **Offers schema-free data storage. You do not need to define attribute columns before you use them. You do not require table-level changes to add or delete attribute columns. You can configure the time to live (TTL) parameter for a table to manage the lifecycle of data, and delete expired data from the table.**
- **Adopts the triplicate technology to keep three copies of data on three servers across three different racks. A cluster can support single storage type instances (SSD only) or mixed storage type instances (SSD and HDD) to meet different budget and performance requirements.**
- **Adopts a fully redundant architecture that prevents single point of failures (SPOFs). With support for smooth online upgrades, hot cluster upgrades, and automatic data migration, you can dynamically add or remove nodes for maintenance without incurring service interruptions. The concurrent read/write throughput and storage capacity can be linearly scaled. Each cluster can have no less than 500 hosts.**
- **Supports highly concurrent read/write operations. Concurrent read/write capabilities can be scaled out as the number of hosts increases. The read/write performance is indirectly related to the amount of data in a single table.**
- **Supports identity authentication and multi-tenancy. Comprehensive access control and isolation mechanisms are provided to safeguard your data. VPC and access over HTTPS are supported. Provides multiple authentication and authorization mechanisms so that you can define access permissions on individual tables and operations.**

5.2 Limits

Before using Table Store, you need to take note of the following precautions and limits.

The following table describes the limits for Table Store. Some of the limit ranges indicate the maximum allowable values instead of the suggested values. For better performance, set the table scheme and data size in a single row properly based on actual conditions, and adjust the following configurations as needed.

Item	Limit	Description
The number of instances under an Apsara Stack tenant account	1,024	To raise the limit, contact the technical support personnel.
The number of tables in an instance	1,024	To raise the limit, contact the technical support personnel.
Instance name length	3 to 16 Bytes	The instance name can contain uppercase and lowercase letters, digits, and hyphens (-). It must start with a letter and cannot end with a hyphen (-).
Table name length	1 to 255 Bytes	The table name can contain uppercase and lowercase letters, digits, and underscores (_). It must start with a letter or underscore (_).
Column name length	1 to 255 Bytes	The column name can contain uppercase and lowercase letters, digits, and underscores (_). It must start with a letter or underscore (_).
The number of columns in a primary key	1 to 4	A primary key can contain one to four columns.
The size of the value in a string type primary key column	1 KB	The size of the value in a string type primary key column cannot exceed 1 KB.
The size of the value in a string type attribute column	2 MB	The size of the value in a string type attribute column cannot exceed 2 MB.
The size of the value in a binary type primary key column	1 KB	The size of the value in a binary type primary key column cannot exceed 1 KB.

Item	Limit	Description
The size of the value in a binary type attribute column	2 MB	The size of the value in a binary type attribute column cannot exceed 2 MB.
The number of attribute columns in a single row	Unlimited	A single row can contain an unlimited number of attribute columns.
The number of attribute columns written by one request	1,024	During a PutRow, UpdateRow, or BatchWriteRow operation, the number of attribute columns written in a row cannot exceed 1,024.
The data size of a row	Unlimited	The total size of all column names and column values for a row is unlimited.

5.3 Quick start

5.3.1 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 ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel. The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.**
- **We recommend that you use the Google Chrome browser.**

Procedure

1. **In the address bar, enter the URL used to access the ASCM console. Press Enter.**
2. **Enter your username and password.**

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

   Note:

When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).

3. Click Login to go to the ASCM console homepage.
4. In the top navigation bar, click Products and select Table Store.

5.3.2 Create instances

An instance is a logical entity in Table Store and is used to manage tables. An instance is the basic unit of the resource management system of Table Store. Table Store implements application access control and resource measurement at the instance level. This topic describes how to create an instance.

Procedure

1. *Log on to the Table Store console*.
2. On the Overview page, click Create Instance.

   Note:
   You can create different instances to manage the tables for different business, or create different instances for development, testing, and production environments of the same business. Table Store allows you to create up to 1,024 instances and up to 1,024 tables in each instance under an Apsara Stack tenant account.

3. On the Create Table Store Instance page, configure parameters.

Parameter	Description
Region	Select a region from the drop-down list for the instance.
Organization	Select an organization from the drop-down list for the instance.
Resource Set	Select a resource set from the drop-down list for the instance.

Parameter	Description
Instance Name	Specify the instance name. Instance naming conventions: The name must be 3 to 16 characters in length and can only contain letters, digits, and hyphens (-). It must start with a letter and cannot start with `ali` or `ots`.
Instance Type	Specify the instance type. Table Store provides high-performance instances and capacity instances. The instance types vary based on the type of cluster you deploy.

4. Click Submit.

   The created instance is displayed on the Overview page.

5.3.3 Create tables

This topic describes how to create a Table Store table.

Procedure

1. *Log on to the Table Store console*.
2. Find the instance in which you want to create the table, and click the instance name to go to the Instance Management page.
3. On the Instance Management page, click Create Table.

   Note:
   You can create a maximum of 1,024 tables in each instance.

4. In the Create Table dialog box that appears, configure Table Name and Primary Key.

   The following table describes the parameters.

Parameter	Description
Table Name	The table name can contain letters, digits, and underscores (_). It must start with a letter or an underscore (_). The table names in an instance must be unique. The primary key configuration and the key order cannot be modified after they are set.

Parameter	Description
Primary Key	A maximum of four primary keys can be set. By default, the first primary key is the partition key. Click Add a Primary Key to add a new primary key. The primary keys can be of the following types: `Integer`, `String`, `Binary`, and `autoIncrement`. The primary key configuration and the key order cannot be modified after they are set.  Note: The partition key cannot be set as an auto-increment column. The primary key name can contain letters, digits, and underscores (_). It must start with a letter or an underscore (_).

5. Optional. To configure advanced settings such as the lifecycle for a table, turn on Advanced Settings in the upper-right corner of the Create Table dialog box.

   The following table describes the parameters.

Parameter	Description
Time To Live	The minimum TTL is 86,400 seconds (one day). A value of -1 indicates that the data does not expire.
Max Versions	A non-zero value. It specifies the maximum number of data versions that can be stored in each attribute column of a table. When the number of versions in an attribute column exceeds the parameter value, the earliest version is deleted. This operation is performed asynchronously.

Parameter	Description
Max Version Offset	The difference between the version number and the write time of the data must be within the value of Max Version Offset. Otherwise, an error will occur when data is written. The valid version range of an attribute column is as follows: `[Data write time - Valid version offset, Data write time + Valid version offset)`.
Reserved Read Throughput	The reserved read/write throughput can be set to 0. When the reserved read/write throughput is greater than 0, Table Store allocates and reserves corresponding resources for the table based on the configuration. Valid values: integers from 0 to 5000. Capacity instances do not support this parameter.
Reserved Write Throughput	

6. Click OK to complete the table creation.

   The system automatically returns to the Instance Management page and displays the table creation result. After the table is created, it is displayed in the Table List section.

5.3.4 Read and write data

This topic describes how to read and write data.

Write data

The procedure is as follows:

1. *Log on to the Table Store console*.
2. Find the table to which you want to write data. Click Data Editor in the Actions column corresponding to the table.
3. On the Data Editor tab that appears, click Insert.

4. In the Insert dialog box that appears, enter the primary key Value. Click Add Column and configure the parameters as described in the following table.

Parameter	Description
Name	The primary key name can contain letters, digits, and underscores (_). It must start with a letter or an underscore (_).
Type	Valid values: `Integer`, `String`, `Binary`, `Float`, and `Boolean`.
Value	None
Version	A non-zero value. The system time is used by default. It specifies the maximum number of data versions that can be stored in each attribute column of a table. When the number of versions in an attribute column exceeds the parameter value, the earliest version is deleted. This operation is performed asynchronously.

5. Click OK. A new row of data is displayed on the Data Editor tab.

Delete data

You can delete rows of data that you no longer need. The procedure is as follows:

1. Select the rows of data that you want to delete and click Delete.
2. In the Delete message that appears, click OK.

Update data

You can update the attribute columns of a data row in the console. The procedure is as follows:

1. Select the data row to be updated, and click Update.
2. In the Update dialog box that appears, click Add Column to add an attribute column to this row. You can also delete or update an existing attribute column.
3. Click OK.

Query data

You can query data in a single row (GetRow) or query data within a specified range (RangeQuery). The procedure is as follows:

Single-row query

1. Find the table from which you want to query data. On the Data Editor tab, click Search. The Search dialog box appears.
2. Select GetRow for Mode.
3. In the Columns to Return section, turn off All Columns and enter the attribute column to be returned in the field that appears. To return all the attribute columns of the row, retain All Columns turned on.
4. Specify the primary key column.

   Enter the complete primary key Value of the row to be read. The integrity and accuracy of the primary key value affect the query results.
5. Set Max Versions to specify the number of data versions to be returned.
6. Click OK.

Range query

1. Find the table from which you want to query data. On the Data Editor tab, click Search. The Search dialog box appears.
2. Select Range Search for Mode.
3. In the Columns to Return section, turn off All Columns and enter the attribute column to be returned in the field that appears. To return all the attribute columns of the row, retain All Columns turned on.
4. Specify the primary key range for the query. You can specify Min Value and Max Value, or select Custom to specify a range.

   Note:

   - Range queries use a combination of joint indexes and leftmost matching. The first primary key value takes priority when the range query mode is used. When the minimum and maximum values of the first primary key are the same, the system will use the second primary key to perform the query.
   - The Custom range is a left-open and right-closed interval.
5. Set Max Versions to specify the number of data versions to be returned.
6. Select Forward Search or Backward Search for Sequence as needed.

5.4 Instances

5.4.1 View instances

This topic describes how to view details of an instance.

Procedure

1. *Log on to the Table Store console*.
2. **On the Overview page, click the name of the instance that you want to view.**
3. **On the Instance Management page that appears, you can view Instance Access URL and Table List.**
4. **Click the Network Management tab to view VPC List.**

5.4.2 Release instances

This topic describes how to release a Table Store instance.

Prerequisites

- **Before releasing an instance, you must delete all tables from the instance.**
- **Before releasing an instance, you must unbind the VPC from the instance.**

Procedure

1. *Log on to the Table Store console*.
2. **On the Overview page, click Release in the Actions column corresponding to the instance to be released.**
3. **In the Release message that appears, click OK.**

5.5 Tables

5.5.1 View table details

This topic describes how to view the basic information and actual usage of a table.

Procedure

1. *Log on to the Table Store console*.
2. **Find the instance to which the target table belongs, and click the instance name to go to the Instance Management page.**
3. **In the Table List section, find the table.**

4. Click Details in the Actions column corresponding to the table.

   The Manage Table page appears. On the Details tab, you can view the table name, time to live (TTL), reserved read throughput, table size, log expiration time, and the primary keys and their types.

5.5.2 Update table attributes

This topic describes how to update the attributes of a table.

Procedure

1. *Log on to the Table Store console*.
2. Find the instance to which the target table belongs, and click the instance name to go to the Instance Management page.
3. In the Table List section, find the table.
4. Click Details in the Actions column corresponding to the table.
5. On the Manage Table page that appears, click Modify Attributes.
6. In the Modify Attributes dialog box that appears, modify Time To Live, Max Version, and Max Version Offset as needed.
7. Click OK. Go back to the Instance Management page, you can view the new parameter values. These values take effect immediately.

5.5.3 Delete tables

This topic describes how to delete a table.

Context

Notice:

After a table is deleted, data in the table cannot be restored.

Procedure

1. *Log on to the Table Store console*.
2. Find the instance to which the target table belongs, and click the instance name to go to the Instance Management page.
3. In the Table List section, find the table.
4. Click More in the Actions column corresponding to the table and choose Delete from the shortcut menu.

5. In the Delete Table message that appears, click OK.

> **Note:**
> After the delete operation is confirmed, the table and its data are permanently deleted.

5.5.4 Manage Stream

Table Store Stream is a data channel used to obtain incremental data from Table Store tables. This topic describes how to enable and disable Stream.

Enable Stream

1. *Log on to the Table Store console*.
2. Find the instance to which the target table belongs, and click the instance name to go to the Instance Management page.
3. In the Table List section, find the table.
4. Click Details in the Actions column corresponding to the table.
5. On the Manage Table page that appears, find the Stream Information section and click Enabled.
6. In the Enable Stream dialog box that appears, set Log Expiration Time. Click Enabled.

   The value of Log Expiration Time is expressed in hours and must be a non-zero integer. The maximum value is 168.

Disable Stream

On the Manage Table page, find the Stream Information section and click Disabled. In the Disable the Stream Function message that appears, click Disabled.

> **Note:**
> If Stream is disabled, all existing Stream records are permanently deleted.

5.6 Bind a VPC

ECS instances in a VPC can access Table Store instances in the same region through the VPC. Before accessing a Table Store instance over a VPC, you must bind the VPC to the Table Store instance.

Prerequisites

- You must create a VPC that is in the same region as the Table Store instance. For more information about how to create a VPC, see theCreate a VPC section in *VPC User Guide* .
- After the VPC is created, create an ECS instance in the VPC.

Procedure

1. *Log on to the Table Store console*.
2. Find the instance to be bound to a VPC, and click the instance name to go to the Instance Management page. Click the Network Management tab.
3. Click Bind VPC.
4. In the Bind VPC dialog box that appears, configure VPC parameters. Click OK.
5. After the instance is bound to the VPC, you can view the information about the bound VPC in the VPC list. You can use the VPC Access Address to access the Table Store instance from the ECS instance in the VPC.
   - Click VPC Instance List in the Actions column corresponding to the VPC to view all Table Store instances bound to the VPC.
   - If the VPC is no longer needed, click Unbind in the Actions column corresponding to the VPC to unbind the VPC from the instance. After the VPC is unbound from the Table Store instance, the ECS instance in the VPC can no longer access the Table Store instance through the preceding access address. To access the Table Store instance, you must bind the VPC to the instance again.

6 ApsaraDB for RDS

6.1 What is ApsaraDB for RDS?

ApsaraDB for RDS is a stable, reliable, and scalable online database service. Based on the distributed file system and high-performance storage, ApsaraDB for RDS allows you to easily perform database operations and maintenance with its set of solutions for disaster recovery, backup, restoration, monitoring, and migration.

Originally based on a branch of MySQL, ApsaraDB RDS for MySQL is a tried and tested solution for handling high-volume concurrent traffic during Double 11, providing excellent performance. ApsaraDB RDS for MySQL provides whitelist configuration, backup and restoration, transparent data encryption, data migration , and management for instances, accounts, and databases.

ApsaraDB RDS for MySQL also provides read-only instances. In scenarios where RDS has a small number of write requests but a large number of read requests, you can create read-only instances to scale the reading capability and increase the application throughput.

6.2 Log on to the ApsaraDB for RDS console

This topic describes how to log on to the ApsaraDB for RDS console.

Prerequisites

- **Before logging on to the ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel . The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.**
- **We recommend that you use the Google Chrome browser.**

Procedure

1. **In the address bar, enter the URL used to access the ASCM console. Press Enter.**
2. **Enter your username and password.**

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator

can create system users and notify the users of the default passwords by SMS or email.

> **Note:**
> When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).

3. Click Login to go to the ASCM console homepage.
4. In the top navigation bar, choose Products > Database Services > ApsaraDB for RDS.

6.3 Quick start

6.3.1 Limits

To ensure instance stability and security, ApsaraDB RDS for MySQL has some service limits, as listed in the following table.

Operation	Description
Instance parameters	Instance parameters can be modified through the RDS console or API operations. Due to security and stability considerations, only specific parameters can be modified.
Root permissions of databases	The root or system administrator permissions are not provided.
Database backup	· Logical backup can be performed through the command line interface (CLI) or graphical user interface (GUI). · Physical backup can only be performed through the RDS console or API operations.
Database restoration	· Logical restoration can be performed through the CLI or GUI. · Physical restoration can only be performed through the RDS console or API operations.

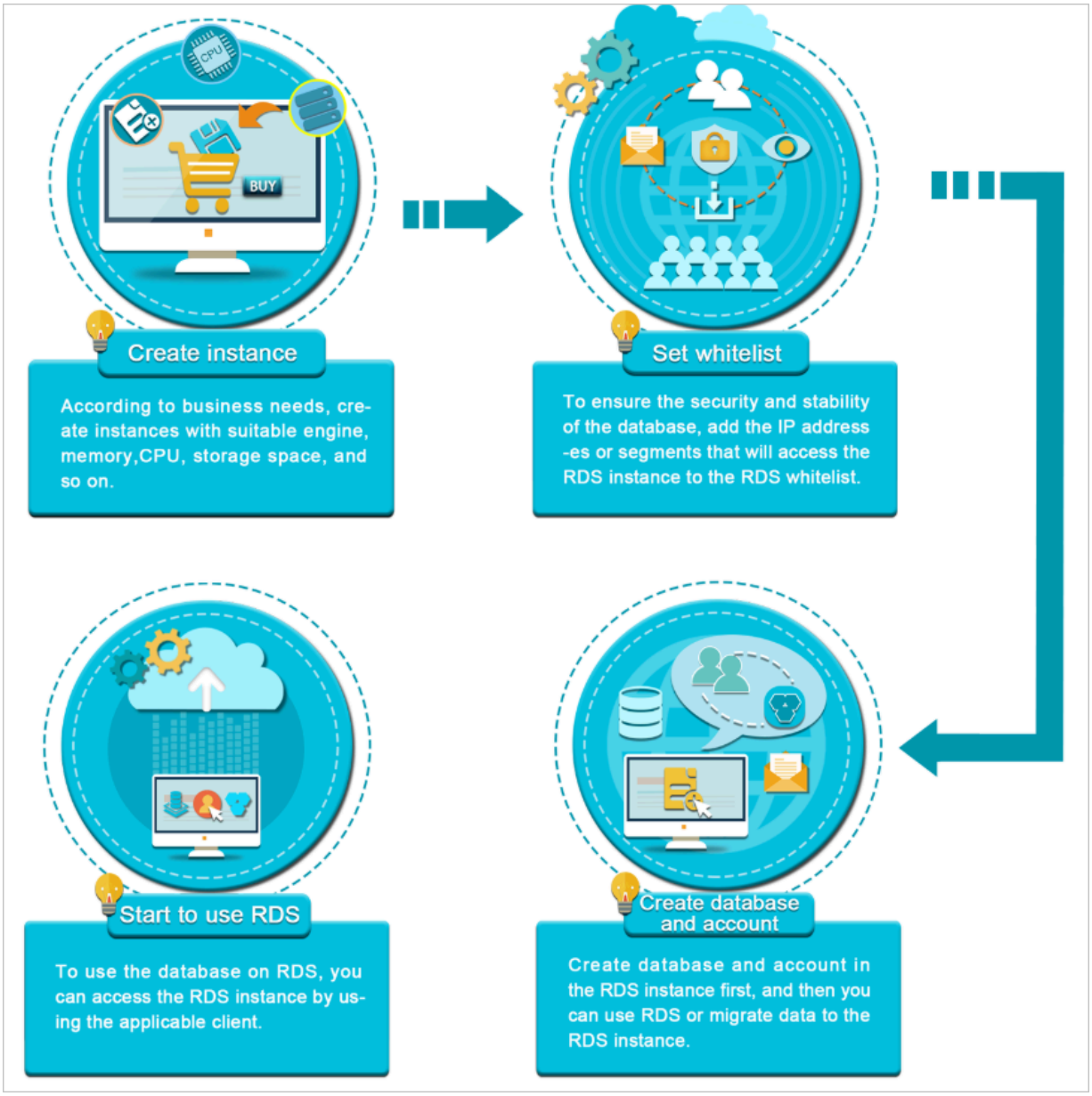
Operation	Description
ApsaraDB RDS for MySQL storage engine	Only InnoDB is supported. · For safety performance and security considerations, we recommend that you use the InnoDB storage engine. · The TokuDB engine is not supported. Percona no longer provides support for TokuDB, leading to bugs that cannot be fixed and can cause business losses in extreme cases. · The MyISAM engine is not supported. Due to the inherent shortcomings of the MyISAM engine, some data may be lost. Only some existing instances use the MylSAM engine . MyISAM engine tables in newly created instances will be automatically converted to InnoDB engine tables. · The Memory engine is not supported. Newly created Memory tables will be automatically converted into InnoDB tables.
Database replicatio n	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.
RDS instance restart	Instances must be restarted through the RDS console or API operations.
Account and database management	ApsaraDB RDS for MySQL uses the RDS console to manage accounts and databases. ApsaraDB RDS for MySQL also allows you to create a privileged account to manage users, passwords, and databases.
Standard account	· Authorization is not allowed. · The RDS console allows you to manage accounts and databases. · Instances that support standard accounts also support privileged accounts.
Privileged account	· Authorization is allowed to standard accounts. · The RDS console does not provide interfaces to manage accounts or databases. These operations can only be performed through code or DMS. · The privileged account cannot be reverted back to a standard account.

6.3.2 Procedure

ApsaraDB for RDS quick start covers the following topics: creating an RDS instance, configuring a whitelist, creating a database, creating an account, and connecting to the instance. This topic uses ApsaraDB RDS for MySQL as an example to describe how to use RDS. It provides all the necessary information to create an RDS instance.

Typically, you must complete several operations after instance creation to make it ready for use, as shown in *Figure 6-1: Quick start flowchart*.

Figure 6-1: Quick start flowchart



- *Create an instance*

  An instance is a virtualized database server on which you can create and manage multiple databases.

- *Configure a whitelist*

  After creating an RDS instance, you must configure its whitelist to allow access from external devices.

  The whitelist improves the access security of your RDS instance. We recommend that you maintain the whitelist on a regular basis. The whitelist configuration process does not affect the normal operations of the RDS instance.

- *Create a database* **and** *Create an account*

  Before using a database, you must first create the database and an account in the RDS instance.

- *Connect to an ApsaraDB RDS for MySQL instance*

  After creating an RDS instance, configuring a whitelist, and creating a database and an account, you can connect to the instance from a database client.

6.3.3 Create an instance

This topic describes how to create an instance in the ApsaraDB for RDS console.

Prerequisites

Before you create an RDS instance, you must apply for an Apsara Stack account.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **On the Instances page, click Create Instance in the upper-right corner.**
3. **Configure the following parameters.**

Category	Parameter	Description
Region	Region	The region where the RDS instance resides. Services in different regions cannot communicate over the internal network. After a region is selected, it cannot be changed.
Basic Settings	Organization	The organization to which the instance belongs.
	Resource Set	The resource set to which the instance belongs.

Category	Parameter	Description
Specifications	Instance Name	The name of the instance. · It must be 2 to 64 characters in length · It must start with a letter. · It can contain underscores (_), hyphens (-), periods (.), colons (:), and commas (,). · It cannot start with http:// or https://.
	Database Engine	The engine of the database, which varies with regions . The available database engines are displayed on the Create ApsaraDB for RDS Instance page.
	Engine Version	The version of the database engine.
	Instance Type	The type of the instance. Memory size determines the maximum number of connections and IOPS. The actual values are displayed on the console. For more information, see *Product Instruction* > *Instance Type* .
	Storage	The storage space of the instance, including the space for data, system files, binlog files, and transaction files. The minimum storage space is 50 GB. You can adjust the storage space in 5 GB increments.
Network Type	Network Type	The network type of the instance. RDS instances support the following network types: · `Classic Network`: Cloud services on 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`: A VPC helps you build an isolated network environment in 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. You can create a VPC in advance, or change the network type to VPC after creating an instance.
	IP Whitelist	You can add IP addresses to allow them to connect to the RDS instance.

4. After you configure the preceding parameters, click Submit.

6.3.4 Initialization settings

6.3.4.1 Configure a whitelist

To ensure database security and reliability, you must modify the whitelist of an ApsaraDB for RDS instance after you create the instance. You must add the IP addresses or CIDR blocks that are used for database access to the whitelist.

Context

The whitelist improves the access security of your RDS instance. We recommend that you maintain the whitelist on a regular basis. The whitelist configuration process does not affect the normal operations of the RDS instance.

Precautions

- **The default whitelist can only be modified or cleared, but cannot be deleted.**
- **You can add up to 1,000 IP addresses or CIDR blocks to a whitelist. If you want to add a large number of IP addresses, we recommend that you merge them into CIDR blocks, such as 192.168.1.0/24.**

Configure a whitelist

1. *Log on to the ApsaraDB for RDS console*.
2. **On the Instances page, click the ID of an instance in the Instance ID/Name column.**
3. **In the left-side navigation pane, click Data Security.**
4. **On the Whitelist Settings tab, click Edit corresponding to the default whitelist, as shown in the following figure.**

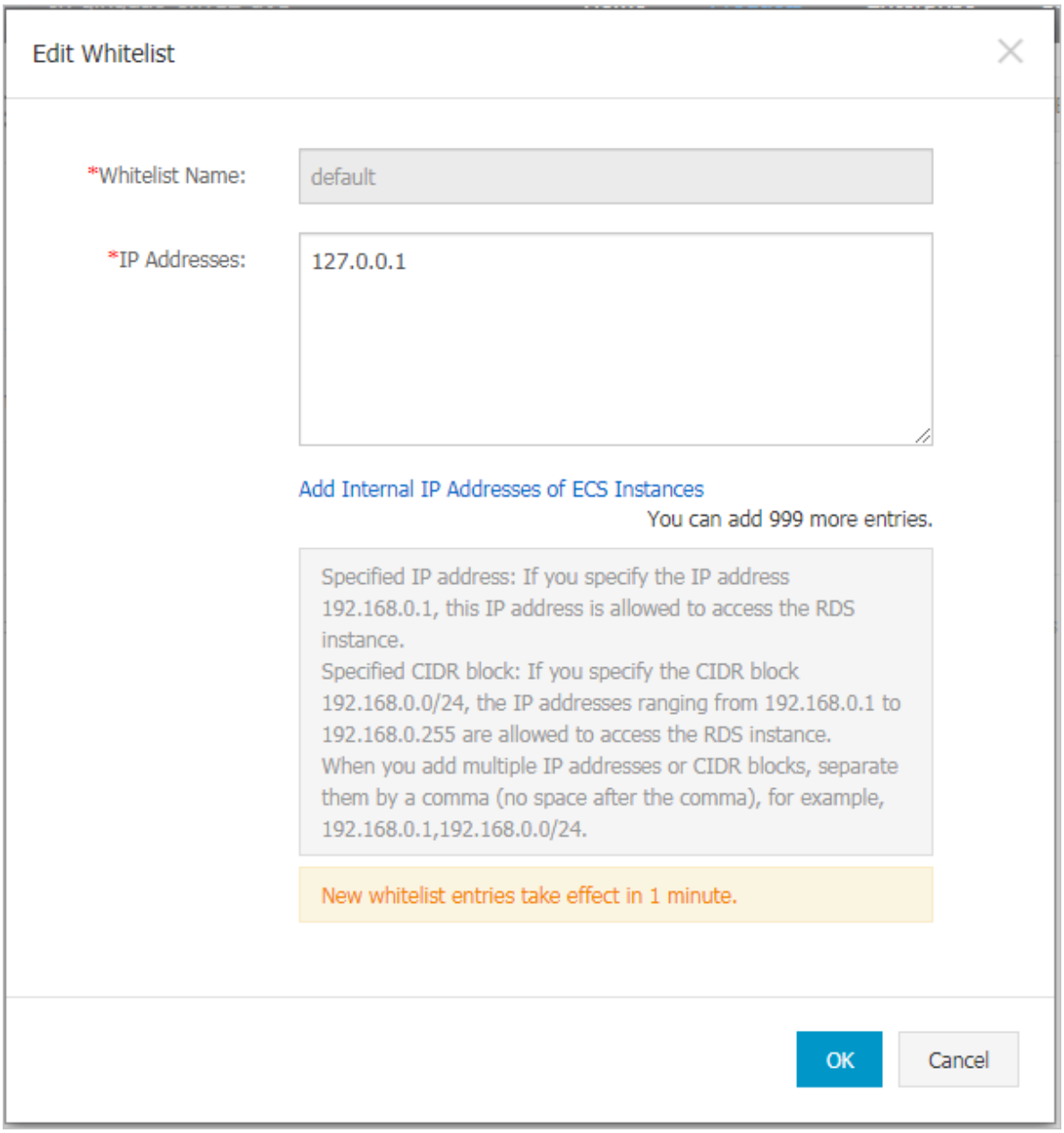
Note:

- **To connect an ECS instance to an RDS instance by using an internal endpoint , you must make sure that the two instances are in the same region and have the same network type. Otherwise, the connection fails.**
- **You can also click Create Whitelist to create a new whitelist.**

5. **In the Edit Whitelist dialog box that appears, specify the IP addresses or CIDR blocks used to access the instance, and then click OK.**
   - **If you specify the CIDR block 10.10.10.0/24, any IP addresses in the 10.10.10.X format are allowed to access the RDS instance.**
   - **If you want to add multiple IP addresses or CIDR blocks, separate each entry with a comma (without spaces), such as 192.168.0.1,172.16.213.9.**
   - **After you click Add Internal IP Addresses of ECS Instances, the IP addresses of all the ECS instances under your Apsara Stack account are displayed. You can select the required IP addresses to add to the whitelist.**

 **Note:**

If you add a new IP address or CIDR block to the default whitelist, the default address 127.0.0.1 is automatically deleted.



6.3.4.2 Create an account

After you create an ApsaraDB for RDS instance and configure its whitelist, you must create a database and an account in the instance. This topic describes how to create privileged and standard accounts.

Context

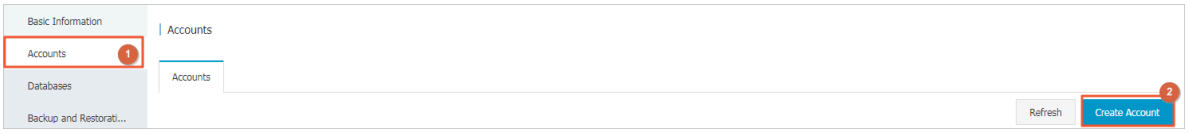
ApsaraDB RDS for MySQL supports two types of database accounts: privileged and standard. You can manage all accounts and databases in the console. For more information about the specific permissions of an account, see *Account permissions*.

Account type	Description
Privileged account	· You can only create and manage privileged accounts in the console or by using API operations. · You can create only one privileged account for each instance. You can use the privileged account to manage all standard accounts and databases. · A privileged account has more permissions, which allows you to perform more fine-grained management operations. For example , you can grant query permissions on different tables to different users. · You can use the privileged account to disconnect any standard accounts from the RDS instance.
Standard account	· You can create and manage standard accounts in the console or by using API operations or SQL statements. · You can create up to 500 databases for a standard account. · You need to manually grant specific database permissions to standard accounts. · You cannot use a standard account to create, manage, or disconnect other accounts.

Account type	Number of created databases	Number of created tables	Number of users
Privileged account	Unlimited	Less than 200, 000	Related to instance kernel parameters
Standard account	500	Less than 200, 000	Related to instance kernel parameters

Create a privileged account

1. *Log on to the ApsaraDB for RDS console*
2. Click the ID of an instance.
3. In the left-side navigation pane, click Accounts.
4. Click Create Account.



5. Configure the following parameters.

Parameter	Description
Database Account	Enter an account name. The requirements are as follows: · The name must be 2 to 16 characters in length. · It must start with a letter and end with a letter or digit. · It can contain lowercase letters, digits, and underscores (_).
Password	Enter an account password. The requirements are as follows: · The password must be 8 to 32 characters in length. · The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. · Special characters include ! @ # $ % ^ & * ( ) _ + - =
Re-enter Password	Enter the password again.

6. Click Create.

Reset the permissions of a privileged account

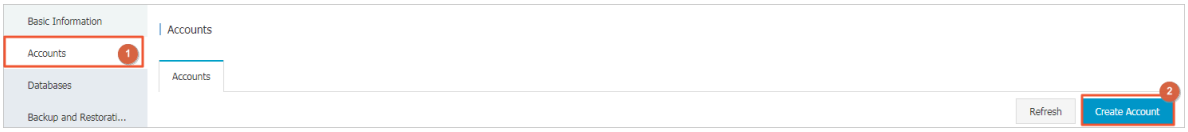
If there is a problem with the privileged account, for example, permissions are unexpectedly revoked, you can reset the permissions of the privileged account by entering the password of the privileged account.

1. *Log on to the ApsaraDB for RDS console*
2. Click the ID of an instance.
3. In the left-side navigation pane, click Accounts.
4. Click Reset Permissions in the Actions column corresponding to a privileged account.
5. Enter the password of the privileged account to reset the account permissions.

Create a standard account

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane, click Accounts.

4. Click Create Account.



5. Configure the following parameters.

Parameter	Description
Database Account	Enter an account name. The requirements are as follows: · The name must be 2 to 16 characters in length. · It must start with a letter and end with a letter or digit. · It can contain lowercase letters, digits, and underscores (_).
Authorize Databases	Grant permissions on one or more databases to the account. You do not have to configure this parameter at this time. You can authorize databases after the account is created. a. Select one or more databases from the left-side section, and click Add to add them to the right-side section. b. In the right-side section, select Read/Write or Read-only for a database. If you want to grant the same permissions on multiple databases to the account, click the button in the upper-right corner of the right-side section. The button may appear as Set All to Read/Write. Note: The button in the upper-right corner changes after you click. For example, after you click Set All to Read/Write, the button changes to Set All to Read-only.
Password	Enter an account password. The requirements are as follows: · The password must be 8 to 32 characters in length. · The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. · Special characters include ! @ # $ % ^ & * ( ) _ + - =
Re-enter Password	Enter the password again.

Parameter	Description
Description	Optional. Enter information about the account to facilitate subsequent management. The description can be up to 256 characters in length.

6. Click Create.

Account permissions

Account type	Authorization type	Permission				
Privileged account	N/A	SELECT	INSERT	UPDATE	DELETE	CREATE
		DROP	RELOAD	PROCESS	REFERENCES	INDEX
		ALTER	CREATE TEMPORARY TABLES	LOCK TABLES	EXECUTE	REPLICATION SLAVE
		REPLICATION CLIENT	CREATE VIEW	SHOW VIEW	CREATE ROUTINE	ALTER ROUTINE
		CREATE USER	EVENT	TRIGGER	N/A	N/A
Standard account	Read-only	SELECT	LOCK TABLES	SHOW VIEW	PROCESS	REPLICATION SLAVE
		REPLICATION CLIENT	N/A	N/A	N/A	N/A
	Read/Write	SELECT	INSERT	UPDATE	DELETE	CREATE
		DROP	REFERENCES	INDEX	ALTER	CREATE TEMPORARY TABLES
		LOCK TABLES	EXECUTE	CREATE VIEW	SHOW VIEW	CREATE ROUTINE
		ALTER ROUTINE	EVENT	TRIGGER	PROCESS	REPLICATION SLAVE
		REPLICATION CLIENT	N/A	N/A	N/A	N/A
	DDL only	CREATE	DROP	INDEX	ALTER	CREATE TEMPORARY TABLES

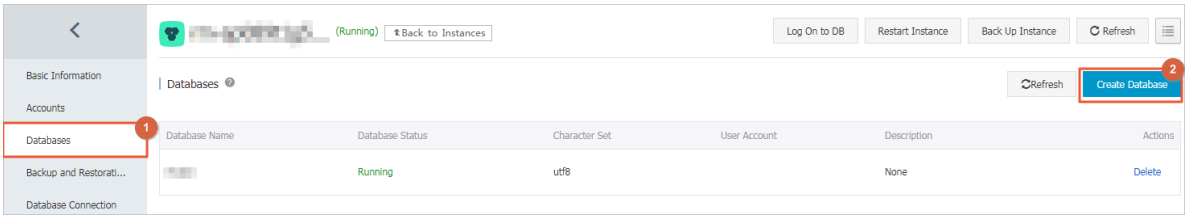
Acc type	Authorization type	Permission				
		LOCK TABLES	CREATE VIEW	SHOW VIEW	CREATE ROUTINE	ALTER ROUTINE
		PROCESS	REPLICATION SLAVE	REPLICATION CLIENT	N/A	N/A
	DML only	SELECT	INSERT	UPDATE	DELETE	CREATE TEMPORARY TABLES
		LOCK TABLES	EXECUTE	SHOW VIEW	EVENT	TRIGGER
		PROCESS	REPLICATION SLAVE	REPLICATION CLIENT	N/A	N/A

6.3.4.3 Create a database

After you create an ApsaraDB for RDS instance and configure its whitelist, you must create a database and an account in the instance.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane, click Databases.
4. Click Create Database.



5. On the Create Database page that appears, configure the following parameters.

Parameter	Description
Database Name	· The database name must be 2 to 64 characters in length. · It must start with a letter and end with a letter or digit. · It can contain lowercase letters, digits, underscores (_), and hyphens (-). · Each database name must be unique in an instance.

Parameter	Description
Supported Character Set	Select utf8, gbk, latin1, or utf8mb4. If you want to use other character sets, select all, and then select the required character set from the list.
Description	Optional. Enter information about the database to facilitate subsequent management. The description can be up to 256 characters in length.

6. Click Create.

6.3.5 Connect to an ApsaraDB RDS for MySQL instance

After you complete the initial configurations, you can use an ECS instance or a database client to connect to an ApsaraDB RDS for MySQL instance.

Context

After you perform operations such as *Create an instance*, *Configure a whitelist*, and *Create an account*, you can use a general database client or configure the endpoint, port, and account information in an application to connect to the MySQL instance.

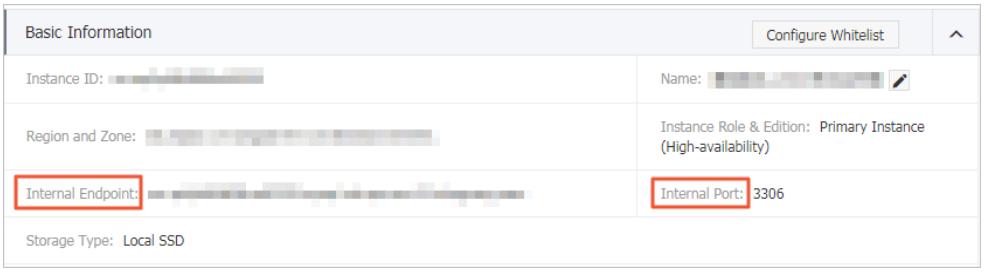
If you need to connect an ECS instance to an ApsaraDB for RDS instance, make sure that both instances are in classic networks or in the same VPC, and the IP address of the ECS instance is correctly configured in the RDS whitelist.

Connect to an instance from a client

ApsaraDB RDS for MySQL is fully compatible with MySQL. You can connect to an ApsaraDB for RDS instance from a general database client in the similar way you connect to a MySQL database. The following example uses the *HeidiSQL* client:

1. Start the HeidiSQL client.
2. In the lower-left corner, click Create.
3. Specify the information about the RDS instance you want to connect. The parameters are as follows.

Paramete	Description
Network type	The network type of the database you want to connect. Select MariaDB or MySQL (TCP/IP).

Parameter	Description
Hostname/IP	Enter the internal or public IP address of the RDS instance. · If your client is deployed in an ECS instance, and the instance is in the same region and has the same network type as the RDS instance you want to access, you can use the internal IP address. For example, if your ECS and RDS instances are both in a VPC located in the China (Hangzhou) region, you can use the internal IP address provided to create a secure connection. · Use the public IP address for other situations. To view the internal and public endpoints together with the corresponding port numbers of the RDS instance, perform the following steps: a. *Log on to the ApsaraDB for RDS console*. b. Find the instance and click its ID. c. On the Basic Information page, you can view the internal endpoint and internal port number of the instance. 
User	Enter the name of the account used to access the RDS instance.
Password	The password of the account.
Port	The port number of the RDS instance. If you connect to the instance over the internal network, enter the internal port number of the instance. If you connect to the instance over the Internet, enter the public port number of the instance.

4. Click Open. If the connection information is correct, you can connect to the instance.

6.4 Instances

6.4.1 Create an instance

This topic describes how to create an instance in the ApsaraDB for RDS console.

Prerequisites

Before you create an RDS instance, you must apply for an Apsara Stack account.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **On the Instances page, click Create Instance in the upper-right corner.**
3. **Configure the following parameters.**

Category	Parameter	Description
Region	Region	The region where the RDS instance resides. Services in different regions cannot communicate over the internal network. After a region is selected, it cannot be changed .
Basic Settings	Organization	The organization to which the instance belongs.
	Resource Set	The resource set to which the instance belongs.
Specifications	Instance Name	The name of the instance. · It must be 2 to 64 characters in length · It must start with a letter. · It can contain underscores (_), hyphens (-), periods (.), colons (:), and commas (,). · It cannot start with http:// or https://.
	Database Engine	The engine of the database, which varies with regions . The available database engines are displayed on the Create ApsaraDB for RDS Instance page.
	Engine Version	The version of the database engine.
	Instance Type	The type of the instance. Memory size determines the maximum number of connections and IOPS. The actual values are displayed on the console. For more information, see *Product Instruction* > *Instance Type* .
	Storage	The storage space of the instance, including the space for data, system files, binlog files, and transaction files. The minimum storage space is 50 GB. You can adjust the storage space in 5 GB increments.

Category	Parameter	Description
Network Type	Network Type	The network type of the instance. RDS instances support the following network types: · `Classic Network`: Cloud services on 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`: A VPC helps you build an isolated network environment in 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. You can create a VPC in advance, or change the network type to VPC after creating an instance.
	IP Whitelist	You can add IP addresses to allow them to connect to the RDS instance.

4. After you configure the preceding parameters, click Submit.

6.4.2 View basic information about an instance

You can view the details of an instance, such as its basic information, internal network connection information, running status, and configurations.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. You can use one of the following methods to access the Basic Information page of an instance:
   - On the RDS Management page, click the ID of the instance to access its Basic Information page.
   - On the RDS Management page, click Management in the Actions column corresponding to the instance to access its Basic Information page.

6.4.3 Restart an instance

If the number of connections exceeds the threshold or any performance issue occurs on an instance, you can manually restart the instance.

Context

Note:
A restart will disconnect the instance. We recommend that you make appropriate service arrangements before you restart an instance. Proceed with caution.

Procedure

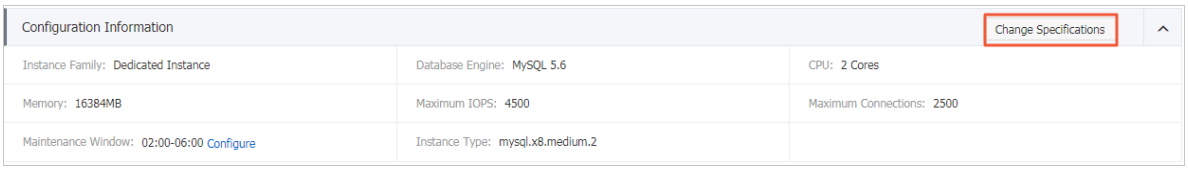
1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the upper-right corner of the page, click Restart Instance.
4. In the Restart Instance message that appears, click Confirm to restart the instance.

6.4.4 Change specifications

You can change specifications of your instance, such as the instance type and storage space, if the specifications do not meet the requirements of your application.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the Configure Information section of the Basic Information page, click Change Specifications.



4. In the Upgrade dialog box that appears, click Next.
5. On the Change Specifications page that appears, select Instance Type and Storage.
6. After you configure the preceding parameters, click Submit.

6.4.5 Set a maintenance window

You can set a maintenance window for an ApsaraDB for RDS instance as needed.

Context

To ensure the stability of ApsaraDB for RDS instances, the backend system performs maintenance of the instances at irregular intervals. The default

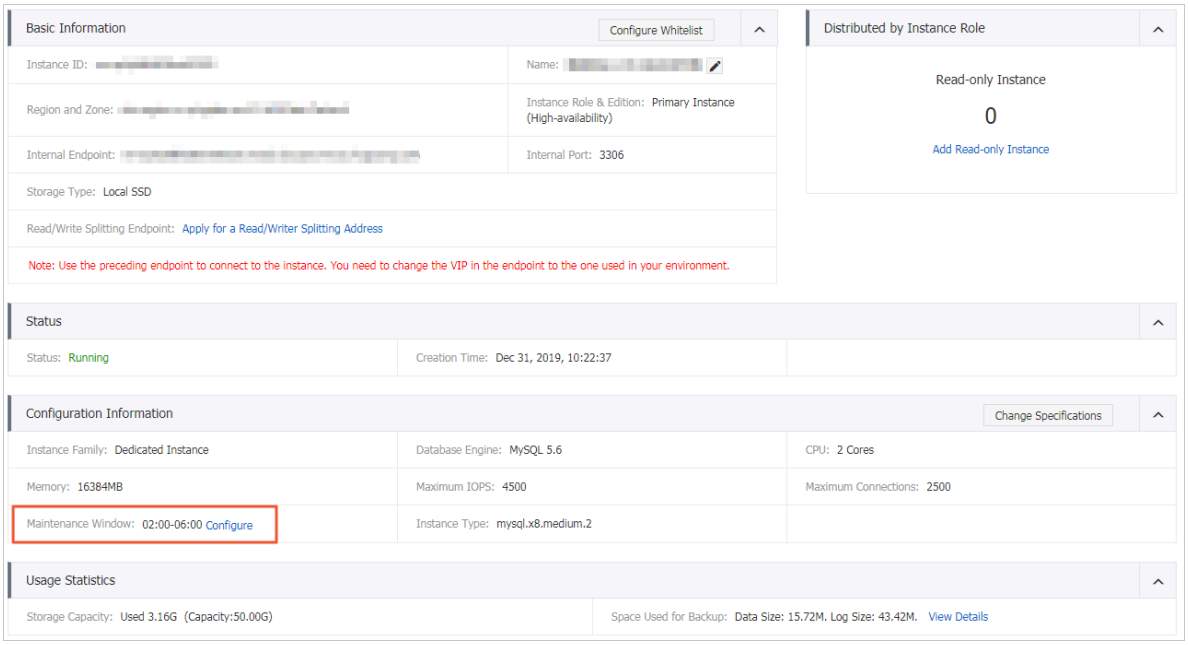
maintenance window is from 02:00 to 06:00. You can set the maintenance window to the off-peak period of your business to avoid impact on business.

Precautions

- **To ensure the stability of the maintenance process, the instance will enter the Maintaining Instance state before the maintenance time. When the instance is in this state, access to data in the database and query operations such as performance monitoring are not affected. However, apart from account and database management and IP address whitelist configuration, modification operations such as upgrade, downgrade, and restart are temporarily unavailable.**
- **During the maintenance window, the instance is disconnected once or twice. Make sure that you configure automatic reconnection policies for your applicatio ns to avoid service disruptions.**

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance or click Manage in the Actions column corresponding to the instance.**
3. **In the Configuration Information section, click Configure to the right of Maintenance Window.**



4. **Select a maintenance window and click Save.**

Note:

The maintenance window is in UTC+8.

6.4.6 Change the data replication mode

You can set the data replication mode between primary and secondary ApsaraDB for RDS instances as needed to improve database availability.

Context

You can change the data replication mode for ApsaraDB RDS for MySQL 5.6 and 5.7 High-availability Edition instances.

Note:
You can only replicate data for ApsaraDB RDS for MySQL 5.6 and 5.7 Enterprise Edition instances in Sync mode.

- **Sync**
  - **After an application-initiated update is completed on the primary instance , logs are synchronized to all secondary instances. This transaction is committed only after the majority of secondary instances have received and stored the logs.**
  - **In Sync mode, instance data replication is always synchronous and will not degrade to Async mode.**
  - **Synchronous replication is supported only when the number of instances is greater than or equal to 3. Therefore, only Enterprise Edition (formerly known as Finance Edition) instances support synchronous replication. The data replication method of Enterprise Edition instances cannot be changed.**
- **Semi-sync**

  After an application-initiated update is completed on the primary instance, logs are synchronized to all secondary instances. This transaction is considered committed after at least one secondary instance has received the logs. This way, there is no need to wait for the logs to be applied.

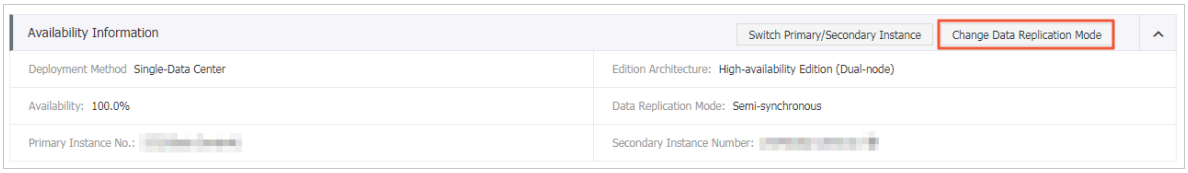
  If the secondary instances are unavailable or a network exception occurs between the primary and secondary instances, semi-synchronous replication will degrade to Async mode.

- **Async**

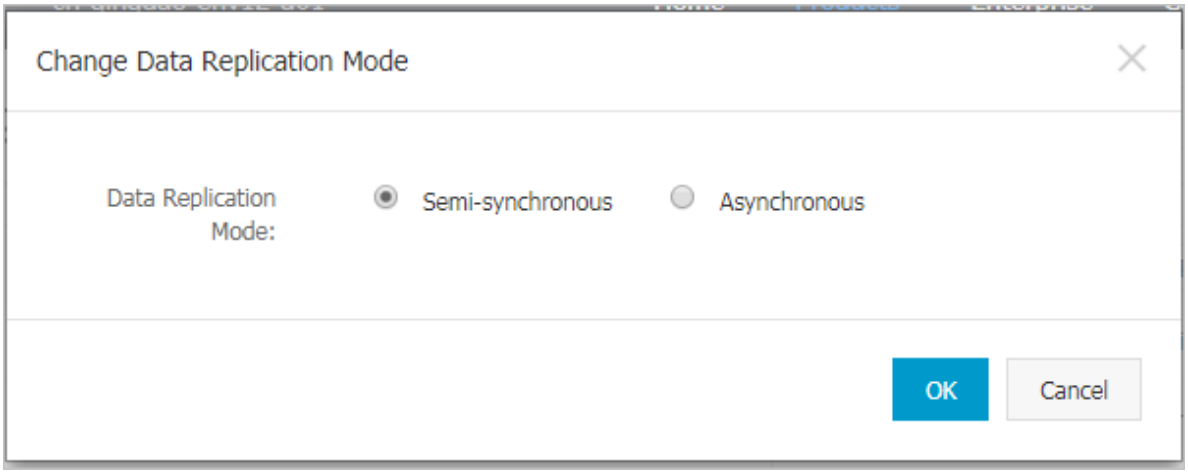
  When an application initiates an update request to add, delete, or modify data, the primary instance responds to the application immediately after completing the operation. The primary instance then replicates data to the secondary instances asynchronously. During asynchronous data replication, the unavailability of secondary instances does not affect the operations on the primary instance. Data will remain consistent even if the primary instance is unavailable.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane, click Service Availability.**
4. **Click Change Data Replication Mode.**



5. **In the dialog box that appears, select a data replication mode and click OK.**



6.4.7 Release an instance

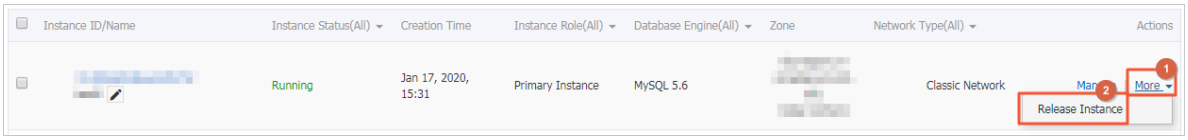
You can manually release instances as needed.

Precautions

- **You can manually release only instances that are in the running state.**
- **After an instance is released, the instance data is immediately cleared. We recommend that you back up your data before you release an instance.**

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. In the Actions column corresponding to the instance you want to release, choose More > Release Instance.



3. In the Release Instance message that appears, click Confirm.

6.5 Accounts

6.5.1 Create an account

After you create an ApsaraDB for RDS instance and configure its whitelist, you must create a database and an account in the instance. This topic describes how to create privileged and standard accounts.

Context

ApsaraDB RDS for MySQL supports two types of database accounts: privileged and standard. You can manage all accounts and databases in the console. For more information about the specific permissions of an account, see *Account permissions*.

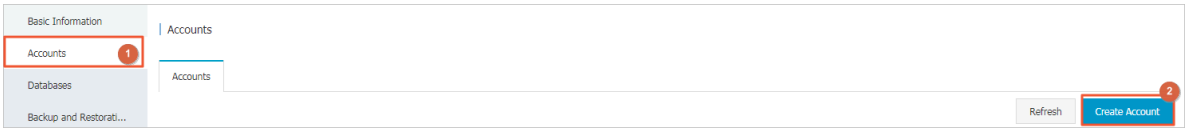
Account type	Description
Privileged account	· You can only create and manage privileged accounts in the console or by using API operations. · You can create only one privileged account for each instance. You can use the privileged account to manage all standard accounts and databases. · A privileged account has more permissions, which allows you to perform more fine-grained management operations. For example , you can grant query permissions on different tables to different users. · You can use the privileged account to disconnect any standard accounts from the RDS instance.

Account type	Description
Standard account	· You can create and manage standard accounts in the console or by using API operations or SQL statements. · You can create up to 500 databases for a standard account. · You need to manually grant specific database permissions to standard accounts. · You cannot use a standard account to create, manage, or disconnect other accounts.

Account type	Number of created databases	Number of created tables	Number of users
Privileged account	Unlimited	Less than 200, 000	Related to instance kernel parameters
Standard account	500	Less than 200, 000	Related to instance kernel parameters

Create a privileged account

1. *Log on to the ApsaraDB for RDS console*
2. Click the ID of an instance.
3. In the left-side navigation pane, click Accounts.
4. Click Create Account.



5. Configure the following parameters.

Parameter	Description
Database Account	Enter an account name. The requirements are as follows: · The name must be 2 to 16 characters in length. · It must start with a letter and end with a letter or digit. · It can contain lowercase letters, digits, and underscores (_).

Parameter	Description
Password	Enter an account password. The requirements are as follows: · The password must be 8 to 32 characters in length. · The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. · Special characters include ! @ # $ % ^ & * ( ) _ + - =
Re-enter Password	Enter the password again.

6. Click Create.

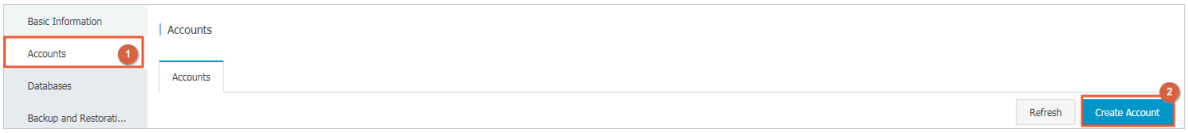
Reset the permissions of a privileged account

If there is a problem with the privileged account, for example, permissions are unexpectedly revoked, you can reset the permissions of the privileged account by entering the password of the privileged account.

1. *Log on to the ApsaraDB for RDS console*
2. Click the ID of an instance.
3. In the left-side navigation pane, click Accounts.
4. Click Reset Permissions in the Actions column corresponding to a privileged account.
5. Enter the password of the privileged account to reset the account permissions.

Create a standard account

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane, click Accounts.
4. Click Create Account.



5. Configure the following parameters.

Parameter	Description
Database Account	Enter an account name. The requirements are as follows: · The name must be 2 to 16 characters in length. · It must start with a letter and end with a letter or digit. · It can contain lowercase letters, digits, and underscores (_).
Authorize Databases	Grant permissions on one or more databases to the account. You do not have to configure this parameter at this time. You can authorize databases after the account is created. a. Select one or more databases from the left-side section, and click Add to add them to the right-side section. b. In the right-side section, select Read/Write or Read-only for a database. If you want to grant the same permissions on multiple databases to the account, click the button in the upper-right corner of the right-side section. The button may appear as Set All to Read/Write. Note: The button in the upper-right corner changes after you click. For example, after you click Set All to Read/Write, the button changes to Set All to Read-only.
Password	Enter an account password. The requirements are as follows: · The password must be 8 to 32 characters in length. · The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. · Special characters include ! @ # $ % ^ & * ( ) _ + - =
Re-enter Password	Enter the password again.
Description	Optional. Enter information about the account to facilitate subsequent management. The description can be up to 256 characters in length.

6. Click Create.

Account permissions

Acc type	Authorization type	Permission				
Privileged account	N/A	SELECT	INSERT	UPDATE	DELETE	CREATE
		DROP	RELOAD	PROCESS	REFERENCES	INDEX
		ALTER	CREATE TEMPORARY TABLES	LOCK TABLES	EXECUTE	REPLICATION SLAVE
		REPLICATION CLIENT	CREATE VIEW	SHOW VIEW	CREATE ROUTINE	ALTER ROUTINE
		CREATE USER	EVENT	TRIGGER	N/A	N/A
Standard account	Read-only	SELECT	LOCK TABLES	SHOW VIEW	PROCESS	REPLICATION SLAVE
		REPLICATION CLIENT	N/A	N/A	N/A	N/A
	Read/Write	SELECT	INSERT	UPDATE	DELETE	CREATE
		DROP	REFERENCES	INDEX	ALTER	CREATE TEMPORARY TABLES
		LOCK TABLES	EXECUTE	CREATE VIEW	SHOW VIEW	CREATE ROUTINE
		ALTER ROUTINE	EVENT	TRIGGER	PROCESS	REPLICATION SLAVE
		REPLICATION CLIENT	N/A	N/A	N/A	N/A
	DDL only	CREATE	DROP	INDEX	ALTER	CREATE TEMPORARY TABLES
		LOCK TABLES	CREATE VIEW	SHOW VIEW	CREATE ROUTINE	ALTER ROUTINE
		PROCESS	REPLICATION SLAVE	REPLICATION CLIENT	N/A	N/A

Acc type	Authorization type	Permission				
	DML only	SELECT	INSERT	UPDATE	DELETE	CREATE TEMPORARY TABLES
		LOCK TABLES	EXECUTE	SHOW VIEW	EVENT	TRIGGER
		PROCESS	REPLICATION SLAVE	REPLICATION CLIENT	N/A	N/A

6.5.2 Reset your password

You can use the ApsaraDB for RDS console to reset the password of your database account.

Context

Notice:
To ensure data security, we recommend that you change your password on a regular basis.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane, click Accounts.**
4. **Find the target account and click Reset Password in the Actions column.**
5. **In the dialog box that appears, enter and confirm the new password, and then click OK.**

Note:
The password must meet the following requirements:

- **The password must be 8 to 32 characters in length.**
- **The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters.**
- **Special characters include ! @ # $ % ^ & * ( ) _ + - =**

6.5.3 Edit account permissions

You can edit the account permissions of your ApsaraDB for RDS instances at any time.

Prerequisites

You can edit the permissions of a standard account as needed. The permissions of privileged accounts can only be reset to the default settings and cannot be changed to a specific set of permissions.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane, click Accounts.**
4. **Find the target account and click Edit Permissions in the Actions column.**
5. **Configure the following parameters.**

Parameter	Description
Authorize Databases	Select a database and click Add or Remove.
Authorized Databases	In the Authorized Databases list, you can set the account permissions to Read/Write or Read-only. You can also click Set All to Read/Write or Set All to Read-only to set the permissions of the account on all authorized databases. · Read-only: grants the database read-only permissions to the account. · Read/Write: grants the database read/write permissions to the account.

6. **Click OK.**

6.5.4 Delete an account

You can delete a database account from the ApsaraDB for RDS console.

Prerequisites

You can use the console to delete privileged and standard accounts that are no longer used.

Procedure

1. *Log on to the ApsaraDB for RDS console*.

2. Click the ID of an instance.
3. In the left-side navigation pane, click Accounts.
4. Find the account you want to delete and click Delete in the Actions column.
5. In the message that appears, click Confirm.

> **Note:**
> Accounts in the processing state cannot be deleted.

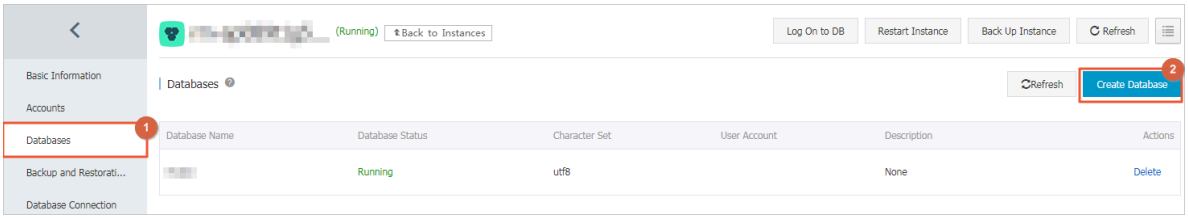
6.6 Databases

6.6.1 Create a database

After you create an ApsaraDB for RDS instance and configure its whitelist, you must create a database and an account in the instance.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane, click Databases.
4. Click Create Database.



5. On the Create Database page that appears, configure the following parameters.

Parameter	Description
Database Name	· The database name must be 2 to 64 characters in length. · It must start with a letter and end with a letter or digit. · It can contain lowercase letters, digits, underscores (_), and hyphens (-). · Each database name must be unique in an instance.
Supported Character Set	Select utf8, gbk, latin1, or utf8mb4. If you want to use other character sets, select all, and then select the required character set from the list.

Parameter	Description
Description	Optional. Enter information about the database to facilitate subsequent management. The description can be up to 256 characters in length.

6. Click Create.

6.6.2 Delete a database

You can delete out-of-date databases in the ApsaraDB for RDS console.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane, click Databases.
4. Find the database you want to delete and click Delete in the Actions column.
5. In the message that appears, click Confirm.

6.7 Database connection

6.7.1 Change the endpoint of an instance

This topic describes how to change the endpoint of an instance.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane, click Database Connection.
4. Click Change Endpoint.
5. In the dialog box that appears, set Connection Type, Endpoint, and Port, and click OK.

Note:

- The prefix of the endpoint must be 8 to 64 characters in length and can contain letters, digits, and hyphens (-). It must start with a lowercase letter.
- The port number must be in the range of 1000 to 65535.

6.7.2 Switch the 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.**

Context

`Standard Mode` **and** `Safe Mode` **have the following differences:**

- **Standard Mode: ApsaraDB for RDS uses SLB to eliminate the impact of high -availability switching between database engines on the application layer. This mode reduces 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 intercepts SQL injection attacks based on semantic analysis. However, it increases the response time by more than 20%.**

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane, click Database Connection.**
4. **Click Switch Access 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 changed.**

6.8 Monitoring and alerts

6.8.1 View resource and engine monitoring data

The ApsaraDB for RDS console provides a variety of performance metrics for you to monitor the status of your instances.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**

3. In the left-side navigation pane, click Monitoring and Alerts.
4. On the Monitoring and Alerts page, select Resource Monitoring or Engine Monitoring, and select a time range to view the corresponding monitoring data. The specific metrics are described as follows.

Category	Metric	Description
Resource Monitoring	Disk Space (MB)	The disk space usage of the instance, including: · Instance size · Data usage · Log size · Temporary file size · Other system file size Unit: MB.
	IOPS (Input/ Output Operations per Second)	The number of I/O requests per second for the instance.
	Total Connections	The total number of connections to the instance, including the number of active connections and the total number of connections.
	CPU and Memory Usage (%)	The CPU and memory usage of the instance (excluding the CPU and memory usage for the operating system).
	Network Traffic (KB)	The inbound and outbound traffic of the instance per second. Unit: KB.
Engine Monitoring	TPS (Transactions per Second)/ QPS (Queries per Second)	The average number of transactions per second and the average number of SQL statements executed per second.
	InnoDB Buffer Pool Read Hit Ratio, Usage Ratio , and Dirty Block Ratio (%)	The read hit ratio, usage ratio, and dirty block ratio of the InnoDB buffer pool.
	InnoDB Read/ Write Volume (KB)	The amount of data that InnoDB reads and writes per second. Unit: KB.
	InnoDB Buffer Pool Read/Write Frequency	The number of read and write operations that InnoDB performs per second.

Category	Metric	Description
	InnoDB Log Read/ Write/fsync	The average frequency of physical writes to log files per second by InnoDB, the log write request frequency, and the average frequency of fsync writes to log files.
	Number of Temporary Tables Created Automatically on the Hard Disk when MySQL Statements Are Being Executed	The number of temporary tables that are automatically created on the hard disk when the database executes SQL statements.
	MySQL_COMDML	The number of SQL statements that the database executes per second. The SQL statements include: · Insert · Delete · Insert_Select · Replace · Replace_Select · Select · Update
	MySQL_RowDML	The number of operations that InnoDB performs per second, including: · The average number of physical writes to log files per second · The number of rows that are read, updated, deleted, and inserted from InnoDB tables per second

6.8.2 Set a monitoring frequency

The ApsaraDB for RDS console provides a variety of performance metrics for which you can set a monitoring frequency.

Context

ApsaraDB for RDS provides the following monitoring frequencies:

- Every 5 seconds for the first seven days. After seven days, performance metrics are monitored every minute.

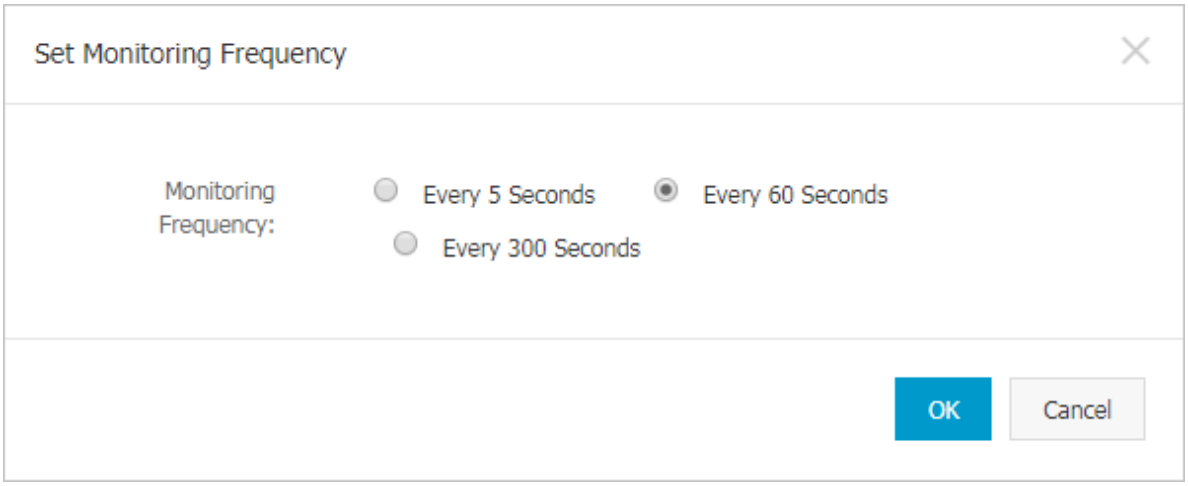
- Every 60 seconds for 30 days.
- Every 300 seconds for 30 days.

The following table lists the monitoring configuration policies in detail.

Instance type	Every 5 seconds	Every 60 seconds	Every 300 seconds
Basic Edition	Not supported	Supported for free	Default configuration
High-availability Edition and Enterprise Edition: less than 8 GB memory	Not supported	Supported for free	Default configuration
High-availability Edition and Enterprise Edition: at least 8 GB memory	Not supported	Default configuration	Supported for free

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane, click Monitoring and Alerts.
4. On the Resource Monitoring tab, click Set Monitoring Frequency.
5. In the Set Monitoring Frequency dialog box, select the monitoring frequency you want.



6. Click OK.

Note:

If your instance does not support the selected monitoring frequency, a prompt is displayed in the Set Monitoring Frequency dialog box. Select a monitoring frequency supported by the instance as prompted.

6.9 Data security

6.9.1 Configure a whitelist

To ensure database security and reliability, you must modify the whitelist of an ApsaraDB for RDS instance after you create the instance. You must add the IP addresses or CIDR blocks that are used for database access to the whitelist.

Context

The whitelist improves the access security of your RDS instance. We recommend that you maintain the whitelist on a regular basis. The whitelist configuration process does not affect the normal operations of the RDS instance.

Precautions

- **The default whitelist can only be modified or cleared, but cannot be deleted.**
- **You can add up to 1,000 IP addresses or CIDR blocks to a whitelist. If you want to add a large number of IP addresses, we recommend that you merge them into CIDR blocks, such as 192.168.1.0/24.**

Configure a whitelist

1. *Log on to the ApsaraDB for RDS console*.
2. **On the Instances page, click the ID of an instance in the Instance ID/Name column.**
3. **In the left-side navigation pane, click Data Security.**
4. **On the Whitelist Settings tab, click Edit corresponding to the default whitelist, as shown in the following figure.**

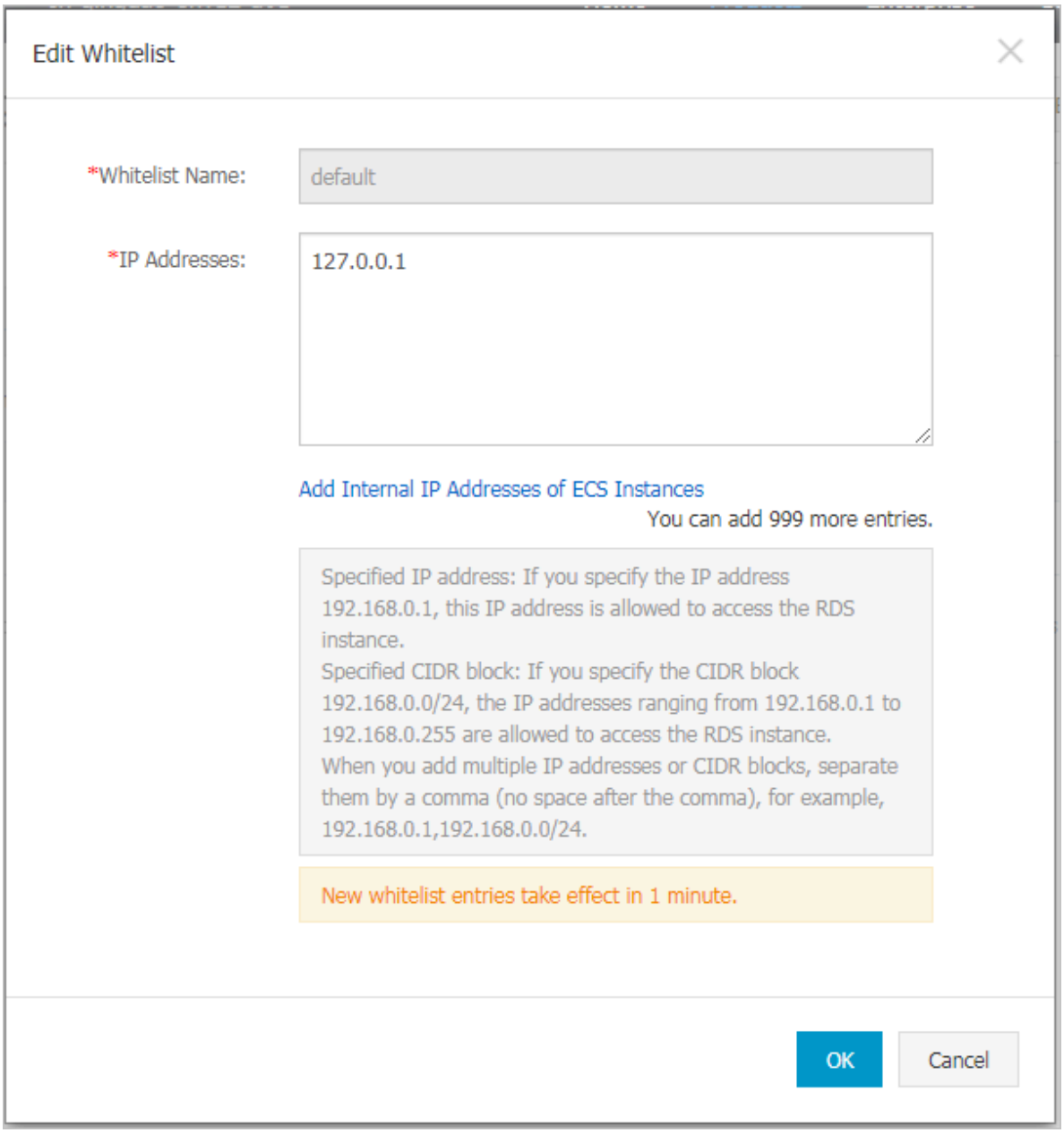
Note:

- To connect an ECS instance to an RDS instance by using an internal endpoint , you must make sure that the two instances are in the same region and have the same network type. Otherwise, the connection fails.
- You can also click Create Whitelist to create a new whitelist.

5. In the Edit Whitelist dialog box that appears, specify the IP addresses or CIDR blocks used to access the instance, and then click OK.
   - If you specify the CIDR block 10.10.10.0/24, any IP addresses in the 10.10.10.X format are allowed to access the RDS instance.
   - If you want to add multiple IP addresses or CIDR blocks, separate each entry with a comma (without spaces), such as 192.168.0.1,172.16.213.9.
   - After you click Add Internal IP Addresses of ECS Instances, the IP addresses of all the ECS instances under your Apsara Stack account are displayed. You can select the required IP addresses to add to the whitelist.

 **Note:**

If you add a new IP address or CIDR block to the default whitelist, the default address 127.0.0.1 is automatically deleted.



6.9.2 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. SSL is used on the transport layer to encrypt network connections. SSL not only increases the

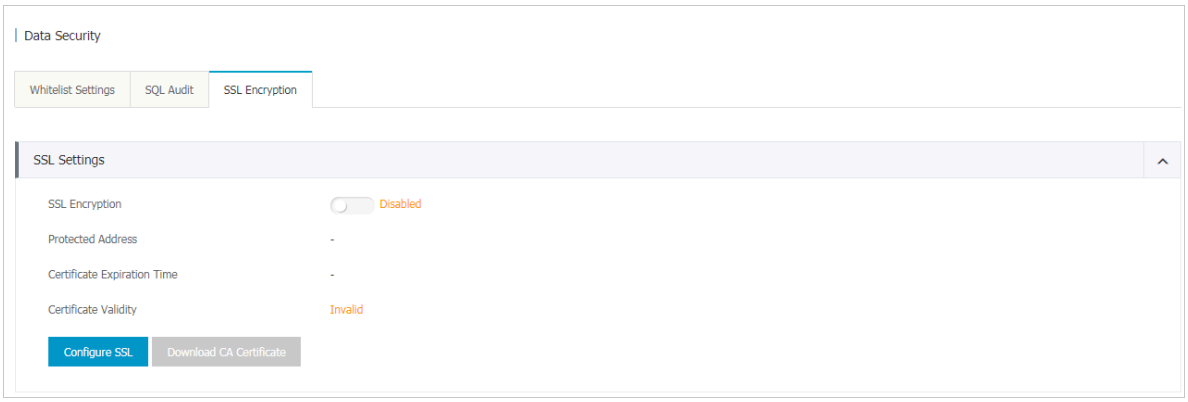
security and integrity of communication data, but also increases the response time for network connection.

Precautions

- **The validity period of an SSL CA certificate is one year. You must renew the validity period of the SSL CA certificate in your application or client within one year. Otherwise, your application or client that uses an encrypted network connection cannot connect to RDS properly.**
- **Due to the inherent defects of SSL encryption, this feature significantly increases the CPU utilization. We recommend that you enable SSL encryption only when external endpoints need to be encrypted. Typically, internal endpoints do not require SSL encryption.**
- **Read/write splitting endpoints do not support SSL encryption.**
- **Disabling SSL encryption will cause the instance to restart. Proceed with caution .**

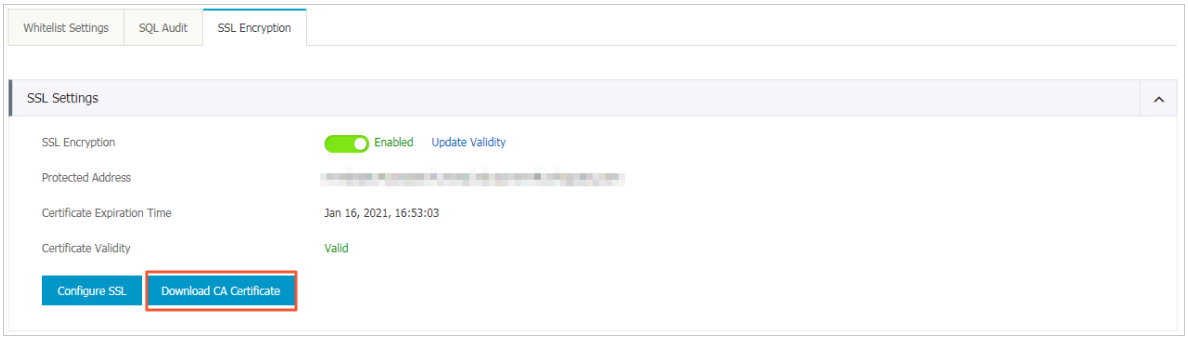
Enable SSL encryption

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane, click Data Security.**
4. **Click the SSL Encryption tab.**



5. **In the SSL Settings section, turn on SSL Encryption.**
6. **In the Configure SSL dialog box that appears, select the endpoint for which you want to enable SSL encryption and click OK.**

7. Click Download CA Certificate to download the SSL CA certificate files in a compressed package.



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 must modify the default JDK security configuration in JDK7 and JDK8. Open the `/jre/lib/security/java.security` file on the host where the database that needs SSL access resides, and modify the following configurations:
>
> ```
> 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. Typically, other similar errors are also caused by Java security configurations.
>
> ```
> javax.net.ssl.SSLHandshakeException: DHPublicKey does not comply
> to algorithm constraints
> ```

Configure SSL CA certificates

After the SSL encryption feature is enabled, configure the SSL CA certificate for your application or client so that the application or client can connect to the ApsaraDB for RDS instance. This section uses MySQL Workbench and Navicat as an example to describe how to install the SSL CA certificate. For other applications or clients, see the instructions for the corresponding product.

Configuration on MySQL Workbench

1. **Start MySQL Workbench.**
2. **Choose Database > Manage Connections.**
3. **Enable Use SSL and import the SSL CA certificate.**

Configuration on Navicat

1. **Start Navicat.**
2. **Right-click the target database and choose Edit Connection from the shortcut menu.**
3. **Click the SSL tab. Select the path of the PEM-formatted CA certificate.**
4. **Click OK.**

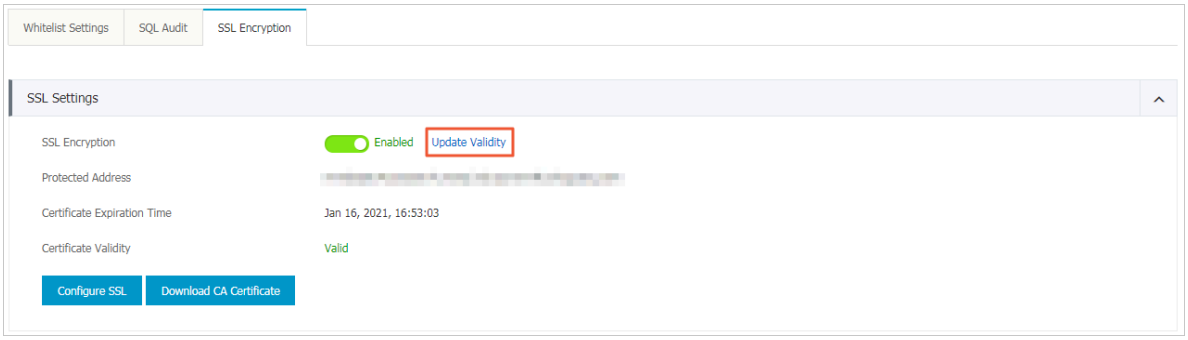
> **Note:**
> **If an error of `connection is being used` is displayed, the previous session is not disconnected. Restart Navicat.**

5. **Double-click the target database to test whether the database is connected.**

Update the validity period of certificates

> **Note:**
> **Updating the validity period will cause the instance to restart. Proceed with caution.**

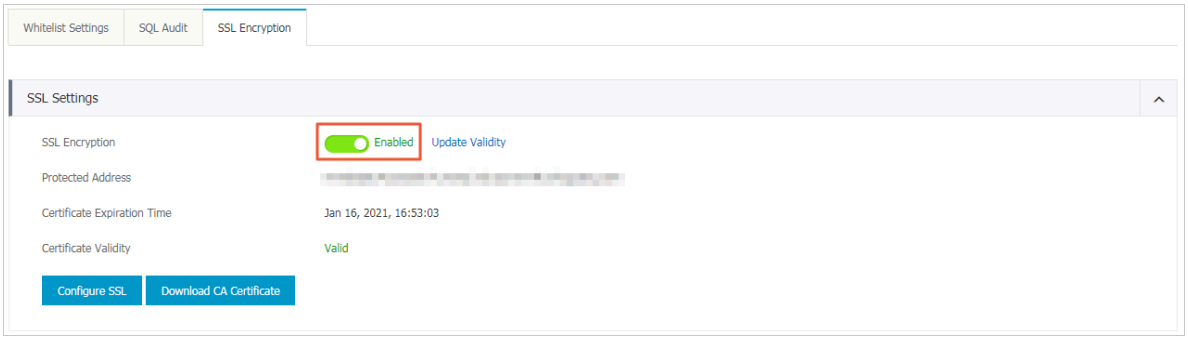
1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane, click Data Security.**
4. **Click the SSL Encryption tab.**
5. **Click Update Validity.**



Disable SSL encryption

> **Note:**
>
> - **Disabling SSL encryption will cause the RDS instance to restart. To reduce impact on your business, the system triggers primary/secondary switchover. We recommend that you disable SSL encryption during off-peak hours.**
> - **Database access features higher performance but lower security after SSL encryption is disabled. We recommend that you disable SSL encryption only in secure environments.**

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane, click Data Security.**
4. **Click the SSL Encryption tab.**
5. **Turn off SSL Encryption. In the message that appears, click OK.**



6.9.3 SQL audit

You can use the SQL audit feature to audit SQL executions and check the details. SQL audit does not affect instance performance.

Context

> **Note:**
>
> **You cannot view the logs that are generated before you enable SQL audit.**

You can view the incremental data of your ApsaraDB RDS for MySQL instance by using SQL audit logs or binlogs. However, these two methods differ in the following aspects:

- **SQL audit logs are similar to audit logs in MySQL and record all DML and DDL operations by using network protocol analysis. SQL audit does not parse the**

actual parameter values. Therefore, a small amount of information may be lost if a large number of SQL statements are executed to query data. The incremental data obtained by using this method may be inaccurate.

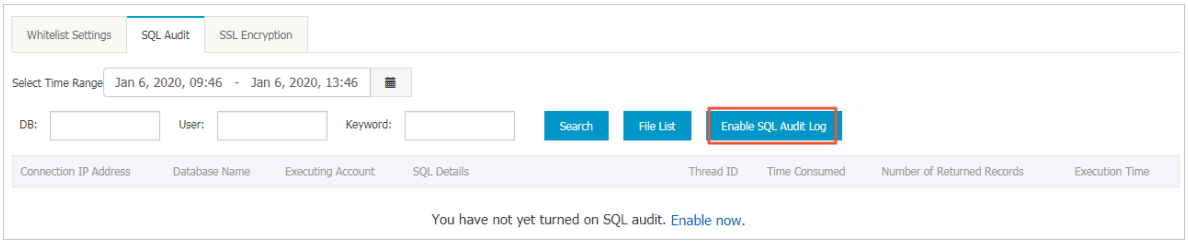
- Binlogs record all add, delete, and modify operations and the incremental data used for data restoration. Binlogs are temporarily stored in your ApsaraDB for RDS instance after they are generated. The system transfers full binlog files to OSS on a regular basis. OSS then stores the files for seven days. However, a binlog file cannot be transferred if data is being written to it. Therefore, you may find that some binlog files fail to be uploaded to OSS after you click Upload Binlogs on the Backup and Restoration page. Binary logs are not generated in real time, but you can obtain accurate incremental data from them.

Precautions

- SQL audit is disabled by default. SQL audit does not affect instance performance.
- SQL audit logs are retained for 30 days.
- Log files exported from SQL audit are retained for two days. The system clears files that are retained for more than two days.

Enable SQL audit

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane, click Data Security.
4. Click the SQL Audit tab.



5. Click Enable SQL Audit Log.
6. In the message that appears, click Confirm.

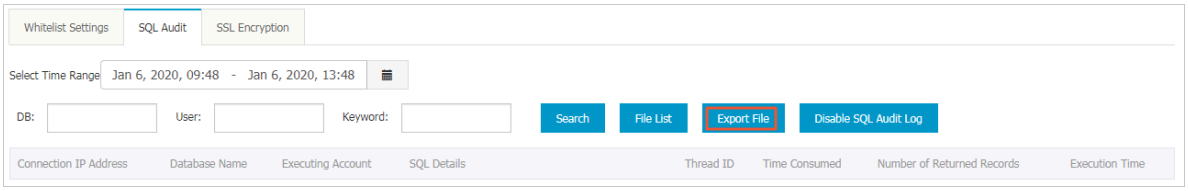
   After enabling SQL audit, you can query SQL information based on conditions such as the time, database, user, and keyword.

Disable SQL audit

> **Note:**
> **If SQL audit is disabled, all the SQL audit logs are cleared. We recommend that you export and store the audit logs locally before you disable SQL audit.**

You can disable SQL audit when you do not need it to avoid charges. To disable SQL audit, follow these steps:

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane, click Data Security.**
4. **Click the SQL Audit tab.**



5. **Click Export File to export and store the SQL audit content locally.**
6. **After the file is exported, click Disable SQL Audit Log.**
7. **In the message that appears, click Confirm.**

6.10 Database backup and restoration

6.10.1 Automatic backup

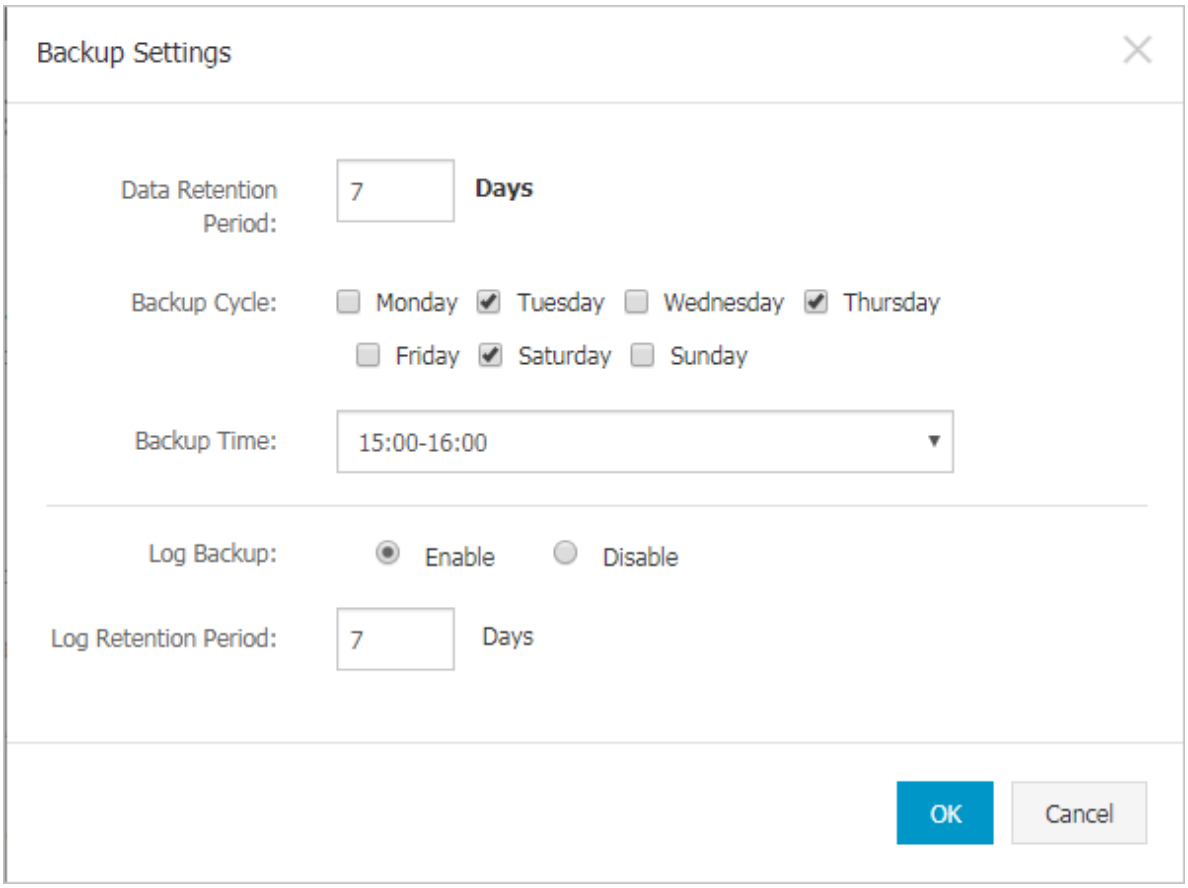
RDS automatic backup supports full physical backups. ApsaraDB for RDS automatically backs up data based on pre-configured policies. This topic describes how to configure a policy for automatic backup.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **In the left-side navigation pane, click Backup and Restoration.**
4. **Click the Backup Settings tab.**

5. Click Edit.

> **Note:**
> To ensure data security, the system compares the new backup cycle and time with the original settings, and selects the most recent time point to back up the data. Therefore, the next backup may still be performed based on the original backup cycle and time. For example, if the backup time is set to 19:00-20:00 every Wednesday and you modify the time to 19:00-20:00 every Thursday before 19:00 of this Wednesday, the system will still back up data during 19:00-20:00 this Wednesday.



6. Configure the following parameters.

Parameter	Description
Data Retention Period (Days)	The number of days that data backup files are retained. Valid values: 7 to 730. Default value: 7.
Backup Cycle	One or multiple days in a week.
Backup Time	Any period of time within a day. Unit: hours. We recommend that you back up data during off-peak hours.

Parameter	Description
Log Backup	Specifies whether to enable log backup. **Notice:** If you disable log backup, all the log files are deleted, and you cannot restore data to a saved point in time.
Log Retention Period (Days)	The number of days that log files are retained. Valid values: 7 to 730. Default value: 7.

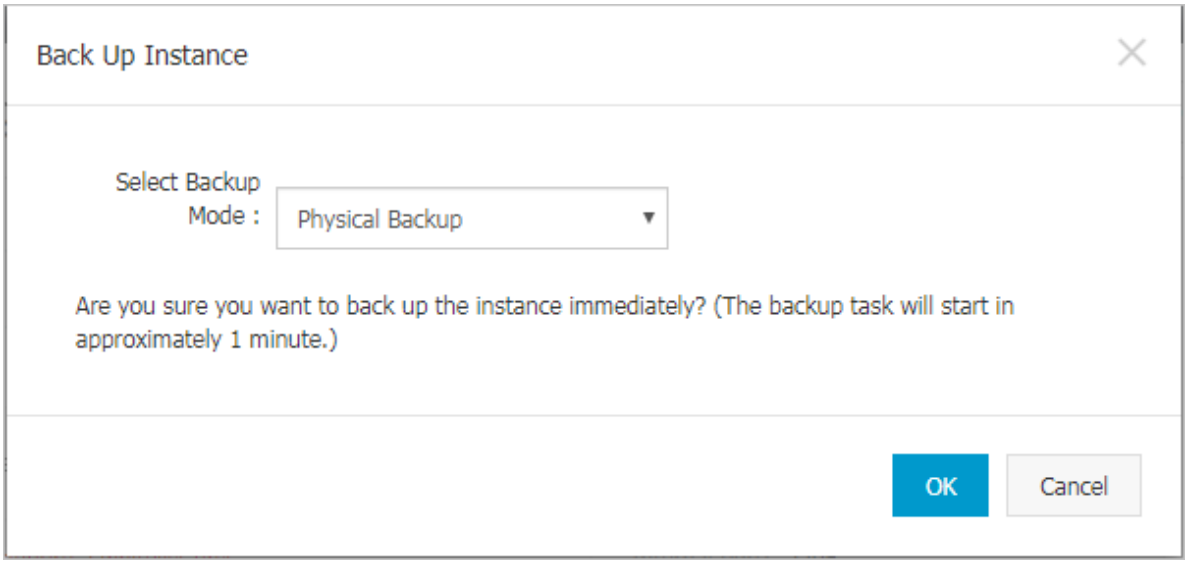
7. After you configure the preceding parameters, click OK.

6.10.2 Manual backup

Manual backup supports both full physical backups and full logical backups. This topic describes how to manually back up ApsaraDB for RDS data.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the upper-right corner, click Back Up Instance.



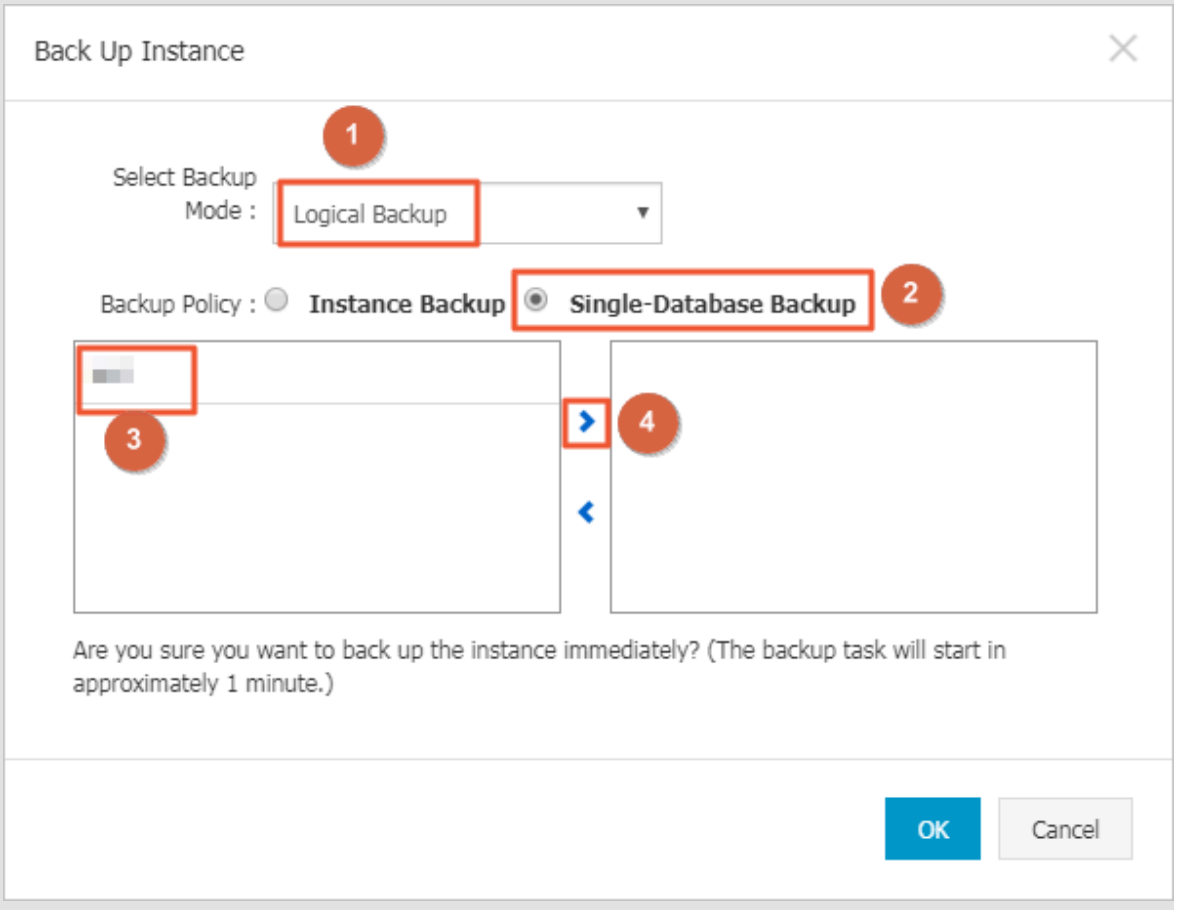
4. Set the backup mode and backup policy, and click OK.

Parameter	Value	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.

Parameter	Value	Description
	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.

Note:

If you choose Logical Backup > Single-Database Backup, select the database to back up on the left, click > to add the database to the list on the right, and then click OK.



6.10.3 Restore data to a new instance (formerly known as cloning 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

The following requirements must be met:

- The primary instance is in the running state.
- The primary instance does not have an ongoing migration task.
- Data backup and log backup are enabled.
- The primary instance has at least one completed backup set before you clone the instance by 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 copies only the data of the primary instance, but not the data of read-only instances. The copied data includes database information, account information, and instance settings such as whitelist settings, backup settings, parameter settings, and alert threshold settings.
- The database engine of a cloned instance must be the same as that of the primary instance. Other settings, such as the instance edition, zone, network type, instance type, and storage space, can be different. If you want to clone an instance to restore the data of a primary instance, we recommend that you select an instance type that has higher specifications and more storage space than that of the primary instance to speed up the data restoration process.
- 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 changed.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. Click the ID of an instance.
3. In the left-side navigation pane, click Backup and Restoration.
4. On the Data Backup tab, find the backup set you want to restore and click Restore in the Actions column.
5. In the dialog box that appears, select Restore Database and click OK.

6. On the Restore RDS Instance page, configure the following parameters.

Category	Parameter	Description
Region	Region	The region where the ApsaraDB for RDS instance resides.
Database Restoration	Restore Mode	The data restore mode of the primary instance. Valid values: · By Time · By Backup Set
	Time	Select the point in time to which you want to restore the database. Note: When Restore Mode is set to By Time, you must specify this parameter.
	Backup Set	Select the backup set for restoration. Note: When Restore Mode is set to By Backup Set, you must specify this parameter.
Specifications	Database Engine	The engine of the database, which cannot be modified.
	Engine Version	The version of the database engine, which cannot be modified.
	Instance Type	The type of the cloned instance. Note: We recommend that you select the instance type and storage space that are higher than those of the primary instance. Otherwise, the data restoration may take a long time due to performance limitations.

Category	Parameter	Description
	Storage	The storage space of the instance, including the space for data, system files, binlog files, and transaction files. You can adjust the storage space in 5 GB increments. **Note:** For dedicated instances with local SSDs, the storage space is associated with the instance type.
Network Type	Network Type	The network type of the instance. RDS instances support the following network types: • `Classic Network`: Cloud services on 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`: A VPC helps you build an isolated network environment in 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.

7. After you configure the preceding parameters, click Submit.

6.11 Read-only instances

6.11.1 Overview

ApsaraDB RDS for MySQL allows you to create read-only instances. In scenarios where RDS has a small number of write requests but a large number of read requests, you can create read-only instances to distribute database access loads away from the primary instance. This topic describes the features and limits of read-only instances.

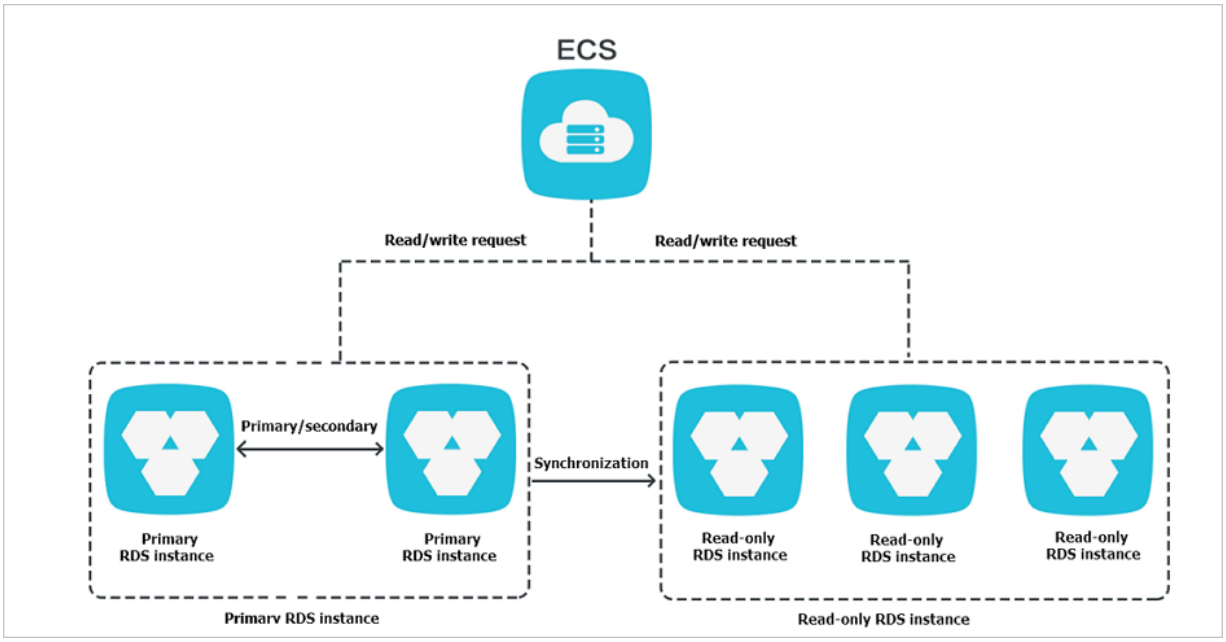
To scale the reading capability and distribute database access loads, you can create one or more read-only instances in a region. Read-only instances allow RDS to increase the application throughput when a large amount of data is being read.

Note:

Currently, only the following MySQL instance engines support read-only instances:

- **MySQL 5.7 High-availability Edition (based on local SSDs)**
- **MySQL 5.6**

A read-only instance with a single physical node and no backup node uses the native replication capability of MySQL to synchronize changes from the primary instance to all its read-only instances. Real-only instances must be in the same region as the primary instance but do not have to be in the same zone as the primary instance. The following figure shows the topology of read-only instances.



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.**
- **Read-only instances do not require account or database maintenance. Account and database information is synchronized from the primary instance.**
- **The whitelists of read-only instances can be configured independently.**
- **System performance monitoring is provided.**

  RDS provides up to 20 system performance monitoring views, including those for disk capacity, IOPS, connections, CPU utilization, and network traffic. You can view the load of instances easily.

- **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.**

6.11.2 Create a read-only instance

You can create read-only instances of different specifications based on your business requirements.

Precautions

- **A maximum of five read-only instances can be created for a primary instance.**
- **Backup settings and temporary backup are not supported.**
- **Instance restoration is not supported.**
- **Data migration to read-only instances is not supported.**
- **Database creation and deletion are not supported.**
- **Account creation, deletion, authorization, and password changes are not supported.**
- **After a read-only instance is created, you cannot restore data by directly overwriting the primary instance with a backup set.**

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **On the Basic Information page, click Add Read-only Instance on the right.**
4. **On the Create Read-only RDS Instance page, configure the read-only instance parameters.**

Category	Parameter	Description
Region	Region	The region where the ApsaraDB for RDS instance resides.
Specifications	Database Engine	The database engine of the read-only instance, which is the same as that of the primary instance and cannot be modified.

Category	Parameter	Description
	Engine Version	The engine version of the read-only instance, which is the same as that of the primary instance and cannot be modified.
	Instance Type	The type of the read-only instance. The type of the read-only instance can be different from that of the primary instance, and can be modified at any time to facilitate flexible upgrade and downgrade.
	Storage	The storage space of the read-only instance. To ensure sufficient I/O throughput for data synchronization, we recommend that you select at least the same instance type and storage space as the primary instance for the read-only instance.
Network Type	Network Type	The network type of the read-only instance, which is the same as that of the primary instance and cannot be modified.
	VPC	Select a VPC if the network type is set to VPC.
	VSwitch	Select a VSwitch if the network type is set to VPC.

5. After you configure the preceding parameters, click Submit.

6.11.3 View details of read-only instances

This topic describes how to view details of read-only instances. You can go to the Basic Information page of a read-only instance from the Instances page or the read-only instance list of the primary instance. Read-only instances are managed in the same way as regular instances. The Basic Information page shows the management operations that can be performed.

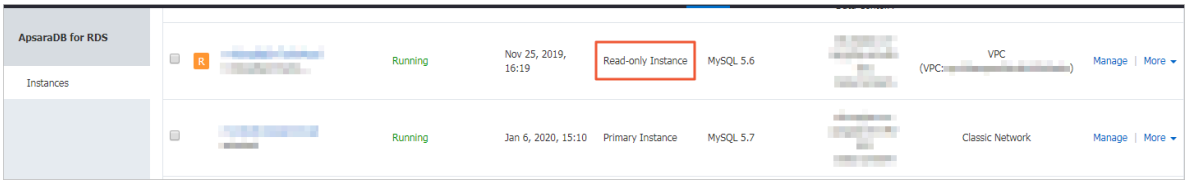
View instance details through a read-only instance

1. *Log on to the ApsaraDB for RDS console*.

2. **On the Instances page, click the ID of a read-only instance. The Basic Information page that appears allows you to manage the read-only instance.**

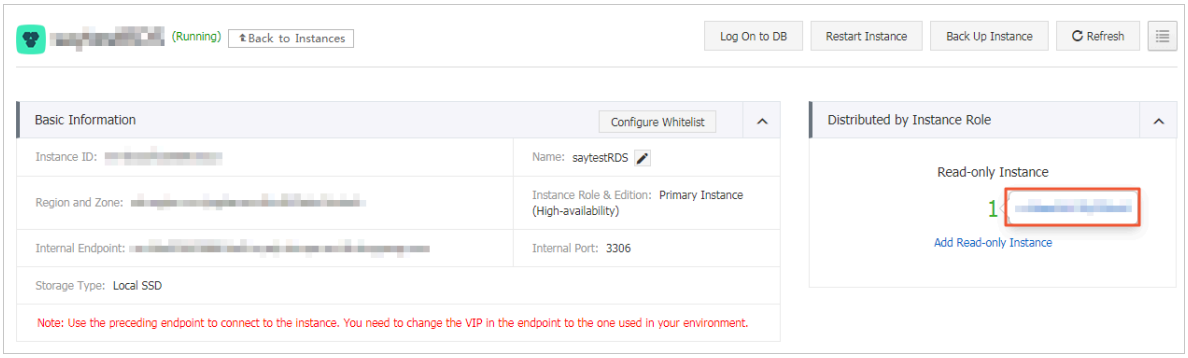
   In the instance list, `Instance Role` **of read-only instances is displayed as** `Read-only Instance`**, as shown in** *Figure 6-2: View read-only instances*.

   Figure 6-2: View read-only instances



View instance details through the primary instance

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**
3. **On the Basic Information page, move the pointer over the number below Read-only Instance in the Distributed by Instance Role section. The ID of the read-only instance is displayed.**



4. **Click the ID of the read-only instance to go to the Basic Information page of the read-only instance.**

6.12 Logs

This topic describes how to manage the error logs, slow query logs, and primary/secondary instance switching logs of an ApsaraDB for RDS instance through the ApsaraDB for RDS console. The logs help you locate faults.

Procedure

1. *Log on to the ApsaraDB for RDS console*.
2. **Click the ID of an instance.**

3. In the left-side navigation pane, click Logs. Click the Error Logs, Slow Log Details, or Primary/Secondary Switching Logs tab, select a time range, and click Search.

Tab	Description
Error Logs	Records database running errors that occurred within the last month.
Slow Log Details	Records SQL statements that took longer than one second to execute from the last month, and deletes redundant SQL statements.  Note: Slow query logs in the ApsaraDB for RDS console are updated once every minute. However, you can query real-time slow query logs from the mysql.slow_log table.
Primary/ Secondary Switching Logs	Records the primary/secondary instance switching logs. This feature is suitable for ApsaraDB RDS for MySQL High-availability Edition instances.

6.13 Migrate data from an on-premises database to an ApsaraDB for RDS instance

6.13.1 Use mysqldump to migrate MySQL data

This topic describes how to use mysqldump to migrate local data to RDS for MySQL.

Prerequisites

An ECS instance has be activated.

Context

mysqldump is easy to use but has long downtimes. The tool is suitable for scenarios where the amount of data is small or long downtimes are allowed.

RDS for MySQL is fully compatible with the native database service. The procedure to migrate the original database to an RDS for MySQL instance is similar to that of migrating data from one MySQL server to another.

Before you migrate data, create a migration account in the local database, and grant the read and write permissions on the database to the migration account.

Procedure

1. **Run the following command to create a migration account in the local database:**

   ```
   CREATE USER 'username'@'host' IDENTIFIED BY 'password';
   ```

   Parameter description:

   - **username: specifies the name of the account to be created.**
   - **host: specifies the database host to which the account logs on. As a local user, you can use `localhost` to log on to the database. To enable the account to log on to any host, you can use wildcard `%`.**
   - **password: specifies the password that is used to log on to the account.**

   The following example creates an account named `William` with password `Changme123`, which is allowed to log on to the local database from any host.

   ```
   CREATE CREATEUSER'William'@'%' IDENTIFIED BY 'Changme123';
   ```

2. **Run the following command to authorize the migration account of the local 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 database name. To grant all database permissions to the account, use wildcard `*`.**
- **Tablename: specifies the table name. 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 a local 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 local database from any host:

```
GRANT ALL ON *. * TO 'William'@'%';
```

3. **Use the data export tool of mysqldump to export data from the database as data files.**

Notice:
Do not update data during the data export. This step exports data only. It does not 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 local database server.**
- **userName: specifies the migration account of the local 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.**

   > **Notice:**
   > **Skip this step if no stored procedures, triggers, or functions are used in the database. When you export stored procedures, triggers, or functions, you must remove the 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 local database server.**
   - **userName: specifies the migration account of the local 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 that is used to connect to the RDS for MySQL instance. In this example, an intranet IP address is used.**
   - **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.**

7 AnalyticDB for PostgreSQL

7.1 What is AnalyticDB for PostgreSQL?

AnalyticDB for PostgreSQL (formerly known as HybridDB for PostgreSQL) is a distributed cloud database that uses multiple compute groups to provide Massively Parallel Processing (MPP) data warehousing service.

AnalyticDB for PostgreSQL is developed based on the Greenplum Open Source Database project and has been enhanced by Alibaba Cloud. This service has the following features:

- **Compatible with Greenplum, allowing you to use all tools that support Greenplum.**
- **Supports OSS, JSON, and HyperLogLog, a probability cardinality estimation algorithm.**
- **Supports flexible hybrid analysis through the SQL:2003 standard and OLAP aggregate functions.**
- **Provides a hybrid mode that supports both column store and row store, enhancing analytics performance.**
- **Supports data compression technology to reduce storage costs.**
- **Provides online expansion and performance monitoring services to free you from managing and maintaining large numbers of MPP clusters. This enables DBAs, developers, and data analysts to focus on improving enterprise productivi ty and creating core business by using SQL.**

7.2 Quick start

7.2.1 Overview

This topic provides a quick start guide about how to perform management tasks for AnalyticDB for PostgreSQL instances such as creating an instance and logging on to a database.

- *Log on to the AnalyticDB for PostgreSQL console*

  This topic describes how to log on to the AnalyticDB for PostgreSQL console.

- *Create an instance*

  You can create an instance in the console and then manage the instance.

- *Configure a whitelist*

  To ensure a secure and stable database, you must add IP addresses or CIDR blocks that are allowed to access the database to a whitelist of the instance before you use the AnalyticDB for PostgreSQL instance.

- *Create an initial account*

  After you create an instance, you must create an initial account to log on to the database.

- *Connect to a database*

  You can use a client that supports PostgreSQL or Greenplum to connect to the database.

7.2.2 Log on to the AnalyticDB for PostgreSQL console

This topic describes how to log on to the AnalyticDB for PostgreSQL console.

Prerequisites

- **Before logging on to the ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel . The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.**
- **We recommend that you use the Google Chrome browser.**

1. **In the address bar, enter the URL used to access the ASCM console. Press Enter.**
2. **Enter your username and password.**

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

   Note:

When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).

3. Click Login to go to the ASCM console homepage.
4. In the top navigation bar, choose Products > AnalyticDB for PostgreSQL.

7.2.3 Create an instance

You can create an instance in the console and then manage the instance.

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. In the upper-right corner of the page, click Create Instance.
3. On the AnalyticDB for PostgreSQL buy page, configure the following parameters.

Section	Parameter	Description
Region	Region	The region of the instance. Note: If you need to access the AnalyticDB for PostgreSQL instance from an ECS instance over VPC, you must deploy the instance in the same region and zone as those of the ECS instance.
	Zone	The zone of the instance.
Basic Settings	Organization	The organization to which the instance belongs.
	Resource Set	The resource set to which the instance belongs.
	Engine	Currently, only the integrated computing and storage version is supported.
	Node Type	The unit of computing resources. Different group types have different storage capacities and computing capabilities.
	Nodes	The number of `compute nodes`. An instance must contain at least two compute nodes. The performance of an instance scales linearly with the number of compute nodes.

Section	Parameter	Description
Network	Network Type	Valid values: · `Classic Network`: Cloud services on a classic network are not isolated from each other. Unauthorized access to a cloud service is blocked only by the security group or whitelist policy of the service. · `VPC`: A VPC helps you to build an isolated network environment in 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. You can create a VPC in advance, or change the network type to VPC after instance creation.
	VPC	The VPC where the AnalyticDB for PostgreSQL instance is located. **Note:** Virtual Private Cloud (VPC): You can use a VPC to build an isolated network environment in Alibaba Cloud. You can customize the route table, IP address range, and gateway in a VPC.
	VSwitch	The VSwitch where the AnalyticDB for PostgreSQL instance is located.
	IP Whitelist	The IP addresses that are allowed to access the instance.

4. After you have configured the preceding parameters, click Submit.

7.2.4 Configure a whitelist

To ensure a secure and stable database, you must add IP addresses or CIDR blocks that are allowed to access the database to a whitelist.

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. Find the target instance and click its ID. The Basic Information page appears.
3. In the left-side navigation pane, click Security Controls. The Security Controls page appears.

4. On the Whitelist Settings tab, click Modify corresponding to the `default` whitelist. The Modify Group page appears.

   > **Note:**
   > You can also click Clear corresponding to the `default` whitelist to delete the IP addresses of the default whitelist, and then click Add Group to create a new whitelist.

5. Delete 127.0.0.1 in the `default` whitelist and enter your IP addresses in the whitelist. The following table lists the parameters.

Parameter	Description
Group Name	Specify the name of the whitelist. The whitelist name must be 2 to 32 characters in length and can contain lowercase letters, digits, and underscores (_). It must start with a lowercase letter and end with a letter or digit. The default whitelist cannot be modified or deleted.
Whitelist	Enter the CIDR blocks or IP addresses that are allowed to access the database. Use commas (,) to separate multiple CIDR blocks or IP addresses. · A whitelist can contain IP addresses such as 10.10.10.1 and CIDR blocks such as 10.10.10.0/24. This CIDR block indicates that any IP addresses in the 10.10.10.X format have access to the database. · The percent sign (%) or 0.0.0.0/0 indicates that any IP addresses are allowed to access the database. **Notice:** This configuration is not recommended because it reduces the security of the database. · Default whitelists of new instances contain the loopback address 127.0.0.1. This configuration allows no access from external IP addresses.

6. Click OK to create a whitelist.

What's next

- We recommend that you regularly maintain the whitelist to ensure secure access for AnalyticDB for PostgreSQL.
- You can click Modify or Delete to modify or delete custom whitelists.

7.2.5 Create an initial account

After you create an instance, you must create an initial account to log on to the database.

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. **Find the target instance and click its ID. The Basic Information page appears.**
3. **In the left-side navigation pane, click Account Management. The Account Management page appears.**
4. **In the upper-right corner of the page, click Create Account. The Create Account page appears.**
5. **Enter the account and password and click OK.**

Parameter	Description
Account	The name of the account must be 2 to 16 characters in length and can contain lowercase letters, digits, and underscores (_). It must start with a letter and end with a letter or digit.
New Password	The password must be 8 to 32 characters in length . It must contain at least three of the following character types: uppercase letters, lowercase letters , digits, and special characters.
Password	Enter the password again.

7.2.6 Obtain the client tool

The interface protocol of AnalyticDB for PostgreSQL is compatible with Greenplum Community Edition and PostgreSQL 8.2. Because of this, you can use the Greenplum or PostgreSQL client to connect to AnalyticDB for PostgreSQL.

Note:

Apsara Stack is an isolated environment. You must deploy software installation packages to the internal environment.

Graphical client tools

AnalyticDB for PostgreSQL users can directly use client tools that support Greenplum, such as *SQL Workbench*, *Navicat Premium*, *Navicat for PostgreSQL*, **and** *pgAdmin III (1.6.3)*.

Command-line client psql (for RHEL 6, RHEL 7, CentOS 6, and CentOS 7)

For Red Hat Enterprise Linux (RHEL) and CentOS 6 or 7, you can download the tools from the following addresses and decompress the packages to use them:

- **For RHEL 6 or CentOS 6, click** *hybriddb_client_package_el6*.
- **For RHEL version 7 or CentOS version 7, click** *hybriddb_client_package_el7*.

Command-line client psql (for other Linux systems)

The compilation methods for client tools applicable to other Linux systems are as follows:

1. **Obtain the source code by using one of the following methods:**
   - **Obtain the git directory. You must first install the git tool.**

     ```
     git clone https://github.com/greenplum-db/gpdb.git
     cd gpdb
     git checkout 5d870156
     ```

   - **Download the code.**

     ```
     wget https://github.com/greenplum-db/gpdb/archive/5d87015609
     abd330c68a5402c1267fc86cbc9e1f.zip
     unzip 5d87015609abd330c68a5402c1267fc86cbc9e1f.zip
     cd gpdb-5d87015609abd330c68a5402c1267fc86cbc9e1f
     ```

2. **Use gcc and other compilers.**

   ```
   ./configure
   make -j32
   make install
   ```

3. **Use psql and pg_dump. The paths of the two tools are as follows:**

   psql: */usr/local/pgsql/bin/psql*

   pg_dump: */usr/local/pgsql/bin/pg_dump*

Command-line client psql (for Windows and other systems)

For client tools for Windows and other systems, go to the Pivotal website to download *HybridDB Client*

7.2.7 Connect to a database

The Greenplum Database and AnalyticDB for PostgreSQL are both developed based on PostgreSQL 8.2 and fully compatible with its messaging protocol. AnalyticDB for

PostgreSQL users can use tools that support the PostgreSQL 8.2 message protocol, such as libpq, JDBC, ODBC, psycopg2, and pgAdmin III.

Context

AnalyticDB for PostgreSQL provides psql, a binary program of Red Hat. For more information about the download link, see *Obtain the client tool*. The Greenplum official website provides an easy-to-install installation package that includes JDBC, ODBC, and libpq. For more information, see *Greenplum official documentation*.

> **Note:**
>
> - Apsara Stack is an isolated environment. To access Apsara Stack, you must prepare the necessary software installation packages in advance.
> - AnalyticDB for PostgreSQL instances can only be accessed by clients deployed on ECS instances within the same region and zone.

psql

psql is a common tool used together with Greenplum, and provides a variety of command functions. Its binary files are located in the `bin` directory of Greenplum. The procedure is as follows:

1. Connect to AnalyticDB for PostgreSQL by using one of the following methods:
   - Connection string

     ```
     psql "host=yourgpdbaddress.gpdb.rds.aliyuncs.com port=3432 dbname=
     postgres user=gpdbaccount password=gpdbpassword"
     ```

   - Specified parameters

     ```
     psql -h yourgpdbaddress.gp.aliyun-inc.com -p 3432 -d postgres -U
     gpdbaccount
     ```

     Parameters:

     - -h: specifies the host address.
     - -p: specifies the port number.
     - -d: specifies the database. The default database is postgres.
     - -U: specifies the user to connect to the database.

     In psql, you can run the `psql --help` command to view more options. You can run the `\?` command to view the commands supported in psql.

2. Enter the password to go to the psql shell interface.

```
postgres=>
```

References

- For more information about the Greenplum psql usage, visit *psql*.
- AnalyticDB for PostgreSQL also supports psql statements of PostgreSQL. Pay attention to the differences between Greenplum psql and PostgreSQL psql. For more information, visit *PostgreSQL 8.3.23 Documentation - psql*.

pgAdmin III

pgAdmin III is a PostgreSQL graphical client and can be directly used to connect to AnalyticDB for PostgreSQL. For more information, click *here*. For more information about other graphical clients, see *Obtain the client tool*.

1. Download pgAdmin III 1.6.3 or earlier versions.

   You can download pgAdmin III 1.6.3 from the *PostgreSQL website*. pgAdmin III 1.6.3 supports various operating systems, such as Windows, macOS, and Linux.

   > **Note:**
   > AnalyticDB for PostgreSQL is compatible with PostgreSQL 8.2. Therefore, you must use pgAdmin III 1.6.3 or earlier to connect to AnalyticDB for PostgreSQL. pgAdmin 4 and later versions are not supported.

2. Choose File > Add Server.
3. In the New Server Registration dialog box that appears, enter the configuration information.
4. Click OK to connect to AnalyticDB for PostgreSQL.

JDBC

JDBC uses the interface provided by PostgreSQL. The download methods are as follows:

Click *PostgreSQL JDBC Driver* to download the official JDBC of PostgreSQL, and then add it to the environment variables.

The sample code is as follows:

```
import java.sql.Connection; import java.sql.DriverManager; import java.
sql.ResultSet; import java.sql.SQLException; import java.sql.Statement
; public class gp_conn { public static void main(String[] args) { try
```

```
 { Class.forName("org.postgresql.Driver"); Connection db = DriverMana
ger.getConnection("jdbc:postgresql://mygpdbpub.gpdb.rds.aliyuncs.com:
3432/postgres","mygpdb","mygpdb"); Statement st = db.createStatement();
ResultSet rs = st.executeQuery("select * from gp_segment_configuration
;"); while (rs.next()) { System.out.print(rs.getString(1)); System.out
.print(" | "); System.out.print(rs.getString(2)); System.out.print(" |
 "); System.out.print(rs.getString(3)); System.out.print(" | "); System
.out.print(rs.getString(4)); System.out.print(" | "); System.out.print(
rs.getString(5)); System.out.print(" | "); System.out.print(rs.getString
(6)); System.out.print(" | "); System.out.print(rs.getString(7)); System
.out.print(" | "); System.out.print(rs.getString(8)); System.out.print
(" | "); System.out.print(rs.getString(9)); System.out.print(" | ");
System.out.print(rs.getString(10)); System.out.print(" | "); System.out.
println(rs.getString(11)); } rs.close(); st.close(); } catch (ClassNotFo
undException e) { e.printStackTrace(); } catch (SQLException e) { e.
printStackTrace(); } } }
```

Python

Python uses psycopg2 to connect to Greenplum and PostgreSQL. The procedure is as follows:

1. **Install psycopg2. There are three installation methods in CentOS:**
   - **Method 1: Run the `yum -y install python-psycopg2` command.**
   - **Method 2: Run the `pip install psycopg2` command.**
   - **Method 3: Run the source code:**

     ```
     yum install -y postgresql-devel*
     wget http://initd.org/psycopg/tarballs/PSYCOPG-2-6/psycopg2-2.6.
     tar.gz
     tar xf psycopg2-2.6.tar.gz
     cd psycopg2-2.6
     python setup.py build
     sudo python setup.py install
     ```

2. **Run the following commands to set PYTHONPATH and reference it:**

   ```
   import psycopg2
    sql = 'select * from gp_segment_configuration;'
    conn = psycopg2.connect(database='gpdb', user='mygpdb', password='
   mygpdb', host='mygpdbpub.gpdb.rds.aliyuncs.com', port=3432)
    conn.autocommit = True
    cursor = conn.cursor()
    cursor.execute(sql)
    rows = cursor.fetchall()
    for row in rows:
        print row
    conn.commit()
    conn.close()
   ```

   A similar output is displayed:

   ```
   (1, -1, 'p', 'p', 's', 'u', 3022, '192.168.2.158', '192.168.2.158',
   None, None)(6, -1, 'm', 'm', 's', 'u', 3019, '192.168.2.47', '192.168
   .2.47', None, None)(2, 0, 'p', 'p', 's', 'u', 3025, '192.168.2.148',
    '192.168.2.148', 3525, None)(4, 0, 'm', 'm', 's', 'u', 3024, '192.168
   .2.158', '192.168.2.158', 3524, None)(3, 1, 'p', 'p', 's', 'u', 3023,
   ```

```
 '192.168.2.158', '192.168.2.158', 3523, None)(5, 1, 'm', 'm', 's', 'u
', 3026, '192.168.2.148', '192.168.2.148', 3526, None)
```

libpq

libpq is the C language interface to AnalyticDB for PostgreSQL. You can use the libpq library to access and manage PostgreSQL databases in a C program. You can locate its static and dynamic libraries under the lib directory.

For the example programs, visit *Example Programs*.

For more information about libpq, visit *PostgreSQL 9.4.17 Documentation - Chapter 31. libpq - C Library*.

ODBC

PostgreSQL ODBC is an open-source version based on the GNU Lesser General Public License (LGPL) protocol. You can download it from the *PostgreSQL website*.

1. **Install the driver.**

```
yum install -y unixODBC.x86_64
yum install -y postgresql-odbc.x86_64
```

2. **View the driver configuration.**

```
cat /etc/odbcinst.ini
# Example driver definitions
# Driver from the postgresql-odbc package
# Setup from the unixODBC package
[PostgreSQL]
Description = ODBC for PostgreSQL
Driver = /usr/lib/psqlodbcw.so
Setup = /usr/lib/libodbcpsqlS.so
Driver64 = /usr/lib64/psqlodbcw.so
Setup64 = /usr/lib64/libodbcpsqlS.so
FileUsage = 1
# Driver from the mysql-connector-odbc package
# Setup from the unixODBC package
[MySQL]
Description = ODBC for MySQL
Driver = /usr/lib/libmyodbc5.so
Setup = /usr/lib/libodbcmyS.so
Driver64 = /usr/lib64/libmyodbc5.so
Setup64 = /usr/lib64/libodbcmyS.so
FileUsage = 1
```

3. **Configure the DSN. Replace the `****` in the following code with the corresponding connection information.**

```
[mygpdb]
Description = Test to gp
Driver = PostgreSQL
Database = ****
Servername = ****.gpdb.rds.aliyuncs.com
UserName = ****
```

```
Password = ****
Port = ****
ReadOnly = 0
```

4. **Test connectivity.**

```
echo "select count(*) from pg_class" | isql mygpdb
+---------------------------------------+
| Connected!                            |
|                                       |
| sql-statement                         |
| help [tablename]                      |
| quit                                  |
|                                       |
+---------------------------------------+
SQL> select count(*) from pg_class
+----------------------+
| count                |
+----------------------+
| 388                  |
+----------------------+
SQLRowCount returns 1
1 rows fetched
```

5. **After ODBC is connected to the instance, connect the application to ODBC. For more information, see** *PostgreSQL ODBC Driver* **and** *psqlODBC HOWTO - C#*.

References

- *Greenplum official documentation*
- *PostgreSQL psqlODBC*
- *Compiling psqlODBC on Unix*
- *Download ODBC connectors*
- *Download JDBC connectors*
- *The PostgreSQL JDBC Interface*

7.3 Instances

7.3.1 Reset the password

If you forget the password of your database account, you can reset the password in the AnalyticDB for PostgreSQL console.

> **Note:**
> **We recommend that you change your password periodically to ensure data security.**

1. *Log on to the AnalyticDB for PostgreSQL console*.

2. Find the target instance and click its ID. The Basic Information page appears.
3. In the left-side navigation pane, click Account Management. The Account Management page appears.
4. Click Reset Password in the corresponding Actions column of the account. The Reset Account Password page appears.
5. After you enter and confirm the new password, click OK.

> **Note:**
> The password must be 8 to 32 characters in length. It must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. We recommend that you do not use a previously used password.

7.3.2 View monitoring information

You can go to the monitoring information page in the console to view the operation status of an instance.

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. Find the target instance and click its ID. The Basic Information page appears.
3. In the left-side navigation pane, click Monitoring and Alarms. The Monitoring and Alarms page appears.

   Specify a duration of time n up to seven days in length to view the metrics for that last n period.

7.3.3 Switch the network type of an instance

The default network type of an instance is Virtual Private Cloud (VPC). After an instance has been created, you can switch its network type between classic network and VPC as needed.

Context

AnalyticDB for PostgreSQL supports two network types: classic network and VPC . Both network types use BGP connections, and are independent of the public network of your service provider. These network types only differ in function, and you can choose a network type based on your requirements. The two network types are applicable to different scenarios:

- **Classic network: IP addresses are allocated by Alibaba Cloud. Classic networks are easy to configure and use. This network type is suitable for users who do not need to perform complex operations, or who only require short deployment cycles.**
- **VPC: a logically isolated private network. You can customize the network topology and IP addresses and connect through a leased line. This network type is suitable for advanced users.**

 **Warning:**
Switching the network type will cause the database service to stop. Proceed with caution.

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. **Find the target instance and click its ID. The Basic Information page appears.**
3. **In the left-side navigation pane, click Database Connection. The Database Connection page appears.**
4. **In the upper-right corner of the page, click Switch to Classic Network or Switch to VPC.**
5. **If you click Switch to VPC, you must select the destination VPC and VSwitch. Click OK.**

    **Note:**
   To switch the network type to VPC, a VPC and VSwitch must exist or be created in the zone where the instance is located.
6. **If you click Switch to Classic Network, click OK in the displayed message.**

 **Note:**
After you switch the network type, it takes 3 to 30 minutes for the instance to enter the running state.

7.3.4 Restart an instance

To better meet your needs, AnalyticDB for PostgreSQL automatically updates the database kernel version. When you create an instance, the latest database kernel is used by default. After a new version is released, you can restart your instance to

update the database kernel and use its extended features. This topic describes how to restart an instance.

Warning:

Restarting an instance will cause the database service to stop. Proceed with caution.

1. *Log on to the AnalyticDB for PostgreSQL console*.
2. **Find the target instance and click its ID. The Basic Information page appears.**
3. **In the upper-right corner of the page, click Restart Instance.**

Note:

The restart process typically takes from 3 to 30 minutes. During the restart period, the instance cannot provide external services. We recommend that you take precautionary measures before restarting instances. After the instance has been restarted and enters the running state, you can access the database.

7.3.5 Import data

7.3.5.1 Import or export data from or to OSS in parallel

AnalyticDB for PostgreSQL can import or export data from or to OSS tables in parallel by using the OSS external table feature, gpossext. AnalyticDB for PostgreSQL also supports GZIP compression for OSS external tables to reduce file size and storage costs. gpossext can read from and write to TEXT and CSV files, even when they are compressed in GZIP packages.

- **Create an OSS external table extension (oss_ext)**

  To use an OSS external table, you must first create an OSS external table extension in AnalyticDB for PostgreSQL. You must create an extension for each database that you need to access.

  - **Creation statement:** `CREATE EXTENSION IF NOT EXISTS oss_ext;`
  - **Deletion statement:** `DROP EXTENSION IF EXISTS oss_ext;`

- **Import data in parallel**

  1. **Distribute data evenly among multiple OSS files for storage. We recommend that you set the number of OSS files to an integer that is the multiple of the number of compute nodes in AnalyticDB for PostgreSQL.**
  2. **Create a READABLE external table in AnalyticDB for PostgreSQL.**
  3. **Execute the following statement to import data in parallel:**

     ```
     INSERT INTO <destination table> SELECT * FROM <external table>
     ```

     Note:

     - **The data import performance depends on the OSS performance and resources of the AnalyticDB for PostgreSQL instance, such as CPU, I/O, memory, and network resources. To ensure the best import performance, we recommend that you use column store and compression when you create a table. For example, you can specify the following clause: `WITH (APPENDONLY =true, ORIENTATION=column, COMPRESSTYPE=zlib, COMPRESSLEVEL=5, BLOCKSIZE=1048576)`. For more information, see *Greenplum Database official documentation on database table creation syntax*.**
     - **We recommend that you configure OSS and AnalyticDB for PostgreSQL instances within the same region to implement the best import performance. For more information, see *Endpoints*.**

- **Export data in parallel**

  1. **Create a WRITABLE external table in AnalyticDB for PostgreSQL.**
  2. **Execute the following statement to export data to OSS in parallel:**

     ```
     INSERT INTO <external table> SELECT * FROM <source table>
     ```

- **Create OSS external tables**

  Note:

  The syntax to create and use external tables is the same as that of Greenplum Database, except for the syntax of location-related parameters.

  ```
  CREATE [READABLE] EXTERNAL TABLE tablename
  ( columnname datatype [, ...] | LIKE othertable )
  LOCATION ('ossprotocol')
  FORMAT 'TEXT'
              [( [HEADER]
                 [DELIMITER [AS] 'delimiter' | 'OFF']
  ```

```
               [NULL [AS] 'null string']
               [ESCAPE [AS] 'escape' | 'OFF']
               [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
               [FILL MISSING FIELDS] )]
           | 'CSV'
            [( [HEADER]
               [QUOTE [AS] 'quote']
               [DELIMITER [AS] 'delimiter']
               [NULL [AS] 'null string']
               [FORCE NOT NULL column [, ...]]
               [ESCAPE [AS] 'escape']
               [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
               [FILL MISSING FIELDS] )]
[ ENCODING 'encoding' ]
[ [LOG ERRORS [INTO error_table]] SEGMENT REJECT LIMIT count
       [ROWS | PERCENT] ]
CREATE WRITABLE EXTERNAL TABLE table_name
( column_name data_type [, ...] | LIKE other_table )
LOCATION ('ossprotocol')
FORMAT 'TEXT'
               [( [DELIMITER [AS] 'delimiter']
               [NULL [AS] 'null string']
               [ESCAPE [AS] 'escape' | 'OFF'] )]
          | 'CSV'
               [([QUOTE [AS] 'quote']
               [DELIMITER [AS] 'delimiter']
               [NULL [AS] 'null string']
               [FORCE QUOTE column [, ...]] ]
               [ESCAPE [AS] 'escape'] )]
[ ENCODING 'encoding' ]
[ DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY ]
ossprotocol:
   oss://oss_endpoint prefix=prefix_name
    id=userossid key=userosskey bucket=ossbucket compressiontype=[
none|gzip] async=[true|false]
ossprotocol:
   oss://oss_endpoint dir=[folder/[folder/]...]/file_name
    id=userossid key=userosskey bucket=ossbucket compressiontype=[
none|gzip] async=[true|false]
ossprotocol:
   oss://oss_endpoint filepath=[folder/[folder/]...]/file_name
```

```
    id=userossid key=userosskey bucket=ossbucket compressiontype=[
none|gzip] async=[true|false]
```

Parameters

Table 7-1: Common parameters

Parameter	Description
Protocol and endpoint	**It is in the `protocol name://oss_endpoint` format. The protocol name is oss. oss_endpoint is the domain name used by users to access OSS in a region.** **Note:** **You can access the database from a VPC host by using an internal endpoint containing "internal" in the name in order not to generate public traffic.**
id	**The AccessKey ID of the OSS account.**
key	**The AccessKey secret of the OSS account.**
bucket	**The bucket where the data file is located. You must use OSS to create the bucket before data import.**

Parameter	Description
prefix	The prefix of the path name corresponding to the data file. Prefixes are directly matched and cannot be controlled by regular expressions. The prefix, filepath, and dir parameters are mutually exclusive and only one parameter can be specified at a time. • If you create a READABLE external table for data import, all OSS files that contain the specified prefix will be imported. - If you set prefix to test/filename, the following files will be imported: ■ test/filename ■ test/filenamexxx ■ test/filename/aa ■ test/filenameyyy/aa ■ test/filenameyyy/bb/aa - If you set prefix to test/filename/, only the following file out of the preceding files will be imported: test/filename/aa • if you create a WRITABLE external table for data export, each exported file will have a unique name based on this parameter. Note: One or more files can be exported for each compute node. The names of exported files are in the `prefix_tablename_uuid.x` format. uuid indicates a timestamp in microseconds as an int64 value. x indicates the node ID. You can use an external table for multiple export operations. Each export operation is assigned a uuid value. The files exported during each operation share a uuid value.

Parameter	Description
dir	The virtual folder path in OSS. The prefix, filepath, and dir parameters are mutually exclusive and only one parameter can be specified at a time. · A folder path must end with a forward slash (/) such as `test/mydir/`. · If you use this parameter when creating an external table for data import, all files under the specified virtual directory (except for its subdirectories and contained files) will be imported. Unlike filepath, dir does not require you to specify the names of files in the directory. · If this parameter is used in creating an external table for data export, all data will be exported to multiple files within the specified directory. The names of exported files are in the `filename.x` format, where x is a digit. The values of x may not be consecutive.

Parameter	Description

filepath	The file name that contains a path in OSS. The prefix , filepath, and dir parameters are mutually exclusive and only one parameter can be specified at a time. You can only specify the filepath parameter when you create a READABLE external table for data import.
	· The file name includes the file path, but not the bucket name.
	· The filename specified for data import must be in the `filename` or `filename.x` format. The values of x must be consecutive digits starting from 1.
	For example, if filepath is set to filename and OSS contains the following files, the imported files include filename, filename.1, and filename.2, but filename.4 is not imported because filename.3 does not exist.
	`filename`
	`filename.1`
	`filename.2`
	`filename.4`

Table 7-2: Import mode parameters

Parameter	Description
async	Specifies whether to load data asynchronously. · Asynchronous data import is enabled by default. You can set async to false or f to disable asynchronous data import. · Enables the worker thread to load data from OSS to accelerate the import performance. The default import mode is asynchronous mode. · Asynchronous data import consumes more hardware resources than normal data import.

Parameter	Description
compressiontype	The compression format of the imported file. Valid values: · none: specifies to import files without compressing them. This is the default value. · gzip: specifies compress imported files in the GZIP format. Only the GZIP format is supported.
compressionlevel	The compression level of the files written to OSS. Valid values: 1 to 9. Default value: 6.

Table 7-3: Export mode parameters

Parameter	Description
oss_flush_block_size	The size of each data block written to OSS. Valid values: 1 MB to 128 MB. Default value: 32 MB.
oss_file_max_size	The maximum size of each file written to OSS. If the limit is exceeded, subsequent data is written to another file. Valid values: 8 MB to 4000 MB. Default value: 1024 MB.
num_parallel_worker	The number of parallel compression threads for data written to OSS. Valid values: 1 to 8. Default value: 3.

Additionally, you must pay attention to the following items for the export mode:

- WRITABLE is the keyword of the external table for data export. You must specify this keyword when creating an external table.
- Only the prefix and dir parameters are supported for data export. The filepath parameter is not supported.
- You can use the DISTRIBUTED BY clause to write data from compute nodes to OSS based on the specified distribution keys.

Other common parameters

The following error-tolerance parameters can be used for data import and export:

Table 7-4: Error-tolerance parameters

Parameter	Description
oss_connect_timeout	The connection timeout period. Unit: seconds. Default value: 10.
oss_dns_cache_timeout	The DNS timeout period. Unit: seconds. Default value : 60.
oss_speed_limit	The minimum rate tolerated. Default value: 1024 bit/s (1 Kbit/s).
oss_speed_time	The maximum amount of time tolerated. Unit: seconds. Default value: 15.

If the default values are used for the preceding parameters, a timeout will occur when the transmission rate is lower than 1 Kbit/s for 15 consecutive seconds. For more information, *see* **Troubleshooting in OSS SDK reference.**

The other parameters are compatible with the original external table syntax of Greenplum Database. For more information about the syntax, see *Greenplum Database official documentation on external table syntax*. **These parameters include:**

- **FORMAT: indicates the supported file format, such as TEXT and CSV.**
- **ENCODING: indicates the data encoding format of a file, such as UTF-8.**
- **LOG ERRORS: indicates that the clause can ignore imported erroneous data and write the data to error_table. You can also use the count parameter to specify the error reporting threshold.**

Examples

```
#Create a READABLE external table of OSS. create readable external table
 ossexample (date text, time text, open float, high float, low float,
volume int) location('oss://oss-cn-hangzhou.aliyuncs.com prefix=osstest
/example id=XXX key=XXX bucket=testbucket compressiontype=gzip') FORMAT
 'csv' (QUOTE '''' DELIMITER E'\t') ENCODING 'utf8' LOG ERRORS INTO
my_error_rows SEGMENT REJECT LIMIT 5;create readable external table
ossexample (date text, time text, open float, high float, low float,
volume int) location('oss://oss-cn-hangzhou.aliyuncs.com dir=osstest
/ id=XXX key=XXX bucket=testbucket') FORMAT 'csv' LOG ERRORS SEGMENT
REJECT LIMIT 5;create readable external table ossexample (date text,
time text, open float, high float, low float, volume int) location('oss
://oss-cn-hangzhou.aliyuncs.com filepath=osstest/example.csv id=XXX key
=XXX bucket=testbucket') FORMAT 'csv' LOG ERRORS SEGMENT REJECT LIMIT
 5;#Create a WRITABLE external table of OSS. create WRITABLE external
 table ossexample_exp (date text, time text, open float, high float,
low float, volume int) location('oss://oss-cn-hangzhou.aliyuncs.com
prefix=osstest/exp/outfromhdb id=XXX key=XXX bucket=testbucket') FORMAT
 'csv' DISTRIBUTED BY (date);create WRITABLE external table ossexample
_exp (date text, time text, open float, high float, low float, volume
```

```
 int) location('oss://oss-cn-hangzhou.aliyuncs.com dir=osstest/exp/
id=XXX key=XXX bucket=testbucket') FORMAT 'csv' DISTRIBUTED BY (date
);#Create a heap table to load data. create table example (date text,
time text, open float, high float, low float, volume int) DISTRIBUTED
 BY (date);#Load data from ossexample to example in parallel. insert
into example select * from ossexample;#Export data from example to OSS
. insert into ossexample_exp select * from example;#Each compute node
 is involved. #Each compute node pulls data from OSS in parallel. The
redistribution motion node calculates the hash value of the data, and
then distributes the hash value to the corresponding compute node. That
 compute node then imports the data to the database through the insert
 node. explain insert into example select * from ossexample; QUERY PLAN
 ------------------------------------------------------------------------
Insert (slice0; segments: 4) (rows=250000 width=92) -> Redistribute
 Motion 4:4 (slice1; segments: 4) (cost=0.00..11000.00 rows=250000
width=92) Hash Key: ossexample.date -> External Scan on ossexample (
cost=0.00..11000.00 rows=250000 width=92)(4 rows)#The compute node
exports the local data to OSS. Data redistribution is not performed.
explain insert into ossexample_exp select * from example; QUERY PLAN
 ------------------------------------------------------------------ Insert
 (slice0; segments: 3) (rows=1 width=92) -> Seq Scan on example (cost=0.
00..0.00 rows=1 width=92)(2 rows)
```

TEXT and CSV format description

The following parameters specify the formats of files read from and written to OSS. You can specify the parameters in the external DDL parameters.

- **\n: a line delimiter or line break for TEXT and CSV files.**
- **DELIMITER: specifies the delimiter of columns.**
  - **If the DELIMITER parameter is specified, the QUOTE parameter must also be specified.**
  - **Recommended column delimiters include commas (,), vertical bars (|), \t, and other special characters.**
- **QUOTE: encloses user data that contains special characters by column.**
  - **Strings that contain special characters will be enclosed by QUOTE to differentiate user data from the control characters.**
  - **To optimize the efficiency, it is unnecessary to enclose data such as integers in QUOTE characters.**
  - **QUOTE cannot be the same string as specified in DELIMITER. The default value of QUOTE is double quotation marks (").**
  - **User data that contains QUOTE characters must also contain ESCAPE characters to differentiate user data from machine code.**

- **ESCAPE: specifies the escape character.**
  - **Place an escape character before a special character that needs to be escaped to indicate that it is not a special character.**
  - **If ESCAPE is not specified, the default value is the same as QUOTE.**
  - **You can also use other characters as ESCAPE characters such as backslashes (\), which is used by MySQL.**

Default control characters for TEXT and CSV files

Table 7-5: Default control characters for TEXT and CSV files

Control character	TEXT	CSV
DELIMITER	\t (tab)	, (comma)
QUOTE	" (double quotation mark)	" (double quotation mark)
ESCAPE	N/A	Same as QUOTE
NULL	\N (backslash-N)	Empty string without quotation marks

Note:
All control characters must be single-byte characters.

SDK troubleshooting

The following *Table 7-6: Error log information* table lists the error logs generated when an error occurs during the import or export process.

Table 7-6: Error log information

Keyword	Description
code	The HTTP status code of the error request.
error_code	The error code returned by OSS.
error_msg	The error message returned by OSS.
req_id	The UUID used to identify the request. If you require assistance in solving a problem, you can submit a ticket containing the req_id of the failed request to OSS developers.

For more information, see *OSS API error responses*. You can handle timeout-related errors by using parameters related to oss_ext.

References

- *Greenplum Database official documentation on external table syntax*
- *Greenplum Database official documentation on table creation syntax*

7.3.5.2 Import data from MySQL

You can use the mysql2pgsql tool to migrate tables from MySQL to AnalyticDB for PostgreSQL, Greenplum Database, PostgreSQL, or PPAS.

Background information

mysql2pgsql connects a source MySQL database to a destination AnalyticDB for PostgreSQL database, queries data to be exported from the MySQL database, and then imports the data to the destination database by using the \COPY statement . The tool supports multi-thread import. Each worker thread imports a part of database tables.

To download the binary installation package of mysql2pgsql, click *here*.

To view instructions on source code compilation of mysql2pgsql, click *here*.

Procedure

1. **Modify the my.cfg configuration file to configure the connection information of source and destination databases.**
   a. **Modify the connection information of the source MySQL database.**

> **Note:**
> **You must have the read permissions on all user tables.**

```
[src.mysql]
host = "192.168.1.1"
port = "3306"
user = "test"
password = "test"
db = "test"
encodingdir = "share"
```

```
encoding = "utf8"
```

b. **Modify the connection information of the destination PostgreSQL, PPAS, or AnalyticDB for PostgreSQL database.**

> **Note:**
> **You must have write permissions on the destination table.**

```
[desc.pgsql]
connect_string = "host=192.168.1.2 dbname=test port=3432  user=
test password=pgsql"
```

2. **Import data by using mysql2pgsql.**

```
./mysql2pgsql -l <tables_list_file> -d -n -j <number of threads> -s
 <schema of target able>
```

Table 7-7: Parameters

Parameter	Description

-l	Optional. Used to specify a text file that contains tables to be synchronized. If you do not specify this parameter, all the tables in the database that is specified in the configuration file will be synchronized. `<tables_list_file>` is the name of a file that contains a collection of tables to be synchronized and conditions for table queries. The content format is as follows:
	`table1 : select * from table_big where column1 < '2016 -08-05'`
	`table2 :`
	`table3`
	`table4: select column1, column2 from tableX where column1 ! = 10`
	`table5: select * from table_big where column1 >= '2016 -08-05'`

-d	Optional. Indicates the table creation DDL statement that creates the destination table but does not synchronize data.
-n	Optional. Must be used along with -d to specify that the table partition definition is not included in the DDL statement.
-j	Optional. Used to specify the number of threads used for data synchronization. If you do not specify this parameter, five concurrent threads will be used by default.

Parameter	Description
-s	Optional. Used to specify the schema of the destination table. Only one schema at a time can be specified by the command. If you do not specify the parameter, the data is imported into the table under the public schema.

Typical usage

Full database migration

1. **Obtain the DDL statements of the corresponding destination table by running the following command:**

   ```
   ./mysql2pgsql -d
   ```

2. **Create a table in the destination database based on these DDL statements with the distribution key information added.**
3. **Run the following command to synchronize all tables:**

   ```
   ./mysql2pgsql
   ```

   This command will migrate the data from all MySQL tables in the database that is specified in the configuration file to the destination database. By default, five concurrent threads are used to read and import data from involved tables.

Partial table migration

1. **Create a new file tab_list.txt and enter the following content:**

   ```
   t1
   t2 : select * from t2 where c1 > 138888
   ```

2. **Run the following command to synchronize the specified t1 and t2 tables (note that for the t2 table, only data that meets the c1 > 138888 condition is migrated):**

   ```
   ./mysql2pgsql -l tab_list.txt
   ```

7.3.5.3 Import data from PostgreSQL

You can use the pgsql2pgsql tool to migrate tables across AnalyticDB for PostgreSQL, Greenplum Database, PostgreSQL, and PPAS.

Context

pgsql2pgsql supports the following features:

- **Full migration across PostgreSQL, PPAS, Greenplum Database, and AnalyticDB for PostgreSQL.**

- **Full migration and incremental migration from PostgreSQL or PPAS (version 9.4 or later) to AnalyticDB for PostgreSQL or ApsaraDB RDS for PPAS.**

You can download the software packages from the *dbsync project* library.

- **To download the binary installation package of pgsql2pgsql, click *here*.**
- **To view instructions on source code compilation of pgsql2pgsql, click *here*.**

Procedure

1. **Modify the my.cfg configuration file to configure the connection information of source and destination databases.**

   a) **Modify the connection information of the source PostgreSQL database.**

   Note:

   In the connection information of the source PostgreSQL database, we recommend that you set the user to the owner of the source database.

   ```
   [src.pgsql]
   connect_string = "host=192.168.1.1 dbname=test port=3432  user=
   test password=pgsql"
   ```

   b) **Modify the connection information of the local temporary PostgreSQL database.**

   ```
   [local.pgsql]
   connect_string = "host=192.168.1.2 dbname=test port=3432  user=
   test2 password=pgsql"
   ```

   c) **Modify the connection information of the destination PostgreSQL database.**

   Note:

   You must have the write permissions on the destination table.

   ```
   [desc.pgsql]
   connect_string = "host=192.168.1.2 dbname=test port=3432  user=
   test3 password=pgsql"
   ```

   Note:

   - **If you need to synchronize incremental data, you must have the permissions to create replication slots in the source database.**
   - **PostgreSQL versions 9.4 and later support logic flow replication, meaning that source databases of the versions support incremental migration. The**

kernel supports logic flow replication only if you configure the following kernel parameters:

wal_level = logical

max_wal_senders = 6

max_replication_slots = 6

2. Use pgsql2pgsql to perform full database migration.

   ```
   ./pgsql2pgsql
   ```

   By default, the migration program migrates the table data of all users from the source PostgreSQL database to the destination PostgreSQL database.

3. View the status information.

   You can view the status information in a single migration process by connecting to the local temporary database. The information is stored in the db_sync_status table, including the start and end time of the full migration, the start time of the incremental migration, and the status of incremental synchronization.

7.3.5.4 Import data by using the \COPY statement

You can use the `\COPY` statement to import the data of local text files into AnalyticDB for PostgreSQL databases. The local text files must be formatted, such as files that use commas (,), semicolons (;), or special characters as delimiters.

Context

- Parallel writing of large amounts of data is not available because the `\COPY` statement writes data in serial using the coordinator node. If you need to import a large amount of data in parallel, you can use the OSS-based data import method.
- The `\COPY` statement is a psql instruction. If you use the database statement `COPY` instead of the `\COPY` statement, you must note that only stdin is supported. This COPY statement does not support file because the root user does not have the superuser permissions to perform operations on files.
- AnalyticDB for PostgreSQL also allows you to use JDBC to execute the COPY statement. The CopyIn method is encapsulated within JDBC. For more information, see *Interface CopyIn*.
- For more information about how to use the COPY statement, see *COPY*.

Procedure

Import data by using the following sample code:

```
\COPY table [(column [, ...])] FROM {'file' | STDIN}
     [ [WITH]
       [OIDS]
       [HEADER]
       [DELIMITER [ AS ] 'delimiter']
       [NULL [ AS ] 'null string']
       [ESCAPE [ AS ] 'escape' | 'OFF']
       [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
       [CSV [QUOTE [ AS ] 'quote']
            [FORCE NOT NULL column [, ...]]
       [FILL MISSING FIELDS]
       [[LOG ERRORS [INTO error_table] [KEEP]
        SEGMENT REJECT LIMIT count [ROWS | PERCENT] ]
\COPY {table [(column [, ...])] | (query)} TO {'file' | STDOUT}
      [ [WITH]
        [OIDS]
        [HEADER]
        [DELIMITER [ AS ] 'delimiter']
        [NULL [ AS ] 'null string']
        [ESCAPE [ AS ] 'escape' | 'OFF']
        [CSV [QUOTE [ AS ] 'quote']
             [FORCE QUOTE column [, ...]] ]
      [IGNORE EXTERNAL PARTITIONS ]
```

7.4 Databases

7.4.1 Overview

The operations based on the Greenplum Database in AnalyticDB for PostgreSQL are the same as those in the Greenplum Database, including schema, supported data types, and user permissions. Except for certain operations exclusive to the Greenplum Database (such as the partition keys and AO tables), you can refer to PostgreSQL for other operations.

References

- *Pivotal Greenplum Official Documentation*
- *Greenplum 4.3 Best Practices*
- *Golden Rules of Greenplum Data Distribution*

7.4.2 Create a database

After you log on to the AnalyticDB for PostgreSQL instance, you can execute SQL statements to create databases.

Similar to PostgreSQL, in AnalyticDB for PostgreSQL you can execute SQL statements to create databases. For example, after psql is connected to Greenplum, execute the following statements:

```
=> create database mygpdb;
CREATE DATABASE
=> \c mygpdb
psql (9.4.4, server 8.3devel)
You are now connected to database "mygpdb" as user "mygpdb".
```

7.4.3 Create a partition key

AnalyticDB for PostgreSQL is a distributed database and data is distributed across all the data nodes. You must create partition keys to distribute the data. The partition keys are vital to query performance. Partition keys are used to ensure even data distribution. Proper selection of keys can significantly improve query performance.

Specify a partition key

In AnalyticDB for PostgreSQL, tables can be distributed across all compute nodes in either hash or random mode. You must specify the partition key when creating a table. Imported data will be distributed to the specific compute node based on the hash value calculated by the partition key.

```
=> create table vtbl(id serial, key integer, value text, shape cuboid,
 location geometry, comment text) distributed by (key);
CREATE TABLE
```

If you do not specify the partition key (that means a statement without the `distributed by (key)` field), AnalyticDB for PostgreSQL will randomly allocate the ID field by using the round-robin algorithm.

Rules for selecting the partition key

- **Select evenly distributed columns or multiple columns to prevent data skew.**
- **Select fields commonly used for connection operations, especially for highly concurrent statements.**
- **Select the condition columns that feature high concurrency queries and high filterability.**

- **Do not use random distribution.**

7.4.4 Construct data

In some test scenarios, you must construct data to fill the database.

1. **Create a function that generates random strings.**

```
CREATE OR REPLACE FUNCTION random_string(integer) RETURNS text AS $
body$
SELECT array_to_string(array
                      (SELECT substring('0123456789ABCDEFGHIJ
KLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz'
                                        FROM (ceil(random()*62))::
int
                                        FOR 1)
                         FROM generate_series(1, $1)), '');
$body$
LANGUAGE SQL VOLATILE;
```

2. **Create a partition key.**

```
CREATE TABLE tbl(id serial, KEY integer, locate geometry, COMMENT
text) distributed by (key);
```

3. **Construct data.**

```
INSERT INTO tbl(KEY, COMMENT, locate)
 SELECT
     KEY,
     COMMENT,
     ST_GeomFromText(locate) AS locate
 FROM
     (SELECT
         (a + 1) AS KEY,
         random_string(ceil(random() * 24)::integer) AS COMMENT,
         'POINT(' || ceil(random() * 36 + 99) || ' ' || ceil(random
() * 24 + 50) || ')' AS locate
     FROM
         generate_series(0, 99999) AS a)
     AS t;
```

7.4.5 Query data

This topic describes the query statements and how to view the query plans.

Query statement sample

```
  => select * from tbl where key = 751;
  | id  | key | value | shape        | locate
           | comment        |
  +-----+-----+-------+-------------
+-----------------------------------------+----------------+
  | 751 | 751 | red   | 0101000000000000000C05B400000000000004A40 |
 B9hPhjeNWPqV |
  (1 row)
```

```
  Time: 513.101 ms
```

View a query plan

```
  => explain select * from tbl where key = 751;
  Gather Motion 1:1  (slice1; segments: 1)  (cost=0.00..1519.28 rows=1
width=53)
    ->  Seq Scan on tbl  (cost=0.00..1519.28 rows=1 width=53)
          Filter: key = 751
  Settings:  effective_cache_size=8GB; gp_statistics_use_fkeys=on
  Optimizer status: legacy query optimizer
```

7.4.6 Manage extensions

You can use extensions to expand database features. AnalyticDB for PostgreSQL enables you to manage extensions.

Extension types

AnalyticDB for PostgreSQL supports the following extensions:

- **PostGIS: supports geographic information data.**
- **MADlib: supports the machine learning function library.**
- **fuzzystrmatch: supports the fuzzy matching of strings.**
- **orafunc: compatible with some Oracle functions.**
- **oss_ext: supports reading data from OSS.**
- **hll: collects statistics by using the HyperLogLog algorithm.**
- **pljava: supports compiling user-defined functions (UDF) in the PL/Java language.**
- **pgcrypto: supports cryptographic hash functions.**
- **intarray: supports integer array-related functions, operators, and indexes.**

Create an extension

Execute the following statements to create an extension:

```
CREATE EXTENSION <extension name>;
CREATE SCHEMA <schema name>;
CREATE EXTENSION IF NOT EXISTS <extension name> WITH SCHEMA <schema
name>;
```

Note:

Before you create the MADlib extension, you must create the plpythonu extension first.

```
CREATE EXTENSION plpythonu;
```

```
CREATE EXTENSION madlib;
```

Delete an extension

Execute the following statements to delete an extension:

```
DROP EXTENSION <extension name>;
DROP EXTENSION IF EXISTS <extension name> CASCADE;
```

Note:
If there are objects dependent on the extension, you must add the CASCADE keyword to delete all dependent objects.

7.4.7 Manage users and permissions

This topic describes how to manage users and permissions in AnalyticDB for PostgreSQL.

Manage users

The system prompts you to specify an initial username and password when you create an instance. This initial user is the root user. After the instance is created , you can use the root user account to connect to the database. The system also creates superusers such as aurora and replicator for internal management.

You can run the `\du+` command to view the information of all the users after you connect to the database by using the client tool of PostgreSQL or Greenplum. Example:

```
postgres=> \du+
                                   List of roles
  Role name   |            Attributes             | Member of |
 Description
--------------+-----------------------------------+-----------
+---------------
 root_user    |                                   |      | rds_superuser
 ...
```

AnalyticDB for PostgreSQL does not provide superuser permissions, but offers a similar role, RDS_SUPERUSER, which is consistent with the permission system of ApsaraDB RDS for PostgreSQL. The root user (such as root_user in the preceding example) has the permissions of the RDS_SUPERUSER role. You can only identify this permission attribute by viewing the user description.

The root user has the following permissions:

- **Can create databases and users and perform actions such as LOGIN, excluding the SUPERUSER permissions.**
- **Can view and modify the data tables of other users and perform actions such as SELECT, UPDATE, DELETE, and changing owners.**
- **Can view the connection information of other users, cancel their SQL statements , and kill their connections.**
- **Can create and delete extensions.**
- **Can create other users with RDS_SUPERUSER permissions. Example:**

```
CREATE ROLE root_user2 RDS_SUPERUSER LOGIN PASSWORD 'xyz' ;
```

Manage permissions

You can manage permissions at the database, schema, and table levels. For example , if you want to grant read permissions on a table to a user and revoke their write permissions, you can execute the following statements:

```
GRANT SELECT ON TABLE t1 TO normal_user1;
REVOKE UPDATE ON TABLE t1 FROM normal_user1;
REVOKE DELETE ON TABLE t1 FROM normal_user1;
```

7.4.8 Manage JSON data

JavaScript Object Notation (JSON) has become a basic data type in the Internet and IoT fields. For more information about JSON, visit *JSON official website*. PostgreSQL support for JSON has been well developed. Optimized by Alibaba Cloud, AnalyticDB for PostgreSQL supports the JSON type based on the PostgreSQL syntax.

Check whether the current version supports JSON

Execute the following statement to check whether the current version supports JSON:

```
=> SELECT '""'::json;
```

If the following output is displayed, it indicates the JSON type is supported and the instance is ready for use. If the operation fails, restart the instance.

```
 json
------
 ""
(1 row)
```

If the following output is displayed, it indicates the JSON type is not supported.

```
ERROR:  type "json" does not exist
```

```
LINE 1: SELECT '""'::json;
               ^
```

The preceding command converts data from the string type to the JSON type. PostgreSQL supports operations on JSON data based on this conversion.

JSON conversion in the database

Database operations include reading and writing. The written data is typically converted from the string type to the JSON type. The contents of a string must meet the JSON standard, such as strings, digits, arrays, and objects. Example:

String

```
=> SELECT '"hijson"'::json;
 json
-------
 "hijson"
(1 row)
```

`::` is used for explicit type conversion in PostgreSQL, Greenplum, and AnalyticDB for PostgreSQL. The database calls the input function in JSON type during the conversion. Therefore, the JSON format check is performed as follows:

```
=> SELECT '{hijson:1024}'::json;
ERROR:  invalid input syntax for type json
LINE 1: SELECT '{hijson:1024}'::json;
               ^
DETAIL:  Token "hijson" is invalid.
CONTEXT:  JSON data, line 1: {hijson...
=>
```

In the preceding example, `hijson` must be enclosed in double quotation marks (") because JSON requires the KEY value to be a string. A syntax error is returned when `{hijson:1024}` is entered.

Apart from explicit type conversion, database records can also be converted to JSON.

Typically, JSON is not used for a string or a digit, but an object that contains one or more key-value pairs. AnalyticDB for PostgreSQL can support most JSON scenarios after data is converted from the string type to objects. Example:

```
=> select row_to_json(row('{"a":"a"}', 'b'));
           row_to_json
---------------------------------
 {"f1":"{\"a\":\"a\"}","f2":"b"}
(1 row)
=> select row_to_json(row('{"a":"a"}'::json, 'b'));
        row_to_json
---------------------------
```

```
 {"f1":{"a":"a"},"f2":"b"}
(1 row)
```

You can see the differences between the string and JSON here. The whole record is conveniently converted into the JSON type.

JSON data types

- **Object**

  The object is the most frequently used data type in JSON. Example:

  ```
  => select '{"key":"value"}'::json;
        json
  -----------------
   {"key":"value"}
  (1 row)
  ```

- **Integer and floating point number**

  JSON only supports three data types for numeric values: integer, floating point number, and constant expression. AnalyticDB for PostgreSQL supports all three types.

  ```
  => SELECT '1024'::json;
   json
  ------
   1024
  (1 row)
  => SELECT '0.1'::json;
   json
  ------
   0.1
  (1 row)
  ```

  The following information is required in some special situations:

  ```
  => SELECT '1e100'::json;
   json
  -------
   1e100
  (1 row)
  => SELECT '{"f":1e100}'::json;
      json
  -------------
   {"f":1e100}
  (1 row)
  ```

  Extra-long numbers are also supported. Example:

  ```
  => SELECT '9223372036854775808'::json;
          json
  ---------------------
   9223372036854775808
  ```

```
  (1 row)
```

- **Array**

```
=> SELECT '[[1,2], [3,4,5]]'::json;
        json
  ------------------
   [[1,2], [3,4,5]]
  (1 row)
```

Operators

Operators supported by JSON

```
=> select oprname,oprcode from pg_operator where oprleft = 3114;
 oprname |          oprcode
---------+---------------------------
 ->      | json_object_field
 ->>     | json_object_field_text
 ->      | json_array_element
 ->>     | json_array_element_text
 #>      | json_extract_path_op
 #>>     | json_extract_path_text_op
(6 rows)
```

Basic usage

```
=> SELECT '{"f":"1e100"}'::json -> 'f';
 ? column?
----------
 "1e100"
(1 row)
=> SELECT '{"f":"1e100"}'::json ->> 'f';
 ? column?
----------
 1e100
(1 row)
=> select '{"f2":{"f3":1},"f4":{"f5":99,"f6":"stringy"}}'::json#>array
['f4','f6'];
 ? column?
-----------
 "stringy"
(1 row)
=> select '{"f2":{"f3":1},"f4":{"f5":99,"f6":"stringy"}}'::json#>'{f4,
f6}';
 ? column?
-----------
 "stringy"
(1 row)
=> select '{"f2":["f3",1],"f4":{"f5":99,"f6":"stringy"}}'::json#>>'{f2
,0}';
 ? column?
----------
 f3
```

```
(1 row)
```

JSON functions

Supported JSON functions

```
postgres=# \df *json*
                                                                List of
functions
   Schema   |           Name            | Result data type |
         Argument data types                    |  Type
------------+---------------------------+------------------
+-----------------------------------------------------------+--------
 pg_catalog | array_to_json             | json             | anyarray
                                                 | normal
 pg_catalog | array_to_json             | json             | anyarray
, boolean                                         | normal
 pg_catalog | json_array_element        | json             | from_json
 json, element_index integer                     | normal
 pg_catalog | json_array_element_text   | text             | from_json
 json, element_index integer                     | normal
 pg_catalog | json_array_elements       | SETOF json       | from_json
 json, OUT value json                            | normal
 pg_catalog | json_array_length         | integer          | json
                                                 | normal
 pg_catalog | json_each                 | SETOF record     | from_json
 json, OUT key text, OUT value json              | normal
 pg_catalog | json_each_text            | SETOF record     | from_json
 json, OUT key text, OUT value text              | normal
 pg_catalog | json_extract_path         | json             | from_json
 json, VARIADIC path_elems text[]                | normal
 pg_catalog | json_extract_path_op      | json             | from_json
 json, path_elems text[]                         | normal
 pg_catalog | json_extract_path_text    | text             | from_json
 json, VARIADIC path_elems text[]                | normal
 pg_catalog | json_extract_path_text_op | text             | from_json
 json, path_elems text[]                         | normal
 pg_catalog | json_in                   | json             | cstring
                                                 | normal
 pg_catalog | json_object_field         | json             | from_json
 json, field_name text                           | normal
 pg_catalog | json_object_field_text    | text             | from_json
 json, field_name text                           | normal
 pg_catalog | json_object_keys          | SETOF text       | json
                                                 | normal
 pg_catalog | json_out                  | cstring          | json
                                                 | normal
 pg_catalog | json_populate_record      | anyelement       | base
anyelement, from_json json, use_json_as_text boolean | normal
 pg_catalog | json_populate_recordset   | SETOF anyelement | base
anyelement, from_json json, use_json_as_text boolean | normal
 pg_catalog | json_recv                 | json             | internal
                                                 | normal
 pg_catalog | json_send                 | bytea            | json
                                                 | normal
 pg_catalog | row_to_json               | json             | record
                                                 | normal
 pg_catalog | row_to_json               | json             | record,
boolean                                          | normal
 pg_catalog | to_json                   | json             |
anyelement                                                | normal
```

```
(24 rows)
```

Basic usage

```
=> SELECT array_to_json('{{1,5},{99,100}}'::int[]);
  array_to_json
-----------------
 [[1,5],[99,100]]
(1 row)
=> SELECT row_to_json(row(1,'foo'));
     row_to_json
---------------------
 {"f1":1,"f2":"foo"}
(1 row)
=> SELECT json_array_length('[1,2,3,{"f1":1,"f2":[5,6]},4]');
 json_array_length
-------------------
                 5
(1 row)
=> select * from json_each('{"f1":[1,2,3],"f2":{"f3":1},"f4":null,"f5
":99,"f6":"stringy"}') q;
 key |   value
-----+-----------
 f1  | [1,2,3]
 f2  | {"f3":1}
 f4  | null
 f5  | 99
 f6  | "stringy"
(5 rows)
=> select json_each_text('{"f1":[1,2,3],"f2":{"f3":1},"f4":null,"f5":"
null"}');
  json_each_text
------------------
 (f1,"[1,2,3]")
 (f2,"{""f3"":1}")
 (f4,)
 (f5,null)
(4 rows)
=> select json_array_elements('[1,true,[1,[2,3]],null,{"f1":1,"f2":[7,
8,9]},false]');
  json_array_elements
-----------------------
 1
 true
 [1,[2,3]]
 null
 {"f1":1,"f2":[7,8,9]}
 false
(6 rows)
create type jpop as (a text, b int, c timestamp);
=> select * from json_populate_record(null::jpop,'{"a":"blurfl","x":43
.2}', false) q;
   a    | b | c
--------+---+---
 blurfl |   |
(1 row)
=>     select * from json_populate_recordset(null::jpop,'[{"a":"blurfl
","x":43.2},{"b":3,"c":"2012-01-20 10:42:53"}]',false) q;
   a    | b |            c
--------+---+--------------------------
 blurfl |   |
        | 3 | Fri Jan 20 10:42:53 2012
```

```
(2 rows)
```

Code examples

Create a table

```
create table tj(id serial, ary int[], obj json, num integer);
=> insert into tj(ary, obj, num) values('{1,5}'::int[], '{"obj":1}', 5
);
INSERT 0 1
=> select row_to_json(q) from (select id, ary, obj, num from tj) as q;
               row_to_json
-------------------------------------------
 {"f1":1,"f2":[1,5],"f3":{"obj":1},"f4":5}
(1 row)
=> insert into tj(ary, obj, num) values('{2,5}'::int[], '{"obj":2}', 5
);
INSERT 0 1
=> select row_to_json(q) from (select id, ary, obj, num from tj) as q;
               row_to_json
-------------------------------------------
 {"f1":1,"f2":[1,5],"f3":{"obj":1},"f4":5}
 {"f1":2,"f2":[2,5],"f3":{"obj":2},"f4":5}
(2 rows)
```

Join multiple tables

```
create table tj2(id serial, ary int[], obj json, num integer);
=> insert into tj2(ary, obj, num) values('{2,5}'::int[], '{"obj":2}',
5);
INSERT 0 1
=> select * from tj, tj2 where tj.obj->>'obj' = tj2.obj->>'obj';
 id |  ary  |    obj    | num | id |  ary  |    obj    | num
----+-------+-----------+-----+----+-------+-----------+-----
  2 | {2,5} | {"obj":2} |   5 |  1 | {2,5} | {"obj":2} |   5
(1 row)
=> select * from tj, tj2 where json_object_field_text(tj.obj, 'obj')
 = json_object_field_text(tj2.obj, 'obj');
 id |  ary  |    obj    | num | id |  ary  |    obj    | num
----+-------+-----------+-----+----+-------+-----------+-----
  2 | {2,5} | {"obj":2} |   5 |  1 | {2,5} | {"obj":2} |   5
(1 row)
```

Use the JSON function index

```
CREATE TEMP TABLE test_json (
       json_type text,
       obj json
);
=> insert into test_json values('aa', '{"f2":{"f3":1},"f4":{"f5":99,"
f6":"foo"}}');
INSERT 0 1
=> insert into test_json values('cc', '{"f7":{"f3":1},"f8":{"f5":99,"
f6":"foo"}}');
INSERT 0 1
=> select obj->'f2' from test_json where json_type = 'aa';
 ? column?
----------
 {"f3":1}
(1 row)
```

```
=> create index i on test_json (json_extract_path_text(obj, '{f4}'));
CREATE INDEX
=> select * from test_json where json_extract_path_text(obj, '{f4}') =
 '{"f5":99,"f6":"foo"}';
 json_type |                 obj
-----------+---------------------------------------
 aa        | {"f2":{"f3":1},"f4":{"f5":99,"f6":"foo"}}
(1 row)
```

Note:

JSON data cannot be used as the partition key and does not support JSON aggregate functions.

Example of using Python to access the database:

```
#! /bin/env python
import time
import json
import psycopg2
def gpquery(sql):
    conn = None
    try:
        conn = psycopg2.connect("dbname=sanity1x2")
        conn.autocommit = True
        cur = conn.cursor()
        cur.execute(sql)
        return cur.fetchall()
    except Exception as e:
        if conn:
            try:
                conn.close()
            except:
                pass
            time.sleep(10)
        print e
    return None
def main(_):
    sql = "select obj from tj;"
    #rows = Connection(host, port, user, pwd, dbname).query(sql)
    rows = gpquery(sql)
    for row in rows:
        print json.loads(row[0])
if __name__ == '__main__':
    main()
```

7.4.9 Use HyperLogLog

AnalyticDB for PostgreSQL is highly optimized by Alibaba Cloud, and not only has the features of Greenplum Database, but also supports HyperLogLog. It is suitable for industries such as Internet advertising and estimation analysis that require quick estimation of business metrics such as PV and UV.

Create a HyperLogLog extension

You can execute the following statement to create a HyperLogLog extension:

```
CREATE EXTENSION hll;
```

Basic types

- **Execute the following statement to create a table containing the hll field:**

  ```
  create table agg (id int primary key,userids hll);
  ```

- **Execute the following statement to convert int to hll_hashval:**

  ```
  select 1::hll_hashval;
  ```

Basic operators

- **The hll type supports** =, !=, <>, ||, **and** #**.**

  ```
  select hll_add_agg(1::hll_hashval) = hll_add_agg(2::hll_hashval);
  select hll_add_agg(1::hll_hashval) || hll_add_agg(2::hll_hashval);
  select #hll_add_agg(1::hll_hashval);
  ```

- **The hll_hashval type supports** =, !=, **and** <>**.**

  ```
  select 1::hll_hashval = 2::hll_hashval;
  select 1::hll_hashval <> 2::hll_hashval;
  ```

Basic functions

- **Hash functions such as Hll_hash_boolean, hll_hash_smallint, and hll_hash_b igint.**

  ```
  select hll_hash_boolean(true);
  select hll_hash_integer(1);
  ```

- **hll_add_agg: converts the int format to the hll format.**

  ```
  select hll_add_agg(1::hll_hashval);
  ```

- **hll_union: aggregates the hll fields.**

  ```
  select hll_union(hll_add_agg(1::hll_hashval),hll_add_agg(2::hll_hashva
  l));
  ```

- **hll_set_defaults: sets the precision.**

  ```
  select hll_set_defaults(15,5,-1,1);
  ```

- **hll_print: displays debug information.**

  ```
  select hll_print(hll_add_agg(1::hll_hashval));
  ```

Examples

```
create table access_date (acc_date date unique, userids hll);
```

```
insert into access_date select current_date, hll_add_agg(hll_hash_i
nteger(user_id)) from generate_series(1,10000) t(user_id);
insert into access_date select current_date-1, hll_add_agg(hll_hash_i
nteger(user_id)) from generate_series(5000,20000) t(user_id);
insert into access_date select current_date-2, hll_add_agg(hll_hash_i
nteger(user_id)) from generate_series(9000,40000) t(user_id);
postgres=# select #userids from access_date where acc_date=current_da
te;
     ? column?
------------------
 9725.85273370708
(1 row)
postgres=# select #userids from access_date where acc_date=current_da
te-1;
     ? column?
------------------
 14968.6596883279
(1 row)
postgres=# select #userids from access_date where acc_date=current_da
te-2;
     ? column?
------------------
 29361.5209149911
(1 row)
```

7.4.10 Use the CREATE LIBRARY statement

AnalyticDB for PostgreSQL introduces the CREATE LIBRARY and DROP LIBRARY statements to allow you to import custom software packages.

Syntax

```
CREATE LIBRARY library_name LANGUAGE [JAVA] FROM oss_location OWNER
ownername
CREATE LIBRARY library_name LANGUAGE [JAVA] VALUES file_content_hex
OWNER ownername
DROP LIBRARY library_name
```

Table 7-8: Parameters

Parameter	Description
library_name	The name of the library to be installed. If the library to be installed has the same name as an existing library, you must delete the existing library before installing the new one.
LANGUAGE [JAVA]	The programming language to be used. Only PL/Java is supported.

Parameter	Description
oss_location	The location of the package. You can specify the OSS bucket and object names. Only one object can be specified and the specified object cannot be a compressed file. The format is as follows: `oss://oss_endpoint filepath=[folder/[folder/]...]/file_name id=userossid key=userosskey bucket=ossbucket`
file_content_hex	The content of the file. The byte stream is in hexadecimal notation. For example, 73656c6563742031 indicates the hexadecimal byte stream of " select 1". You can use this syntax to import packages without using OSS.
ownername	Specifies the user.
DROP LIBRARY	Deletes a library.

Examples

- Example 1: Install a JAR package named analytics.jar.

```
create library example language java from 'oss://oss-cn-hangzhou.aliyuncs.com filepath=analytics.jar id=xxx key=yyy bucket=zzz';
```

- Example 2: Import the file content with the byte stream in hexadecimal notation.

```
create library pglib LANGUAGE java VALUES '73656c6563742031' OWNER "myuser";
```

- Example 3: Delete a library.

```
drop library example;
```

- Example 4: View installed libraries.

```
select name, lanname from pg_library;
```

7.4.11 Create and use the PL/Java UDF

AnalyticDB for PostgreSQL allows you to compile and upload JAR software packages written in PL/Java language, and use these JAR packages to create user-defined functions (UDFs). The PL/Java language supported by AnalyticDB for PostgreSQL is Community Edition PL/Java 1.5.0 and the JVM version is 1.8. This topic describes how to create a PL/Java UDF. For more PL/Java examples, see *PL/Java code*. For the compiling method, see *PL/Java documentation*.

Procedure

1. **In AnalyticDB for PostgreSQL, execute the following statement to create a PL/Java extension. You only need to execute the statement once for each database.**

```
create extension pljava;
```

2. **Compile the UDF based on your business needs. For example, you can use the following code to compile the Test.java file:**

```
public class Test
 {
     public static String substring(String text, int beginIndex,
                int endIndex)
             {
                     try {
                         Process process = null;
                         process = Runtime.getRuntime().exec("echo
Test running");
                     } catch (Exception e) {
                        return "" + e;
                     }
                     return text.substring(beginIndex, endIndex);
             }
 }
```

3. **Compile the manifest.txt file.**

```
 Manifest-Version: 1.0
 Main-Class: Test
 Specification-Title: "Test"
 Specification-Version: "1.0"
 Created-By: 1.7.0_99
 Build-Date: 01/20/2016 21:00 AM
```

4. **Run the following commands to compile and package the program.**

```
javac Test.java
 jar cfm analytics.jar manifest.txt Test.class
```

5. **Upload the analytics.jar file generated in step 4 to OSS by using the following OSS console command.**

```
osscmd put analytics.jar oss://zzz
```

6. **In AnalyticDB for PostgreSQL, execute the CREATE LIBRARY statement to import the file to AnalyticDB for PostgreSQL.**

```
create library example language java from 'oss://oss-cn-hangzhou.
aliyuncs.com filepath=analytics.jar id=xxx key=yyy bucket=zzz';
```

Note:

You can only use the filepath variable in the CREATE LIBRARY statement to import files one at a time. Additionally, the CREATE LIBRARY statement also supports byte streams to import files without using OSS. For more information, see *Use the CREATE LIBRARY statement*.

7. **In AnalyticDB for PostgreSQL, execute the following statements to create and use the UDF.**

```
create table temp (a varchar) distributed randomly;
 insert into temp values ('my string');
 create or replace function java_substring(varchar, int, int)
returns varchar as 'Test.substring' language java;
 select java_substring(a, 1, 5) from temp;
```

7.5 Table

7.5.1 Create a table

You can create tables within your databases.

Syntax

The complete syntax for creating a table is as follows. Depending on your business needs, not all clauses will be required. Use the clauses that can fulfill your business needs.

```
CREATE [[GLOBAL | LOCAL] {TEMPORARY | TEMP}] TABLE table_name (
[ { column_namedata_type [ DEFAULT default_expr ]
   [column_constraint [ ... ]
[ ENCODING ( storage_directive [,...] ) ]
]
   | table_constraint
   | LIKE other_table [{INCLUDING | EXCLUDING}
                      {DEFAULTS | CONSTRAINTS}] ...}
   [, ... ] ]
   )
   [ INHERITS ( parent_table [, ... ] ) ]
   [ WITH ( storage_parameter=value [, ... ] )
   [ ON COMMIT {PRESERVE ROWS | DELETE ROWS | DROP} ]
   [ TABLESPACE tablespace ]
   [ DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY ]
   [ PARTITION BY partition_type (column)
       [ SUBPARTITION BY partition_type (column) ]
          [ SUBPARTITION TEMPLATE ( template_spec ) ]
       [...]
    ( partition_spec )
        | [ SUBPARTITION BY partition_type (column) ]
          [...]
    ( partition_spec
      [ ( subpartition_spec
           [(...)]
       ) ]
```

```
    )
```

The `column_constraint` **clause can be defined as follows:**

```
[CONSTRAINT constraint_name]
   NOT NULL | NULL
   | UNIQUE [USING INDEX TABLESPACE tablespace]
            [WITH ( FILLFACTOR = value )]
   | PRIMARY KEY [USING INDEX TABLESPACE tablespace]
                 [WITH ( FILLFACTOR = value )]
   | CHECK ( expression )
   | REFERENCES table_name [ ( column_name [, ... ] ) ]
            [ key_match_type ]
            [ key_action ]
```

The `storage_directive` **clause of columns can be defined as follows:**

```
   COMPRESSTYPE={ZLIB | QUICKLZ | RLE_TYPE | NONE}
    [COMPRESSLEVEL={0-9} ]
    [BLOCKSIZE={8192-2097152} ]
```

The `storage_parameter` **clause of tables can be defined as follows:**

```
   APPENDONLY={TRUE|FALSE}
   BLOCKSIZE={8192-2097152}
   ORIENTATION={COLUMN|ROW}
   CHECKSUM={TRUE|FALSE}
   COMPRESSTYPE={ZLIB|QUICKLZ|RLE_TYPE|NONE}
   COMPRESSLEVEL={0-9}
   FILLFACTOR={10-100}
   OIDS[=TRUE|FALSE]
```

The `table_constraint` **clause can be defined as follows:**

```
   [CONSTRAINT constraint_name]
   UNIQUE ( column_name [, ... ] )
          [USING INDEX TABLESPACE tablespace]
          [WITH ( FILLFACTOR=value )]
   | PRIMARY KEY ( column_name [, ... ] )
                 [USING INDEX TABLESPACE tablespace]
                 [WITH ( FILLFACTOR=value )]
   | CHECK ( expression )
   | FOREIGN KEY ( column_name [, ... ] )
            REFERENCES table_name [ ( column_name [, ... ] ) ]
            [ key_match_type ]
            [ key_action ]
            [ key_checking_mode ]
```

Valid values of `key_match_type`**:**

```
    MATCH FULL
  | SIMPLE
```

Valid values of `key_action`**:**

```
    ON DELETE
  | ON UPDATE
```

```
   | NO ACTION
   | RESTRICT
   | CASCADE
   | SET NULL
   | SET DEFAULT
```

Valid values of `key_checking_mode`**:**

```
     DEFERRABLE
   | NOT DEFERRABLE
   | INITIALLY DEFERRED
   | INITIALLY IMMEDIATE
```

Valid values of `partition_type`**:**

```
     LIST
   | RANGE
```

The `partition_specification` **clause can be defined as follows:**

```
partition_element [, ...]
```

The `partition_element` **clause can be defined as follows:**

```
    DEFAULT PARTITION name
   | [PARTITION name] VALUES (list_value [,...] )
   | [PARTITION name]
      START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE]
      [ END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] ]
      [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]
   | [PARTITION name]
      END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE]
      [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]
[ WITH ( partition_storage_parameter=value [, ... ] ) ]
[ TABLESPACE tablespace ]
```

The `subpartition_spec` **or** `template_spec` **clause can be defined as follows:**

```
subpartition_element [, ...]
```

The `subpartition_element` **clause can be defined as follows:**

```
    DEFAULT SUBPARTITION name
   | [SUBPARTITION name] VALUES (list_value [,...] )
   | [SUBPARTITION name]
      START ([datatype] 'start_value') [INCLUSIVE | EXCLUSIVE]
      [ END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE] ]
      [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]
   | [SUBPARTITION name]
      END ([datatype] 'end_value') [INCLUSIVE | EXCLUSIVE]
      [ EVERY ([datatype] [number | INTERVAL] 'interval_value') ]
[ WITH ( partition_storage_parameter=value [, ... ] ) ]
```

```
[ TABLESPACE tablespace ]
```

The `storage_parameter` clause of partitions can be defined as follows:

```
    APPENDONLY={TRUE|FALSE}
    BLOCKSIZE={8192-2097152}
    ORIENTATION={COLUMN|ROW}
    CHECKSUM={TRUE|FALSE}
    COMPRESSTYPE={ZLIB|QUICKLZ|RLE_TYPE|NONE}
    COMPRESSLEVEL={1-9}
    FILLFACTOR={10-100}
    OIDS[=TRUE|FALSE]
```

Parameters

The *Table 7-9: Table creation parameters* table describes the key parameters for creating a table.

Table 7-9: Table creation parameters

Parameter	Description
TABLE_NAME	The name of the table to be created.
column_name	The name of a column to be created in the new table.
data_type	The data type of the column. For columns that contain textual data, set the data type to `VARCHAR` or `TEXT`. We do not recommend the `CHAR` type.
DEFAULT default_expr	Specifies a default value for the column. The system will assign this default value to all columns that do not have a value. The default values can be any variable-free expression . Subqueries or cross-references to other columns in the table are not allowed. The data type of the default expression must match the data type of the column. If a column does not have a default value, the default value is null.
ENCODING storage_directive	Specifies the type of compression and block size for the column data. This clause is valid only for append-optimized, column-oriented tables. Column compression settings are inherited from the table level to the partition level to the sub-partition level. The lowest-level settings have priority over inherited settings.

Parameter	Description
INHERITS	Specifies that all columns in the new table automatically inherit a parent table. You can use `INHERITS` to create a persistent relationship between the new child table and its parent table. Schema modifications to the parent table are applied to the child table as well. When the parent table is also scanned, the data of the child table is scanned as well.
LIKE other_table	Specifies a table from which the new table automatically copies all column names, data types, NOT NULL constraints, and distribution policies. Storage properties such as append-optimized or partition structure are not copied. Unlike `INHERITS`, the new table is completely decoupled from the original table after the new table is created.
CONSTRAINT constraint_name	Configures a column or table constraint. When a constraint is violated, the constraint name will be displayed in the error message. Constraint names can be used to communicate helpful information to client applications. Constraint names that contain spaces must be enclosed by double quotation marks (").
WITH (storage_option=value)	Configures storage options for the table or its indexes.
ON COMMIT	The operation that the system performs on the temporary tables at the end of a transaction. Valid values: · PRESERVE ROWS: No special action is taken. The data will be retained after the transaction is completed. The data will only be released when the session is disconnected. · DELETE ROWS: All rows in the temporary table are deleted. · DROP: The temporary table is deleted.
TABLESPACE tablespace	Specifies the name of the tablespace in which the new table is to be created. If not specified, the default tablespace of the database is used.

Parameter	Description
DISTRIBUTED BY	Specifies the distribution policy for the database. • `DISTRIBUTED BY (column, [ ... ] )`: specifies the partition key. The system uses hash distribution based on the distribution key. To evenly distribute data, you must set the partition key to the primary key of the table or a unique column or a set of columns. • `DISTRIBUTED RANDOMLY`: distributes data randomly. **Note:** We recommend that you do not use random distribution.
PARTITION BY	Configures the partition key to partition the table. Partitioning large tables improves data access efficiency. To partition a table is to create a top-level (parent) table and multiple lower-level (child) tables. A parent table is always empty when the partition table is created. Data is stored in the lowest-level child tables. In a multi-level partition table, data is only stored in the lowest-level sub-partitions. Valid values: RANGE, LIST, and a combination of the two.
SUBPARTITION BY	Configures a multi-level partitioned table.
SUBPARTITION TEMPLATE	You can specify a sub-partition template to create sub-partitions (lower-level child tables). This sub-partition template is applied to all parent partitions to ensure the same sub-partition structure.

Examples

Create a table and configure the partition key. The primary key is the default partition key in AnalyticDB for PostgreSQL.

```
CREATE TABLE films (
code        char(5) CONSTRAINT firstkey PRIMARY KEY,
title       varchar(40) NOT NULL,
did         integer NOT NULL,
date_prod   date,
kind        varchar(10),
len         interval hour to minute
);
```

```
CREATE TABLE distributors (
did    integer PRIMARY KEY DEFAULT nextval('serial'),
name   varchar(40) NOT NULL CHECK (name <> '')
);
```

Create a compressed table and configure the partition key.

```
CREATE TABLE sales (txn_id int, qty int, date date)
WITH (appendonly=true, compresslevel=5)
DISTRIBUTED BY (txn_id);
```

Use sub-partition templates of each level and the default partition to create a three-level partition table.

```
CREATE TABLE sales (id int, year int, month int, day int,
region text)
DISTRIBUTED BY (id)
PARTITION BY RANGE (year)

  SUBPARTITION BY RANGE (month)
    SUBPARTITION TEMPLATE (
       START (1) END (13) EVERY (1),
       DEFAULT SUBPARTITION other_months )

  SUBPARTITION BY LIST (region)
    SUBPARTITION TEMPLATE (
       SUBPARTITION usa VALUES ('usa'),
       SUBPARTITION europe VALUES ('europe'),
       SUBPARTITION asia VALUES ('asia'),
       DEFAULT SUBPARTITION other_regions)

( START (2008) END (2016) EVERY (1),
```

```
  DEFAULT PARTITION outlying_years);
```

7.5.2 Principles and scenarios of row store, column store, heap tables, and AO tables

AnalyticDB for PostgreSQL supports row store, column store, heap tables, and AO tables. This topic describes their principles and scenarios.

Row store and column store

Table 7-10: Comparison

Dimension	Row store	Column store
Definition	Row store stores data in the form of rows. Each row is a tuple. To read a column, you must deform all of the columns that precede the target column. Because of this, the costs for accessing the first and the last columns are different.	Column store stores data as columns corresponding to a file or a batch of files. The cost of reading any column is the same. However, if you need to read multiple columns, you must access multiple files. The more columns you access, the higher the overheads are.
Compression ratio	Low.	High.
Cost of reading any column	Columns with larger column numbers cost more.	Same.
Vector computing and JIT architecture	Not suitable. Not suitable for batch computation.	Suitable. More efficient when accessing and obtaining statistics of a batch of data.

Dimension	Row store	Column store
Scenarios	If you need to perform a large number of update and delete operations due to OLTP requirements such as when querying table details where multiple columns are returned, you can use row store. You can use partition tables if you have diversified requirements. For example, if you need to partition the data based on time, you can use row store to query the details of recent data and use column store to obtain more statistics from historical data.	You can use column store if you need data statistics because of the OLAP requirements. If you need a higher compression ratio, you can use column store.

Heap tables

A heap table is heap storage. All changes to the heap table generate redo logs that can be used to restore data by time point. However, heap tables cannot implement logical incremental backup because any data block in the table may be changed and it is not convenient to record the position by using the heap storage.

Commit and redo logs are used to ensure reliability when transactions are finished. You can also implement redundancy by building secondary nodes through redo logs.

Append-optimized (AO) tables

AO tables are used to append data for storage. When you delete the updated data, you can use another bitmap file to mark the row to be deleted and use the bit and offset to determine whether a row is deleted.

When the transaction is finished, you must call the fsync function to record the offset of the data block that performs the last write operation. Even if the data block only contains one record, a new data block will be appended for the next transactio n. The data block is synchronized to the secondary node for data redundancy.

AO tables are not suitable for small transactions because the fsync function is called at the end of each transaction, and this data block will not be reused even if there is space left.

AO tables are suitable for OLAP scenarios, batch data writing, high compression ratio, and logical backup that supports incremental backup. During backup, you only need to record the offset from the backup and the bitmap deletion mark for each full backup.

Usage scenarios of heap tables

- **When multiple small transactions are handled, use a heap table.**
- **When you need to restore data by time point, use a heap table.**

Usage scenarios of AO tables

- **When you need to use column store, use an AO table.**
- **When data is written in batches, use an AO table.**

7.5.3 Enable the column store and compression features

If you want to improve performance, speed up data import, or reduce costs for tables with infrequent updates and multiple fields, we recommend that you use column store and compression. This will increase the compression ratio threefold to ensure faster performance and import speed.

To enable the column store and compression features, you must specify the column store and compression options when creating a table. For example, you can add

the following clause to the CREATE statement to enable the two features. For more information about the table creation syntax, see *Create a table*.

```
with (APPENDONLY=true, ORIENTATION=column, COMPRESSTYPE=zlib,
COMPRESSLEVEL=5, BLOCKSIZE=1048576, OIDS=false)
```

Note:

AnalyticDB for PostgreSQL only supports zlib and RLE_TYPE compression algorithms. If you specify the quicklz algorithm, it is automatically converted to zlib.

7.5.4 Add a field to a column store table and set the default value

This topic describes how to add a field to a column store table and set the default value for the field, and how to use the ANALYZE statement to view the impact of updated data on the size of the column store table.

Context

In a column store table, each column is stored as a file, and two columns in the same row correspond to each other by using the offset. For example, if you add two fields of the INT8 type, you can quickly locate column B from column A by using the offset.

When you add the field, AO tables are not rewritten. If an AO table contains the records of deleted data, the added field must be filled with the deleted records before using the relative offset.

Procedure

1. **Create three AO column store tables.**

```
postgres=# create table tbl1 (id int, info text) with (appendonly=
true, blocksize=8192, compresstype=none, orientation=column);
NOTICE:  Table doesn't have 'DISTRIBUTED BY' clause -- Using column
named 'id' as the Greenplum Database data distribution key for this
table.
HINT:  The 'DISTRIBUTED BY' clause determines the distribution of
data. Make sure column(s) chosen are the optimal data distribution
key to minimize skew.
CREATE TABLE

postgres=# create table tbl2 (id int, info text) with (appendonly=
true, blocksize=8192, compresstype=none, orientation=column);
NOTICE:  Table doesn't have 'DISTRIBUTED BY' clause -- Using column
named 'id' as the Greenplum Database data distribution key for this
table.
HINT:  The 'DISTRIBUTED BY' clause determines the distribution of
data. Make sure column(s) chosen are the optimal data distribution
key to minimize skew.
```

```
CREATE TABLE

postgres=# create table tbl3 (id int, info text) with (appendonly=
true, blocksize=8192, compresstype=none, orientation=column);
NOTICE:  Table doesn't have 'DISTRIBUTED BY' clause -- Using column
named 'id' as the Greenplum Database data distribution key for this
table.
HINT:  The 'DISTRIBUTED BY' clause determines the distribution of
data. Make sure column(s) chosen are the optimal data distribution
key to minimize skew.
CREATE TABLE
```

2. **Insert 10 million entries to the first two tables and 20 million entries to the third one.**

```
postgres=# insertinto tbl1 select generate_series(1,10000000),'test
';
INSERT 0 10000000
postgres=# insert into tbl2 select generate_series(1,10000000),'test
';
INSERT 0 10000000
postgres=# insert into tbl3 select generate_series(1,20000000),'test
';
INSERT 0 20000000
```

3. **Analyze the tables and display their sizes.**

```
postgres=# analyze tbl1;
ANALYZE
postgres=# analyze tbl2;
ANALYZE
postgres=# analyze tbl3;
ANALYZE

postgres=# select pg_size_pretty(pg_relation_size('tbl1'));
 pg_size_pretty
----------------
 88 MB
(1 row)
postgres=# select pg_size_pretty(pg_relation_size('tbl2'));
 pg_size_pretty
----------------
 88 MB
(1 row)
postgres=# select pg_size_pretty(pg_relation_size('tbl3'));
 pg_size_pretty
----------------
 173 MB
(1 row)
```

4. **Update all the data in the first table. Display the table size after the update. The size is twice as large as the size before the update.**

```
postgres=# update tbl1 set info='test';
UPDATE 10000000
postgres=# analyze tbl1;
ANALYZE
postgres=# select pg_size_pretty(pg_relation_size('tbl1'));
 pg_size_pretty
----------------
```

```
 173 MB
(1 row)
```

5. **Add fields to the three tables and set the default values.**

```
postgres=# alter table tbl1 add column c1 int8 default 1;
ALTER TABLE
postgres=# alter table tbl2 add column c1 int8 default 1;
ALTER TABLE
postgres=# alter table tbl3 add column c1 int8 default 1;
ALTER TABLE
```

6. **Analyze the tables and view the table sizes.**

```
postgres=# analyze tbl1;
ANALYZE
postgres=# analyze tbl2;
ANALYZE
postgres=# analyze tbl3;
ANALYZE

postgres=# select pg_size_pretty(pg_relation_size('tbl1'));
 pg_size_pretty
----------------
 325 MB
(1 row)

postgres=# select pg_size_pretty(pg_relation_size('tbl2'));
 pg_size_pretty
----------------
 163 MB
(1 row)

postgres=# select pg_size_pretty(pg_relation_size('tbl3'));
 pg_size_pretty
----------------
 325 MB
(1 row)
```

When you add fields to the AO tables, the number of entries in the existing files will prevail. Even if all the entries are deleted, you must initialize the original data in the newly added fields.

7.5.5 Configure the table partition

For fact tables or large-sized tables in the database, we recommend that you configure table partitions.

Configure the table partition

You can use the table partitioning feature to delete data by using the `ALTER TABLE DROP PARTITION` statement to delete all the data in a partition, and import data by using the `ALTER TABLE EXCHANGE PARTITION` statement to add a new data partition on a regular basis.

AnalyticDB for PostgreSQL supports range partitioning, list partitioning, and composite partitioning. Range partitioning only supports partitioning by fields of the numeric or datetime data types.

The following example shows a table that uses range partitioning.

```
CREATE TABLE LINEITEM (
L_ORDERKEY              BIGINT NOT NULL,
L_PARTKEY               BIGINT NOT NULL,
L_SUPPKEY               BIGINT NOT NULL,
L_LINENUMBER    INTEGER,
L_QUANTITY              FLOAT8,
L_EXTENDEDPRICE FLOAT8,
L_DISCOUNT              FLOAT8,
L_TAX                   FLOAT8,
L_RETURNFLAG    CHAR(1),
L_LINESTATUS    CHAR(1),
L_SHIPDATE              DATE,
L_COMMITDATE    DATE,
L_RECEIPTDATE   DATE,
L_SHIPINSTRUCT  CHAR(25),
L_SHIPMODE              CHAR(10),
L_COMMENT               VARCHAR(44)
)  WITH  (APPENDONLY=true, ORIENTATION=column, COMPRESSTYPE=zlib
, COMPRESSLEVEL=5, BLOCKSIZE=1048576, OIDS=false) DISTRIBUTED BY (
l_orderkey)
PARTITION BY RANGE (L_SHIPDATE) (START (date '1992-01-01') INCLUSIVE
END (date '2000-01-01') EXCLUSIVE EVERY (INTERVAL '1 month' ));
```

Principles of table partitioning

The purpose of partitioning is to minimize the amount of data that needs to be scanned during a query, so partitions must be associated with the query conditions .

- **Principle 1: Select the fields related to the query conditions to configure partitions and reduce the amount of data to be scanned.**
- **Principle 2: When multiple query conditions exist, configure sub-partitions to further reduce the amount of data to be scanned.**

7.5.6 Configure the sort key

A sort key is an attribute of a table. Data on disks is stored in the order of the sort key.

Context

Sort keys have two major advantages:

- **Speed up and optimize column-store operations. The min and max meta information the system collects seldom overlaps with each other, which features good filterability.**
- **Eliminate the need to perform ORDER BY and GROUP BY operations. The data directly read from the disk is ordered as required by the sorting conditions.**

Create a table

```
Command:     CREATE TABLE
Description: define a new table
Syntax:
CREATE [[GLOBAL | LOCAL] {TEMPORARY | TEMP}] TABLE table_name (
[ { column_name data_type [ DEFAULT default_expr ]     [column_con
straint [ ... ]
[ ENCODING ( storage_directive [,...] ) ]
]
   | table_constraint
   | LIKE other_table [{INCLUDING | EXCLUDING}
                      {DEFAULTS | CONSTRAINTS}] ...}
   [, ... ] ]
   [column_reference_storage_directive [, ] ]
   )
   [ INHERITS ( parent_table [, ... ] ) ]
   [ WITH ( storage_parameter=value [, ... ] )
   [ ON COMMIT {PRESERVE ROWS | DELETE ROWS | DROP} ]
   [ TABLESPACE tablespace ]
   [ DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY ]
   [ SORTKEY (column, [ ... ] )]
   [ PARTITION BY partition_type (column)
       [ SUBPARTITION BY partition_type (column) ]
          [ SUBPARTITION TEMPLATE ( template_spec ) ]
       [...]
    ( partition_spec )
        | [ SUBPARTITION BY partition_type (column) ]
          [...]
    ( partition_spec
      [ ( subpartition_spec
           [(...)]
        ) ]
```

```
)
```

Examples:

```
create table test(date text, time text, open float, high float, low
 float, volume int) with(APPENDONLY=true,ORIENTATION=column) sortkey (
volume);
```

Sort the table

```
VACUUM SORT ONLY [tablename]
```

Modify the sort key

This statement only modifies the catalog and does not sort data. You must execute the `VACUUM SORT ONLY` statement to sort the table.

```
ALTER [[GLOBAL | LOCAL] {TEMPORARY | TEMP}] TABLE table_name SET
SORTKEY (column, [ ... ] )
```

Examples:

```
alter table test set sortkey (high,low);
```

7.6 Best practices

7.6.1 Configure memory and load parameters

You must configure memory and load parameters to improve database stability.

Background information

AnalyticDB for PostgreSQL is an MPP database with high computational and resource requirements. It consumes all of the resources provided to it, allowing AnalyticDB for PostgreSQL to have higher processing speeds but making it very easy to reach its limits.

The worst-case scenario in the event of the CPU, network, or hard disk exceeding its limits is a hardware bottleneck. However, in the event that memory is completely consumed, the database may crash.

How to avoid OOM errors

Out of memory (OOM) indicates that the system is unable to provide sufficient memory requested by a process. The following prompt appears when OOM errors occur:

```
Out of memory (seg27 host.example.com pid=47093) VM Protect failed to
allocate 4096 bytes, 0 MB available
```

Causes

Possible causes of the OOM error include:

- **The memory of the database node is insufficient.**
- **Kernel parameters related to the memory of the operating system are incorrectly configured.**
- **Data skew has occurred, causing a compute node to request a large amount of memory.**
- **Query skew has occurred. For example, if the grouping fields of some aggregate or window functions are not distribution keys, the data must be redistributed. After redistribution, data will be skewed in a certain computer node and result in the node requesting a large amount of memory.**

Solutions

1. **Modify the queries to request less memory.**
2. **Use the resource queue provided by AnalyticDB for PostgreSQL to limit the number of concurrent queries. Reduce the number of queries executed within the cluster at the same time to reduce the overall memory requested by the system.**
3. **Reduce the number of compute nodes deployed on a host. For example, deploy 8 compute nodes instead of 16 compute nodes on a host with 128 GB of memory . This allows each compute node to use twice the amount of memory compared with the latter.**
4. **Increase the memory of a host.**
5. **Set the gp_vmem_protect_limit parameter to limit the maximum VMEM that can be used by a single compute node. The memory size of a single host and the number of compute nodes deployed on the host determine the maximum memory size that a single compute node can use on average.**
6. **For SQL statements that have unpredictable memory usage, you can set the statement_mem parameter in the session to limit the memory usage of a single SQL statement, so as to prevent a single SQL statement from consuming all available memory.**
7. **Set the statement_mem parameter at the database level to apply to all the sessions in the database.**

8. **Use the resource queue to limit the maximum memory usage of the resource group. Add database users to the resource group to limit the overall memory used by these users.**

Configure memory-related parameters

Properly configuring the operating system, database parameters, and resource queue can effectively reduce the probability of OOM.

When calculating the average memory usage of a single compute node on a single host, you must consider both the primary and secondary compute nodes. When the cluster encounters a host failure, the system will switch the service from primary compute nodes to the corresponding secondary compute nodes. During this time , the number of compute nodes on the host will be greater than usual. Therefore, you must consider the amount of resources that will be occupied by the secondary compute nodes during failover.

The following tables describe how to configure parameters of the operating system kernel and database to avoid OOM.

Table 7-11: Operating system kernel parameters **describes the parameter configuration of the operating system kernel.**

Table 7-11: Operating system kernel parameters

Parameter	Description
huge page	Do not configure the huge page parameter of the system . AnalyticDB for PostgreSQL does not support the latest version of PostgreSQL and therefore does not support the huge page feature. The huge page parameter locks a part of the allocated memory. Database nodes will not be able to use this part of the memory.

Parameter	Description
vm.overcommit_memory	If you use the swap space, set this parameter to 2. If you do not use the swap space, set this parameter to 0. Valid values: · 0: The requested memory space cannot exceed the difference between the total memory and the resident set size (RSS). An error is returned only when the memory has been exceeded. · 1: Most processes use the malloc function to apply for the memory, but do not use all the memory applied. When this parameter is set to 1, the memory requested by the malloc function will be allocated under any circumstances unless there is not sufficient memory. · 2: The swap space is also considered when the system calculates the memory space that can be applied for. You can apply for a large amount of memory even if the swap space is triggered.
overcommit_ratio	The larger the value, the more memory that process can apply for and the less that will be reserved for the operating system. For the formula used to calculate the memory parameters, see *Examples to calculate the memory parameters*. When this parameter is set to 2, the memory address that can be applied for cannot exceed swap + memory × overcommit_ratio.

Database parameters describes the parameter configuration of the database.

Table 7-12: Database parameters

Parameter	Description
gp_vmem_protect_limit	Specifies the maximum amount of memory that all processes can apply for on each node. If the value is too large, it may result in a system OOM error or even more serious problems. If the value is too small, SQL statements may not be executed even when the system has enough memory.

Parameter	Description
runaway_detector_activation_percent	Default value: 90. This value is specified as a percentage. When the memory used by any compute node exceeds runaway_detector_activation_percent × gp_vmem_protect_limit/100, the query is terminated to prevent OOM. The termination starts from the query that occupies the maximum memory until the memory reaches a value lower than runaway_detector_activation_percent × gp_vmem_protect_limit/100. You can use the gp_toolkit.session_level_memory_consumption view to observe the memory usage of each session and runaway information.

Parameter	Description
statement_mem	Specifies the maximum amount of memory that a single SQL statement can apply. When the maximum memory is exceeded, spill files are created. Default value: 125. Unit : MB. We recommend that you set this parameter according to the following formula: `(gp_vmem_protect_limit × 0.9)/max_expected_concurrent_queries` **Note:** · You can specify the statement_mem parameter in a session. If the current concurrency is low and a session needs to run a query that requires a large amount of memory, you must specify this parameter in the session. · Statement_mem is suitable for limiting memory usage in low concurrency scenarios. If you use statement_mem to limit the memory for high concurrency scenarios, each query is allocated with a very small amount of memory. As a result, the performance of a small number of queries with high memory requirements in high concurrency scenarios is affected. We recommend that you use the resource queue to limit the maximum memory usage in high concurrency scenarios.
gp_workfile_limit_files_per_query	Specifies the maximum number of spill files that can be created by each query. When the memory requested by the query exceeds the statement_mem limit, spill files (also known as work files) are created, which is similar to the swap space of the operating system. When the number of spill files used exceeds the limit, the query will be terminated. Default value: 0, which indicates that an unlimited number of spill files can be created.

Parameter	Description
gp_workfile_compress_algorithm	Specifies the compression algorithm for spill files. Valid values: none and zlib. Specifies the compression algorithm. The values optimize storage space or I/O by sacrificing CPU. You can set this parameter when the disk is insufficient or the spill files meet a write bottleneck.

Examples to calculate the memory parameters

The environment is as follows:

- **Host configuration:**

```
Total RAM = 256 GB
SWAP = 64 GB
```

- **Four hosts, each deployed with eight primary compute nodes and eight secondary compute nodes.**

  When a host fails, the eight primary compute nodes are distributed to the remaining three hosts. A single host can be deployed with at most three extra primary compute nodes from the failed host. A single host can be deployed with at most 11 primary compute nodes.

1. **Calculate the total memory allocated to AnalyticDB for PostgreSQL by the operating system.**

   Reserve 7.5 GB and 5% of memory for the operating system and calculate the available memory for all applications, and divide the available memory by the empirical coefficient of 1.7.

```
gp_vmem = ((SWAP + RAM) – (7.5 GB + 0.05 × RAM))/1.7
        = ((64 + 256) – (7.5 + 0.05 × 256))/1.7
        = 176
```

2. **Use the empirical coefficient of 0.026 to calculate overcommit_ratio.**

```
vm.overcommit_ratio = (RAM – (0.026 × gp_vmem))/RAM
                    = (256 – (0.026 × 176))/256
                    =. 982

Set vm.overcommit_ratio to 98.
```

3. **Calculate gp_vmem_protect_limit (the protection parameter of the maximum memory usage for each compute node), and divide gp_vmem by maximum_acting_primary_segments (the number of primary compute nodes to be run on each other host after one host fails).**

```
gp_vmem_protect_limit calculation
gp_vmem_protect_limit = gp_vmem/maximum_acting_primary_segments
                      = 176/11
                      = 16 GB
                      = 16384 MB
```

Configure the resource queue

You can use resource queues to limit the number of concurrent queries and the total memory usage. When a query is running, it is added to the corresponding queue and the resources used are recorded in the queue. The resource limit of the queue is applied to all sessions in the queue.

The resource queue in AnalyticDB for PostgreSQL is similar to cgroup in Linux.

The syntax to create a resource queue is as follows:

```
Command:     CREATE RESOURCE QUEUE
Description: create a new resource queue for workload management
Syntax:
CREATE RESOURCE QUEUE name WITH (queue_attribute=value [, ... ])
where queue_attribute is:
   ACTIVE_STATEMENTS=integer
        [ MAX_COST=float [COST_OVERCOMMIT={TRUE|FALSE}] ]
        [ MIN_COST=float ]
        [ PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX} ]
        [ MEMORY_LIMIT='memory_units' ]
|  MAX_COST=float [ COST_OVERCOMMIT={TRUE|FALSE} ]
        [ ACTIVE_STATEMENTS=integer ]
        [ MIN_COST=float ]
        [ PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX} ]
        [ MEMORY_LIMIT='memory_units' ]
```

Table 7-13: Resource queue creation parameters **describes the parameters for creating the resource queue.**

Table 7-13: Resource queue creation parameters

Parameter	Description
ACTIVE_STATEMENTS	**The number of SQL statements that are allowed to run (in the active state) concurrently.** **The value -1 indicates an unlimited number of SQL statements can run concurrently.**

Parameter	Description

MEMORY_LIMIT ' memory_units kB, MB or GB'	Specifies the maximum memory usage allowed by all SQL statements in the resource queue. The value -1 indicates unlimited memory usage, but it is easy to trigger OOM errors because it is limited by the database or system parameters mentioned in the preceding sections.
	The memory usage of SQL statements is limited by resource queues and parameters.
	· When the gp_resqueue_memory_policy parameter is set to none, the limit is the same as that in the Greenplum databases earlier than version 4.1.
	· When the gp_resqueue_memory_policy parameter is set to auto and you have specified the statement_mem parameter for a session or at the database level, the allowed memory of a single query will exceed the MEMORY_LIMIT of the resource queue.
	Example:
	`=> SET statement_mem='2GB';`
	`=> SELECT * FROM my_big_table WHERE column=' value' ORDER BY id;`
	`=> RESET statement_mem;`
	· The system parameter max_statement_mem can limit the maximum memory usage at the compute node level. The memory requested by a single query cannot exceed max_statement_mem.
	You can modify the statement_mem parameter at the session level, but do not modify the max_statem ent_mem parameter. We recommend that you specify max_statement_mem as follows:
	`(seghost_physical_memory) / (average_nu mber_concurrent_queries)`
	· When the gp_resqueue_memory_policy parameter is set to eager_free, it indicates that the query is divided into several stages and that the database allocates the memory requested in the current stage. For example, if a query requests 1 GB of memory in total but only needs 100 MB during each stage, the database will allocate 100 MB of memory to the query. You can use eager_free to reduce the possibility of insufficient memory for the query.

Parameter	Description
MAX_COST float	The maximum cost of the queries that are allowed to execute concurrently by the resource group. The cost is the estimated total cost in the SQL execution plan. The value of the parameter can be specified as a floating-point number (such as 100.0) or an exponent (such as 1e+2). A value of -1 indicates the cost is unlimited.
COST_OVERCOMMIT boolean	Specifies whether the limit of max_cost can be exceeded when the system is idle. The value TRUE indicates the limit can be exceeded.
MIN_COST float	When the resources requested exceed the limit, the queries are queued. However, when the cost of a query is lower than the min_cost, the query can run without queuing.
PRIORITY={MIN|LOW|MEDIUM|HIGH|MAX}	The priority of the current resource queue. When resources are insufficient, CPU resources are allocated to the resource queue with a higher priority. The SQL statements in the resource queue with a higher priority can obtain CPU resources first. We recommend that you allocate users that initiate queries with high real-time requirements to resource queues with higher priority. This parameter is similar to the CPU resource group in the Linux cgroup and the time slice policy of real-time and common tasks.

Example of modifying resource queue limits:

```
ALTER RESOURCE QUEUE myqueue WITH (MAX_COST=-1.0, MIN_COST= -1.0);
```

Example of putting the user in the resource queue:

```
ALTER ROLE sammy RESOURCE QUEUE poweruser;
```

The following table describes the parameters of resource queues.

Table 7-14: Resource queue parameters

Parameter	Description
gp_resqueue_memory_policy	Specifies the memory management policy of the resource queue.
gp_resqueue_priority	Specifies whether to enable query prioritization. Valid values: · On · Off If this parameter is disabled, existing priority settings are not evaluated.
gp_resqueue_priority_cpucores_per_segment	Specifies the number of CPU cores allocated to each compute node. For example, if an 8-core host is configured with two primary compute nodes, you can set the parameter to 4. If there are no other nodes on the primary node, set the parameter to 8. When the CPU is preempted, the SQL statements running in the resource group with higher priority are allocated with CPU resources first.
gp_resqueue_priority_sweeper_interval	Specifies the interval at which CPU utilization is recalculated for all active statements. The share value is calculated when the SQL statement is executed. You can calculate the share value based on the priority and gp_resqueue_priority_cpucores_per_segment. The smaller the value and the more frequent the calculation, the better the result brought by the priority settings and the larger the overhead.

Tips for configuring resource queues:

- **We recommend that you create a resource queue for each user.**

  The default resource queue of AnalyticDB for PostgreSQL is pg_default. If no queue is created, all users are assigned to pg_default. This operation is not recommended. We recommend that you create a resource queue for each user. Typically, a database user corresponds to a business. Different database users may correspond to different businesses or users, such as business users, analysts , developers, and DBAs.

- **We do not recommend that you use superusers to execute queries.**

  Queries initiated by superusers are only limited by the preceding parameters and not by the resource queue. We do not recommend that you use superusers to execute queries if you want to use resource queues to limit the use of resources.
- **ACTIVE_STATEMENTS indicates the SQL statements that can be executed concurrently within the resource queue. When the cost of a query is lower than the min_cost, the query can run without queuing.**
- **You can specify the MEMORY_LIMIT parameter to set the allowed maximum memory usage of all the SQL statements in a resource queue. The statement_mem parameter has higher priority that can break through the limit of resource queues.**

  Note:
  The memory of all resource queues cannot exceed gp_vmem_protect_limit.
- **You can distinguish businesses by configuring the priorities of resource queues.**

  For example, assume that report forms have top priority, while common businesses and analysts have lower priorities. In this case, you can create three resource queues with the max, high, and medium priorities, respectively.
- **If the amount of resources requested at different times vary, you can use the crontab command to adjust the limits of resource queues periodically based on usage patterns.**

  For example, the queue of analysts has top priority during the day, while the queue of forms has lower priority at night. AnalyticDB for PostgreSQL does not support resource limits by time period. Therefore, you can only deploy tasks externally by using the ALTER RESOURCE QUEUE statement.
- **You can use the view provided by gp_toolkit to observe the resource usage of the resource queues.**

```
gp_toolkit.gp_resq_activity

gp_toolkit.gp_resq_activity_by_queue

gp_toolkit.gp_resq_priority_backend

gp_toolkit.gp_resq_priority_statement

gp_toolkit.gp_resq_role

gp_toolkit.gp_resqueue_status
```

8 KVStore for Redis

8.1 What is KVStore for Redis?

KVStore for Redis is an online key-value storage service compatible with open-source Redis protocols. KVStore for Redis supports various data types such as strings, lists, sets, sorted sets, and hash tables. The service also supports advanced features such as transactions, message subscription, and message publishing.

You can easily deploy and manage KVStore for Redis databases in the KVStore for Redis console.

- **You can create an instance to initialize a database environment.**
- **Before using a KVStore for Redis instance, add IP addresses or CIDR blocks used to access the database to the instance whitelist.**
- **You can manage instances in the KVStore for Redis console.**
- **To secure data, you can periodically or immediately back up or restore databases in the KVStore for Redis console.**
- **You can log on to a database by using a client and then execute SQL statements to perform database operations.**

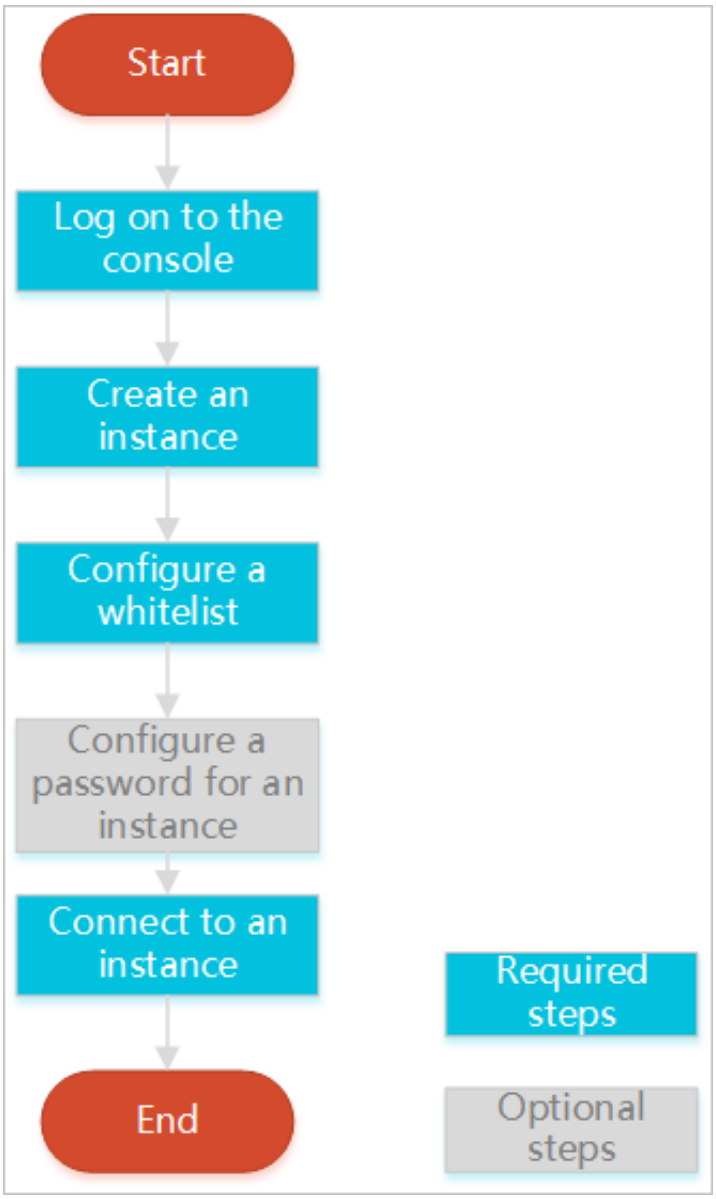
8.2 Quick Start

8.2.1 Get started with KVStore for Redis

This topic describes a series of operations from creating a KVStore for Redis instance to logging on to a database. This helps you understand the procedures to use KVStore for Redis instances.

The flowchart to use KVStore for Redis instances is as follows.

Figure 8-1: Flowchart for the KVStore for Redis instance



- *Log on to the KVStore for Redis console*

  This topic describes how to log on to the KVStore for Redis console.

- *Create an instance*

  KVStore for Redis supports two network types: classic network and VPC. You can create KVStore for Redis instances of different network types.

- *Configure a whitelist*

  Before using a KVStore for Redis instance, add IP addresses or CIDR blocks used to access the database to the instance whitelist to improve database security and stability.

- **If you have not specified a password when creating the instance, set the password of the instance on the Instance Information page.**

- *Connect to the instance*

  You can use a client that supports Redis protocols or use the Redis command-line interface (redis-cli) program to connect to the KVStore for Redis instance.

8.2.2 Log on to the KVStore for Redis console

This topic describes how to log on to the KVStore for Redis console.

Prerequisites

- **Before logging on to the ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel . The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.**
- **We recommend that you use the Google Chrome browser.**

Procedure

1. **In the address bar, enter the URL used to access the ASCM console. Press Enter.**
2. **Enter your username and password.**

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

   Note:
   When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8

to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).

3. Click Login to go to the ASCM console homepage.
4. In the top navigation bar, choose Products > KVStore for Redis.

8.2.3 Create an instance

This topic describes how to create an instance in the KVStore for Redis console.

Prerequisites

To use the Virtual Private Cloud (VPC) service, you must create a VPC in the same region as KVStore for Redis.

Note:
You must specify the network type when you create the instance, and cannot modify the network type later.

Procedure

1. *Log on to the KVStore for Redis console*.
2. Click Create Instance in the upper-right corner.
3. Set the following parameters:

Table 8-1: KVStore for Redis instance parameters

Section	Parameter	Description
Basic Settings	Organization	Specifies the project that the target KVStore for Redis instance belongs to.
	Resource Set	Specifies the zone where the target KVStore for Redis instance is located.  **Notice:** After a project is selected, the KVStore for Redis instance is accessible only to the members of the selected project.
Region	Region	Specifies the region where the target KVStore for Redis instance is located.
	Zone	Specifies the zone where the target KVStore for Redis instance is located.

Section	Parameter	Description
Specifications	Engine Version	The following Redis versions are supported: · Redis 2.8 · Redis 4.0
	Architecture Type	Specifies an architecture type for the target KVStore for Redis instance. KVStore for Redis provides the cluster and standard architectures. The cluster architecture can meet high capacity and performance requirements. Because Redis only supports single-thread processing, we recommend that you use a standard architecture if your business requires QPS performance of 100,000 or less. For higher performance, select a cluster architecture.
	Node Type	Specifies a node type for the target KVStore for Redis instance. KVStore for Redis supports the dual-copy structure.
	Service Plan	Specifies a standard or premium plan.
	Instance Type	The instance specifications. The maximum connections and maximum internal network bandwidth vary depending on the instance specifications.

Section	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: Cloud services in a classic network are not isolated. Unauthorized access to a cloud service is only blocked by security groups or the service whitelist policy. · VPC: A VPC helps you build an isolated network environment in Alibaba Cloud. You can customize the routing table, CIDR blocks, and gateway of a VPC. You can also smoothly migrate applications to the cloud by using a leased line or virtual private network (VPN) to integrate on-premise data centers and cloud resources. **Note:** Before you select the VPC type, create a VPC. For more information, see Create a default VPC and VSwitch in *VPC User Guide* .
Password	Instance Name	Specifies the name of the target instance. · The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), and hyphens(-). · The name must start with an uppercase or lowercase letter.
	Set Password	You can select Now or Later.
	Password	Set a password used to access the instance. A password must follow these rules: · It must be 8 to 30 characters in length. · It must contain uppercase and lowercase letters as well as digits. Special characters are not supported.
	Confirm Password	Enter the password again.

4. After you set the parameters, click Create.

8.2.4 Configure a whitelist

Before using a KVStore for Redis instance, add IP addresses or CIDR blocks used to access the database to the instance whitelist to improve database security and stability.

Context

> **Note:**
> **A properly configured whitelist can guarantee the highest-level security protection for your KVStore for Redis instance. We recommend that you maintain the whitelist on a regular basis.**

Procedure

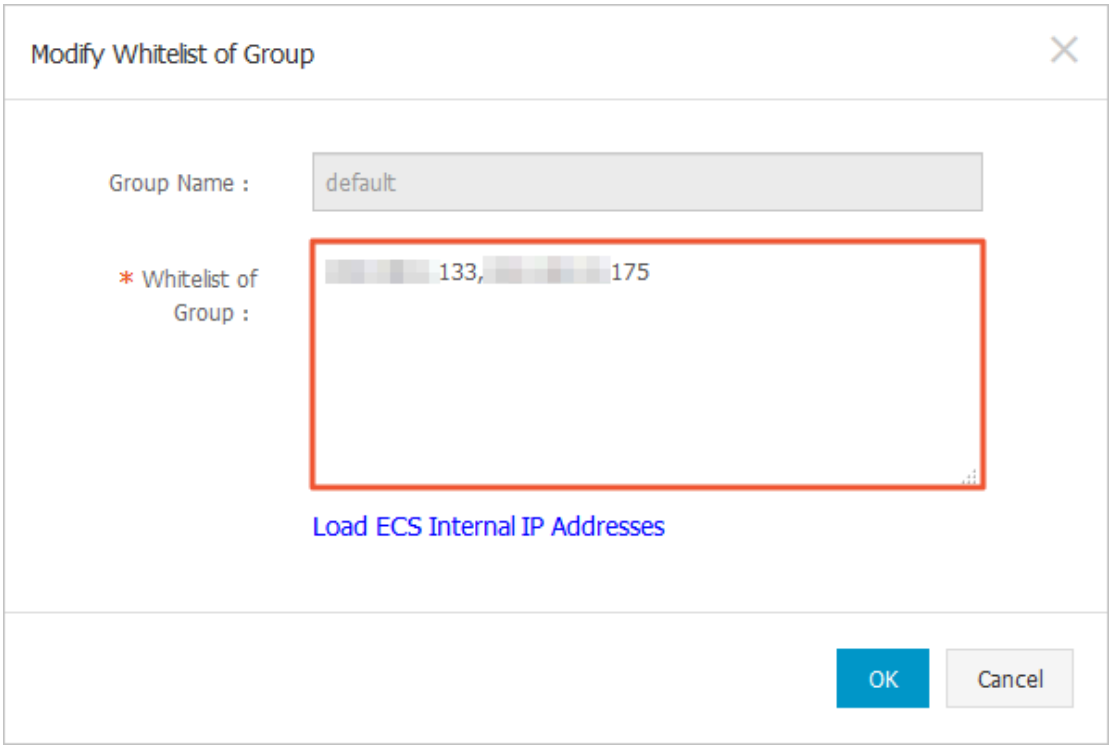
1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. On the Instance Information page, click Whitelist Settings in the left-side navigation pane.
4. On the Whitelist Settings page, proceed in either of the following ways:
   - To customize the whitelist group name, create a new whitelist group:
     a. Click Add a Whitelist Group in the upper-right corner.
     b. In the Add a Whitelist Group dialog box that appears, set Group Name.

        > **Note:**
        > **A group name must be 2 to 32 characters in length and contain lowercase letters, digits, or underscores (_). The group name must start with a lowercase letter and end with a letter or digit. You cannot change this name after you create the whitelist group.**

   - If you do not require a custom whitelist group, click Modify next to the target whitelist group.

5. **In the Add a Whitelist Group or Modify Whitelist of Group dialog box that appears, proceed in either of the following ways:**
   - **Manually modify the Whitelist of Group field:**
     a. **In the Whitelist of Group field, enter the IP addresses or CIDR blocks that you can use to connect to the ApsaraDB for Redis instance.**

Figure 8-2: Manually modify the whitelist group



Note:

- **Set the whitelist to `0.0.0.0/0` to allow connections from all IP addresses.**
- **Set the whitelist to `127.0.0.1` to block connections from all IP addresses.**
- **Set the whitelist to a CIDR block to allow connections from the IP addresses within the CIDR block, such as `10.10.10.0/24`.**
- **When you enter multiple IP addresses or CIDR blocks, separate them with commas (,) and leave no space before or after each comma.**

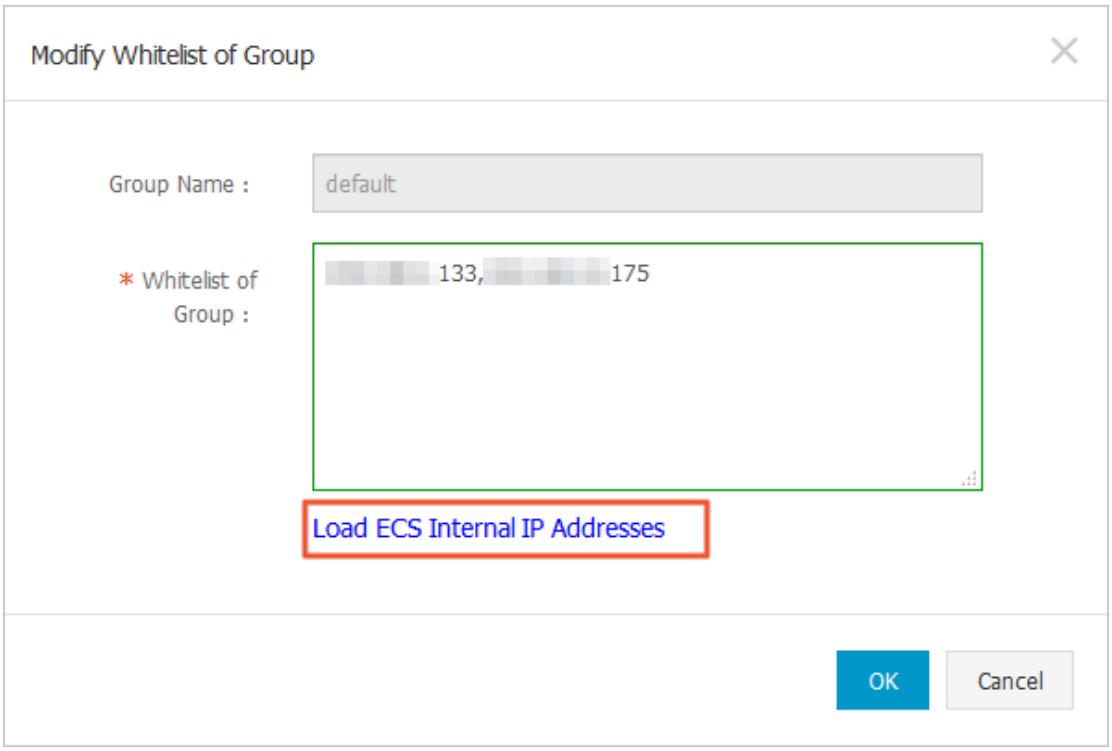
- **You can add 1,000 or fewer IP addresses or CIDR blocks to each whitelist group.**

b. Click OK.

- **Load internal IP addresses of target ECS instances under the current Alibaba Cloud account:**

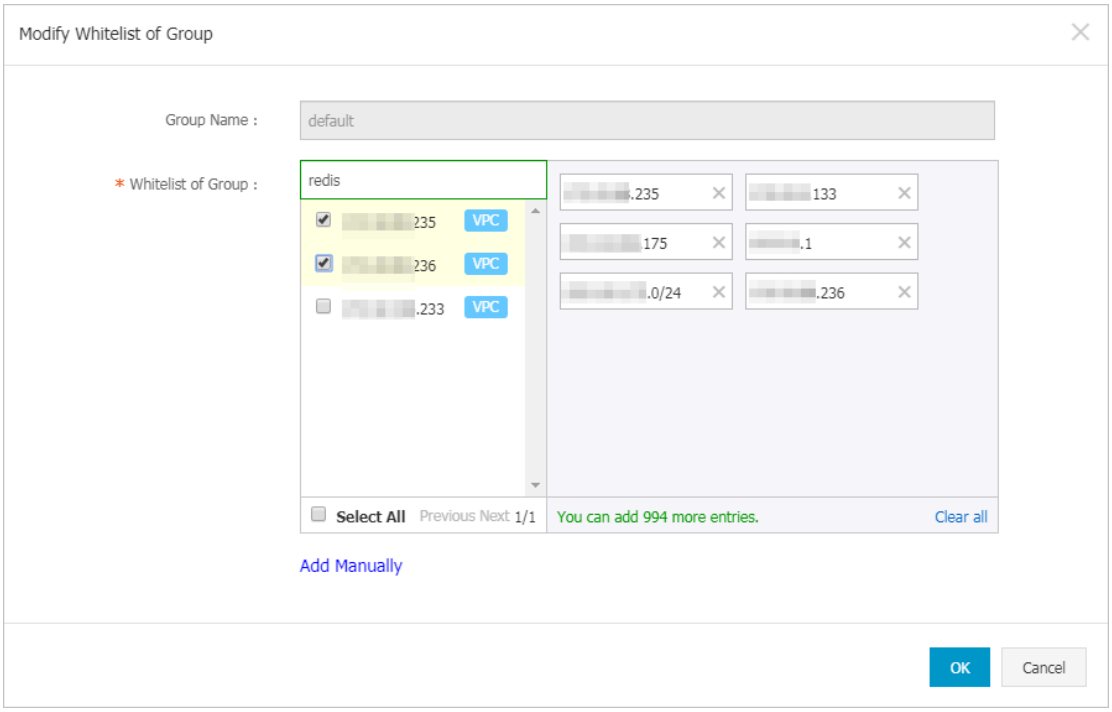
  a. Click Load ECS Internal IP Addresses.

Figure 8-3: Load internal IP addresses of target ECS instances



b. **Select internal IP addresses of target ECS instances.**

Figure 8-4: Select internal IP addresses of target ECS instances



Note:
You can perform a fuzzy search by ECS instance name, ID, or IP address on the search bar above the list of ECS internal IP addresses.

c. **Click OK.**

8.2.5 Connect to an instance

8.2.5.1 Use a Redis client

You can connect to the KVStore for Redis instance by using clients of several programming languages.

The database service of KVStore for Redis is fully compatible with Redis database service. Therefore, you can connect to both database services in similar ways. All clients that are compatible with Redis protocols support connections to KVStore for Redis. You can use any of these clients according to your application features.

For more information about Redis clients, visit *http://redis.io/clients*.

Prerequisites

- **The internal IP address of the ECS instance or the public IP address of the local host has been added to the whitelist of the KVStore for Redis instance. For more information about how to configure a whitelist, see *Configure a whitelist*.**
- **If you use a custom account to connect to the KVStore for Redis instance, the connection password must be in the format of `<user>:<password>`. For example, if the username of a custom account is `admin` and the password is `password`, the password used to connect to the KVStore for Redis instance must be `admin:password`.**

Jedis client

You can use a Jedis client to connect to KVStore for Redis in any of the following ways:

- **Single Jedis connection**
- **JedisPool-based connection**

To use a Jedis client to connect to an KVStore for Redis instance, follow these steps:

1. **Download and install the Jedis client. For more information, see *Jedis*.**
2. **Example of single Jedis connection**

   a. **Open the Eclipse client, create a project, and then enter the following code:**

```
import redis.clients.jedis.Jedis;
public class jedistest {
public static void main(String[] args) {
try {
     String host = "xx.kvstore.aliyuncs.com";//You can view the
endpoint in the 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. Default value: 0.
     jedis.select(1);
     //Set 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();
 }
}
}
```

b. **Run the project. You have connected to KVStore for Redis if you view the following result in the Eclipse console.**

```
Set Key redis Value aliyun-redis
Get Key redis ReturnValue aliyun-redis
```

Afterward, you can use your local Jedis client to manage your KVStore for Redis instance. You can also connect to your KVStore for Redis instance by using JedisPool.

3. **Example of JedisPool-based connection**

a. **Open the Eclipse client, create a project, and then configure the pom file as follows:**

```
<dependency>
<groupId>redis.clients</groupId>
<artifactId>jedis</artifactId>
<version>2.7.2</version>
<type>jar</type>
<scope>compile</scope>
</dependency>
```

b. **Add the following application to the project:**

```
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;
```

c. **If your Jedis client version is Jedis-2.7.2, enter the following code in the project:**

```
JedisPoolConfig config = new JedisPoolConfig();
//Maximum number of idle connections. You can customize this
parameter. Make sure that the specified maximum number of idle
connections does not exceed the maximum number of connections that
 the KVStore for Redis instance supports.
config.setMaxIdle(200);
//Maximum number of connections. You can customize this parameter
. Make sure that the specified maximum number of connections does
not exceed the maximum number of connections that the KVStore for
Redis instance supports.
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();
```

d. If your Jedis client version is Jedis-2.6 or Jedis-2.5, enter the following code in the project:

```
JedisPoolConfig config = new JedisPoolConfig();
//Maximum number of idle connections. You can customize this
parameter. Make sure that the specified maximum number of idle
connections does not exceed the maximum number of connections that
 the KVStore for Redis instance supports.
config.setMaxIdle(200);
//Maximum number of connections. You can customize this parameter
. Make sure that the specified maximum number of connections does
not exceed the maximum number of connections that the KVStore for
Redis instance supports.
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);
 }
```

```
}
```

e. **Run the project. You have connected to KVStore for Redis if you view the following result in the Eclipse console.**

```
Set Key redis Value aliyun-redis
Get Key redis ReturnValue aliyun-redis
```

Afterward, you can use your local Jedis client to manage your KVStore for Redis instance.

phpredis client

To use a phpredis client to connect to an KVStore for Redis instance, follow these steps:

1. **Download and install the phpredis client. For more information, see** *phpredis*.
2. **In any editor that supports PHP editing, enter the following code:**

```
<? php
 /* Replace the following parameter values with the host name and
port number of the target instance. */
 $host = "localhost";
 $port = 6379;
 /* Replace the following parameter values with the ID and password
of the target 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($redis->getLastError());
  }
  /* You can perform database operations after authentication. For
more information, visit https://github.com/phpredis/phpredis. */
 if ($redis->set("foo", "bar") == false) {
         die($redis->getLastError());
 }
 $value = $redis->get("foo");
 echo $value;
 ? >
```

3. **Run the code. Afterward, you can use your local phpredis client to connect to your KVStore for Redis instance. For more information, visit** *https://github.com/phpredis/phpredis*.

redis-py client

To use a redis-py client to connect to an KVStore for Redis instance, follow these steps:

1. **Download and install the redis-py client. For more information, see** *redis-py*.

2. **In any editor that supports Python editing, enter the following code. Afterward, you can use the local redis-py client to connect to the KVStore for Redis instance and perform database operations.**

```
#! /usr/bin/env python
#-*- coding: utf-8 -*-
import redis
#Replace the following parameter values with the host name and port
number of the target instance.
host = 'localhost'
port = 6379
#Replace the following parameter value with the password of the target
 instance.
pwd = 'test_password'
r = redis.StrictRedis(host=host, port=port, password=pwd)
#You can perform database operations after you establish a connection
. For more information, visit https://github.com/andymccurdy/redis-py.
r.set('foo', 'bar');
print r.get('foo')
```

C or C++ client

To use a C or C++ client to connect to an KVStore for Redis instance, follow these steps:

1. **Download, compile, and install the C client by using the following code:**

```
git clone https://github.com/redis/hiredis.git
cd hiredis
make
sudo make install
```

2. **Enter the following code in the C or C++ editor:**

```
#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 (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;
```

```
    }
```

3. **Compile the code.**

```
gcc -o example -g example.c -I /usr/local/include/hiredis -lhiredis
```

4. **Test the code.**

```
 example xxx.kvstore.aliyuncs.com 6379 instance_id password
```

Now, the C or C++ client is connected to the KVStore for Redis instance.

.NET client

To use a .NET client to connect to an KVStore for Redis instance, follow these steps:

1. **Download and use the .NET client.**

```
 git clone https://github.com/ServiceStack/ServiceStack.Redis
```

2. **Create a .NET project on the .NET client.**
3. **Add the reference file stored in the library file directory ServiceStack.Redis/lib/ tests to the client.**
4. **Enter the following code in the .NET project to connect to the KVStore for Redis instance. For more information about API operations, visit** *https://github.com/ServiceStack/ServiceStack.Redis*.

```
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 host that you
 want to connect to*/
         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 that you want to connect to:Port*/
 };
 RedisConfig.VerifyMasterConnections = false;//You must set the
parameter.
 PooledRedisClientManager redisPoolManager = new PooledRedi
sClientManager(10/*Number of connections in the pool*/, 10/*
Connection pool timeout value*/, testReadWriteHosts);
 for (int i = 0; i < 100; i++){
         IRedisClient redisClient = redisPoolManager.GetClient();//
Obtain the connection.
         RedisNativeClient redisNativeClient = (RedisNativeClient)
redisClient;
         redisNativeClient.Client = null;  //KVStore for Redis does
not support the CLIENT SETNAME command. 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();
 }
 }
 }
```

node-redis client

To use a node-redis client to connect to an KVStore for Redis instance, follow these steps:

1. **Download and install a node-redis client.**

```
npm install hiredis redis
```

2. **Enter and run the following code on the node-redis client to connect to the KVStore for Redis instance.**

```
 var redis = require("redis"),
```

```
 client = redis.createClient(<port>, <"host">, {detect_buffers: true
});
 client.auth("password", redis.print)
```

Note:

In the code, the port field specifies the port of the KVStore for Redis instance. Default value: 6379. The host field specifies the endpoint of the KVStore for Redis instance. The following example shows the settings of the port and host fields:

```
client = redis.createClient(6379, "r-abcdefg.redis.rds.aliyuncs.com
", {detect_buffers: true});
```

3. **Use the KVStore for Redis instance.**

```
   // Write data to the instance.
 client.set("key", "OK");
 // Query data on the instance. The instance returns data of String
type.
 client.get("key", function (err, reply) {
 console.log(reply.toString()); // print `OK`
 });
 // If you specify a buffer, the instance returns a buffer.
 client.get(new Buffer("key"), function (err, reply) {
 console.log(reply.toString()); // print `<Buffer 4f 4b>`
 });
 client.quit();
```

C# client StackExchange.Redis

To use the C# client StackExchange.Redis to connect to an KVStore for Redis instance, follow these steps:

1. **Download and install** *StackExchange. Redis*.
2. **Add a reference.**

```
using StackExchange.Redis;
```

3. **Initialize ConnectionMultiplexer.**

   ConnectionMultiplexer is the core of StackExchange.Redis, and shared in the entire application. You must use ConnectionMultiplexer as a singleton. ConnectionMultiplexer is initialized in the following way:

```
 // redis config
 private static ConfigurationOptions configurationOptions =
ConfigurationOptions.Parse("127.0.0.1:6379,password=xxx,connectTim
eout=2000");
  //the lock for singleton
 private static readonly object Locker = new object();
  //singleton
 private static ConnectionMultiplexer redisConn;
```

```
//singleton
public static ConnectionMultiplexer getRedisConn()
{
    if (redisConn == null)
    {
        lock (Locker)
        {
            if (redisConn == null || ! redisConn.IsConnected)
            {
                redisConn = ConnectionMultiplexer.Connect(
configurationOptions);
            }
        }
    }
    return redisConn;
}
```

Note:

ConfigurationOptions contains multiple options, such as keepAlive, connectRetry, and name. For more information, see StackExchange.Redis.ConfigurationOptions.

4. **GetDatabase() returns a lightweight object. You can obtain this object from the object of ConnectionMultiplexer.**

```
redisConn = getRedisConn();
var db = redisConn.GetDatabase();
```

5. **The following examples show five types of data structures. The API operations used in these examples are different from their usage in the native Redis service. These data structures include: string, hash, list, set, and sortedset.**

   - **string**

```
//set get
string strKey = "hello";
string strValue = "world";
bool setResult = db.StringSet(strKey, strValue);
Console.WriteLine("set " + strKey + " " + strValue + ", result is
 " + setResult);
//incr
string counterKey = "counter";
long counterValue = db.StringIncrement(counterKey);
Console.WriteLine("incr " + counterKey + ", result is " +
counterValue);
//expire
db.KeyExpire(strKey, new TimeSpan(0, 0, 5));
Thread.Sleep(5 * 1000);
Console.WriteLine("expire " + strKey + ", after 5 seconds, value
is " + db.StringGet(strKey));
//mset mget
KeyValuePair<RedisKey, RedisValue> kv1 = new KeyValuePair<RedisKey
, RedisValue>("key1", "value1");
KeyValuePair<RedisKey, RedisValue> kv2 = new KeyValuePair<RedisKey
, RedisValue>("key2", "value2");
```

```
db.StringSet(new KeyValuePair<RedisKey, RedisValue>[] {kv1,kv2});

RedisValue[] values = db.StringGet(new RedisKey[] {kv1.Key, kv2.
Key});
Console.WriteLine("mget " + kv1.Key.ToString() + " " + kv2.Key.
ToString() + ", result is " + values[0] + "&&" + values[1]);
```

- **hash**

```
string hashKey = "myhash";
//hset
db.HashSet(hashKey,"f1","v1");
db.HashSet(hashKey,"f2", "v2");
HashEntry[] values = db.HashGetAll(hashKey);
//hgetall
Console.Write("hgetall " + hashKey + ", result is");
for (int i = 0; i < values.Length;i++)
{
  HashEntry hashEntry = values[i];
  Console.Write(" " + hashEntry.Name.ToString() + " " + hashEntry.
Value.ToString());
}
Console.WriteLine();
```

- **list**

```
//list key
string listKey = "myList";
//rpush
db.ListRightPush(listKey, "a");
db.ListRightPush(listKey, "b");
db.ListRightPush(listKey, "c");
//lrange
RedisValue[] values = db.ListRange(listKey, 0, -1);
Console.Write("lrange " + listKey + " 0 -1, result is ");
for (int i = 0; i < values.Length; i++)
{
 Console.Write(values[i] + " ");
}
Console.WriteLine();
```

- **set**

```
//set key
string setKey = "mySet";
//sadd
db.SetAdd(setKey, "a");
db.SetAdd(setKey, "b");
db.SetAdd(setKey, "c");
//sismember
bool isContains = db.SetContains(setKey, "a");
Console.WriteLine("set " + setKey + " contains a is " + isContains
 );
```

- **sortedset**

```
string sortedSetKey = "myZset";
//sadd
db.SortedSetAdd(sortedSetKey, "xiaoming", 85);
db.SortedSetAdd(sortedSetKey, "xiaohong", 100);
db.SortedSetAdd(sortedSetKey, "xiaofei", 62);
db.SortedSetAdd(sortedSetKey, "xiaotang", 73);
```

```
//zrevrangebyscore
RedisValue[] names = db.SortedSetRangeByRank(sortedSetKey, 0, 2,
Order.Ascending);
Console.Write("zrevrangebyscore " + sortedSetKey + " 0 2, result
is ");
for (int i = 0; i < names.Length; i++)
{
  Console.Write(names[i] + " ");
}
Console.WriteLine();
```

8.2.5.2 Use redis-cli

You can use the Redis command-line interface (redis-cli) tool to connect to a KVStore for Redis instance.

Notice:

Because KVStore for Redis only supports connections over an internal network, only the ECS instances that run on the same node as KVStore for Redis and that are installed with redis-cli can connect to KVStore for Redis and perform data operations.

Install the redis-cli utility

Install the Redis software distribution that includes the redis-cli utility in Linux. For more information about the detailed procedure, see *Redis community*.

Prerequisites

Connection over the internal network

- **If the network types for the ECS and KVStore for Redis instances are both classic network, the two instances must reside in the same region.**
- **You have added the private IP address of an ECS instance to the whitelist of a KVStore for Redis instance.**
- **The operating system of the local host must be Linux.**
- **You have installed the Redis software distribution on the ECS instance.**
- **If you use a custom account to connect to the KVStore for Redis instance, the connection password must be in the format of `<user>:<password>`. For example, if the username of a custom account is `admin` and the password is `password`, the password used to connect to the KVStore for Redis instance must be `admin:password`.**

Connection over the Internet

- **The KVStore for Redis instance has a public connection string. For more information, see**
- **You have added the public IP address of the local host to the whitelist of the KVStore for Redis instance.**
- **The operating system of the local host must be Linux.**
- **You have installed the Redis software distribution on the local host.**
- **If you use a custom account to connect to the KVStore for Redis instance, the connection password must be in the format of `<user>:<password>`. For example, if the username of a custom account is `admin` and the password is `password`, the password used to connect to the KVStore for Redis instance must be `admin:password`.**

Connect to a KVStore for Redis instance

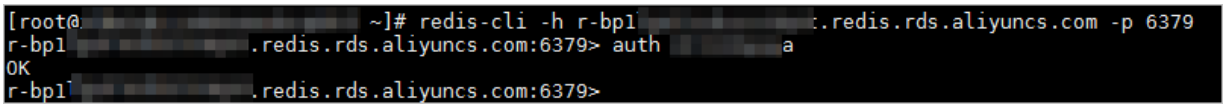
You can use the following command to connect to a KVStore for Redis instance.

```
redis-cli -h <host> -p <port> -a <password>
```

Table 8-2: Parameters

Parameter	Description
-h	The connection string of the KVStore for Redis instance.
-p	The service port of the KVStore for Redis instance. The default port number is 6379 and cannot be changed.
-a	The password used to connect the KVStore for Redis instance. You can skip this parameter to avoid revealing the password over plaintext and enhance security. After running the preceding command, you can enter `auth <password>` to complete the authentication. The following figure shows an example.

Figure 8-5: Command example



8.3 Instance management

8.3.1 Change the password

If you forget your password, need to change your password, or have not set a password for an instance, you can set a new password for the instance.

Procedure

1. *Log on to the KVStore for Redis console*.
2. **On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.**
3. **In the upper-right corner of the Basic Information page, click Modify Password.**
4. **In the Change Password dialog box that appears, set Old Password, New Password, Confirm Password.**

> **Note:**
>
> - **If you forget your password, you can click Forgot password? in the Change Password dialog box. In the Reset Password dialog box that appears, you can set a new password.**
> - **The password must be 8 to 32 characters in length.**
> - **The password must contain characters from at least three of the following categories: uppercase letters, lowercase letters, digits, and special characters . Special characters include ! @ # $ % ^ & * ( ) _ + - =**

8.3.2 Configure a whitelist

Before using a KVStore for Redis instance, add IP addresses or CIDR blocks used to access the database to the instance whitelist to improve database security and stability.

Context

> **Note:**
> **A properly configured whitelist can guarantee the highest-level security protection for your KVStore for Redis instance. We recommend that you maintain the whitelist on a regular basis.**

Procedure

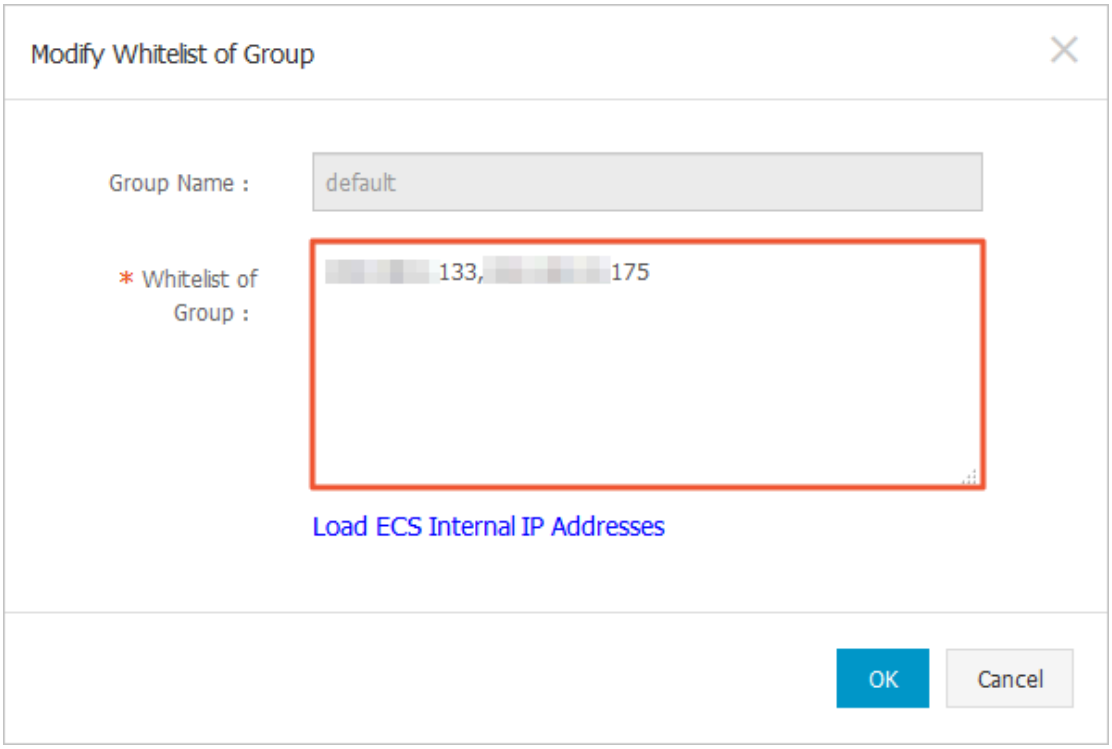
1. *Log on to the KVStore for Redis console*.

2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. On the Instance Information page, click Whitelist Settings in the left-side navigation pane.
4. On the Whitelist Settings page, proceed in either of the following ways:
   - To customize the whitelist group name, create a new whitelist group:
     a. Click Add a Whitelist Group in the upper-right corner.
     b. In the Add a Whitelist Group dialog box that appears, set Group Name.

         **Note:**
        A group name must be 2 to 32 characters in length and contain lowercase letters, digits, or underscores (_). The group name must start with a lowercase letter and end with a letter or digit. You cannot change this name after you create the whitelist group.
   - If you do not require a custom whitelist group, click Modify next to the target whitelist group.

5. **In the Add a Whitelist Group or Modify Whitelist of Group dialog box that appears, proceed in either of the following ways:**
   - **Manually modify the Whitelist of Group field:**
     a. **In the Whitelist of Group field, enter the IP addresses or CIDR blocks that you can use to connect to the ApsaraDB for Redis instance.**

Figure 8-6: Manually modify the whitelist group



Note:

- **Set the whitelist to `0.0.0.0/0` to allow connections from all IP addresses.**
- **Set the whitelist to `127.0.0.1` to block connections from all IP addresses.**
- **Set the whitelist to a CIDR block to allow connections from the IP addresses within the CIDR block, such as `10.10.10.0/24`.**
- **When you enter multiple IP addresses or CIDR blocks, separate them with commas (,) and leave no space before or after each comma.**

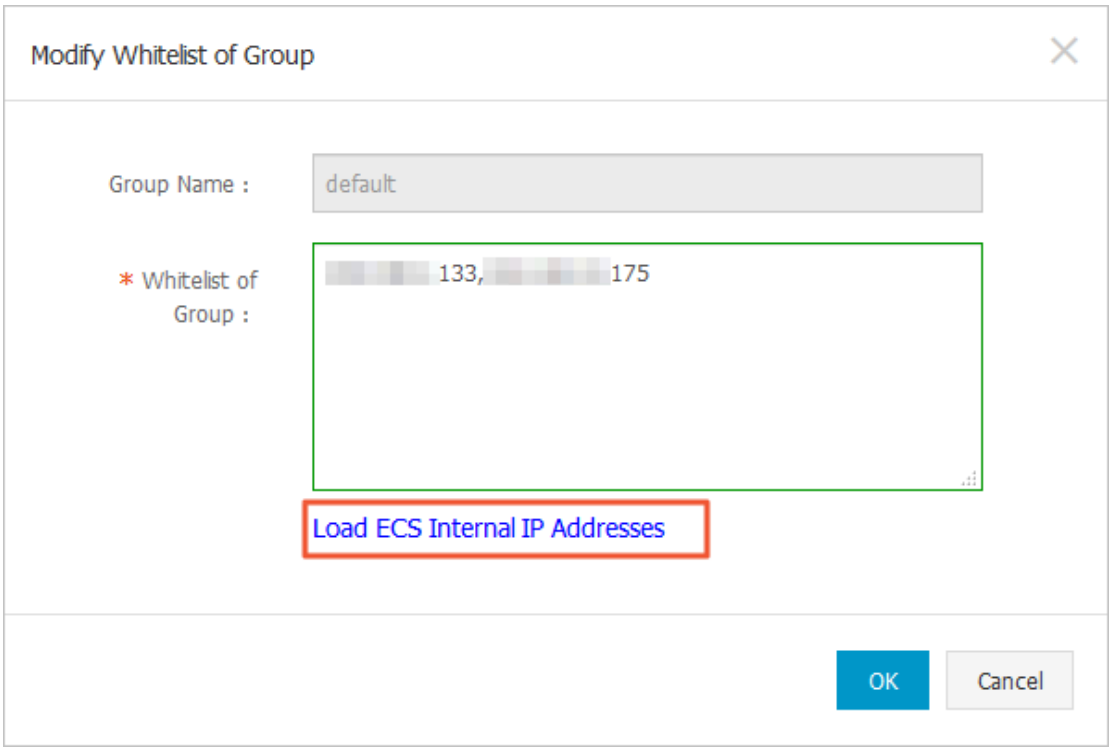
- **You can add 1,000 or fewer IP addresses or CIDR blocks to each whitelist group.**

b. Click OK.

- **Load internal IP addresses of target ECS instances under the current Alibaba Cloud account:**

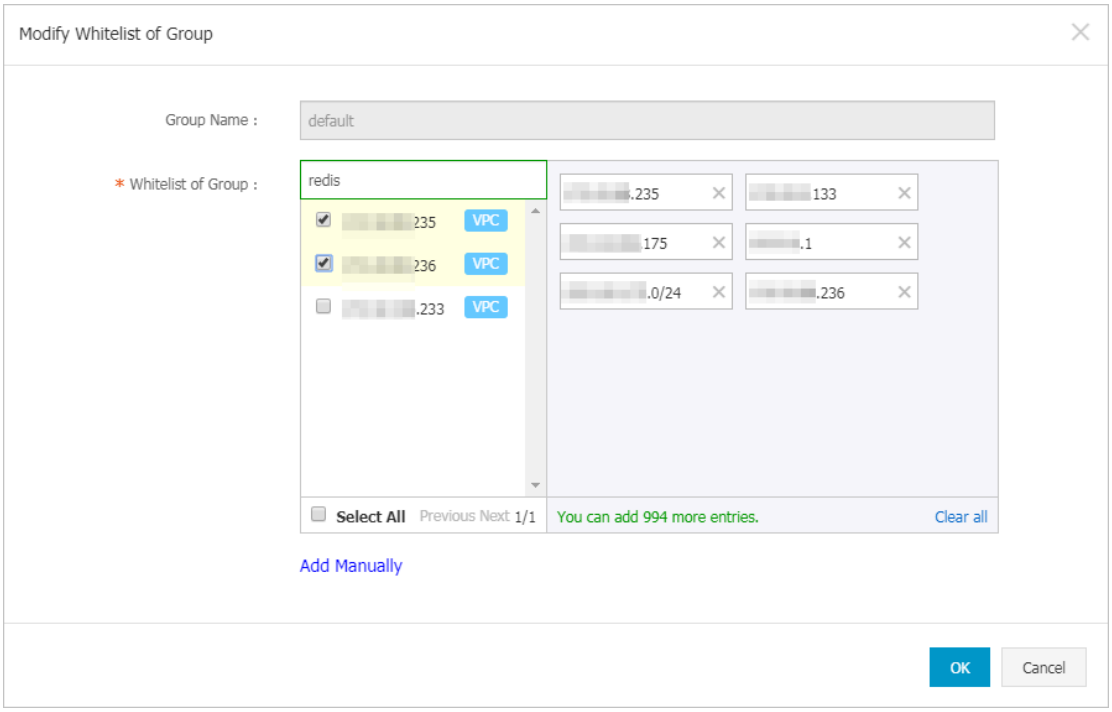
  a. Click Load ECS Internal IP Addresses.

Figure 8-7: Load internal IP addresses of target ECS instances



b. Select internal IP addresses of target ECS instances.

Figure 8-8: Select internal IP addresses of target ECS instances



Note:
You can perform a fuzzy search by ECS instance name, ID, or IP address on the search bar above the list of ECS internal IP addresses.

c. Click OK.

8.3.3 Change the instance configuration

This topic describes how to change the configuration of a KVStore for Redis instance.

Context

Note:
After configuration changes have been completed, the system will migrate data and experience transient disconnection for a few seconds during this process. We recommend that you upgrade or downgrade the instance configuration during off-peak hours.

Procedure

1. *Log on to the KVStore for Redis console*.

2. On the Instance List page, find the target instance. Then, click the instance ID or click Change Configurations in the Actions column.
3. On the Modify Instance page, change the configuration and click Submit.

8.3.4 Set a maintenance window

You can modify the default maintenance window to perform maintenance on KVStore for Redis during off-peak hours.

Context

To ensure the stability of KVStore for Redis instances on the Alibaba Cloud platform , the backend system performs maintenance on instances and servers occasionally.

To guarantee the stability of the maintenance process, instances will enter the Maintaining Instance status before the preset maintenance window on the day of maintenance. While an instance is in this state, data in the database can still be accessed and query operations such as performance monitoring are still available . However, change operations such as configuration change are temporarily unavailable for this instance in the console.

 **Note:**
During the maintenance process, instances may be disconnected in the process of maintenance. We recommend that you set the maintenance window to a period during off-peak hours.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. On the Instance Information page, click Settings to the right of the Maintenance Window field in the Basic Information section.
4. Select time periods and click Save.

 **Note:**
The time periods are in UTC+8.

8.3.5 Upgrade the minor version

Alibaba Cloud has continuously optimized the kernel of KVStore for Redis to fix security vulnerabilities and provide more stable services. You can upgrade the kernel version (minor version) of a KVStore for Redis instance with one click in the console.

Context

Note:

- **We recommend that you upgrade instance versions during off-peak hours and ensure that your application supports automatic reconnection.**
- **The system automatically checks the kernel version of an instance. If the current version is the latest, the Minor Version Upgrade button in the upper-right corner of the Basic Information section for this instance will appear dimmed.**

Procedure

1. *Log on to the KVStore for Redis console*.
2. **On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.**
3. **On the Instance Information page, click Minor Version Upgrade in the upper-right corner of the Basic Information section.**
4. **In the Minor Version Upgrade dialog box that appears, click Upgrade Now.**

   On the Instance Information page, the instance status will become Upgrading a minor version. When the instance status returns to Available, the upgrade has been completed.

8.3.6 Configure SSL encryption

The standard and cluster Instances of Redis 2.8 and the cluster instances of Redis 4.0 support secure sockets layer (SSL) encryption. You can enable SSL encryption to ensure more secure data transmission.

Context

Note:

SSL encryption may increase the network response time of instances. We recommend that you enable this feature only when necessary.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the left-side navigation pane of the Instance Information page, click SSL Settings.
4. In the upper-right corner of the SSL Settings page, click Configure SSL.
5. In the Configure SSL dialog box that appears, turn on the Enable switch. The switch will turn from green to gray when it is enabled. Click OK.
   - If an error message is displayed to indicate that the instance is in an abnormal state, click OK in the message that appears.
   - If an error message is displayed to indicate that the feature is not supported in this version, upgrade the minor version of the instance. For more information, see *Upgrade the minor version*.
   - After the operation, you must wait for a short period of time before the system displays the operation result.
   - You can also click Update Validity and Download CA Certificate in the upper-right corner of the SSL Settings page to perform relevant operations.

8.3.7 Clear data

You can clear the data of a KVStore for Redis instance in the console.

Context

> **Warning:**
> This operation will delete all data contained on an instance. Deleted data cannot be restored. Proceed with caution.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the upper-right corner of the Instance Information page, click Clear Data.
4. In the Clear Data message that appears, click OK.

8.3.8 Release an instance

You can release a KVStore for Redis instance at any time based on your business needs. This topic describes how to release a KVStore for Redis instance.

Procedure

1. *Log on to the KVStore for Redis console*.
2. **On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.**
3. **On the Instance List page, find the target instance. Then, click Release in the Actions column.**

   Warning:

   After an instance is released, it cannot be restored. Proceed with caution. We recommend that you back up your data before releasing the instance.
4. **In the Release Instance message that appears, click OK.**

8.3.9 Manage a database account

KVStore for Redis allows you to create multiple database accounts for an instance. You can grant permissions to these accounts based on the actual usage to flexibly manage your instance and minimize misoperation.

Prerequisites

The engine version of the instance is Redis 4.0 or later.

Note:

The engine version of the instance is not Redis 4.0, only the default account is available. The default account is created when you create the instance. For more information about how to change the password of the default account, see *Change the password*.

Context

You can create accounts, delete accounts, reset the password, and change the permissions. After an account is created, you can use this account to log on to the database and use the command-line tool to perform operations on the database with the account and granted permissions.

Create an account

1. *Log on to the KVStore for Redis console*.
2. **On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.**
3. **On the Instance Information page, click Account Management in the left-side navigation pane.**

   Note:
   If Account Management is unavailable in the left-side navigation pane of an instance with Redis 4.0 or later, you can try to *Upgrade the minor version*.
4. **In the upper-right corner of the Account Management page, click Create.**
5. **In the Create Account dialog box that appears, configure the following parameters and click OK.**

8.3.10 Use a Lua script

KVStore for Redis instances of all editions support Lua commands.

Support for Lua commands

Lua scripts improve the performance of KVStore for Redis. With support for the Lua environment, KVStore for Redis is able to perform check-and-set (CAS) operations, allowing you to combine and run multiple commands in an efficient manner.

Note:
If the Eval command cannot be executed, such as when the "`ERR command eval not support for normal user`" message is displayed, you can try to *Upgrade the minor version*. During the upgrade, the instance may be disconnected and become read-only for a few seconds. We recommend that you upgrade the version of an instance during off-peak hours.

Limits on Lua scripts

To ensure that all operations in a Lua script are performed within the same hash slot, the cluster edition of KVStore for Redis sets the following limits on a Lua script :

Note:

If you want to break the Lua limits of Redis Cluster and can ensure that all operations are performed in the same hash slot in the code, you can set the script_check_enable parameter to 0 in the console to disable the backend script check.

8.3.11 Restart an instance

You can restart an instance from the Instance List page of the console.

Procedure

1. *Log on to the KVStore for Redis console*.
2. **On the Instance List page, find the target instance. Then, click Restart in the Actions column.**

   Notice:
   During the restart, the instance may be disconnected for a few seconds. We recommend that you restart instances during off-peak hours and ensure that your application supports automatic reconnection.
3. **In the dialog box that appears, select a restart time and click OK.**
   - **Restart Immediately: restarts the instance immediately.**
   - **Restart Within Maintenance Window: restarts the instance within the preset** *maintenance window*.

8.3.12 Export the instance list

You can export the list of KVStore for Redis instances from the KVStore for Redis console for offline management.

Procedure

1. *Log on to the KVStore for Redis console*.
2. **In the upper-right corner of the Instance List page, click the Export Instances icon.**
3. **In the Export Instance List dialog box that appears, select the columns to export and click OK.**

   Note:
   After you click OK, the browser begins to download the CSV file. You can use Excel or a text editor to view this file.

8.4 Connection management

8.4.1 View connection strings

You can view the internal and public endpoints of instances in the KVStore for Redis console.

Context

Note:

- **The virtual IP address of a KVStore for Redis instance may change when you maintain or modify the service. To ensure connection availability, we recommend that you use a connection string to access the KVStore for Redis instance.**
- **For more information about how to apply for a public connection string, see** *Applies for a public connection string*.

Procedure

1. *Log on to the KVStore for Redis console*.
2. **On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.**
3. **On the Instance Information page, view Internal Connection Address (Host) and Public Endpoint (Host) in the Connection Information section.**

8.4.2 Applies for a public connection string

This topic describes how to apply for a public endpoint.

Procedure

1. *Log on to the KVStore for Redis console*.
2. **On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.**
3. **On the Instance Information page, click Apply for External IP Address in the Connection Information section.**

4. In the Apply for External IP Address dialog box that appears, enter an endpoint and port number, and click OK.

   Note:

   - The custom endpoint prefix must be 8 to 64 characters in length and can contain lowercase letters and digits. It must start with a lowercase letter.
   - The custom port range is 1024 to 65535. The default value is 6379.
   - After applying for a public endpoint, you must add the public IP address to the whitelist before the instance can be accessed over the Internet. For more information about how to configure a whitelist, see *Configure a whitelist*.

5. On the Instance Information page, view the Public Endpoint in the Connection Information section.

   Note:
   If a public endpoint is no longer needed, you can click Release Public Endpoint next to the Public Endpoint to release the endpoint.

8.4.3 Change the connection string of an instance

KVStore for Redis allows you to modify internal and public endpoints for instances. When changing the KVStore for Redis instance, you can change the endpoint of the new instance to the endpoint of the original instance without modifying the application.

Prerequisites

The instance is in the Running state.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. On the Instance Information page, click Modify Public Endpoint in the Connection Information section.
4. In the Modify Public Endpoint dialog box that appears, set Connection Type, Endpoint, and Port. Click OK.

   Note:

- **The custom endpoint prefix must be 8 to 64 characters in length and can contain lowercase letters and digits. It must start with a lowercase letter.**
- **The custom port range is 1024 to 65535. The default value is 6379.**
- **If Connection Type is set to Internal Address, you cannot set Port.**

8.5 Parameter configuration

KVStore for Redis allows you to customize certain instance parameters. This topic describes parameters and the common methods to modify them in the KVStore for Redis console.

Context

KVStore for Redis is completely compatible with the native database services of Redis. The method to set parameters for KVStore for Redis is similar to that of an on-premises Redis database. You can set the parameters described in this topic in the KVStore for Redis console.

Parameters

Table 8-3: Parameters

Parameter	Description
#no_loose_check-whitelist-always	Specifies whether to check whether the client IP address is in the whitelist of the KVStore for Redis instance after password-free access is enabled in Virtual Private Cloud (VPC). Default value: no. If you set this parameter to yes, the whitelist will still take effect in password-free access mode for VPC. Valid values: · yes · no
#no_loose_disabled-commands	Specifies the disabled commands. Separate multiple commands with commas (,). You can disable the following commands: `FLUSHALL`, `FLUSHDB`, `KEYS`, `HGETALL`, `EVAL`, `EVALSHA`, and `SCRIPT`.

Parameter	Description
#no_loose_ssl-enabled	Specifies whether to enable SSL encryption. Default value: no. Valid values: · yes · no
#no_loose_sentinel-enabled	Specifies whether to enable Sentinel-compatible mode. Default value: no. Valid values: · yes · no
client-output-buffer-limit pubsub	Limits the size of output buffers for Pub/Sub clients. This parameter can contain options in the following format: `<hard limit> <soft limit> <soft seconds>`. · Hard limit: If the output buffer of a Pub/Sub client reaches or exceeds the number of bytes specified by hard limit, the client is immediately disconnected. · soft limit and soft seconds: If the output buffer of a Pub/Sub client reaches or exceeds the size in bytes specified by soft limit for a period of time in seconds specified by soft seconds, the client will be disconnected.
dynamic-hz	Specifies whether to enable dynamic frequency control for background tasks. Default value: yes. Valid values: · yes · no
hash-max-ziplist-entries	Specifies the maximum size of each key-value pair stored within a hash in bytes. A hash is encoded using ziplist when it meets the following conditions: 1. The maximum size of each key-value pair stored within the hash in bytes must be less than the value of the hash-max-ziplist-value parameter. 2. The number of key-value pairs stored within the hash must be less than the value of the hash-max-ziplist-entries parameter.

Parameter	Description
hash-max-ziplist-value	Specifies the maximum size of each key-value pair stored within a hash in bytes. A hash is encoded using ziplist when it meets the following conditions: 1. The maximum size of each key-value pair stored within the hash in bytes must be less than the value of the hash-max-ziplist-value parameter. 2. The number of key-value pairs stored within the hash must be less than the value of the hash-max-ziplist-entries parameter.
hz	Specifies the execution frequency for background tasks, such as tasks to evict expired keys. Valid values : 1 to 500. Default value: 10. The larger the value of the hz parameter, the more frequently background tasks are performed and the more precisely timeout events are handled, but the more CPU KVStore for Redis consumes. We recommend that you do not set the hz parameter to a value greater than 100.
lazyfree-lazy-eviction	Specifies whether to enable lazyfree for the eviction feature. Default value: no. Valid values: · yes · no
lazyfree-lazy-expire	Specifies whether to enable lazyfree to delete expired keys. Default value: yes. Valid values: · yes · no
lazyfree-lazy-server-del	Specifies whether to enable lazyfree to asynchronously delete data with the `DEL` command. Default value: yes. Valid values: · yes · no

Parameter	Description
list-compress-depth	Specifies the number of nodes that are not compressed at each side in a list. Default value: 0. Valid values: · 0: does not compress any nodes in the list. · 1: does not compress the first node from each side of the list, but compresses all nodes in between. · 2: does not compress the first two nodes from each side of the list, but compresses all nodes in between . · 3: does not compress the first three nodes from each side of the list, but compresses all nodes in between. · And so on up to 65535.
list-max-ziplist-size	· Specifies the maximum size of each ziplist in a quicklist. A positive number indicates the maximum number of elements in each ziplist of a quicklist. For example, if you set this parameter to 5 , each ziplist of a quicklist can contain a maximum of five elements. · A negative number indicates the maximum number of bytes in each ziplist of a quicklist. Default value: -2. Valid values: - -5: indicates that each ziplist of a quicklist cannot exceed 64 KB (1 KB = 1,024 bytes). - -4: indicates that each ziplist of a quicklist cannot exceed 32 KB. - -3: indicates that each ziplist of a quicklist cannot exceed 16 KB. - -2: indicates that each ziplist of a quicklist cannot exceed 8 KB. - -1: indicates that each ziplist of a quicklist cannot exceed 4 KB.

Parameter	Description
maxmemory-policy	Specifies the policy used to evict keys if the memory is fully occupied. Valid values: LRU means least recently used. LFU means least frequently used. LRU, LFU, and TTL are implemented by using approximated randomized algorithms. · volatile-lru: evicts the approximated least recently used (LRU) keys among keys with a preset expiration time. · allkeys-lru: evicts the approximated LRU keys. · volatile-lfu: evicts the approximated least frequently used (LFU) keys among keys with a preset expiration time. · allkeys-lfu: evicts the approximated LFU keys. · volatile-random: evicts random keys among keys with a preset expiration time. · allkeys-random: evicts random keys. · volatile-ttl: evicts keys with the nearest time to live (TTL) among keys with a preset expiration time. · noeviction: does not evict any keys, but returns an error on write operations.

Parameter	Description
notify-keyspace-events	Specifies the events that the Redis server can notify clients of. The value of this parameter is any combination of the following characters, each of which specifies a type of event to be notified: · · K: keyspace events, published with the __keyspace@<db>__ prefix. · E: keyevent events, published with the __keyevent@<db>__ prefix. · g: generic commands that are non-type specific, such as DEL, EXPIRE, and RENAME. · l: list commands. · s: set commands. · h: hash commands. · z: sorted set commands. · x: expired key events. An expired key event is generated when a key expires. · e: evicted key events. An evicted key event is generated when a key is evicted due to the policy specified by the maxmemory-policy parameter. · A: the alias for g$lshzxe.
set-max-intset-entries	Specifies the maximum number of data entries in a set. A set is encoded by using intset when it meets the following conditions: 1. The set is composed of just strings. The number of strings is less than the value of this parameter. 2. All strings are integers in radix 10 in the range of 64-bit signed integers.
slowlog-log-slower-than	Specifies whether to log slow queries. · Negative number: does not log slow queries. · 0: logs all queries. · Positive number: logs queries that exceed an execution time specified by this positive number, in microseconds. Valid values: 0 to 10,000,000. Default value: 10,000.

Parameter	Description
slowlog-max-len	Specifies the maximum number of slow query log entries that can be stored. Valid values: 100 to 10,000. Default value: 1,024.
stream-node-max-bytes	Specifies the maximum memory that can be used by each macro node in streams. Valid values: 0 to 999,999,999,999,999. If you set the parameter to 0, each macro node can use an unlimited amount of memory.
stream-node-max-entries	Specifies the maximum number of stream entries that can be stored within each macro node. Valid values: 0 to 999,999,999,999,999. If you set the parameter to 0, each macro node can store unlimited stream entries.
timeout	Specifies a timeout period for client connections. Unit: seconds. Valid values: 0 to 100,000. 0 indicates that client connections never time out.
zset-max-ziplist-entries	Specifies the maximum size of each key-value pair stored within a sorted set in bytes. A sorted set is encoded using ziplist when it meets the following conditions: 1. The maximum size of each key-value pair stored within the sorted set in bytes must be less than the value of the zset-max-ziplist-value parameter. 2. The number of key-value pairs stored within the sorted set must be less than the value of the zset-max-ziplist-entries parameter.
zset-max-ziplist-value	Specifies the maximum size of each key-value pair stored within a sorted set in bytes. A sorted set is encoded using ziplist when it meets the following conditions: 1. The maximum size of each key-value pair stored within the sorted set in bytes must be less than the value of the zset-max-ziplist-value parameter. 2. The number of key-value pairs stored within the sorted set must be less than the value of the zset-max-ziplist-entries parameter.

Parameter	Description
list-max-ziplist-entries	Specifies the maximum size of each key-value pair stored within a list in bytes. A list is encoded using ziplist when it meets the following conditions: 1. The maximum size of each key-value pair stored within the list in bytes must be less than the value of the list-max-ziplist-value parameter. 2. The number of elements stored within the list is less than the value of the list-max-ziplist-entries parameter.
list-max-ziplist-value	Specifies the maximum size of each key-value pair stored within a list in bytes. A list is encoded using ziplist when it meets the following conditions: 1. The maximum size of each key-value pair stored within the list in bytes must be less than the value of the list-max-ziplist-value parameter. 2. The number of elements stored within the list is less than the value of the list-max-ziplist-entries parameter.
cluster_compat_enable	Specifies whether to enable compatibility with the syntax of Redis Cluster. Default value: 1. Valid values: · 0: no · 1: yes
script_check_enable	Specifies whether to confirm that all the keys used in a Lua script are in the same hash slot. Default value: 1. Valid values: · 0: no · 1: yes

Note:

The maxclients parameter, which is used to specify the maximum number of connections to Redis data nodes, is fixed to 10,000. You cannot modify the value of this parameter.

Configure parameters in the KVStore for Redis console

1. *Log on to the KVStore for Redis console*.

2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the left-side navigation pane of the Instance Information page, click System Parameters.
4. Find the target parameter and click Modify in the Action column.
5. In the dialog box that appears, modify the parameter value and click OK.

8.6 Backup and recovery

8.6.1 Back up data automatically

An increasing number of applications use KVStore for Redis for persistent storage. Because of this, KVStore for Redis supports routine backup mechanisms to restore data after misoperations occur. Alibaba Cloud provides secondary nodes to back up .rdb files as snapshots. Backup operations do not affect the performance of your instance. You can customize the backup operation in the console.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the left-side navigation pane of the Instance Information page, click Backup and Recovery.
4. On the Backup and Recovery page, click the Backup Settings tab.
5. Click Edit to customize the automatic backup cycle and backup time.

    **Note:**
   Backup data is retained for seven days. You cannot modify this configuration.
6. Click OK.

8.6.2 Back up data manually

You can initiate a manual backup task in the console at any time.

Procedure

1. *Log on to the KVStore for Redis console*.

2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the left-side navigation pane of the Instance Information page, click Backup and Recovery.
4. In the upper-right corner of the Data Backup tab, click Create Backup.
5. In the message that appears, click Confirm.

> **Note:**
> On the Data Backup tab, you can select a time range to query historical backup data. Backup data is retained for seven days, so you can query historical backup data in the past seven days.

8.6.3 Download backup files

To archive these backup files for a longer period, you can copy the link in the console and manually download the database backup files for local storage.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the left-side navigation pane of the Instance Information page, click Backup and Recovery.
4. On the Data Backup tab, select the backup set to be archived and click Download.

8.6.4 Restore data

You can use backup files to restore data in the console.

Context

> **Notice:**
> - Data restoration is highly risky. Check the data to be restored before performing this operation. Proceed with caution.
> - This feature is not applicable to non-cluster KVStore for Redis instances.

Procedure

1. *Log on to the KVStore for Redis console*.

2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the left-side navigation pane of the Instance Information page, click Backup and Recovery.
4. On the Data Backup tab, select the target backup file and click Restore Data.
5. In the Restore Data dialog box that appears, click Continue.

   You can apply backup files to a new instance by *cloning an instance*.

8.6.5 Clone an instance

You can apply backup files to a new instance by cloning an instance.

Context

> **Note:**
> This feature is applicable to non-cluster KVStore for Redis instances.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the left-side navigation pane of the Instance Information page, click Backup and Recovery.

8.7 Performance monitoring

8.7.1 View monitoring data

You can query the monitoring data of a KVStore for Redis instance for a specified period within the last month.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the left-side navigation pane of the Instance Information page, click Performance Monitor.

4. On the Historical Monitoring Data page, click the calendar icon next to Query Time.
5. Set a query time range and click OK.

> **Note:**
> For more information about metrics, see *Metrics*.

8.7.2 Customize metrics

You can select the metrics to be displayed on the Historical Monitoring Data page of the KVStore for Redis console as needed.

Context

KVStore for Redis supports more than 10 groups of monitoring metrics. By default, the Performance Monitor page displays the monitoring metrics of the basic monitoring group. You can click the Custom Metrics button to switch to the metrics of other monitoring groups. The following table describes the monitoring groups.

Monitoring group	Description
Basic monitoring group	The basic instance monitoring information, such as the QPS, bandwidth, and memory usage.
Key monitoring group	The monitoring information on the use of key-value related commands, such as the number of times DEL and EXITS are called.
String monitoring group	The monitoring information on the use of string-related commands, such as the number of times APPEND and MGET are called.
Hash monitoring group	The monitoring information on the use of hash-related commands, such as the number of times HGET and HDEL are called.
List monitoring group	The monitoring information on the use of list-related commands, such as the number of times BLPOP and BRPOP are called.
Set monitoring group	The monitoring information on the use of set-related commands, such as the number of times SADD and SCARD are called.

Monitoring group	Description
Zset monitoring group	The monitoring information on the use of zset-related commands, such as the number of times ZADD and ZCARD are called.
HyperLog monitoring group	The monitoring information on the use of HyperLogLog-related commands, such as the number of times PFADD and PFCOUNT are called.
Pub/Sub monitoring group	The monitoring information on the use of publication and subscription-related commands, such as the number of times PUBLISH and SUBSCRIBE are called.
Transaction monitoring group	The monitoring information on the use of transaction-related commands, such as the number of times WATCH, MULTI, and EXEC are called.
Lua script monitoring group	The monitoring information on the use of Lua script-related commands, such as the number of times EVAL and SCRIPT are called.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the left-side navigation pane of the Instance Information page, click Performance Monitor.
4. On the Historical Monitoring Data page, click Customize Metrics in the Data Index section.
5. In the Customize Metrics dialog box that appears, select the new monitoring group and click OK.

8.7.3 Modify monitoring frequency

KVStore for Redis console allows you to set the frequency at which monitoring data is collected.

Context

You can set the monitoring frequency to either 5 or 60 seconds to specify how often monitoring data to be collected by KVStore for Redis. The default monitoring time of 60 seconds is sufficient to meet common monitoring requirements. If you

need to observe certain metrics at a higher frequency and lower latency, you can change the monitoring frequency to 5 seconds as described in the following section. Monitoring data does not occupy instance storage space, and collection of monitoring data does not affect normal running of the instance.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. In the left-side navigation pane of the Instance Information page, click Performance Monitor.
4. In the upper-right corner of the Historical Monitoring Data page, click Monitoring Frequency.
5. In the Monitoring Frequency dialog box that appears, select the new monitoring frequency and click OK.

8.8 Alert settings

You can set alert rules for an KVStore for Redis instance based on the performance monitoring data of the instance. If a performance data change triggers an alert rule, you can immediately receive an alert notification.

Context

Monitoring and alerting are implemented through CloudMonitor. CloudMonit or enables you to set metrics and contacts. When the alert rules of a metric are triggered, CloudMonitor will notify all members of the alert contact group. You can maintain contact groups for alert metrics to ensure that the contacts are notified in a timely manner when alerts are reported.

Procedure

1. *Log on to the KVStore for Redis console*.
2. On the Instance List page, find the target instance. Click the instance ID or click Manage in the Actions column.
3. On the Instance Information page, click Alarm Settings in the left-side navigation pane.

4. On the Alarm Settings page, click Alarm Settings in the upper-right corner to go to the CloudMonitor console.
5. On the Alarm Rules page, click Create Alarm Rule.
6. Configure required parameters and click OK.

> **Note:**
> After an alert rule has been created, you can modify, disable, and delete the alert rule on the Alarm Rules page of the CloudMonitor console. You can also view the alert history on the page.

9 Data Transmission Service (DTS)

9.1 What is DTS?

Data Transmission Service (DTS) is a data service provided by Alibaba Cloud that supports data exchange between relational databases, OLAP databases, and other data sources.

DTS supports data migration, real-time data subscription, and real-time data synchronization. DTS can be used in multiple business scenarios, including interruption-free data migration, geo-disaster recovery, cross-border data synchronization, and cache update policies, helping you build a secure, scalable, and highly available data architecture.

- **DTS aims to help you with complex data interactions so that you can focus on upper-layer service development.**
- **DTS supports the following data sources:**
  - **Relational databases: MySQL and Oracle**
  - **OLAP databases: MaxCompute**

9.2 Log on to the DTS console

This topic uses the Google Chrome browser as an example to describe how to log on to the DTS console.

Prerequisites

- **Before logging on to the ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel . The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.**
- **We recommend that you use the Google Chrome browser.**

Procedure

1. **In the address bar, enter the URL used to access the ASCM console. Press Enter.**

2. **Enter your username and password.**

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

    **Note:**
   When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).

3. **Click Login to go to the ASCM console homepage.**
4. **In the top navigation bar, choose Products > Data Transmission Service.**
5. **On the DTS page that appears, select the organization and region, and then click DTS.**

9.3 Data migration

9.3.1 Overview

Data migration allows you to quickly migrate data between multiple data sources. Typical scenarios include data migration to the cloud, data migration between instances within Alibaba Cloud, and database split and scale-out. This section introduces the instance types and data source types supported by the data migration feature of DTS.

Instance types supported by data migration

Table *Data source types supported by data migration* **lists the instance types supported by the data migration feature.**

Data source types supported by data migration

Table *Data source types supported by data migration* **lists the data source types supported by the data migration feature.**

Table 9-1: Data source types supported by data migration

Source database	Schema migration	Full data migration	Incremental data migration
MySQL > ApsaraDB RDS for MySQL	Supported	Supported	Supported
Oracle > ApsaraDB RDS for MySQL	Supported	Supported	Supported

Data migration supports the following source instance types:

- **RDS instances**
- **Oracle instances**
- **User-created databases**

Data migration supports the following destination instance types:

- **RDS instances**

9.3.2 Create a data migration task

The data migration feature provided by DTS allows you to configure a migration task by using three steps. This topic describes how to configure a task to migrate data from a MySQL instance to an ApsaraDB RDS for MySQL instance. You can follow a similar procedure to configure a task to migrate data to or from databases with other storage engines.

Prerequisites

- **You have created the destination database in the destination RDS instance.**

  If the destination database does not exist in the destination RDS instance, DTS automatically creates a destination database during data migration. However, in either of the following two cases, you must manually create the destination database in the RDS console before configuring a migration task:

  - **The database name does not meet the following requirements of the RDS instance: A database name must be 1 to 64 characters in length and can contain lowercase letters, digits, underscores (_), and hyphens (-). It must start with a letter and end with a letter or digit.**
  - **The destination database has a different name from the source database.**

- You have created the migration accounts.

  When configuring a migration task, you must provide the migration accounts of the source and destination instances. For more information about the database permissions required by each storage engine, see the *DTS documentation* .

  If you have not created a migration account for your source MySQL instance, follow the instructions in *Grant syntax* to create a migration account that meets the permission requirements.

  If you have not created a migration account for your destination ApsaraDB RDS for MySQL instance, follow the instructions *in* the Account management section of the ApsaraDB for RDS User Guide. You need to create a migration account and grant this account the read/write permission on the source and destination databases.

After the destination database and migration account are created, you can configure a migration task. Perform the following steps to configure a migration task:

Procedure

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner.
3. In the Create DTS Instances dialog box that appears, select a region, enter the number of data migration instances to be created, and click Create.
4. In the message that appears, click OK.
5. In the migration task list, find the migration task that you created and click Configure Migration Task.

6. **Configure the source and destination database instances, and click Set Whitelist and Next in the lower-right corner.**

   In this step, you need to configure the migration task name and the source and destination database instances. *Table 9-2: Parameters for configuring the migration task* **describes related parameters.**

   Table 9-2: Parameters for configuring the migration task

Type	Parameter	Description
Common parameter	Task Name	DTS automatically generates a name for each task. We recommend that you replace the default name with an informative name for easy identification.
Source database information	Instance Type	Valid values: · RDS Instance · User-Created Database with Public IP Address
	Database Type	Valid values: MySQL and Oracle.
	Hostname/IP Address	The IP address of your on-premises MySQL instance.
	Port Number	The port on which the MySQL instance is listening.
	Database Account	The account of the source database instance.
	Database Password	The password of the source database instance.
Destination database information	Instance Type	Valid values: · RDS Instance · User-Created Database with Public IP Address · PetaData
	RDS Instance ID	The ID of the destination database instance.
	Database Account	The account of the destination database instance.

Type	Parameter	Description
	Database Password	The password of the destination database instance.

7. Select migration types and objects to be migrated.

   In this step, you need to select migration types and objects to be migrated. The parameters are described as follows:

   - Migration types

     Available migration types are schema migration, full data migration, and incremental data migration.

     To perform a full data migration, select both Schema Migration and Full Data Migration for Migration Type.

     To perform a zero downtime migration, select Schema Migration, Full Data Migration, and Incremental Data Migration.
   - Objects to be migrated

     Select the objects to be migrated. Click the right arrow to add the selected objects to the Selected section on the right. An object to be migrated can be a database, table, or column.

     After objects are migrated to the destination database, the object names remain the same as that in the source database by default. If the object you migrate has different names in the source and destination instances, you must use the object name mapping feature provided by DTS. For more information, see *Database, table, and column name mapping*.

8. **Perform a precheck.**

   A precheck is required before you can start the migration task. A migration task can only be started after it passes the precheck.

   If the migration task fails the precheck, click the information icon correspond ing to each failed check item to view details, fix the issue, and run the precheck again.

   Click the information icon corresponding to each failed check item to view the cause and solution.

   After troubleshooting, select the task from the task list and perform the precheck again.

9. **Start the migration task.**

   After the migration task passes the precheck, you can start the migration task and check the migration status and progress in the task list.

Result

Note:
This is the complete procedure for creating the data migration task. You can follow a similar procedure to configure a task for migrating data to or from other types of instances or databases with other storage engines.

9.3.3 Precheck items

9.3.3.1 Source database connectivity

This check item checks whether the DTS server can connect to the source database for migration. DTS creates a connection to the source database by using the JDBC protocol. If the connection fails, the check item fails.

The source database connectivity check may fail for the following reasons:

- **An incorrect account or password is provided when a migration task is created.**

  Diagnostics:

  On any network-ready server that can connect to the source database, use the account and password specified for creating the migration task to connect to the source database through client software. Check whether the connection

succeeds. If an error is reported for the connection and the error message contains Access deny, the account or password is incorrect.

Troubleshooting:

Modify the migration task in the DTS console. Correct the account and password . Then re-run the precheck.

- The migration account of the source database implements access control based on source IP addresses.

  Diagnostics:

  - On any network-ready server that can connect to the source database, use the account and password specified for creating the migration task to connect to the source database through client software. Check whether the connection succeeds. If the connection succeeds, the source database has access restrictions based on IP addresses. Only allowed servers can connect to it. The IP address of the DTS server is not included in the whitelist of the source database, so the DTS server cannot connect to the source database.
  - If the source database is a MySQL database, you can access the source database by using the MySQL client. Run the `select host from mysql.user where user='Migration account', password='Migration account password'` command. If the query result is not %, the IP address of the DTS server is not included in the whitelist of the source database, which results in the connection failure.

  Troubleshooting:

  - If the source database is a MySQL database, run the `grant all on . to 'Migration account'@'%' identified by 'Migration account password';` command in the source database to re-authorize the migration account. Replace the migration account and password in this command with the real ones. After the account is authorized, re-run the precheck.
- A firewall is configured on the source database server.

  Diagnostics: If the source database is installed on a Linux server, run `iptables -L` in the shell to check whether a firewall has been configured on the server. If the source database is installed on a Windows server, perform the following operations: Open the Control Panel, click System and Security. On the System and

Security window that appears, click Windows Firewall. Check whether a firewall has been configured on the server.

Troubleshooting:

Disable the firewall and perform the precheck again.

- **There is no connectivity between the DTS server and the source database.**

  If none of the preceding cases applies, the check item may fail because there is no connectivity between the DTS server and the source database. In this case, contact the DTS engineers on duty.

9.3.3.2 Check the destination database connectivity

This check item checks whether the DTS server can connect to the destination database for migration. DTS creates a connection to the destination database by using the JDBC protocol. If the connection fails, the check item fails.

The destination database connectivity precheck may fail for the following reasons:

- **An incorrect account or password is provided when a migration task is created.**

  Diagnostics:

  On any network-ready server that can connect to the destination database, use the account and password specified for creating the migration task to connect to the destination database through client software. Check whether the connection succeeds. If an error is reported for the connection and the error message contains Access deny, the account or password is incorrect.

  Troubleshooting:

  Modify the migration task in the DTS console, correct the account and password, and perform the precheck again.

- **There is no connectivity between the DTS server and destination database.**

  If you check that the password and account are correct, the check item may fail because there is no connectivity between the DTS server and the destination database. In this case, contact the DTS engineers on duty.

9.3.3.3 Binlog configurations in the source database

Check whether binlogging is enabled for the source database

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether binlogging is enabled for the source database. If this check item fails, binlogging is not enabled for the source database.

Troubleshooting: Set `log_bin=mysql_bin` in the configuration file of the source database to enable binlogging. Restart the source database and re-run the precheck.

Check whether the binlog format is ROW in the source database

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether the binlog format is ROW in the source database. If this check item fails, the binlog format is not ROW in the source database.

Troubleshooting: Run the `set global binlog_format=ROW` command in the source database. Then, re-run the precheck. We recommend that you restart the source MySQL database after the modification. Otherwise, connected sessions may continue to be written in non-ROW mode, resulting in data loss.

Check whether a specified binlog file has been deleted from the source database

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether a specified binlog file has been deleted from the source database. If this check item fails, the binlog file does not exist in the source database.

Troubleshooting: Run the `PURGE BINARY LOGS TO "The name of the binlog file ranking the first place among all binlog files that have not been deleted"` command in the source database. Then, re-run the precheck.

For specific purge file names, see the precheck troubleshooting.

Check whether the binlog_row_image value of the MySQL source database is FULL

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether the binlog_row_image of the

source database is FULL, or whether the full image is recorded. If this check item fails, the binlog file of the source database does not record the full image.

Troubleshooting: Run the `set global binlog_row_image=FULL` command in the source database. Then, re-run the precheck.

9.3.3.4 Referential integrity constraint

This check item checks whether all the parent-child tables with foreign key dependencies among the objects to be migrated have been migrated, to avoid damaging the integrity of foreign key constraints.

If this check item fails, the failure cause is that the "parent table name" parent table on which the "child table name" table to be migrated is dependent has not been migrated.

Troubleshooting:

- **Do not migrate the child tables involved in the failed referential integrity constraint check. Modify the migration task and delete these child tables from the list of objects to be migrated. Then re-run the precheck.**
- **Migrate the parent tables for the child tables involved in the failed referential integrity constraint check. To do so, modify the migration task and add these parent tables to the list of objects to be migrated. Then re-run the precheck.**
- **Delete the foreign key dependencies of the child tables involved in the failed referential integrity constraint check. Modify the source database and delete the foreign key dependencies of these child tables. Then, re-run the precheck.**

9.3.3.5 Existence of Federated tables

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether any storage engines not supported by incremental data migration exist in the source database. Currently, incremental data migration does not support the Federated and the MRG_MyISAM storage engines.

If this check item fails and the error message "The Federated engine is used for the following source tables:" is displayed, the storage engine of some tables in the source database is Federated.

If this check item fails and the error message "The MRG_MyISAM engine is used for the following source tables:" is displayed, the storage engine of some tables in the source database is MRG_MyISAM.

Troubleshooting:

Modify the migration task by deleting the tables with the Federated or MRG_MyISAM storage engine from the list of objects to be migrated. Then create a separate migration task to implement schema migration and full data migration for these tables.

9.3.3.6 Permissions

Check the permissions granted to the migration account of the source database

This check item checks whether the migration account of the source database has the required permissions for data migration. For the migration permissions required by each type of database, see the Data Migration chapter.

Check the permissions granted to the migration account of the destination database

This check item checks whether the migration account of the source database has the required permissions for data migration. For the migration permissions required by each type of database, see the data migration chapter.

9.3.3.7 Object name conflict

This check item checks for duplicate object names in the destination and source database. If this check item fails, an object in the destination RDS instance has the same name as an object to be migrated. This causes the migration to fail.

When this check item fails, an error message is displayed indicating that an object in the destination database has the same name as an object to be migrated from the source database.

Troubleshooting:

- **Use the database and table name mapping feature provided by DTS to migrate the object to be migrated to another object with a different name in the destination database.**
- **In the destination database, delete or rename the object that has the same name as the object to be migrated.**

- **Modify the migration task and delete that object to be migrated from the list of objects to be migrated. Do not migrate this object.**

9.3.3.8 Schema existence

This check item checks whether the database to be migrated exists in the destination RDS instance. If no, DTS creates one automatically. However, under the following circumstances, the automatic database creation fails, and this check item prompts a failure:

- **The database name contains characters other than lowercase letters, digits, underscores (_), and hyphens (-).**

  The cause of the precheck failure is that the name of the source database does not comply with the requirements of RDS.

  Troubleshooting: On the database management page of the RDS console, create a database that complies with the requirements of RDS and grant the migration account the read and write permissions on the new database. Use the database name mapping feature provided by DTS to map the source database to the new database. Then, perform the precheck again.
- **The character set of the database is not UTF8, GBK, Latin1, or UTF-8MB4.**

  The cause of the precheck failure is that the character set of the source database does not comply with the requirements of RDS.

  Troubleshooting: On the database management page of the RDS console, create a database that complies with the requirements of RDS and grant the migration account the read and write permissions on the new database. If the new database and the database to be migrated have different names, you can use the database name mapping feature of DTS to map the database to be migrated to the new database. Then re-run the precheck.
- **The migration account of the destination database has no read and write permissions on the database to be migrated.**

  The cause of the precheck failure is that you are not authorized to operate on the source database.

  Troubleshooting: On the database management page of the RDS console, click the Account Management tab. Grant the migration account the read and write permissions on the source database. Then, perform the precheck again.

9.3.3.9 Source database server_id

This check item is run only when incremental data is to be migrated between MySQL instances. This check item checks whether server-id of the source database is set to an integer greater than 1.

If this check item fails, run the `set global server_id='an integer greater than 1'` **command in the source database. Then run the precheck again.**

9.3.3.10 Source database version

This check item checks whether the version of the source database is supported by DTS. *Table 9-3: Source database types and versions* **lists the source database versions supported by DTS.**

Table 9-3: Source database types and versions

Source database type	Supported version
MySQL	5.0, 5.1, 5.5, and 5.6. Only 5.1, 5.5, and 5.6 are supported for incremental data migration.

When the version check fails, you can only upgrade or downgrade the source database to the versions supported by DTS. Then re-run the precheck.

9.3.4 Migrate data from a local MySQL instance to an ApsaraDB RDS for MySQL instance

You can use DTS to migrate data from a local database to an ApsaraDB RDS for MySQL instance without interrupting the services of applications. This section describes how to migrate data from a local database with a private IP address to an ApsaraDB RDS for MySQL instance.

Background

DTS allows you to perform schema migration, full data migration, and incremental data migration on MySQL databases.

- **Schema migration**

  DTS migrates the schema definition of a local database to the destination instance. Currently, DTS supports schema migration for the following objects: tables, views, triggers, stored procedures, and storage functions.

- **Full data migration**

  DTS migrates all existing data of objects from a local database to the destination instance. If you also select incremental data migration, non-transaction tables without primary keys are locked during the full data migration process. Data cannot be written to these locked tables, and the locking duration depends on the data volume of the tables. The locks are released only after these tables are migrated. In this way, data consistency is guaranteed.

- **Incremental data migration**

  In incremental data migration, data changes made during the migration are updated to the destination instance. If DDL operations are performed during migration, the schema changes are not migrated to the destination instance.

Migration restrictions

Migrating data from a local database to an ApsaraDB RDS for MySQL instance is subject to the following restrictions:

- **DDL operations are not supported during migration.**
- **Event migration is not supported in schema migration.**
- **If you use the object name mapping feature when adding an object to be migrated, other objects associated with this object may fail to be migrated.**
- **When incremental data migration is selected, binlogging must be enabled and binlog_format must be set to ROW for the local MySQL instance. If the local MySQL version is 5.6, binlog_row_image must be set to FULL.**

Prerequisites

The ApsaraDB RDS for MySQL instance has been created, and a whitelist has been configured for it. For more information, see the "Set a whitelist" section of the *ApsaraDB for RDS User Guide* **.**

Prepare local data

Before the migration, create the migration accounts in the local database and the RDS for MySQL instance. You also need to create the database to be migrated in the RDS for MySQL instance, and grant the read and write permissions of the database to the migration account. *Table 9-4: Migration types and required permissions* **lists the permissions required by the migration accounts of the source and destination instances when different migration types are used.**

Table 9-4: Migration types and required permissions

Migration type	Schema migration	Full data migration	Incremental data migration
Local database	select	select	· select · replication slave · replication client
ApsaraDB RDS for MySQL instance	Read and write permissions	Read and write permissions	Read and write permissions

1. Run the following command to create a migration account in the local database:

   ```
   CREATE USER 'username'@'host' IDENTIFIED BY 'password';
   ```

   Parameters:

   - username: The migration account that you want to create.
   - host: The host from which you log on to the database by using the account. As a local user, you can use `localhost` to log on to the database. To log on from any other hosts, you can use the wildcard value `%`.
   - password: The logon password for the account.

   For example, if you want to create account `William` with password `Changme123` for logging on to the local database from any hosts, run the following command:

   ```
   CREATE USER 'William'@'%' IDENTIFIED BY 'Changme123';
   ```

2. **Grant permissions to the migration account in the local database.** *Table 9-4: Migration types and required permissions* **lists the permissions required for the migration account of the local database.**

```
GRANT privileges ON databasename.tablename TO 'username'@'host' WITH
 GRANT OPTION;
```

Parameters:

- **privileges: The operation permissions granted to the account, such as SELECT, INSERT, and UPDATE. If you want to grant all permissions to the account, set the value to `ALL`.**
- **databasename: The database name. If you want to grant all database permissions to the account, set the value to the wildcard value `*`.**
- **tablename: The table name. If you want to grant all table permissions to the account, set the value to the wildcard value `*`.**
- **username: The account to which you want to grant the permissions.**
- **host: The host from which the account is authorized to log on to the database. As a local user, you can use `localhost` to log on to the database. To log on from any other hosts, you can use the wildcard value `%`.**
- **WITH GRANT OPTION: Optional. This parameter enables the account to use the GRANT command.**

For example, if you want to grant all of the database and table permissions to account `William` and use the account to log on to the local database from any hosts, run the following command:

```
GRANT ALL ON *. * TO 'William'@'%';
```

Note:
If you want to perform incremental data migration, follow these steps to enable binlogging for the local database and configure this feature correctly.

3. **Run the following command to check whether binlogging has been enabled:**

   ```
   show global variables like "log_bin";
   ```

   If the query result is `log_bin=OFF`, binlogging has not been enabled for the local database. For synchronous migration of the incremental data generated in the migration process, modify the following parameters in configuration file *`my.cnf`*.

   ```
   log_bin=mysql_binbinlog_format=rowserver_id = integer greater than
   1binlog_row_image=full //When the local MySQL version is later than 5.
   6, this item must be set.
   ```

4. **After the parameters are set, run the following commands to restart the MySQL process:**

   ```
   $ Mysql_dir/bin/mysqladmin-u root-P Shutdown
   $ Mysql_dir/bin/maid &
   ```

   ***`mysql_dir`* is the installation directory of MySQL.**

Procedure

Perform migration after data preparation is completed.

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner. In the Create DTS Instance dialog box that appears, create an instance as prompted.**

   Table 9-5: Parameters in the Create DTS Instance dialog box **describes the related parameters.**

   Table 9-5: Parameters in the Create DTS Instance dialog box

Parameter	Description
Function	The instance function. It is specified by the system. Current value: Data Migration.
Region	The region where the instance is located.
Created Instances	The number of instances to be created.

3. **After setting the parameters, click Create.**

4. **In the migration task list, find the instance that you created and click Configure Migration Task on the right. On the Create Migration Task page that appears, complete the configurations as prompted.**

   Table 9-6: Parameters for creating a migration task **describes the related parameters.**

   Table 9-6: Parameters for creating a migration task

Category	Parameter	Description
-	Task Name	DTS automatically generates a task name for each task by default. You can change the default name to an informative one for easy task identification.
Source database information	Instance Type	The type of the source instance. Select User-Created Database with Public IP Address.
-	Source Instance Region	The region where the source instance is located.
-	Database Type	The type of the database to be migrated. Select MySQL.
-	Hostname or IP Address	The connection address of the database to be migrated.
-	Port	The port number of the database to be migrated. The default port number for a MySQL database is 3306.
	Database Account	The account used to log on to the database to be migrated.
	Database Password	The password of the account used to log on to the database to be migrated.
Destination database information	Instance Type	The type of the destination instance. Select RDS Instance.
	Destination Instance Region	The region where the ApsaraDB RDS for MySQL instance is located. It is the same region as that of the source instance.
	RDS Instance ID	The ID of the ApsaraDB RDS for MySQL instance.

Category	Parameter	Description
	Database Account	The account used to log on to the ApsaraDB RDS for MySQL instance.
	Database Password	The password of the account used to log on to the ApsaraDB RDS for MySQL instance.

Note:
After configuring the source and destination databases, you can click Test Connection to test the connectivity.

5. After setting the parameters, click Set Whitelist and Next to go to the Migration Types and Tasks page.
6. Select migration types. Select the objects to be migrated in the Object to Be Migrated area, and click the right arrow to add the selected objects to the Selected area.

Note:
To modify the name of an object to be migrated in the destination database, move the pointer over the database to be modified in the Selected area. The Edit button is displayed.

7. Click Precheck.

Note:

- A precheck is required before you can start the migration task. A migration task can be started only after it passes the precheck.
- If the precheck fails, click the info icon corresponding to each failed check item to view the failure details, troubleshoot the faults, and re-run the precheck.
- After troubleshooting, select the task from the task list and restart the precheck.

8. After the precheck succeeds, you can start the migration task. After the task starts, you can check the migration status and progress on the Migration Tasks page.

Subsequent operations

The migration accounts have been granted the read and write permissions. For security considerations, we recommend that you delete the accounts from the local database and the ApsaraDB RDS for MySQL instance after the data migration.

9.3.5 Migrate data between RDS instances

This topic describes how to configure a task to use DTS for migrating data between two RDS instances.

DTS allows you to migrate data between two RDS instances. For storage engines that support incremental data migration, you can also migrate data to the destination RDS instance without stopping the services of the source RDS instance. DTS only supports migration between homogeneous databases. For example, you can migrate data between two ApsaraDB RDS for MySQL databases. However, migration between heterogeneous databases is not supported.

Permission requirements

When you use DTS to migrate data between RDS instances, the permissions required for the migration accounts vary with migration types. *Table 9-7: Migration types and required permissions* **lists the permissions required for the migration accounts of the source and destination instances when different migration types are used.**

Table 9-7: Migration types and required permissions

Migration type	Schema migration	Full data migration	Incremental data migration
Source RDS instance	Read/write permissions	Read/write permissions	Read/write permissions
Destination RDS instance	Read/write permissions	Read/write permissions	Read/write permissions

Procedure

This section describes how to use DTS for migrating data between two RDS instances. The source and destination RDS instances can be the same or different, indicating that you can use DTS to migrate data within an RDS instance or between two RDS instances.

Create an RDS instance database

If the destination database does not exist in the destination RDS instance, DTS automatically creates the database during data migration. However, in either of the following two cases, you must manually create the destination database in the RDS console before configuring a migration task:

- The database name does not meet the requirements of RDS: A database name must be 1 to 64 characters in length and can contain lowercase letters, digits, underscores (_), and hyphens (-). It must start with a letter and end with a letter or digit.
- The destination database has a different name from the source database.

In either case, you must create the destination database in the RDS console before configuring the migration task. For more information, see the "Create an RDS instance" section in the RDS User Guide.

Create a migration account

When configuring a migration task, you must provide the migration accounts of the source and destination RDS instances. For the permissions required for the migration accounts, see *Permission requirements*. If you have not created a migration account for your source or destination RDS instance, follow the procedure for creating an RDS instance account. You need to create migration accounts for the source and destination RDS instances, and grant the created accounts read/write permissions for the databases and tables to be migrated.

Configure a migration task

After all of the prerequisites are met, you can start to configure a migration task. To configure a migration task, follow these steps:

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner.
3. In the Create DTS Instances dialog box that appears, select a region, enter the number of data migration instances to be created, and click Create.
4. In the message that appears, click OK.
5. In the migration task list, find the migration task that you created and click Configure Migration Task.

6. **Configure the source and destination database instances, and click Set Whitelist and Next in the lower-right corner.**

   In this step, configure the migration task name, source RDS instance, and destination RDS instance. Parameters are described as follows:

   - **Task name**

     DTS automatically generates a task name for each task. Task names are not required to be unique. You can modify the task name as required. We recommend that you use an informative name for easy identification.
   - **Source instance information**
     - **Instance Type: Select RDS Instance.**
     - **RDS Instance ID: Select the ID of the source RDS instance. DTS supports RDS instances in classic networks and VPCs.**
     - **Database Name: Select the default database name that is used to connect to the source RDS instance.**
     - **Database Account: Enter the account that is used to access the source RDS instance.**
     - **Database Password: Enter the password of the account that is used to access the source RDS instance.**
   - **Destination instance information**
     - **Instance Type: Select RDS Instance.**
     - **Instance Region: Select the region where the destination instance resides.**
     - **RDS Instance ID: Select the ID of the destination RDS instance. DTS supports RDS instances in classic networks and VPCs.**
     - **Database Account: Enter the account that is used to access the destination RDS instance.**
     - **Database Password: Enter the password of the account that is used to access the destination RDS instance.**

7. **Select migration types and objects to be migrated.**

   In this step, you need to select migration types and objects to be migrated. Parameters are described as follows:

   - **Migration types**

     Available migration types are schema migration, full data migration, and incremental data migration.

     To perform a full data migration, select both Schema Migration and Full Data Migration for Migration Type.

     To perform a zero downtime migration, select Schema Migration, Full Data Migration, and Incremental Data Migration.

   - **Objects to be migrated**

     Select the objects to be migrated. Click the right arrow to add the selected objects to the Selected section on the right. An object to be migrated can be a database, table, or column.

     After objects are migrated to the destination database, the object names remain the same as those in the source database by default. If the object you migrate has different names in the source and destination instances, you must use the object name mapping feature provided by DTS. For more information about using this feature, see *Database, table, and column name mapping***.**

8. **Perform a precheck.**

   A precheck is required before you can start the migration task. A migration task can only be started after it passes the precheck.

   If the migration task fails the precheck, click the information icon correspond ing to each failed check item to view details, fix the issue, and run the precheck again.

   Click the info icon corresponding to each failed check item to view the cause and solution.

   After troubleshooting, select the task from the task list and perform the precheck again.

9. **Start the migration task.**

   After the migration task passes the precheck, you can start the migration task and check the migration status and progress in the task list.

   Incremental data migration is a dynamic synchronization process. We recommend that you verify the services in the destination database when data is consistent between the source and destination instances during incremental data migration. If the verification is successful, you can stop the migration task and switch services to the destination database.

   At this point, you have configured the task for migrating data between two RDS instances.

9.3.6 Migrate data from a local Oracle instance to an ApsaraDB RDS for MySQL instance

You can use DTS to migrate data from a local Oracle instance to an ApsaraDB RDS for MySQL instance. DTS supports schema migration, full data migration, and incremental data migration. You can use these three migration types in combination to migrate data from the Oracle instance to the destination instance without interrupting normal services of the source Oracle database. This section describes how to configure a task to use DTS to migrate data from an Oracle instance to an ApsaraDB RDS for MySQL instance without service interruptions.

Background

For data migration from an Oracle instance to an ApsaraDB RDS for MySQL instance, DTS supports schema migration, full data migration, and incremental data migration. The restrictions on each migration type are as follows:

- **Schema migration**

  DTS migrates the schema definitions of objects to the destination instance. Currently, DTS supports schema migration only for tables. For other objects such as views, synonyms, triggers, stored procedures, stored functions, packages, and user-defined data types, schema migration is not supported.

- **Full data migration**

  DTS migrates all existing data of objects from the source database to the destination ApsaraDB RDS for MySQL instance. If you perform only a full data migration, data changes to the source Oracle database during the migration

may not be migrated to the destination ApsaraDB RDS for MySQL instance. Therefore, if you only want to perform a full data migration without migrating the incremental data, we recommend that you stop writing data to the source Oracle instance during the migration to ensure data consistency.

- Incremental data migration

  During an incremental data migration, DTS polls and captures the redo logs generated by the source Oracle instance due to data changes. Then, DTS synchronizes the incremental data (or changed data) to the destination ApsaraDB RDS for MySQL instance in real time. Incremental data migration enables real-time data synchronization from the source Oracle instance to the destination ApsaraDB RDS for MySQL instance.

Permission requirements for migration

When you use DTS to migrate data from an Oracle instance to an ApsaraDB RDS for MySQL instance, the permissions required for the migration accounts vary with migration types. The following table lists the permissions required for the migration accounts of the source and destination instances when different migration types are used.

Migration type	Schema migration	Full data migration	Incremental data migration
Local Oracle instance	Schema owner	Schema owner	SYSDBA
Destination ApsaraDB RDS for MySQL instance	Read and write permissions on the database to be migrated	Read and write permissions on the database to be migrated	Read and write permissions on the database to be migrated

Prerequisites

- The version of the Oracle database to be migrated is 10g, 11g, or 12c.
- Supplemental logging has been enabled for the Oracle instance, and supplemental_log_data_pk and supplemental_log_data_ui have been enabled.
- Archive logging has been enabled for the Oracle instance, and archived logs can be accessed over a specific period.

Data type mappings

Oracle and MySQL have different data types. DTS needs to map the data types of the source and destination instances during schema migration. The following table lists the data type mappings between the source and destination instances.

Oracle data type	MySQL data type	Supported by DTS
varchar2(n [char/byte])	varchar(n)	Yes
nvarchar2[(n)]	national varchar[(n)]	Yes
char[(n [byte/char])]	char[(n)]	Yes
nchar[(n)]]	national char[(n)]	Yes
number[(p[,s])]	decimal[(p[,s])]	Yes
float(p)]	double	Yes
long	longtext	Yes
date	datetime	Yes
binary_float	decimal(65,8)	Yes
binary_double	double	Yes
timestamp[(fractional _seconds_precision)]	datetime[(fractional _seconds_precision)]	Yes
timestamp[(fractional _seconds_precision)]with local time zone	datetime[(fractional _seconds_precision)]	Yes
timestamp[(fractional _seconds_precision)]with local time zone	datetime[(fractional _seconds_precision)]	Yes
clob	longtext	Yes
nclob	longtext	Yes
blob	longblob	Yes
raw	varbinary(2000)	Yes
long raw	longblob	Yes
bfile	—	No
interval year(year_precision) to mongth	—	No

Oracle data type	MySQL data type	Supported by DTS
interval day(day_precision) to second[(fractional_seconds_precision)]	—	No

- For char(n), when the definition length n exceeds 255, DTS automatically converts the type to varchar(n).
- MySQL does not support the following data types in Oracle: bfile, interval year to month, and interval day to second. DTS does not convert these three data types during schema migration. You need to exclude columns of these three data types when you specify the objects to be migrated. If any of these three data types are included in the table to be migrated, the schema migration fails.
- The timestamp data type in MySQL does not contain time zone information, while the following data types in Oracle contain time zone information: timestamp with time zone and timestamp with local time zone. DTS converts these two types of data to the UTC time zone format before storing the data to the ApsaraDB RDS for MySQL instance during the migration.

SQL operations supported for synchronization

During incremental data migration, SQL operations that are supported for synchronization include:

- INSERT, DELETE, and UPDATE
- CREATE TABLE // Partitioned tables or tables with built-in functions are not supported.
- ALTER TABLE ADD COLUMN, ALTER TABLE DROP COLUMN, ALTER TABLE RENAME COLUMN, and ALTER TABLE ADD INDEX
- DROP TABLE
- RENAME TABLE, TRUNCATE TABLE, and CREATE INDEX

Create a migration account

When configuring a migration task, you must provide the migration accounts of the local Oracle instance and the destination ApsaraDB RDS for MySQL instance. For permissions required for the migration accounts, see the *Permission requirements for migration* section.

If you have not created a migration account for your source Oracle instance, follow the instructions in *Oracle GRANT Syntax* to create a migration account that meets the requirements.

For more information about how to create a migration account for the destination ApsaraDB RDS for MySQL instance and grant permissions to the account, see the " Create an account" section of the ApsaraDB for RDS User Guide.

Procedure

The following part describes how to configure a task to use DTS to migrate data from a local Oracle database to an ApsaraDB RDS for MySQL instance.

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner. In the Create DTS Instance dialog box that appears, create an instance as prompted.

   The following table describes the parameters in the Create DTS Instance dialog box.

   Table 9-8: Parameters in the Create DTS Instance dialog box

Parameter	Description
Function	The instance function. It is specified by the system. Current value: Data Migration.
Region	The region where the instance is located.
Created Instances	The number of instances to be created.

3. **In the migration task list, find the instance that you created and click Configure Migration Task on the right. On the Create Migration Task page that appears, complete the configurations as prompted.**

   Configurations:

   - **Task Name**

     DTS automatically generates a task name for each task. Task names do not have to be unique. You can modify the task name as required. We recommend that you use an informative name for easy task identification.
   - **Source instance information**
     - **Instance Type: Select User-Created Database with Public IP Address.**
     - **Database Type: Select Oracle.**
     - **Hostname or IP address: Enter the address for accessing the Oracle instance . This address must allows access from public networks.**
     - **Port: Specify the listening port of the Oracle instance.**
     - **Instance Type: Select Non-RAC Instance or RAC Instance based on the local Oracle data type.**
     - **SID: Specify the SID of the Oracle instance.**

       Note:
       This parameter is displayed only when you select Non-RAC Instance as the instance type.
     - **Service Name: Specify the server name of the instance.**

       Note:

This parameter is displayed only when you select RAC Instance as the instance type.

- **Database Account: Enter the account used to connect to the Oracle instance.**
- **Database password: Enter the password of the account used to connect to the Oracle instance.**

· **Destination instance information**

- **Instance Type: Select RDS Instance.**
- **RDS Instance ID: Select the ID of the destination ApsaraDB RDS for MySQL instance. DTS supports RDS instances in classic networks and VPCs.**
- **Database Account: Enter the account used to connect to the ApsaraDB RDS for MySQL instance.**
- **Database Password: Enter the password of the account used to connect to the ApsaraDB RDS for MySQL instance.**

After completing the configurations, click Set Whitelist and Next in the lower-right corner to configure the whitelist. In this step, DTS adds the IP address of the DTS server to the whitelist of the destination ApsaraDB RDS for MySQL instance. This prevents the connection issue where the DTS server cannot connect to the destination ApsaraDB RDS for MySQL instance for data migration.

4. **Select migration types and objects to be migrated.**

· **Migration types**

- **Schema migration**
- **Full data migration**
- **Incremental data migration**

Note:

- **To perform a zero downtime migration, select schema migration, full data migration, and incremental data migration.**
- **To perform only full data migration, select schema migration and full data migration.**

· **Objects to be migrated**

Select the objects to be migrated. An object to be migrated can be a database, a table, or a column. By default, after an object is migrated to the ApsaraDB RDS for MySQL instance, the object name remains the same as that in the local

Oracle instance. If the object you migrate has different names in the source and destination instances, use the object name mapping feature provided by DTS. For more information, see *Database, table, and column name mapping*.

After selecting migration types and objects to be migrated, click Precheck in the lower-right corner to proceed with a precheck.

5. Perform a precheck.

   A precheck is required before you can start the migration task. A migration task can be started only after it passes the precheck. For more information about the precheck items, see *Precheck items*.

   If the precheck fails, click the info icon corresponding to each failed check item to view the failure details, troubleshoot the faults, and re-run the precheck.

   After troubleshooting, select the task from the task list and restart the precheck.

6. Start the migration task.

   After the precheck succeeds, you can start the migration task and check the migration status and progress in the task list.

   When a migration task enters the incremental data migration stage, it does not automatically stop. Incremental data written to the Oracle instance is automatically synchronized to the destination ApsaraDB RDS for MySQL instance. Incremental data migration is a dynamic synchronization process. We recommend that you verify the services in the destination database when data is consistent between the source and destination instances during the incremental data migration. If the verification is successful, you can stop the migration task and switch the services to the destination database.

   Now, you have configured the task for migrating data from a local Oracle instance to an ApsaraDB RDS for MySQL instance.

Subsequent operations

The migration accounts have been granted the read and write permissions. For security considerations, we recommend that you delete the accounts from the local database and the ApsaraDB RDS for MySQL instance after the data migration.

9.3.7 Migrate data from an on-premises Oracle database to another on-premises Oracle database

Data Transmission Service (DTS) allows you to migrate data from an on-premises Oracle database to another on-premises Oracle database. DTS supports schema migration, full data migration, and incremental data migration. To migrate data between on-premises Oracle databases, you can select all of the supported migration types to ensure service continuity. This topic describes how to configure a task for migrating data from an Oracle database to another Oracle database.

Prerequisites

- **The source and destination Oracle databases are created. Database accounts are created. For more information about permissions required for the database accounts, see** *Permissions required for database accounts*.
- **Supplemental logging, including SUPPLEMENTAL_LOG_DATA_PK and SUPPLEMENTAL_LOG_DATA_UI, is enabled for the source Oracle database. For more information, see** *Supplemental Logging*.
- **The ARCHIVELOG mode is enabled for the source Oracle database. Archived log files are accessible and a suitable retention period is set for archived log files. For more information, see** *Managing Archived Redo Log Files*.
- **The available storage space of the destination Oracle database is larger than the total size of the data in the source Oracle database.**

Migration types

- **Schema migration**

  DTS migrates the schemas of the required objects to the destination database. DTS supports schema migration only for tables. DTS does not support schema migration for the following types of objects: view, synonym, trigger, stored procedure, function, package, and user-defined type.

- **Full data migration**

  DTS migrates historical data of the required objects from the source Oracle database to the destination Oracle database.

  > **Note:**
  > **To ensure data consistency, do not write new data into the source Oracle database during full data migration.**

- **Incremental data migration**

  DTS retrieves redo log files from the source Oracle database. Then, DTS synchronizes incremental data from the source Oracle database to the destination Oracle database in real time.

  > **Note:**
  > **To ensure successful migration of incremental data, you must run the `COMMIT` statement after performing an operation in the source Oracle database.**

Permissions required for database accounts

When you use DTS to migrate data between Oracle databases, the permissions required for the database accounts vary with migration types. The following table lists the permissions that are required for the accounts of the source and destination databases.

Database	Schema migration	Full data migration	Incremental data migration
Source Oracle database	The owner permission for schemas	The owner permission for schemas	The SYSDBA permission
Destination Oracle database	The owner permission for schemas	The owner permission for schemas	The owner permission for schemas

> **Note:**
> **For more information about how to create and authorize an Oracle database account, see *CREATE USER* and *GRANT*.**

Procedure

1. *Log on to the DTS console*.

2. **On the Migration Tasks page, click Create Migration Task in the upper-right corner. In the Create DTS Instances dialog box, configure the required parameters.**

   The following table describes the required parameters.

   Table 9-9: Parameters

Parameter	Description
Feature	The feature specified by the system. In this case, the value is Data Migration.
Region	The region where the source instance resides.
Instances to Create	The number of instances that you want to create.

3. **Find the data migration task and click Configure Migration Task in the Actions column.**
4. **Configure the source and destination databases.**

Section	Parameter	Description
N/A	Task Name	· DTS automatically generates a task name. You do not need to use a unique task name. · We recommend that you use an informative name for easy identification.
Source Database	Instance Type	Select User-Created Database with Public IP Address.
	Instance Region	The region where the source Oracle database is located.
	Database Type	Select Oracle.
	Hostname or IP Address	Enter the endpoint that is used to connect to the source Oracle database.
	Port Number	Enter the port number of the source Oracle database. The default port number is 1521.

Section	Parameter	Description
	Instance Type	Select Non-RAC Instance or RAC Instance based on the architecture of the source Oracle database.
	SID	Enter the system ID (SID) of the source Oracle database.  Note: This parameter is required if you select Non-RAC Instance as the instance type.
	Service Name	Enter the server name of the instance.  Note: This parameter is required if you select RAC Instance as the instance type.
	Database Account	Enter the account that is used to connect to the source Oracle database.
	Database Password	Enter the password for the account that is used to connect to the source Oracle database.  Note: After the source database information is specified, click Test Connectivity next to Database Password to verify whether the specified information is valid. If the specified information is valid, the Passed message appears. If the Failed message appears, click Check in the Failed message. Modify the source database information based on the instructions.
Destination Database	Instance Type	Select User-Created Database with Public IP Address.
	Instance Region	The region where the destination Oracle database is located.
	Database Type	Select Oracle.

Section	Parameter	Description
	Hostname or IP Address	Enter the endpoint that is used to connect to the destination Oracle database.
	Port Number	Enter the port number of the destination Oracle database. The default port number is 1521.
	Instance Type	Select Non-RAC Instance or RAC Instance based on the architecture of the destination Oracle database.
	SID	Enter the system ID (SID) of the destination Oracle database.  Note: This parameter is required if you select Non-RAC Instance as the instance type.
	Service Name	Enter the server name of the instance.  Note: This parameter is required if you select RAC Instance as the instance type.
	Database Account	Enter the account that is used to connect to the destination Oracle database.

Section	Parameter	Description
	Database Password	Enter the password for the account that is used to connect to the destination Oracle database. **Note:** After the destination database information is specified, click Test Connectivity next to Database Password to verify whether the specified information is valid. If the specified information is valid, the Passed message appears. If the Failed message appears, click Check in the Failed message. Modify the destination database information based on the instructions.

5. In the lower-right corner of the page, click Set Whitelist and Next.
6. Select the migration types and objects to be migrated.

Item	Description
Migration types	· To perform only full data migration, select Schema Migration and Full Data Migration. **Note:** To ensure data consistency, do not write new data into the source Oracle database during full data migration. · To migrate data with minimal downtime, select Schema Migration, Full Data Migration, and Incremental Data Migration. **Note:** - Incremental data migration supports only tables that have primary keys or UNIQUE NOT NULL indexes. - Incremental data migration does not support the long data type.

Item	Description
Objects to be migrated	Select the source database from the Available section and click the right arrow icon to add the database to the Selected section. **Note:** · You can select columns, tables, or databases as the objects to be migrated. · After an object is migrated to the destination Oracle database, the name of the object remains the same as that in the source Oracle database. If you want an object to have a different name after the object is migrated to the destination Oracle database, you can use the object name mapping feature provided by DTS. For more information about how to use this feature, see *Database, table, and column name mapping*. · If you use the object name mapping feature on an object, other objects that are dependent on the object may fail to be migrated.
Name batch change	No is selected by default. **Note:** You cannot select Yes for data migration between Oracle databases. To change the names of destination Oracle databases and tables, you can use the object name mapping feature. For more information about how to use this feature, see *Database, table, and column name mapping*.

7. In the lower-right corner of the page, click Precheck.

Note:

- A precheck is performed before you can start the migration task. You can start the data migration task only after the task passes the precheck.
- If the task fails the precheck, click the info icon next to each failed item to view details. Fix the issues based on the cause of failure and run the precheck again.

8. **Wait until the precheck is completed. Then, click Next.**

   After the task is started, you can check the migration status and progress on the Migration Tasks page.

   A migration task does not automatically stop after it reaches the incremental data migration process. Incremental data written to the source Oracle database is automatically synchronized to the destination Oracle database. Incremental data migration is a dynamic synchronization process. When data is consistent between the source and destination databases, we recommend that you verify your business running on the destination database. If the verification is successful, you can stop the migration task and switch your workloads to the destination database.

   You have finished migrating data from an on-premises Oracle database to another on-premises Oracle database.

9.3.8 Migrate data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database

You can use DTS to migrate data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database. DTS supports full data migration and incremental data migration to ensure that the source ApsaraDB RDS for MySQL instance can still provide services during data migration. This topic describes how to configure a task to use DTS for migrating data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database without disrupting your businesses.

Prerequisites

Schema migration is not supported when you migrate data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database. Before migration, you must create a destination Oracle database with the same schema as the source database in the ApsaraDB RDS for MySQL instance.

Permission requirements

When you use DTS to migrate data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database, the permissions required for the migration accounts vary with migration types. The following table lists the permissions required for the migration accounts of the source and destination instances when different migration types are used.

Migration type	Full data migration	Incremental data migration
Source ApsaraDB RDS for MySQL instance	Read/write permissions	Read/write permissions
Destination Oracle instance	Database read/write permissions	Database read/write permissions

Procedure

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Data Migration. On the page that appears, click Create Migration Task in the upper-right corner. In the Create DTS Instances dialog box that appears, create an instance as prompted.

   The following table describes the parameter settings.

   Table 9-10: Parameter descriptions

Parameter	Description
Feature	The feature specified by the system. In this case, the value is Data Migration.
Region	The region where the instance resides.
Instances to Create	The number of instances to be created.

3. In the migration task list, find the instance that you created and click Configure Migration Task.
4. (Optional) Create a name for the task.

   DTS automatically generates a name for every task. Task names are not required to be unique. You can modify the task name. We recommend that you use an informative name for easy identification.
5. Enter information about the source and destination databases. The following table describes the parameter settings.

Database type	Parameter	Description
Source database information	Instance Type	Instance Type: Select RDS Instance as the type of the source instance.
	Instance Region	Select the region where the source instance resides.

Database type	Parameter	Description
	RDS Instance ID	Select the ID of the source database.
	Database Account	Enter an account that has read/write permissions on the source database.
	Database Password	Enter the password of the source database account.
Destination database information	Instance Type	Instance Type: Select User-Created Database with Public IP Address as the type of the destination database.
	Instance Region	Select the region where the destination instance resides.
	Database Type	Select Oracle.
	Hostname or IP Address	Enter the hostname or IP address for accessing the Oracle instance.
	Port Number	The default value is 1521.
	Instance Type	Select Non-RAC Instance or RAC Instance based on the Oracle instance type.
	SID	Specify the system ID (SID) of the Oracle database.
	Database Account	Enter an account that has read/write permissions on the destination database.
	Database Password	Enter the password of the destination database account.

6. Click Test Connectivity and confirm that the test results for both the source and destination databases are Passed.
7. Click Set Whitelist and Next in the lower-right corner of the page.
8. Select a migration type. In the Available section, select the source database and click the right arrow to add the database to the Selected section.
   - To migrate data without stopping the running services, you must select Full Data Migration and Incremental Data Migration.
   - To perform full data migration, select Full Data Migration.

9. Click Precheck and wait until the precheck is complete.

   > **Note:**
   > For more information about the precheck items, see *Precheck items*. If the migration task fails the precheck, click the info icon corresponding to each failed check item to view details, fix the issue, and run the precheck again.

10. Click Next to start the migration task.

    After the task is started, you can check the migration status and progress in the task list.

    > **Note:**
    > A migration task does not automatically stop after it reaches the incremental data migration stage. Incremental data written to the ApsaraDB RDS for MySQL instance is automatically synchronized to the destination Oracle instance. Incremental data migration is a dynamic synchronization process. We recommend that you verify the services running on the destination database when data is consistent between the source and destination instances during incremental data migration. If the verification is successful, you can stop the migration task and switch services to the destination database.

11. At this point, you have configured the task for migrating data from an ApsaraDB RDS for MySQL instance to an on-premises Oracle database.

9.3.9 Database, table, and column name mapping

This topic describes how to use the object name mapping feature when you configure a data migration task.

The data migration feature provided by DTS supports object name mapping. Objects to be migrated, such as databases, tables, and columns, can have different names in the source and destination instances.

Database name mapping

If a database you migrate has different names in the source and destination instances, you can map the database names by using the object name mapping feature of DTS.

You can configure the database name mapping feature in Step 2 "Select migration types and objects to be migrated" when you configure the migration task. Perform the following steps to configure the database name mapping feature:

1. **In the Selected area, move the pointer over the row of the object that requires database name mapping. The Edit button appears on the right.**
2. **Modify the database name.**

   If you want the database name to change to jiangliutest after the database is migrated to the destination instance, click Edit to open the Edit Database Name dialog box.

   In the Edit Database Name dialog box, modify the database name directly. The database is stored under the new name in the destination instance.

   Assume that the original database name is amptest.

   In the Edit Database Name dialog box, change amptest to jiangliutest, so that the database name changes to jiangliutest after the database is migrated to the destination instance.

   The database uses the new name in the destination instance.

Table name mapping

If a table you migrate has different names in the source and destination instances, you can map the table names by using the object name mapping feature of DTS.

If you want to use the table name mapping feature, do not select the entire database as the object to be migrated. Instead, select a specific table.

Besides tables, other schema objects such as views, stored procedures, stored functions, and synonyms are also available for object name mapping in the similar way.

You can configure the table name mapping feature in Step 2 "Select migration types and objects to be migrated" when you configure the migration task. Perform the following steps to configure the table name mapping feature:

1. **In the Selected area, move the pointer over the row of the object that requires table name mapping. The Edit button appears on the right.**

2. **Modify the table name.**

   If you want the table name to change from amptest to jiangliutest after the table is migrated to the destination instance, click Edit to open the Edit Table Name dialog box.

   In the Edit Table Name dialog box, modify the table name directly. The table is stored under the new name in the destination instance.

   Assume that the original table name is amptest.

   Change amptest to jiangliutest, so that the table name changes to jiangliutest after the table is migrated to the destination instance.

Column name mapping

If columns of a table that you migrate have different names in the source and destination instances, you can use the object name mapping feature provided by DTS.

You can configure the column name mapping feature in Step 2 "Select migration types and objects to be migrated" when you configure the migration task. If you want to modify the name of a column to be migrated, do not select the entire database as the object to be migrated. Instead, select the table that the column belongs to. Perform the following steps to configure the column name mapping feature:

1. **Assume that you want to change the name of a column in the sbtest1 table. In the Selected area, move the pointer over the row of the sbtest1 table. The Edit button appears on the right.**
2. **Click Edit to open the Edit Table Name dialog box.**

   Modify the column name. After the modification, the column is stored under the new name in the destination database.

   Now, you have configured the column name mapping feature.

9.3.10 Configure an SQL filter for filtering the data to be migrated

This section describes how to configure an SQL filter for filtering migration data when you create a migration task.

DTS allows you to configure an SQL filter to filter the table data to be migrated. The SQL filter applies only to the configured table. DTS filters the data in the table

of the source database based on this filter. Only data that meets this filter can be migrated to the destination database. This feature is applicable to multiple scenarios such as regular incremental data migration and table partitioning.

Functional restrictions

The SQL filter applies only to full data migration. If you select Incremental Data Migration as the migration type, the SQL filter does not apply.

Configure an SQL filter

You can configure an SQL filter in the Migration Types and Tasks step of migration task configuration.

If you want to configure an SQL filter for table migration, you must select a specific table instead of the entire database as the object to be migrated. The following part describes how to configure an SQL filter.

Configure an SQL filter

1. **In the Migration Types and Tasks step, move the pointer over the table for which you want to create an SQL filter in the Selected area. The Edit button appears.**
2. **Click Edit to configure a filter.**

Modify an SQL filter

Filters in DTS are the same as the standard SQL WHERE conditions for databases and support calculation and simple functions.

Enter an SQL filter in the text box as needed.

Now, you have configured an SQL filter.

9.3.11 Troubleshoot migration errors

DTS provides the feature of online troubleshooting in multiple stages to fix migration errors. These stages include:

- **Schema migration**

  DTS supports data migration between heterogeneous data sources. If you import data of unsupported types to the destination instance during a schema migration , the migration fails.

- **Full data migration**

  During full data migration, the migration task may fail because the destination RDS instance does not have sufficient space or required IP addresses have been deleted from the whitelist. In this case, you can modify the task configurations and then restart the task.

DTS provides the online troubleshooting feature that allows you to resume a failed task when an error occurs during migration. The following sections describe how to troubleshoot errors that occur during schema migration and full data migration.

Troubleshoot errors occurred during schema migration

If a schema migration task fails, the task status changes to Migration Failed and the Rectify button appears.

Click Rectify next to a failed object.

Click Rectify next to each failed object. A troubleshooting dialog box appears.

Modify the schema definition based on the cause of failure. Click Rectify after you complete the modification and re-import the modified definition to the destination instance.

If the error persists after you click Rectify, the cause of failure changes to Troubleshooting Failed and the cause of troubleshooting failure is displayed. You need to continue troubleshooting based on the cause of troubleshooting failure until the troubleshooting is successful.

The details page of the schema migration appears after troubleshooting is successful, and the status of the object changes to Finished.

The task resumes after issues with all objects are rectified. For example, the task resumes by proceeding to the full data migration stage.

Troubleshoot errors occurred during full data migration

DTS provides the troubleshooting and retry feature for the following causes of failures:

- **If you fail to connect to the source or destination database, retry the task after you ensure that the network connection is established.**
- **If a connection to the source or destination database times out, retry the task after you ensure that the network connection is established.**

- **If the destination RDS instance does not have sufficient space or the instance is locked, retry the task after you scale up the RDS instance or clean up the instance log space.**
- **If MyISAM of the source database is corrupted, retry the task after troublesho oting.**

For other circumstances, if full data migration fails, DTS only offers the Ignore option. You can ignore the failed object and continue the migration of other objects .

If a full data migration task fails, the status of the task changes to Migration Failed and the Rectify button appears.

When a migration task fails, click Rectify next to a failed object.

If you encounter the preceding failures and the migration tasks can be retried, troubleshoot the errors as prompted. Then, click the Retry button on the full data migration details page to continue the data transfer in the task.

For other causes of failures, DTS only supports the Ignore operation to ignore the full data migration of the object. After you click Ignore, data of this object is not migrated, but data of other objects is migrated to the destination instance.

9.4 Data synchronization

9.4.1 Create a real-time synchronization task

DTS provides a user-friendly real-time data synchronization feature. You can configure a subscription channel with only three steps. This section describes how

to use DTS to quickly create a synchronization task between two ApsaraDB RDS for MySQL instances for real-time synchronization of RDS incremental data.

Synchronization restrictions

- **Synchronization mode**

  Currently, DTS supports only the following modes for real-time synchronization between ApsaraDB RDS for MySQL instances:

  - **From A to B: unidirectional synchronization between two instances. The synchronized objects must be read-only in instance B. Otherwise, a synchronization channel exception may occur.**
  - **From A to B, C, and D: one-to-many distributed synchronization mode. This synchronization mode poses no restrictions on the number of destination RDS instances, but requires that the synchronized object be read-only in the destination instance to avoid synchronization channel exceptions.**
  - **From B, C, and D to A: many-to-one data convergence mode. In the many-to-one mode, the objects to be synchronized through each synchronization channel must be different to guarantee full synchronization.**

  DTS does not support the following modes:

  - **From A to B to C: cascading synchronization mode.**
  - **Between A and B: bidirectional synchronization mode between two instances.**

    If you select any other unsupported synchronization modes during the synchronization channel configuration, the complicated topologies check item in the precheck fails.

- **Incompatible triggers**

  When the object to be synchronized is an entire database and the database contains a trigger that updates the synchronization table, the synchronized data may be inconsistent.

  Suppose that the source instance contains table A and table B. Table A has a trigger that inserts a row of data into table B after the row of data is inserted into table A. The synchronization task synchronizes data from the source instance to the destination instance, where table A and table B are respectively represented as table A' and table B'. During synchronization, the destination instance continuously replicates a full amount of data from the source instance

, including table A and table B. If you insert a row (1) into table A, the trigger in table A inserts a row (2) into table B. The row you inserted (1) and the row inserted by the trigger (2) are synchronized respectively to table A' and table B ' in the destination instance. When the row you inserted (1) is updated to table A', the trigger in table A' inserts another row (2') in table B'. As a result, one row insertion (1) in the source table A triggers two row insertions (2 and 2') in the destination table B'. Therefore, the data between the source and destination instances is inconsistent.

To solve this problem, you must delete the trigger in the destination instance and synchronize table B from the source instance.

Prerequisites

Before configuring a synchronization task, make sure that the source and destination RDS instances exist. If the instances are unavailable, create them first.

Procedure

The following part describes the procedure for creating a synchronization channel between RDS instances:

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Data Synchronization.
3. On the Data Synchronization page, click Create Synchronization Task in the upper-right corner to configure a synchronization task.
4. In the dialog box that appears, set the parameters and click Create.

   Note:
   Currently, DTS only supports the following instance types: MySQL and MaxCompute.
5. In the message that appears, click OK.
6. On the page where synchronization tasks are listed, select the region to which the synchronization task you created belongs.
7. Find the synchronization task you created and click Configure Synchronization Channel.

8. In the Select Source and Destination Instances for Synchronization Channel step, configure the source and destination instances to be connected through the synchronization channel.

   Configure the following information:

   - Synchronization Task Name

     The synchronization task name does not have to be unique. We recommend that you use an informative name for easy task identification and management .

   - Source instance information
     - Instance Type: The type of the source instance. Select RDS Instance.
     - Instance Region: The region where the source instance is located.
     - Instance ID: The ID of the source instance. For a MaxCompute instance, set the value to a project name.
   - Destination instance information
     - Instance Type: The type of the destination instance. Select RDS Instance.
     - Instance Region: The region where the destination instance is located.
     - Instance ID: The ID of the destination instance. For a MaxCompute instance , set the value to a project name.

   > **Note:**
   > The source and destination RDS instances must be different. When you select a value for Instance ID, only the IDs of the ApsaraDB RDS for MySQL instances under the current logon account are displayed in the drop-down list.

   After completing the configuration, click Set Whitelist and Next.
9. Create a synchronization account and click Next.
10. Select the objects to be synchronized in the Source Database Objects area, and click the right arrow to add the selected objects to the Selected area. Click Next.

    DTS allows you to select the objects to be synchronized at the granularity down to table level. You can choose to synchronize certain databases or tables.

    If you select an entire database as the object to be synchronized, the schema change operations, such as CREATE TABLE and DROP VIEW operations,

performed on all the objects in the database are synchronized to the destination database.

If you select a table as the object to be synchronized, only the `DROP TABLE, ALTER TABLE, TRUNCATE TABLE, RENAME TABLE, CREATE INDEX, and DROP INDEX` operations performed on the table are synchronized to the destination database.

Note that the RENAME TABLE operation may result in data inconsistency between the source and destination instances. Suppose that only table A needs to be synchronized from the source instance. If you perform the `rename A to B` operation in the source instance during the synchronization, the subsequent operations performed on the renamed table B are not synchronized to the destination instance. To solve this problem, you can synchronize the entire source database where table A and table B are located.

11. Configure the initial synchronization.

    DTS first performs the initial synchronization when you start a synchronization channel. During initial synchronization, the schemas and data of the objects to be synchronized are replicated from the source instance to the destination instance. These schemas and data are then used as the baseline for subsequent incremental data synchronization.

    Two options are available for initial synchronization: Initial Schema Synchronization and Initial Full Data Synchronization. You must select both Initial Schema Synchronization and Initial Full Data Synchronization by default.

12. ClickPrecheck to start a precheck.

    After the precheck is successful, you can click Start to start the synchronization task.

    After the synchronization task is started, the synchronization task list is displayed. The newly started synchronization task is now in the Initial Synchronization state. The duration of the initial synchronization depends on the data size of the objects to be synchronized from the source instance. After initial synchronization is completed, the synchronization channel enters the Synchronizing state. Now, the synchronization channel between the source and destination instances is established.

9.4.2 Synchronize data between RDS instances in real time

This section describes how to configure a task to use DTS to synchronize data between two Real-time data synchronization between ApsaraDB RDS for MySQL instances under the same account ApsaraDB RDS for MySQL instances in real time.

Supported functions

Real-time data synchronization between ApsaraDB RDS for MySQL instances under the same account.

Synchronization restrictions

- **Synchronization mode**

  Currently, DTS supports only the following modes for real-time synchronization between ApsaraDB RDS for MySQL instances:

  - **From A to B: unidirectional synchronization between two instances. The synchronized objects must be read-only in instance B. Otherwise, a synchronization channel exception may occur.**
  - **From A to B, C, and D: one-to-many distributed synchronization mode. This synchronization mode poses no restrictions on the number of destination RDS instances, but requires that the synchronized object be read-only in the destination instance to avoid synchronization channel exceptions.**
  - **From B, C, and D to A: many-to-one data convergence mode. In the many-to-one mode, the objects to be synchronized through each synchronization channel must be different to guarantee full synchronization.**

  DTS does not support the following modes:

  - **From A to B to C: cascading synchronization mode.**
  - **Between A and B: bidirectional synchronization mode between two instances.**

    If you select any other unsupported synchronization modes during the synchronization channel configuration, the complicated topologies check item in the precheck fails.

- **Functional restrictions**
  - **Incompatible trigger**

    When the object to be synchronized is an entire database and the database contains a trigger that updates the table to be synchronized, the synchronized data may be inconsistent.

    Suppose that source instance A contains table A and table B. Table A has a trigger that inserts a row of data into table B after the row of data is inserted into table A. The synchronization task synchronizes data from the source instance to the destination instance, where table A and table B are respectively represented as table A' and table B'. During synchronization, the destination instance continuously replicates a full amount of data from the source instance, including table A and table B. If you insert a row (1) into table A, the trigger in table A inserts a row (2) into table B. The row you inserted (1) and the row inserted by the trigger (2) are synchronized respectively to table A' and table B' in the destination instance. When the row you inserted (1) is updated to table A', the trigger in table A' inserts another row (2') in table B'. As a result, one row insertion (1) in the source table A triggers two row insertions (2 and 2') in the destination table B'. Therefore, the data between the source and destination instances is inconsistent.

    To solve this problem, you must delete the trigger from the destination instance and synchronize table B from the source instance.
  - **Restrictions on the RENAME TABLE operation**

    The RENAME TABLE operation may cause data inconsistency between the source and destination instances. Suppose that only table A needs to be synchronized from the source instance. If you perform the `rename A to B` operation in the source instance during the synchronization, the subsequent operations performed on the renamed table B will not be synchronized to the destination instance. To solve this problem, you can synchronize the entire source database where table A and table B are located.

Prerequisites

Before configuring a synchronization task, make sure that the source and destination RDS instances exist. If the instances are unavailable, create them first.

Procedure

The following part describes the procedure for creating a synchronization channel between RDS instances:

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **On the Data Synchronization page that appears, click Create Synchronization Task in the upper-right corner to configure a synchronization task.**
4. **In the dialog box that appears, set the parameters and click Create.**

   Note:
   Currently, DTS only supports the following instance types: MySQL and MaxCompute.
5. **In the message that appears, click OK.**
6. **On the page where synchronization tasks are listed, select the region to which the synchronization task you created belongs.**
7. **Find the synchronization task you created and click Configure Synchronization Channel.**

8. **In the Select Source and Destination Instances for Synchronization Channel step, configure the source and destination instances to be connected through the synchronization channel.**

   Configure the following information:

   - **Synchronization Task Name**

     The synchronization task name does not have to be unique. We recommend that you use an informative name for easy task identification and management .

   - **Source instance information**
     - **Instance Type: The type of the source instance. Select RDS Instance.**
     - **Instance Region: The region where the source instance is located.**
     - **Instance ID: The ID of the source instance. For a MaxCompute instance, set the value to a project name.**
   - **Destination instance information**
     - **Instance Type: The type of the destination instance. Select RDS Instance.**
     - **Instance Region: The region where the destination instance is located.**
     - **Instance ID: The ID of the destination instance. For a MaxCompute instance , set the value to a project name.**

   Note:
   The source and destination RDS instances must be different. When you select a value for Instance ID, only the IDs of the ApsaraDB RDS for MySQL instances under the current logon account are displayed in the drop-down list.

   After completing the configuration, click Set Whitelist and Next in the lower-right corner.

9. **Create a synchronization account and click Next.**

10. Select the objects to be synchronized in the Source Database Objects area, and click the right arrow to add the selected objects to the Selected area. Click Next.

    DTS allows you to select the objects to be synchronized at the granularity down to table level. You can choose to synchronize certain databases or tables.

    If you select an entire database, all schema change operations, such as CREATE TABLE and DROP VIEW, performed on all the objects in the database are synchronized to the destination database.

    If you select a table, only the `DROP TABLE, ALTER TABLE, TRUNCATE TABLE, RENAME TABLE`, `CREATE INDEX, and DROP INDEX` operations performed on this table are synchronized to the destination database.

    Note that the RENAME TABLE operation may result in data inconsistency between the source and destination instances. Suppose that only table A needs to be synchronized from the source instance. If you perform the `rename A to B` operation in the source instance during the synchronization, the subsequent operations performed on the renamed table B will not be synchronized to the destination instance. To solve this problem, you can synchronize the entire source database where table A and table B are located.

11. Configure the initial synchronization.

    DTS first performs the initial synchronization when you start a synchroniz ation channel. During initial synchronization, the existing schemas and data of the objects to be synchronized are replicated from the source instance to the destination instance. These schemas and data are then used as the baseline for subsequent incremental data synchronization.

    Two options are available for initial synchronization: Initial Schema Synchronization and Initial Full Data Synchronization. You must select both Initial Schema Synchronization and Initial Full Data Synchronization by default.

12. Click Precheck to start a precheck.

    After the precheck is successful, you can click Start to start the synchronization task.

    After the synchronization task is started, the synchronization task list is displayed. The newly started synchronization task is now in the Initial Synchronization state. The duration of initial synchronization depends on

the data size of the objects to be synchronized from the source instance. After initial synchronization is completed, the synchronization channel enters the Synchronizing state. Now, the synchronization channel between the source and destination instances is established.

9.4.3 Synchronize data from an RDS instance to a MaxCompute instance in real time

This topic describes how to use Data Transmission Service (DTS) to create a task for real-time data synchronization from an RDS instance to a MaxCompute instance. This facilitates real-time data analysis.

Feature

- **DTS supports real-time data synchronization from an ApsaraDB RDS for MySQL instance to a MaxCompute instance under the same Alibaba Cloud account.**
- **DTS supports RDS instances in the classic network and VPCs.**

Objects to be synchronized

DTS only supports the synchronization of tables. Other types of objects are not supported.

Synchronization process

The synchronization process consists of two stages:

1. **Initial full data synchronization.**

   In this stage, all data stored in the ApsaraDB RDS for MySQL instance is replicated to the MaxCompute instance. Data in each synchronized table is replicated and stored independently to a full data table in the MaxCompute instance. The default table name is <source table name>_base. For example , if the source table is named t1, then the destination full data table in the MaxCompute instance is named t1_dts_base. When configuring the synchroniz ation task, you can modify the name prefixes of full data tables.
2. **Incremental data synchronization.**

   In this stage, all the incremental data in the ApsaraDB RDS for MySQL instance is synchronized to the MaxCompute instance in real time. The incremental data is stored in incremental data tables, and each synchronized table corresponds to an incremental data table. The default name of an incremental data table stored

in MaxCompute is <source table name>_log. When configuring the synchronization task, you can modify the name prefixes of incremental log tables.

Incremental log tables store both the incremental data and specific metadata. *Table 9-11: Schema* defines the schema of an incremental log table.

Table 9-11: Schema

record_id	operation_flag	utc_timestamp	before_flag	after_flag	col1	…	colN
1	I	1476258462	N	Y	1	…	JustInsert
2	U	1476258463	Y	N	1	…	JustInsert
2	U	1476258463	N	Y	1	…	JustUpdate
3	D	1476258464	Y	N	1	…	JustUpdate

Where:

- record_id: the unique ID of an incremental log entry. This field auto-increments for each new log entry. If an UPDATE operation is performed on a data record, two log entries are generated for an INSERT operation and a DELETE operation, separately. The two log entries have the same record_id.
- operation_flag: the operation type of the incremental log entry.

  Valid values:

  - I: an INSERT operation.
  - D: a DELETE operation.
  - U: an UPDATE operation.
- dts_utc_timestamp: the operation timestamp of the incremental log. It is also the timestamp of the binary log file for the UPDATE record. The timestamp is in the UTC format.
- before_flag: indicates whether the column values that follow the incremental log entry are pre-update values. Valid values: Y and N. The value of before_flag is Y if the column values that follow the log entry are pre-update values. The value of before_flag is N if the column values are post-update values.
- after_flag: indicates whether the column values that follow the incremental log entry are post-update values. Valid values: Y and N. The value of after_flag

is N if the column values that follow the log entry are pre-update values. The value of after_flag is Y if the column values are post-update values.

For different operation types, before_flag and after_flag of an incremental log entry are defined as follows:

- INSERT operation

record_i	operation_flag	utc_timestamp	before_flag	after_flag	col1	…	colN
1	I	1476258462	N	Y	1	…	JustInsert

For an INSERT operation, the column values that follow an incremental log entry are the newly inserted record values, that is, post-update values. Therefore, the value of before_flag is N and the value of after_flag is Y.

- UPDATE operation

record_i	operation_flag	utc_timestamp	before_flag	after_flag	col1	…	colN
2	U	1476258463	Y	N	1	…	JustInsert
2	U	1476258463	N	Y	1	…	JustUpdate

When an UPDATE operation is performed, two incremental log entries are generated. The two incremental log entries have the same record_id, operation_flag, and dts_utc_timestamp.

The second log entry records the pre-update values, so the value of before_flag is Y and the value of after_flag is N.

The second log entry records the post-update values, so the value of before_flag is N and the value of after_flag is Y.

- DELETE operation

record_i	operation_flag	dts_utc_timestamp	before_flag	after_flag	col1	…	colN
3	D	1476258464	Y	N	1	…	JustUpdate

For a DELETE operation, the column values that follow an incremental log entry are the deleted record values, that is, pre-update values. Therefore, the value of before_flag is Y and the value of after_flag is N.

For each table synchronized from an RDS instance to a MaxCompute instance, a full data table and an incremental data table are generated in the MaxCompute instance. To retrieve the full data of a specific table at a specific time, you must merge the corresponding full data table and incremental data table in the MaxCompute instance. The merging procedure will be described later.

Configure a synchronization task

This section describes how to configure a task for synchronizing data from an RDS instance to a MaxCompute instance in real time.

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **On the Data Synchronization page that appears, click Create Synchronization Task in the upper-right corner to configure a synchronization task.**
4. **In the dialog box that appears, set the parameters and click Create.**

   Note:
   DTS only supports the following instance types: MySQL, AnalyticDB, and MaxCompute.
5. **In the dialog box that appears, click OK.**
6. **On the Synchronization Tasks page, select the region where the synchronization instance resides.**
7. **Find the created synchronization task and click Configure Synchronization Channel in the Actions column.**

8. **In the Configure Source and Destination Instances in Synchronization step, configure the source and destination instances to be connected through the synchronization channel.**

   The parameters are described as follows:

   - **Synchronization Task Name**

     The synchronization task name is not required to be unique. We recommend that you use an informative name to help you identify and manage the synchronization channel.
   - **Source Instance Details**
     - **Instance Type: the type of the source RDS instance. Only ApsaraDB RDS for MySQL is supported. In this example, select RDS Instance.**
     - **Instance Region: the region where the source RDS instance resides.**
     - **Instance ID: the ID of the source RDS instance.**
   - **Destination Instance Details**
     - **Instance Type: the type of the destination instance. RDS for MySQL, MaxCompute (formerly ODPS), and DataHub are supported. When you configure a synchronization channel from an RDS instance to a MaxCompute instance, select MaxCompute as the destination instance type.**
     - **Instance Region: the region where the destination instance resides.**
     - **Instance ID: the ID of the destination instance.**

   After configuring the preceding parameters, click Set Whitelist and Next in the lower-right corner.

9. **Authorize the DTS synchronization account.**

   In this step, you need to grant the DTS synchronization account the write permission on the MaxCompute instance. This allows DTS to replicate data to the MaxCompute instance.

   Grant the DTS synchronization account the following permissions on a project in the MaxCompute instance:

   - **CreateTable**
   - **CreateInstance**
   - **CreateResource**
   - **CreateJob**
   - **List**

   To ensure the synchronization task stability, we recommend that you do not revoke the write permission during the synchronization process. Click Next to create a synchronization account.

   After the account is authorized, you can select the objects to be synchronized.

10. **Click Next to select the objects to be synchronized.**

    After the required permissions on the MaxCompute instance are granted to the DTS synchronization account, proceed with the initial synchronization configuration and select the tables to be synchronized.

    In this step, you need to configure the initial synchronization and select the tables to be synchronized. Where:

    a. **Initial synchronization**

       Two options are available for initial synchronization: Initial Schema Synchronization and Initial Full Data Synchronization.

       During initial schema synchronization, DTS creates a table that has the same schema as the table to be synchronized in the MaxCompute instance. During initial full data synchronization, DTS replicates all the existing data in the table to be synchronized to the MaxCompute instance. We recommend

that you select both Initial Schema Synchronization and Initial Full Data Synchronization when configuring the synchronization task.

b. Select the tables to be synchronized

You can select only tables as objects to be synchronized rather than the entire database. Each table to be synchronized corresponds to a full data table and an incremental data table in the MaxCompute instance. To modify the name of a table, click Edit next to the table in the Selected area to open the Edit Table Name dialog box.

11. Click Precheck to start a precheck.

After the precheck is successful, click Start Task to start the synchronization task.

After the synchronization task is started, it is displayed in the synchronization task list. The new synchronization task appears in the Performing Initial Sync state. The duration of the initial synchronization depends on the data volume of the objects to be synchronized in the source instance. After the initial synchronization is complete, the synchronization channel changes to the Synchronizing status. At this point, the synchronization channel between the source and destination instances is established.

When the synchronization channel is in the Synchronizing state, you can query the full data table and incremental data table in MaxCompute.

At this point, you have configured the task for synchronizing data from an RDS instance to a MaxCompute instance in real time.

Full data merging

This section describes how to retrieve the full data of a table at a specific time from the full data table and incremental data table in the MaxCompute instance. DTS supports full data merging by running SQL statements in MaxCompute.

You can run SQL statements in MaxCompute to merge the full data table and incremental data table to retrieve the full data at the time (t). The following code example shows the SQL statements to run in MaxCompute:

```
insert overwrite table result_storage_table
select col1,
       col2,
       colN
  from(
select row_number() over(partition by t.primary_key_column
```

```
 order by record_id desc, after_flag desc) as row_number, record_id,
operation_flag, after_flag, col1,col2,colN
  from(
select incr.record_id, incr.operation_flag, incr.after_flag, incr.col1
, incr.col2,incr.colN
  from table_log incr
 where utc_timestamp< timestmap
 union all
select 0 as record_id, 'I' as operation_flag, 'Y' as after_flag, base.
col1, base.col2,base.colN
  from table_base base) t) gt
where record_num=1
  and after_flag='Y'
```

The variables in the preceding code are described as follows:

- **result_storage_table: the name of the table that stores the result set of full data merging.**
- **col1, col2, colN: the column names of the synchronized table.**
- **primary_key_column: the name of the primary key column of the synchronized table.**
- **table_log: the name of the incremental data table.**
- **table_base: the name of the full data table.**
- **timestamp: the time when the full data will be merged.**

In the preceding example, the testdb_20161010_base table is the full data table that corresponds to the testdb table. The testdb_20161010_log table is the incremental data table that corresponds to the testdb table.

You can run the following SQL statements in MaxCompute to query the full data of the testdb table at the 1476263486 time:

```
insert overwrite table testdb_1476263486
select id,
       name
  from(
select row_number() over(partition by t.id
 order by record_id desc, after_flag desc) as row_number, record_id,
operation_flag, after_flag, id, name
  from(
select incr.record_id, incr.operation_flag, incr.after_flag, incr.id,
incr.name
  from testdb_20161010_log incr
 where utc_timestamp< 1476263486
 union all
select 0 as record_id, 'I' as operation_flag, 'Y' as after_flag, base.
id, base.name
  from testdb_20161010_base base) t) gt
 where gt.row_number= 1
```

```
    and gt.after_flag= 'Y' ;
```

You can also use DataWorks to add full data merging nodes before you perform further computing and analysis. After full data is merged, downstream computing and analysis nodes can be automatically triggered. You can set an interval for periodic offline data analysis.

At this point, you have configured the task for synchronizing data from an RDS instance to a MaxCompute instance in real time and merged the full data of tables.

9.4.4 Synchronize data from an ApsaraDB RDS for MySQL instance to an AnalyticDB for PostgreSQL instance

This topic describes how to synchronize data from an ApsaraDB RDS for MySQL instance to an AnalyticDB for PostgreSQL instance by using Data Transmission Service (DTS). The data synchronization feature provided by DTS allows you to transfer and analyze data with ease.

Prerequisites

- **The tables to be synchronized from the source database contain primary keys.**
- **An AnalyticDB for PostgreSQL instance is created.**

Notes

- **If the object you want to synchronize is a table or multiple tables (not a database), do not use gh-ost or pt-online-schema-change to perform DDL operations during data synchronization. Otherwise, data synchronization may fail.**
- **If the source database does not have PRIMARY KEY or UNIQUE constraints and all fields are not unique, duplicate data records may exist in the destination database.**
- **Only one-way synchronization is supported.**

Limits

- **Only table data can be synchronized.**
- **Initial schema synchronization is not supported.**
- **Data of the following types cannot be synchronized: JSON, GEOMETRY, CURVE , SURFACE, MULTIPOINT, MULTILINESTRING, MULTIPOLYGON, GEOMETRYCO LLECTION, and BYTEA.**

Supported synchronization statements

- **DML operations: INSERT, UPDATE, and DELETE**
- **DDL operations: ALTER TABLE, ADD COLUMN, DROP COLUMN, and RENAME COLUMN**

> **Note:**
> **The CREATE TABLE and DROP TABLE operations are not supported. To synchronize data from a new table, you must add the table to the selected objects. For more information, see** *Add objects to be synchronized*.

Term mappings

Term in ApsaraDB RDS for MySQL	Term in AnalyticDB for PostgreSQL
Database	Schema
Table	Table

Create a data structure in the destination instance

Create a database, schema, and table in the destination AnalyticDB for PostgreSQL instance based on the data structure of the source RDS instance. For more information, see *AnalyticDB for PostgreSQL User Guide*.

Configure a data synchronization task

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **On the Synchronization Tasks page, click Create Synchronization Task in the upper-right corner.**
4. **In the Create DTS Instances dialog box, configure the required parameters.**

Parameter	Description
Source Instance Region	Select the region where the source RDS instance resides.
Source Instance Type	Select MySQL.
Destination Instance Region	Select the region where the destination AnalyticDB for PostgreSQL instance resides.
Destination Instance Type	Select AnalyticDB for PostgreSQL.

Parameter	Description
Synchronization Mode	Select One-Way Synchronization.
Instances to Create	Enter the number of synchronization instances that you want to create.

5. Click Create.
6. On the Synchronization Tasks page, find the synchronization instance and click Configure Synchronization Channel in the Actions column.
7. Configure the source and destination instances.

Section	Parameter	Description
N/A	Synchronization Task Name	· DTS automatically generates a task name. You do not need to use a unique task name. · We recommend that you use an informative name for easy identification.
Source Instance Details	Instance Type	Select RDS Instance.
	Instance Region	The region where the source RDS instance resides.
	Instance ID	Select the ID of the source RDS instance.
	Encryption	Select Non-encrypted or SSL-encrypted. Note: If you select SSL-encrypted, you must enable SSL encryption for the RDS instance before configuring the data synchronization task.
Destination Instance Details	Instance Type	The type of the destination instance is AnalyticDB for PostgreSQL.
	Instance Region	The region where the destination instance resides.
	Instance ID	Select the ID of the destination AnalyticDB for PostgreSQL instance.
	Database Name	Enter the name of the destination database.

Section	Parameter	Description
	Database Account	Enter the database account of the destination AnalyticDB for PostgreSQL instance. **Note:** The database account must have the SELECT, INSERT, UPDATE, DELETE, COPY, TRUNCATE, and ALTER TABLE permissions.
	Database Password	Enter the password for the database account.

8. In the lower-right corner of the page, click Set Whitelist and Next.
9. Wait until the synchronization account is created. Then, click Next.
10. Configure the synchronization policy and objects.

Section	Parameter	Description
Synchronization policy	Initial Synchronization	Select Initial Full Data Synchronization. **Note:** DTS synchronizes the historical data of the required objects from the source instance to the destination instance. The historical data is the basis for subsequent incremental synchronization.

Section	Parameter	Description
	Processing Mode In Existed Target Table	· Pre-check and Intercept (Selected by default) Checks the Schema Name Conflict item and generates an error message if the destination table contains data. · Clear Target Table Skips the Schema Name Conflict item during the precheck. Clears the data in the destination table before initial full data synchronization. You can select this mode if you want to synchronize your business data after testing the data synchronization task. · Ignore Skips the Schema Name Conflict item during the precheck. Adds new data to the existing data during initial full data synchronization. You can select this mode if you want to synchronize data from multiple tables to one table.
	Synchronization Type	· Insert · Update · Delete · Alter Table Note: Select the types of operations that you want to synchronize based on your business requirements.

Section	Parameter	Description
Objects to be synchronized	N/A	You can select tables as the objects to be synchronized. If you want to specify different column names in the destination table, you can use the object name mapping feature provided by DTS. For more information about how to use this feature, see *Database, table, and column name mapping*. **Note:** The CREATE TABLE operation is not supported. To synchronize data from a new table, you must add the table to the selected objects. For more information, see *Add objects to be synchronized*.

11. In the lower-right corner of the page, click Precheck.

Note:

- A precheck is performed before you can start the data synchronization task . You can start the data synchronization task only after the task passes the precheck.
- If the task fails the precheck, click the info icon next to each failed item to view details. Fix the issues based on the cause of failure and run the precheck again.

12. Close the Precheck dialog box after the following message is displayed: The precheck is passed.
13. Wait until the initial synchronization is complete and the data synchronization task is in the Synchronizing state.

    You can view the status of the data synchronization task on the Synchronization Tasks page.

9.4.5 Configure two-way data synchronization between RDS instances

9.4.5.1 Overview

DTS supports two-way real-time data synchronization between RDS instances on any two clouds. This section describes how to use DTS to create a two-way synchronization task between two ApsaraDB RDS for MySQL instances for active geo-redundancy, geo-disaster recovery, and other scenarios.

9.4.5.2 Supported synchronization statements

Two-way synchronization between ApsaraDB RDS for MySQL instances supports all DML updates (including INSERT, UPDATE, and DELETE) and the following DDL updates:

- **ALTER TABLE, ALTER VIEW, ALTER FUNCTION, and ALTER PROCEDURE**
- **CREATE DATABASE, CREATE SCHEMA, CREATE INDEX, CREATE TABLE, CREATE PROCEDURE, CREATE FUNCTION, CREATE TRIGGER, CREATE VIEW, and CREATE EVENT**
- **DROP FUNCTION, DROP EVENT, DROP INDEX, DROP PROCEDURE, DROP TABLE, DROP TRIGGER, and DROP VIEW**
- **RENAME TABLE and TRUNCATE TABLE**

Note:

To ensure the stability of a two-way synchronization channel, you can synchronize DDL updates on the same table in only one direction.

For example, for two-way synchronization, you must enable DDL synchronization in either the A-to-B or B-to-A direction. If DDL synchronization is configured in one direction, it is not supported in the reverse direction. You can only perform DML synchronization.

9.4.5.3 Detect and resolve conflicts

To ensure data consistency, for two-way synchronized instances, make sure that records with the same primary key, business primary key, or unique key are updated only on one of the instances. If you unexpectedly update a record with the same primary key, business primary key, or unique key on both instances that are two-way synchronized, a synchronization conflict occurs. To maximize the stability

of two-way synchronized instances, DTS supports detecting and resolving data conflicts.

Considerations

During two-way synchronization, the system time of the source and destination instances may not be the same. Additionally, synchronization delays may occur . For these reasons, DTS cannot guarantee that its conflict detection mechanism can completely prevent data conflicts. You must refactor certain business logic to ensure that records of the same primary key, business primary key, or unique key are updated only on one of the instances that are two-way synchronized.

Supported conflict types

Currently, DTS supports detecting the following conflict types:

- **Uniqueness conflicts caused by INSERT operations**

  A uniqueness conflict occurs when the synchronization of an inserted row violates the unique constraint. For example, if two instances in two-way synchronization insert a record with the same primary key value at almost the same time, one of the inserted records fails to be synchronized because a record with the same primary key value already exists in the destination instance.
- **Inconsistent records caused by UPDATE operations**

  Update conflicts occur in the following scenarios:

  - **The records to be updated do not exist in the destination instance. If the records to be updated do not exist, DTS automatically changes the UPDATE operation to the INSERT operation and inserts these records to the destination instance. In this case, duplicate unique key values may occur.**
  - **The primary keys or unique keys of the records to be updated conflict with each other.**
- **A DELETE operation is made on non-existent records**

  A delete conflict occurs when the records to be deleted do not exist in the destination instance.

  In this case, DTS automatically ignores the DELETE operation regardless of the conflict resolution policy that you have configured.

Supported conflict resolution policies

For the preceding synchronization conflicts, DTS provides the following resolution policies. You can select a conflict resolution policy as required when configuring two-way synchronization.

- **TaskFailed: The synchronization task reports an error and automatically exits the process in case of a conflict.**

  When the synchronization encounters a conflict of the preceding types, the synchronization task reports an error and automatically exits the process. The task enters a failed state and you must manually resolve the conflict. This method is the default conflict resolution policy.
- **Ignore: The records in the destination instance are used in case of a conflict.**

  When the synchronization encounters a conflict of the preceding types, the synchronization task skips the current synchronization statement and continues the process. The records in the destination instance are used.
- **Overwrite: The conflict records in the destination instance are overwritten in case of a conflict.**

  When the synchronization encounters a conflict of the preceding types, the conflict records in the destination instance are overwritten.

9.4.5.4 Synchronization restrictions

This section describes the restrictions in cross-cloud data synchronization using DTS.

Restrictions in data sources

Currently, only ApsaraDB RDS for MySQL instances support two-way synchronization. Other heterogeneous data sources do not support two-way synchronization.

The destination instance cannot be an RDS instance that runs in standard access mode and has only a public network address.

Restrictions in synchronization architecture

Currently, DTS only supports two-way synchronization between two ApsaraDB RDS for MySQL instances. Two-way synchronization between more than two instances is not supported.

Feature restrictions

- **Incompatible with triggers**

  When you synchronize an entire database and the database contains a trigger that updates the synchronization table, the synchronized data may be inconsiste nt.

  For example, the object to be synchronized is database A that contains table a and table b. Table a has a trigger that inserts a row to table b after the row is inserted to table a. In this case, if an INSERT operation is performed on table a in the source instance during synchronization, the data in table b is inconsistent between the source and destination instances.

  To resolve this problem, you must delete the trigger in the destination instance, so that the data in table b is only synchronized from the source instance.
- **Restrictions in the RENAME TABLE operation**

  The RENAME TABLE operation may result in inconsistent synchronization data. For example, if the object to be synchronized only includes table a and the rename a to b command is executed in the source instance during synchronization, subsequent operations to the renamed table b are not synchronized to the destination database. To solve this problem, you can synchronize the entire database where table a and table b are stored.
- **Restrictions in DDL synchronization direction**

  To ensure the stability of a two-way synchronization channel, you can synchronize DDL updates on the same table in only one direction. For example, in A-to-B and B-to-A synchronization, you can implement DDL synchronization in either the A-to-B or B-to-A direction. If DDL synchronization is configured in one direction, it is not supported in the reverse direction.

9.4.5.5 Configure two-way data synchronization between RDS instances across IDCs

This topic describes how to configure two-way data synchronization between RDS instances across IDCs.

Prerequisites

To configure a synchronization task, ensure that the source and destination ApsaraDB RDS for MySQL instances are available for two-way synchronization. You must first create the required instances if they do not exist.

Procedure

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **On the Data Synchronization page, click Create Synchronization Task in the upper-right corner to start the task configuration.**

   Note:

   Source Instance Region: Select the region where the source RDS instance resides.

   Source Instance Type: Select the type of the source instance. In this example, select MySQL.

   Destination Instance Region: Select the region where the destination RDS instance resides.

   Destination Instance Type: Select the type of the destination instance. In this example, select MySQL.

   Sync Mode: Select the synchronization mode. In this example, select Two-Way Synchronization.

   Instances to Create: Set the number of instances that you want to create.
4. **After you configure the preceding information, click Create.**

   After you create a synchronization instance, go back to the Synchronization Tasks page. The new synchronization instance is in the Not Configured state and contains two synchronization tasks. You can configure two-way synchronization for the tasks.
5. **Find the required synchronization task and click Set Sync Channel.**

6. **Configure the required information for connecting to the synchronization channel.**

   Parameters are described as follows:

   - **Synchronization task name**

     The synchronization task name is not required to be unique. We recommend that you set an informative name to help you easily identify and manage the synchronization channel.

   - **RDS instance ID of the synchronization task**

     You must specify the ID of the Apsara Stack tenant account to which the destination RDS instance belongs. You can then select an RDS instance ID from the Instance ID drop-down list.

     After you complete the preceding configurations, click Set Whitelist and Next to configure the RDS instance whitelists.

7. **Configure the RDS instance whitelists.**

   In this step, add the IP addresses of the DTS servers to the whitelists of the source and destination RDS instances. This helps you avoid failure in creating a synchronization task when the DTS server cannot connect to the RDS instances because of the whitelist mechanism.

   We recommend that you do not remove the server IP addresses from the whitelists of the RDS instances. This ensures the stability of the synchronization task.

   Click Next to create a synchronization account.

8. **Create a synchronization account for connecting to the destination database.**

   In this example, create a synchronization account named dtssyncwriter in the destination RDS instance. During the synchronization, the account cannot be deleted. Otherwise, an interruption occurs.

9. Configure the objects to be synchronized and the synchronization policies.

   After you create a synchronization account for connecting to the destination RDS instance, you can start configuring the objects and synchronization policies.

   - Exclude DDL Statements

     This field determines whether to synchronize DDL statements in a specific direction. To include DDL statements, select No. To exclude DDL statements, select Yes. After you select No, the same table does not support synchronizing DDL statements in the other direction.
   - DML Statements for Synchronization

     This field determines the DML statements to be synchronized. By default, the INSERT, UPDATE, and DELETE statements are selected. You can select the DML statement types based on your business requirements.
   - Conflict Resolution Policy

     This field determines the resolution policy in case of a synchronization conflict. By default, TaskFailed is selected. You can select a conflict resolution policy based on your business requirements.

     For example, if Node A is the primary business center and Node B is a secondary business center, you must give the priority to Node A. Specifically, you need to set the conflict resolution policy in the A-to-B direction to Overwrite and that in the B-to-A direction to Ignore.
   - Objects to Be Synchronized

     The objects to be synchronized in real time include databases and tables.

     If you select an entire database, all schema update operations (such as CREATE TABLE and DROP VIEW) performed on all the objects in the database are synchronized to the destination database.

     If you select a table, only the DROP TABLE, ALTER TABLE, TRUNCATE TABLE, RENAME TABLE, CREATE INDEX, and DROP INDEX operations to this table are synchronized to the destination database.

10. Configure initial synchronization.

    Initial synchronization is the first step to start the synchronization channel. It synchronizes the schema and data of the objects to be synchronized in the

source instance to the destination instance. The schema and data are used as the baseline data for subsequent incremental data synchronization.

Initial synchronization includes Initial Schema Synchronization and Initial Full Data Synchronization. You must select both Initial Schema Synchronization and Initial Full Data Synchronization by default.

If some tables to be synchronized in one direction are also included in the objects to be migrated in the other direction, these tables do not go through the initial synchronization process.

11. Precheck.

    After you complete the preceding configurations, perform the precheck before starting the synchronization task.

    After the precheck is passed, click Start to start the synchronization task.

    After the synchronization task is started, the synchronization task list is displayed. The newly started synchronization task changes to the Performing Initial Sync state. The duration of the initial synchronization depends on the data volume of the objects to be synchronized in the source instance. After initial synchronization, the synchronization channel changes to the Synchronizing state and the synchronization channel between the source and destination instances is established.

    After the synchronization task is configured in this direction, the source and destination RDS instances of the synchronization task in the other direction are fixed and cannot be changed.

12. After completing the synchronization task configurations in one direction, you can configure the synchronization task in the other direction. For more information about the steps, see step 6 to step 12 in the preceding section.

9.4.6 Troubleshoot precheck failures

Before a real-time synchronization channel is started, a precheck is performed. This topic describes the precheck items and how to troubleshoot precheck failures.

Source database connectivity

- **Description**

  This item checks the connectivity between the DTS server and the source RDS instance. DTS creates a connection to the source RDS instance by using the JDBC protocol. If the connection fails, the precheck fails.
- **Cause of failure**
  - **DTS does not support real-time synchronization between RDS instances in the region where the source instance resides.**
  - **The source instance account or password is incorrect.**
- **Solution**

  Submit a ticket and contact Alibaba Cloud technical support.

Destination database connectivity

- **Description**

  This item checks the connectivity between the DTS server and the destination RDS instance. DTS creates a connection to the destination RDS instance by using the JDBC protocol. If the connection fails, the precheck fails.
- **Cause of failure**
  - **DTS does not support real-time synchronization of RDS instances in the region where the destination instance resides.**
  - **The destination instance account or password is incorrect.**
- **Solution**

  Submit a ticket and contact Alibaba Cloud technical support.

Source database version

- **Description**

  This item checks whether:

  1. **The version of the source RDS instance is supported by the real-time synchronization feature.**
  2. **The version of the destination RDS instance is the same as the version of the source RDS instance.**

- **Cause of failure**

  - **The version of the source RDS instance is earlier than the versions supported by DTS. The source instance version must be MySQL 5.1, 5.5, or 5.6 for a real-time synchronization task.**
  - **The version of the destination RDS instance is earlier than the version of the source RDS instance.**

- **Solution**

  - **If the version of the source RDS instance is earlier than the versions supported by DTS, upgrade the source RDS instance to MySQL 5.6 in the RDS console. Then, re-create the synchronization channel.**
  - **If the version of the destination RDS instance is earlier than the version of the source RDS instance, upgrade the destination RDS instance to MySQL 5.6 in the RDS console. Then, re-create the synchronization channel.**

Database existence

This item checks whether the database to be synchronized already exists in the destination instance. If the database to be synchronized does not exist in the destination instance, DTS automatically creates a database. However, DTS fails to create the database and reports a failure under the following circumstances:

- **The database name contains characters other than lowercase letters, digits, underscores (_), and hyphens (-).**
- **The character set of the database is not UTF-8, GBK, Latin1, or UTF-8MB4.**
- **The migration account of the destination database does not have the read/write permission on the source database.**

  If the source database is an RDS instance, the precheck does not fail.

Source database permissions

This item checks whether the synchronization account of the source database has the required permissions. If the synchronization account does not have the required permissions, the precheck fails. If the source database is an RDS instance, the precheck does not fail.

Destination database permissions

- **Description**

  This item checks whether the synchronization account of the destination database has the required permissions. If the synchronization account does not have the required permissions, the precheck fails.
- **Cause of failure**
  - **DTS fails to create an account in the destination RDS instance.**
  - **DTS fails to grant the read/write permission to the synchronization account of the destination RDS instance.**
- **Solution**

  Submit a ticket and contact Alibaba Cloud technical support.

Object name conflict

- **Description**

  This check item applies only when you have configured initial synchroniz ation for a synchronization channel. The item checks whether an object to be synchronized has the same name as an object in the destination RDS instance.
- **Cause of failure**

  If an object in the destination RDS instance has the same name as the object to be synchronized, the precheck fails.
- **Solution**
  - **Remove the object that has the same name as the object to be synchronized from the destination database.**
  - **Then, re-create a synchronization channel. Two initial synchronization options are available: initial schema synchronization and initial full data synchronization.**

Source database server_id

This item checks whether server_id of the source database is set to an integer greater than or equal to 2. If the source database is an RDS instance, the precheck does not fail.

Whether binlogging is enabled for the source database

This item checks whether binlogging is enabled for the source database. If binlogging is not enabled for the source database, the precheck fails. If the source database is an RDS instance, the precheck does not fail.

Whether the binlog format is ROW in the source database

This item checks whether the binlog format is ROW in the source database. If the binlog format is not ROW in the source database, the precheck fails. If the source database is an RDS instance, the precheck does not fail.

Referential integrity constraint

- **Description**

  This item checks whether all the parent-child tables that have foreign key dependencies with the objects to be synchronized have been synchronized. This protects the integrity of foreign key constraints.
- **Cause of failure**

  Some of the objects to be synchronized are child tables with foreign key dependencies, but their parent tables have not been synchronized. This impairs the integrity of foreign key constraints.
- **Solution**

  The following solutions are available:

  - **Re-create the synchronization task and do not synchronize the child tables that failed the referential integrity constraint check.**
  - **Re-create the synchronization task and add the following tables to the list of objects to be synchronized: the parent tables for the child tables that failed the referential integrity constraint check.**
  - **Modify the source database and delete the foreign key dependencies of the child tables that failed the referential integrity constraint check. Then, re-create the synchronization task.**

Storage engine

- **Description**

  This item checks whether the objects to be synchronized use storage engines that are not supported by the real-time synchronization feature, such as Federated, MRG_MYISAM, and TokuDB.

- **Cause of failure**

  If a table to be synchronized uses the storage engine Federated, MRG_MyISAM, or TokuDB, the precheck fails.

- **Solution**

  Change the unsupported storage engines to InnoDB and re-create the synchronization task.

Character set

- **Description**

  This item checks whether the objects to be synchronized use character sets that are not supported by the real-time synchronization feature, such as the UCS-2 character set.

- **Cause of failure**

  If the character sets used by the objects to be synchronized are not supported by the real-time synchronization feature, the precheck fails.

- **Solution**

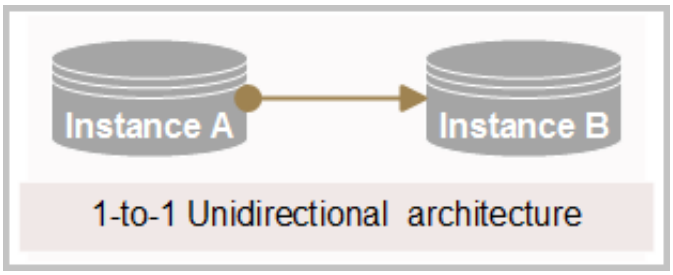
  Change the unsupported character sets to UTF-8, GBK, or Latin1. Then, re-create the synchronization task.

Complicated topologies

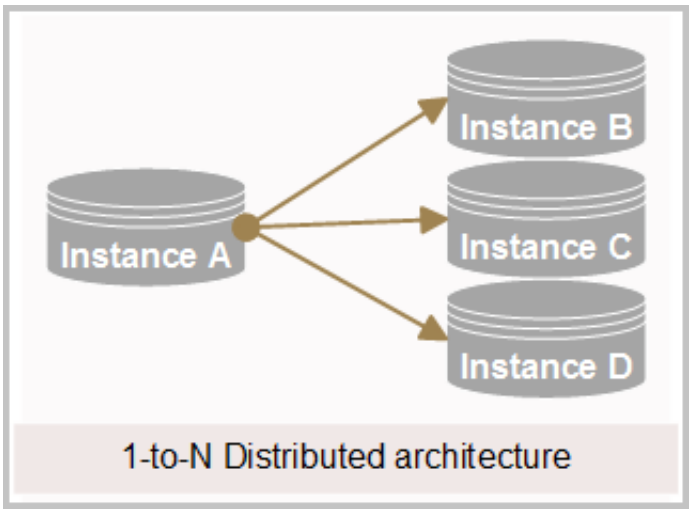
- **Description**

  This item checks whether the source and destination RDS instances of the synchronization task use unsupported synchronization modes. The real-time synchronization feature only supports two synchronization modes:

  - **One-to-one**

  - **One-to-many**

  The real-time synchronization feature does not support the many-to-one, cascading, and two-way synchronization modes.

- **Cause of failure**

  - **The destination RDS instance in the current synchronization task is the source RDS instance in another synchronization task.**
  - **Another synchronization task already runs on the destination RDS instance.**
  - **A migration task is in progress between the source and destination RDS instances, and the objects to be migrated in the migration task overlap the objects in the synchronization task to be created.**

- **Solution**
  - **If a synchronization task already runs between the source and destination RDS instances and this task is the same as the current task that you want to create for synchronizing new objects, do as follows: Modify the existing synchronization channel and do not create a new task. Then, add the expected objects to the list of objects to be synchronized.**
  - **If the synchronization channel conflicts with an existing migration task, wait until the migration task is complete before you can re-create the synchronization task.**
  - **If the new and original synchronization tasks constitute a cascading, two-way, or many-to-one synchronization mode, the mode is not supported for the moment.**

MySQL old password format

This item checks whether the password used by the source instance is an old password. If the source database is an RDS instance, the precheck does not fail.

9.4.7 Check the synchronization performance

DTS provides trend charts on synchronization latency, TPS, and traffic, allowing you to check the running performance of synchronization tasks in real time.

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **In the synchronization task list, click the ID of the synchronization task you want to check.**

   The task details page appears.
4. **On the task details page, click Synchronization Performance in the left-side navigation pane.**

5. **View the trend charts on synchronization performance.**

   DTS provides trend charts on synchronization latency, TPS, and traffic.

   - **Synchronization traffic: The data traffic that the data writing module pulls from the data pulling module per second in DTS. The unit is MB/s.**
   - **Synchronization TPS: The number of queries that DTS synchronizes to the destination RDS instance per second.**
   - **Synchronization latency: The difference between the timestamp of the latest synchronized data in the destination RDS instance and the current timestamp in the source RDS instance. The unit is seconds.**

9.4.8 Add objects to be synchronized

DTS allows you to dynamically modify the objects to be synchronized during the synchronization process. This section describes how to add objects to be synchronized during a synchronization process.

Restrictions on object modifications

You can modify the objects to be synchronized only when the synchronization task is in the Synchronizing or Synchronization Failed state.

Synchronization start time

After an object to be synchronized is added, the synchronization start time depends on whether initial synchronization has been configured for the synchronization task.

- **If initial synchronization has been configured for the synchronization task, DTS performs initial synchronization on the new object before starting incremental synchronization.**
- **If initial synchronization has not been configured for the synchronization task, DTS starts synchronization immediately after incremental data is generated for the objects to be synchronized on the source instance.**

Procedure

1. *Log on to the DTS console*.
2. **In the left-side navigation pane, click Data Synchronization.**
3. **Find the synchronization task to be modified, and click View More > Modify Object to Be Synchronized to modify the objects to be synchronized.**

4. On the Select Object to Be Synchronized page, add objects to be synchronized as needed.
5. Click Next to start a precheck.

   After the precheck is successful, click Start.

   After an object to be synchronized is added, if the synchronization task requires initial synchronization, the task state changes from Synchronizing to Synchronizing (Initial synchronization is being performed on the new object...). At this time, the backend restarts the synchronization channel and the synchronization latency changes to -1 second. After the synchronization channel is restarted, the synchronization latency and speed return to normal.

   In the synchronization task list, you can click View Details to view the initial synchronization progress of the new objects. After the initial synchronization on the new objects is completed, the synchronization task returns to the Synchronizing state.

9.4.9 Remove objects to be synchronized

DTS allows you to dynamically modify the objects to be synchronized during the synchronization process. This section describes how to remove objects to be synchronized during a synchronization task.

Restrictions on object modifications

You can modify the objects to be synchronized only when the synchronization task is in the Synchronizing or Synchronization Failed state.

Procedure

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Data Synchronization.
3. Find the synchronization task to be modified, and click View More > Modify Object to Be Synchronized to modify the objects to be synchronized.
4. Click Configure Synchronization Instance in the Actions column corresponding to a synchronization task. On the Select Object to Be Synchronized page, remove objects to be synchronized as required.

   Now, you have deleted certain objects to be synchronized.

9.5 Change tracking

9.5.1 Overview

Change tracking is a feature provided by DTS that allows you to track data changes in ApsaraDB RDS for MySQL instances in real time. With this feature, you can complete lightweight cache update, asynchronous service decoupling, and real-time synchronization of data by using the Extract, Transform, Load (ETL) logic.

Objects for change tracking

Objects for change tracking include databases and tables. You can specify one or more tables for which you want to track data changes.

In change tracking, incremental data includes data manipulation language (DML) operations and data definition language (DDL) operations. When you configure change tracking, you must select operation types.

Change tracking channels

A change tracking channel is the basic unit of incremental data tracking and consumption. To track data changes in an RDS instance, you must create a change tracking channel in the DTS console for the RDS instance. The change tracking channel pulls incremental data from the RDS instance in real time and locally stores the data. You can use the DTS SDK to consume incremental data in the channel. You can also create, manage, or delete change tracking channels in the DTS console.

9.5.2 Create an RDS change tracking channel

Change tracking is a feature provided by DTS that allows you to track data changes in RDS instances in real time. With this feature, you can complete lightweight cache update, asynchronous service decoupling, and real-time synchronization of data by using the Extract, Transform, Load (ETL) logic.

Prerequisites

Limits on change tracking are listed as follows:

- **The change tracking feature only applies to ApsaraDB RDS for MySQL instances.**
- **The binlog_row_image value of MySQL 5.6 binlog must be full.**
- **Only the InnoDB and MyISAM storage engines are supported.**

- Only the following MySQL character sets are supported: latin1, gbk, utf8, utf8mb4, and binary.

Context

To use the change tracking feature to consume the incremental data of an RDS instance in real time, follow these two steps:

1. Create a change tracking channel for the RDS instance in the DTS console.
2. Use the SDK provided by DTS to access the change tracking channel and consume the incremental data in real time.

This topic describes how to create a change tracking channel in the DTS console. You can create a change tracking channel with only three steps. For more information about how to manage a change tracking channel and use the SDK, see the DTS Product Manual.

The following section describes the procedure for creating a change tracking channel.

Procedure

1. *Log on to the DTS console*.
2. In the left-side navigation pane, click Change Tracking.
3. On the Change Tracking page, click Create Change Tracking Task.
4. In the Create DTS instances dialog box that appears, select a region, enter the number of change tracking channels to be created, and click Create.
5. In the message that appears, click OK.
6. In the change tracking channel list, find the change tracking channel that you created, and click Set Channel in the Actions column.
7. Configure the RDS instance for change tracking.

   In this step, you need to configure the name of the change tracking channel and the ID of the RDS instance. Parameters are described as follows:

   - Task Name: the alias of the change tracking channel. It is not required to be unique. By default, DTS automatically generates a name for each change

tracking channel. You can set the name to an informative one for easy identification of the channel.

- RDS Instance ID: the ID of the RDS instance that contains the incremental data you want to consume.

  After completing the configuration, click Set Whitelist and Next in the lower-right corner.

8. Select the data type for the change tracking channel. Then, select the required objects in the left-side section, and click the right arrow to add the selected objects to the right-side Selected section.

   In this step, select the data types and objects required for change tracking. Parameters are described as follows:

   - Required Data Types

     DTS provides two types of data changes that can be tracked: Data Updates and Schema Updates. Data updates refer to any data changes made by DML operations, such as INSERT, DELETE, and UPDATE operations. Schema updates refer to the schema changes made by DDL operations, such as CREATE TABLE, DROP TABLE, and ALTER TABLE operations.

     If you select Schema Updates, DTS pulls all schema updates in the RDS instance. If you only want the schema updates made by certain DDL operations, you can set filters when you use the DTS SDK to consume data.

   - Objects for change tracking

     DTS allows you to select databases and tables as the objects for change tracking. You can track data changes to specific databases or tables.

9. Click Save and Precheck in the lower-right corner.

   After the precheck is passed, DTS starts the change tracking channel.

   DTS requires about one minute to perform initial change tracking.

   At this point, you have configured the change tracking channel.

9.5.3 Change consumption checkpoints

This section describes how to change consumption checkpoints in the DTS console.

Context

DTS allows you to change consumption checkpoints at any time during the consumption process. After a consumption checkpoint is changed, only data generated after the new consumption checkpoint can be pulled by the downstream SDKs as incremental data. The new consumption checkpoint must be within the data range of the subscription channel. Currently, you can change consumption checkpoints only in the DTS console, and cannot specify consumption checkpoints in the SDK.

Procedure

1. Stop the SDK consumption process.

   Before you change a consumption checkpoint, make sure that all the downstream SDKs connected to the subscription channel have been stopped. You can view the consumer sources (IP addresses) of the subscription channel in the DTS console to check whether all the downstream SDKs have been stopped.

   If the consumer sources are empty, all the downstream SDKs of the subscription channel have been stopped.

2. Change a consumption checkpoint.

   Currently, you can change a consumption checkpoint only in the DTS console.

   If you want to change a consumption checkpoint of the subscription channel, move the pointer over the checkpoint. A pen-like edit icon appears. Click the icon. The dialog box for changing the consumption checkpoint appears.

   > **Note:**
   > The consumption checkpoints configured here must be within the range of the current data tunnel.

3. Restart the SDK consumption process.

   After the consumption checkpoint is changed, restart the local SDK consumption process. Then the SDK subscribes to the incremental data from the new consumption checkpoint.

9.5.4 Modify objects for change tracking

This section describes how to modify objects for change tracking in the DTS console.

Context

DTS allows you to add or remove objects for change tracking in the consumption process. After you add an object, the change tracking channel pulls the incremental data of the new object from the time when the modification takes effect. After you remove an object, the SDK no longer subscribes to the data of the removed object from the time when the modification takes effect.

Procedure

1. Go to the Change Tracking Tasks page to modify the objects for change tracking.

   You can only modify the objects for change tracking in the DTS console.

   Find the change tracking channel for which you want to modify the required objects. Click View More in the Actions column, and select Modify Required Objects.

2. Modify objects for change tracking.

   After you click Modify Required Objects, the Select Required Objects page appears.

   On this page, you can add and remove objects for change tracking, or change the data type. After the objects for change tracking are modified, DTS starts a precheck.

   After the precheck is passed, click Start. Initial change tracking is performed on the change tracking channel.

   After initial change tracking is complete, the change tracking channel switches to the Active state and starts to work as expected. You can now use the SDK to track data changes.

9.5.5 Methods provided by SDK

SDK defines multiple classes. This topic describes the methods provided by these classes.

The DTS SDK is required for tracking and consuming incremental data.

Before using the SDK for data consumption, you must log on to the DTS console and create a change tracking channel for the RDS instance to which you want to subscribe.

After the change tracking channel is created, you can use the SDK to track data changes in real time.

- **DTS provides only the Java version of the SDK.**
- **The data in one change tracking channel can be consumed by only one SDK client . If multiple SDK clients are connected to the same change tracking channel, only one SDK process can pull the incremental data from the channel. If multiple downstream SDK clients need to subscribe to the incremental data in the same RDS instance, you must create a change tracking channel for each downstream SDK client.**

Methods of the RegionContex class

- **setAccessKey(AccessKey)**

  Specifies the AccessKey ID. Set the AccessKey parameter to the AccessKey ID of the account that subscribes to the change tracking channel.
- **setSecret(AccessKeySecret)**

  Specifies the AccessKey Secret. Set the AccessKeySecret parameter to the AccessKey Secret of the account that subscribes to the change tracking channel. You can go to the AccessKey page to create and obtain an AccessKey Secret.
- **setUsePublicIp(usePublicIp)**

  Specifies whether the server where the SDK is running subscribes to data changes over the public network. If the public network is used, set the usePublicIp parameter to True. If the public network is not used, set the usePublicIp parameter to False.

  Data changes can be tracked over internal networks. Before establishing a change tracking channel, the SDK communicates with the DTS control system over the Internet to obtain the physical connection address of the change tracking channel. If data changes must be tracked over internal networks, you need to attach a public IP address to the server where the SDK is deployed.

Methods of the ClusterClient class

- **void addConcurrentListener(ClusterListener arg0)**

  Adds downstream listeners. A listener can subscribe to incremental data in the change tracking channel only after the listener is added to a ClusterClient object.

  The ClusterListener arg0 parameter specifies an object of the ClusterListener class.

- **void askForGUID(String arg0)**

  Requests the incremental data from a specified change tracking channel. The String arg0 parameter specifies the ID of the change tracking channel. You need to obtain the ID in the DTS console.

- **List<ClusterListener> getConcurrentListeners()**

  Obtains the list of listeners in a ClusterClient object. The return type is List< ClusterListener>.

- **void start()**

  Starts the SDK client to subscribe to incremental data.

- **void stop()**

  Stops the SDK client to stop subscribing to incremental data. Data pulling and notification callback are performed in the same thread in the SDK. If the consumption code of the notify() method contains a function that prevents signal interruptions, the stop function may fail to terminate the client.

Methods of the ClusterListener class

void notify(List<ClusterMessage> arg0)

Defines the consumption of incremental data. After receiving data, the SDK client uses the notify() method to instruct a ClusterListener object to consume data. For example, the consumption mode in the SDK demo indicates that tracked data changes are displayed on the screen.

The input parameter type of this method is List<ClusterMessage>, in which ClusterMessage is the schema of tracked data changes. For more information, see *Methods of the ClusterMessage class*.

Methods of the ClusterMessage class

Each ClusterMessage object stores the data record of an RDS transaction. Each record is stored by using a Record object. This section introduces methods of the ClusterMessage class.

- **Record getRecord()**

  Obtains a change record from the ClusterMessage object. The change record indicates each log entry in the RDS binlog file, such as BEGIN, COMMIT, UPDATE , and INSERT operations.

- **void ackAsConsumed**

  To simplify the disaster recovery process of downstream SDK clients, the change tracking server supports consumption checkpoint storage for SDK clients. After a downstream SDK client encounters abnormal downtime and restarts, it automatically subscribes to and consumes data from the last consumption checkpoint that is recorded before downtime occurred.

  After message consumption is complete, you must call this method to send an ACK packet to instruct the DTS server to update the consumption checkpoints for the downstream SDK client. This ensures the integrity of the consumed data after an abnormal SDK client restarts.

Methods of the Record class

A Record object indicates a log entry in the RDS binlog file, such as BEGIN, COMMIT , and UPDATE operations.

- **String getAttribute(String key)**

  Obtains the main attribute values in a Record object. If the input parameter is an attribute name, the value of this attribute is returned.

  Table 9-12: Attribute names **describes the attributes that you can obtain by calling this method.**

  Table 9-12: Attribute names

Key	Description
record_id	The record ID. The ID does not ascend during the change tracking process.
instance	The database endpoint of the record. The format is <IP address>:<Port number>.
source_type	The database engine type of the record. Valid value: mysql.
source_category	The record type. Valid value: full_recorded.
timestamp	The time the record was written to the binlog. It is also the time the SQL statement was run in RDS.

Key	Description
checkpoint	The binlog file checkpoint of the record. The format is file_offset@file_name. The file_name parameter indicates the numeric suffix of the binlog file.
record_type	The operation type of the record. Valid values: insert, update, delete, replace, ddl, begin, commit, and heartbeat.
db	The database name of the table that is updated in the record.
table_name	The name of the table that is updated in the record.
record_recording	The encoding of the record.
primary	The primary key column value of the table that is updated in the record.
fields_enc	The encoded field values in the record. The fields are separated with commas (,). The value of a non-character field is null.

- Type getOpt()

  Obtains the operation type of the record, Valid values: insert, delete, update, replace, ddl, begin, commit, and heartbeat.

  The heartbeat is an exclusively defined indicator to reflect the health status of a change tracking channel. A heartbeat record is generated each second.
- String getCheckpoint()

  Obtains the checkpoint of the change record in the binlog file. The format of the returned checkpoint is binlog_offset@binlog_fid.

  binlog_offset indicates the offset of the change record in the binlog file, and binlog_fid indicates the numeric suffix of the binlog file. For example, if the binlog file name is mysql-bin.0008, the value of binlog_fid is 8.
- String gettimestamp()

  Obtains the timestamp of the change record in the binlog file.
- String getDbname()

  Obtains the database name of the table modified in the change record.
- String getTablename()

  Obtains the name of the table modified in the change record.

- **String getPrimaryKeys()**

  Obtains the primary key column value of the data entry that is modified in the change record. If the primary key is a composite key, the primary key column values are separated with commas (,).
- **DBType getDbType()**

  Obtains the database type of the change tracking instance. The value is MySQL because DTS only supports ApsaraDB RDS for MySQL.
- **String getServerId()**

  Obtains the IP address and port number that the ApsaraDB RDS for MySQL instance uses to run the process corresponding to the change record. The format is <IP address>:<Port number>.
- **int getFieldCount()**

  Obtains the number of fields that are changed in the record.
- **List<Field> getFieldList()**

  Obtains a list of fields. The data type of the returned value is List<Field>.

  List<Field> contains the definitions of all fields that are changed in the record and the image values before and after the change. For more information about the Field class, see *Methods of the Field class*.
- **Boolean isFirstInLogevent()**

  Checks whether the record is the first transaction log entry in a large volume of database changes. If yes, True is returned. If no, False is returned.

Methods of the Field class

The Field class defines the attributes of each field, such as the code, type, name, value, and whether the field is a primary key field. This section defines the methods of the Field class.

- **String getEncoding()**

  Obtains the encoding format of the field value.
- **String getFieldname()**

  Obtains the name of this field.
- **Type getType()**

  Obtains the data type of the field.

- **ByteString getValue()**

  Obtains the value of this field. The return type is ByteString. If the field is not specified, NULL is returned.

- **Boolean isPrimary()**

  Checks whether the field is a primary key field of the table. If yes, True is returned. If no, False is returned.

9.5.6 SDK quick start

This section describes how to use the DTS Java SDK to perform some basic operations.

Initialize RegionContext

RegionContext is used to save and set security authentication information and the network access mode. The following code is used to initialize RegionContext to set the security authentication credentials and network access mode.

```
import java.util.List;
import com.aliyun.drc.clusterclient.ClusterClient;
import com.aliyun.drc.clusterclient.DefaultClusterClient;
import com.aliyun.drc.clusterclient.RegionContext;
public class MainClass
{
      public static void main(String[] args) throws Exception {
        // Create a RegionContext.
          RegionContext context = new RegionContext();
          context.setAccessKey("<AccessKey>");
          context.setSecret("<AccessKeySecret>");
          context.setUsePublicIp(true);
          // Create a subscription client.
          final ClusterClient client = new DefaultClusterClient(
context);
          // The following is other invocation code:
          ...
    }
}
```

Initialize Listeners

The functions of data consumption are implemented through the Listener class. After ClusterClient is initialized, add a Listener class, which defines the notify function to receive and consume subscribed data. The following code demonstrates the most basic consumption logic, and is used to display the subscribed data on the screen.

```
import com.aliyun.drc.clusterclient.ClusterClient;
import com.aliyun.drc.clusterclient.ClusterListener;
import com.aliyun.drc.clusterclient.DefaultClusterClient;
```

```
import com.aliyun.drc.clusterclient.RegionContext;
import com.aliyun.drc.clusterclient.message.ClusterMessage;
public class MainClass
{
    public static void main(String[] args) throws Exception {
        // Initialize a RegionContext object
        ...
        //Initialize a ClusterClient object
        ...
        ClusterListener listener = new ClusterListener(){
             @Override
             public void notify(List<ClusterMessage> messages) throws
Exception {
                  for (ClusterMessage message : messages) {
                    // Display the subscribed incremental data.
                      System.out.println(message.getRecord() + ":" +
message.getRecord().getTablename() + ":"
                      + message.getRecord().getOpt());
                      // After data consumption is completed, send an
ACK packet to DTS by calling
                      message.ackAsConsumed();
               }
      }
     }
}
```

DTS saves the consumption checkpoints of the SDK to the DTS server. This simplifies disaster recovery during the use of the SDK. The askAsConsumed() interface in the preceding sample code reports the consumption checkpoint of the last data record consumed by the SDK to the DTS server. When the SDK restarts after an unexpected downtime, it automatically obtains the consumptio n checkpoint from the DTS server and restarts from this checkpoint to avoid data duplication.

Start ClusterClient

Use the following code:

```
import java.util.List;
import com.aliyun.drc.clusterclient.ClusterClient;
import com.aliyun.drc.clusterclient.ClusterListener;
import com.aliyun.drc.clusterclient.DefaultClusterClient;
import com.aliyun.drc.clusterclient.RegionContext;
import com.aliyun.drc.clusterclient.message.ClusterMessage;
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;
public class MainClass
{
   public static void main(String[] args) throws Exception {
    // Initialize RegionContext.
    ...
    // Initialize ClusterClient.
    ...
    // Initialize ClusterListener.
    ...
    // Add listeners.
      client.addConcurrentListener(listener);
```

```
      // Set the requested subscription channel ID.
      client.askForGUID("dts_rdsrjiei2u2afnb_DSF");
      // Start a background thread. Note that there will be no
blocking and the main thread cannot exit.
      client.start();
}
```

In the preceding code, the askForGUID() interface sets the subscription channel ID requested by the client. The ID of this subscription channel is obtained from the DTS console. Once the subscription channel ID is configured, the SDK can obtain the incremental data through the subscription channel.

You must add a Listener class to a client before starting the client. In this way, when the client pulls incremental data from the subscription channel, it starts data consumption by calling the notify method of the Listener synchronously.

9.5.7 Use SDK to track data changes

You can use the SDK to track data changes. DTS records the tracked data changes in a custom format. This topic describes how to parse various types of SQL statements.

Parse a DDL statement

If a record is a DDL statement, the operation type of this record is DDL. The DDL statement is stored in the value of the first column. You can use the following sample code to obtain the DDL statement:

```
String ddl_string;
Record.Type type=record.getOpt();
if(type.equals(Record.Type.DDL)){
    List<DataMessage.Record.Field> fields = record.getFieldList();
    ddl_string = fields.get(0).getValue().toString();
}
```

Parse an INSERT statement

If a record is an INSERT statement, the operation type of this record is INSERT. You can use the following sample code to obtain the complete INSERT statement:

```
StringBuilder insert_string=new StringBuilder();
Record.Type type=record.getOpt();
DataMessage.Record.Field field;
StringBuilder FieldName=new StringBuilder();
StringBuilder FieldValue = new StringBuilder();
if(type.equals(Record.Type.INSERT)){
       int i=0;
       List<DataMessage.Record.Field> fields = record.getFieldList();

       for (; i < fields.size(); i++) {
              field = fields.get(i);
  FieldName.append('`'+field.getFieldname().toLowerCase()+'`');
              FieldValue.append(field.getValue());
```

```
                if (i ! = fields.size() - 1) {
                        FieldName.append(',');
                        FieldValue.append(',');
                }
        }
        insert_string.append("insert "+ record.getTablename()+"("+
FieldName.toString()+") values("+FieldValue.toString()+");");
}
```

Parse an UPDATE statement

If a record is an UPDATE statement, the operation type of this record is UPDATE.

The field values prior to the UPDATE operation are stored in Record.getFieldList() entries with even indexes. The field values after the UPDATE operation are stored in Record.getFieldList() entries with odd indexes.

You can use the following sample code to obtain the complete UPDATE statement if the updated table has a primary key:

```
StringBuilder update_string=new StringBuilder();
Record.Type type=record.getOpt();
DataMessage.Record.Field field;
StringBuilder SetValue = new StringBuilder();
StringBuilder WhereCondition = new StringBuilder();
String ConditionStr;
boolean hasPk=false;
boolean pkMode=false;
boolean hasSet=false;
if(type.equals(Record.Type.UPDATE)){
    int i=0;
    DataMessage.Record.Field OldField = null;
    DataMessage.Record.Field NewField = null;
    List<DataMessage.Record.Field> fields = record.getFieldList();
    for (; i <fields.size() ; i++) {
        if (i % 2 == 0) {
            OldField = fields.get(i);
            continue;
        }
    NewField = fields.get(i);
    if (field.isPrimary()) {
        if (hasPk) {
            WhereCondition.append(" and ");
        }
        //where old value
        ConditionStr = getFieldValue(OldField);
        if(ConditionStr==null){
  WhereCondition.append("`"+field.getFieldname().toLowerCase()+"`" + "
 " + "is null");
        }else{
               WhereCondition.append("`"+field.getFieldname().
toLowerCase()+"`"+" = "+ NewField.getValue());
         }
        hasPk = true;
    }
    if (hasSet) {
        SetValue.append(COMMA);
    }
```

```
      SetValue.append("`"+field.getFieldname().toLowerCase()+"`" + " =
 " + field.getValue());
    String setStr = getFieldValue(field);
    hasSet = true;
    }
    update_string.append("Update "+record.getTablename() +" Set " +
SetValue + " Where "+WhereCondition +";");
}
```

Parse a DELETE statement

If a record is a DELETE statement, the operation type of this record is DELETE. You can use the following sample code to obtain the complete DELETE statement if the deleted table has a primary key:

```
StringBuilder delete_string=new StringBuilder();
Record.Type type=record.getOpt();
DataMessage.Record.Field field;
StringBuilder FieldName=new StringBuilder();
StringBuilder FieldValue = new StringBuilder();
StringBuilder DeleteCondition = new StringBuilder();
boolean hasPk=false;
boolean pkMode=false;
if(type.equals(Record.Type.DELETE)){
   int i=0;
   List<DataMessage.Record.Field> fields = record.getFieldList();

   delete_string.append("Delete From" + record.getTablename() + "where
");
   // Check whether the table has a primary key.
   if (record.getPrimaryKeys() ! = null) {
             pkMode = record.getPrimaryKeys().length() > 0 ? true :
false;
   }
   for (; i < fields.size(); i++) {
            if ((pkMode && ! field.isPrimary())) {
                    continue;
            }
            if (hasPk) {
                    delete_string.append(" and ");
            }
            delete_string.append(field.getFieldname() + "=" + field.
getValue());
            hasPk = true;
    }
    delete_string.append(";");
}
```

Parse a REPLACE statement

If a REPLACE statement has been executed for the source database, the operation type of this record is UPDATE or INSERT. If the value specified in the REPLACE statement does not exist, the record operation type is INSERT. If the value specified in the REPLACE statement already exists, the record operation type is UPDATE.

Parse a BEGIN statement

If a record is a BEGIN statement, the operation type of this record is BEGIN. You do not need to perform any operations on fields because the BEGIN statement does not modify fields. You only need to determine that the operation is a BEGIN operation. You can use the following sample code to obtain the BEGIN statement:

```
StringBuilder sql_string = new StringBuilder();
Record.Type type = record.getOpt();
if(type.equals(Record.Type.BEGIN)){
        sql_string.append("Begin");
}
```

Parse a COMMIT statement

If a record is a COMMIT statement, the operation type of this record is COMMIT. You do not need to perform any operations on fields because the COMMIT statement does not modify fields. You only need to determine that the operation is a COMMIT operation. You can use the following sample code to obtain the COMMIT statement:

```
StringBuilder sql_string = new StringBuilder();
Record.Type type = record.getOpt();
if(type.equals(Record.Type.COMMIT)){
        sql_string.append("commit");
}
```

9.5.8 Run the SDK demo code

This section describes how to run the demo code provided by the DTS console.

1. **Create an AccessKey.**

   Your account must pass the AccessKey authentication before you can use an SDK to connect to a subscription channel. Therefore, before using the SDK, you must obtain an AccessKey. For more information, see the "Obtain an AccessKey" section of the *DTS Developer Guide* **.**

2. **Install the Java SDK.**

   The development environment supported by the DTS Java SDK is J2SE Development Kit (JDK) V1.5 or later.

   For an Eclipse project, you can follow these steps to install the Java SDK:

   a. **Click View Example Code and download the SDK package** *consumer.jar*.
   b. **Import the JAR package to an Eclipse project as follows:**

      In Eclipse, right-click your project and choose Properties > Java Build Path > Libraries > Add External JARs. Select the path for storing the *consumer.jar* **package** consumer.jar.
   c. **Select the** *consumer.jar* **package and click OK.**

      Then you can use the DTS Java SDK in the project.

3. **Run the demo code.**

   DTS provides the SDK demo code. You can copy the demo code by using the View Demo Code option in the DTS console. For an Eclipse project, you can follow these steps to run the demo code:

   a. **Create a class named MainClass in the src directory of the Eclipse project.**
   b. **Open the generated Java file** *MainClass* **and delete the code template.**
   c. **Paste the demo code into the** *MainClass* **file.**
   d. **Modify the AccessKeyId, AccessKeySecret, and subscription channel ID in the demo code.**

      Change the marked parts in the preceding demo code to the AccessKeyId, AccessKeySecret, and subscription channel ID of your account.

      You can obtain the subscription channel ID from the *DTS console*.
   e. **In Eclipse, right-click the demo file and choose Run as > Java Application to run the demo code.**

10 Data Management (DMS)

10.1 What is DMS?

Data Management (DMS) is an integrated database solution that includes data, schema, and server management, access control, BI insights, data trend analysis, data tracking, and performance optimization.

Supported database types

- **ApsaraDB RDS for MySQL**

Supported database operations

- **SQL operations**
  - **Use of SQL editor**
  - **Use of SQL command line interface**
  - **Saving of work environment settings**
  - **SQL execution**
  - **SQL optimization**
  - **SQL formatting (SQL statement improvement)**
  - **Viewing of execution plans**
  - **Smart SQL completion**
- **Operations on database objects**
  - **Operations on data tables**
  - **Operations on schemas: creation and deletion of tables and modification of schemas**
  - **Changes to table data: insertion, update, and deletion of data**
  - **Table data query and visualized editing**
- **Operations on views and programmable objects such as functions, stored procedures, triggers, and events**
  - **Creation**
  - **Modification**
  - **Deletion**
  - **Enabling and disabling**

- **Data processing**
  - **Data import**
  - **Data export**
- **Performance and diagnostics**
  - **Lock wait analysis**
- **Use of data processing tools**
  - **Drawing of ER diagrams**
  - **Collection of statistics on table data volumes**
  - **Batch operations on tables**

User-friendly interactions

DMS provides user-friendly tips. When an error occurs, DMS displays suggestions to guide you to complete your goal.

10.2 Log on to an ApsaraDB for RDS instance through DMS

This topic describes how to log on to an ApsaraDB for RDS instance through DMS.

Context

The ApsaraDB for RDS console provides the option to log on to ApsaraDB for RDS instances through DMS. You can use DMS to manage database data and schemas.

Procedure

1. **Log on to the ApsaraDB for RDS console. For more information, see the** *Log on to the ApsaraDB for RDS console* **topic in** *ApsaraDB for RDS User Guide* **.**
2. **Click the ID of the target instance or click Manage in the Actions column corresponding to the instance. The Basic Information page appears.**
3. **Click Log On to DB to go to the logon page of the DMS console.**

4. Enter the logon information.

Parameter	Description
IP address:Port	The internal or public endpoint and corresponding port number of the target ApsaraDB for RDS instance, such as `rm-txxxxxxxxx xxxxxx.mysql.aliyun-inc.com:3306`. To obtain the internal or public endpoint and its port number, perform the following steps: a. Log on to the ApsaraDB for RDS console. b. Click the ID of the target instance or click Manage in the Actions column corresponding to the instance. c. In the Basic Information section of the page that appears, query the endpoint and port number.
Database Username	The account that connects to the ApsaraDB for RDS instance. **Note:** The account is created in the ApsaraDB for RDS instance. For more information about how to create an account for a MySQL instance, see the *Create a standard account* topic in *ApsaraDB for RDS User Guide* .
Enter your password	The password of the account. **Note:** The password is specified when you create the account in the ApsaraDB for RDS instance.

5. Click Login.

Note:
If you want the browser to remember the password, select `Remember your password` and then click Login.

10.3 SQL operations

10.3.1 Use the command window

This topic briefly describes how to use the DMS command window.

Context

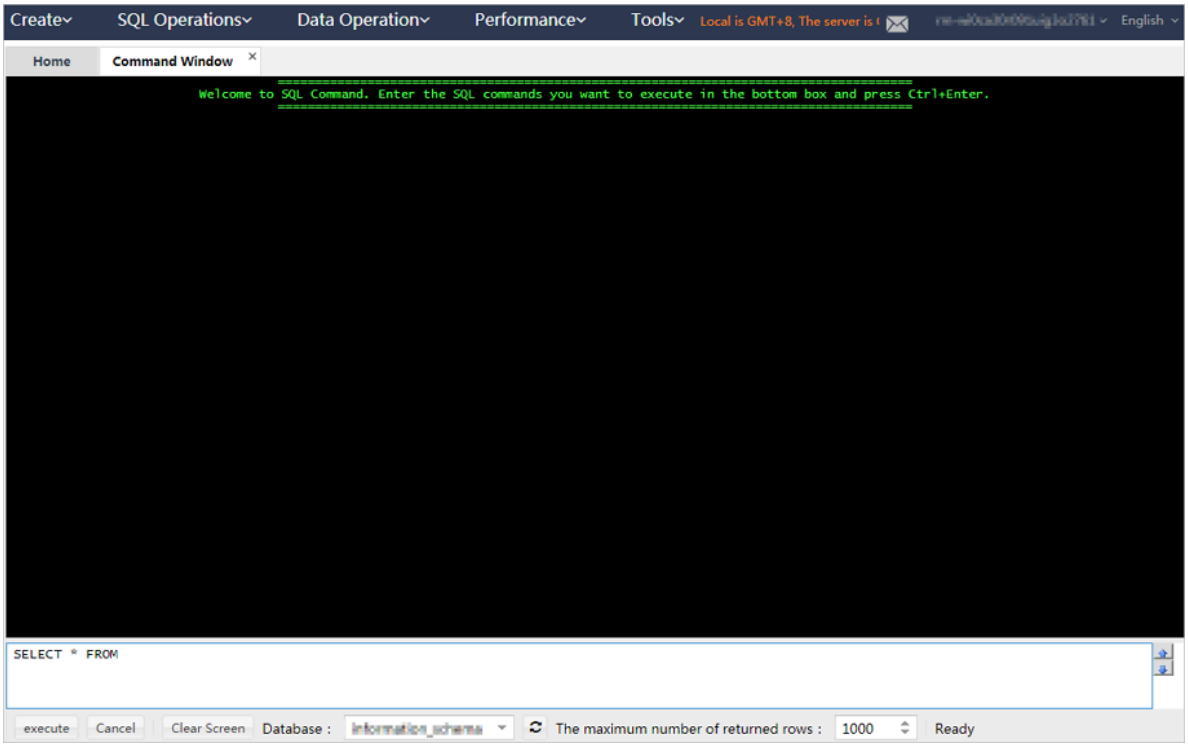
A MySQL database is used as an example.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. **In the top navigation bar, choose SQL Operations > Command Window.**

   An empty command window appears, as shown in *Figure 10-1: Command window*.

   Figure 10-1: Command window



3. **Enter an SQL statement in the command window, click Execute, and view the execution result, as shown in** *Figure 10-2: Execution of a SQL statement*.

   Figure 10-2: Execution of a SQL statement

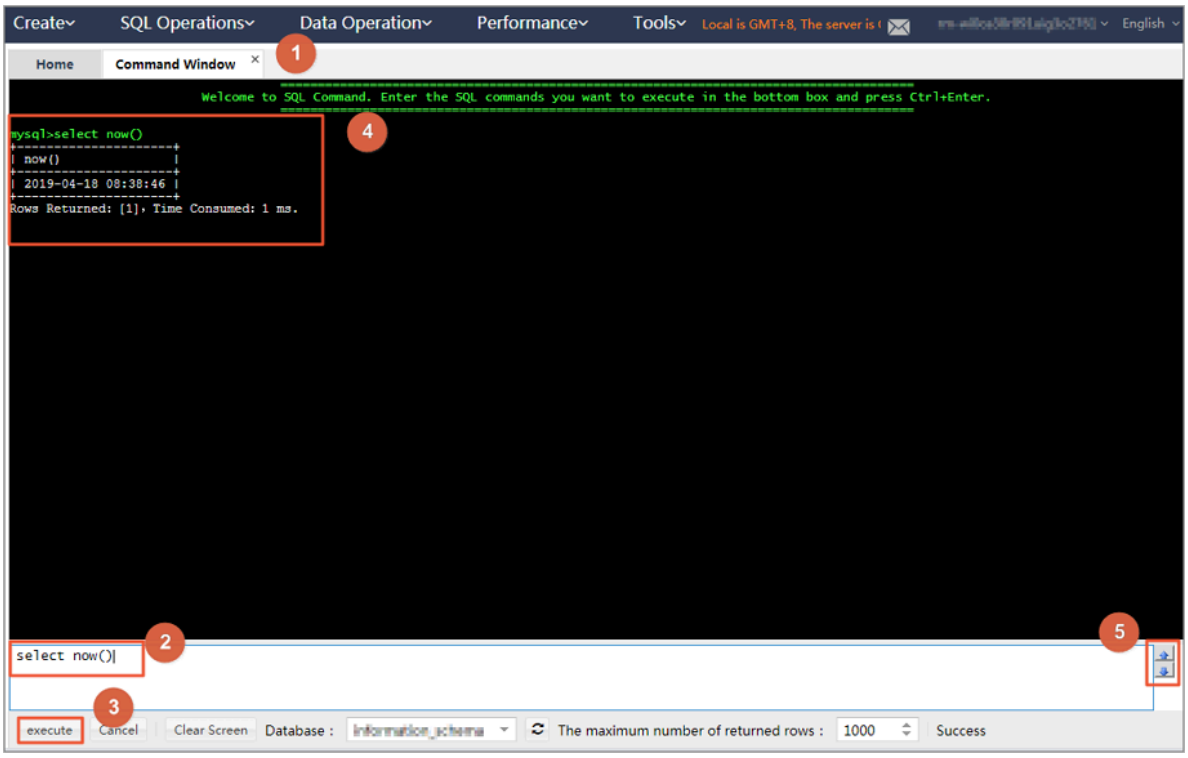


Table 10-1: Description of the numbered items **describes the numbered items.**

Table 10-1: Description of the numbered items

No.	Name	Description
1	Command window	Displays the execution results of SQL statements.
2	SQL statement input area	Offers an area to enter SQL statements.
3	`Execute` button	Executes the entered SQL statements.
4	Result display area	Adds execution results to `Result Area`.
5	`Up and Down` arrows	You can click the up or down arrow to view an executed SQL statement and execute it again.

4. **Optional: If the execution process takes longer time than expected, you can click Cancel to abort the execution.**

5. **Optional: Click Clear Screen to clear the results for proper display of subsequent results.**

   To switch to another database, select the new database from the `Database` **drop-down list.**

10.3.2 Use the SQL window

10.3.2.1 Open an empty SQL window

This topic describes how to use SQL windows.

Context

- **A MySQL example is used as an example.**
- **A maximum of 20 SQL windows (including the homepage) can be opened at the same time in DMS. We recommend that you open no more than five SQL windows .**

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.

2. **In the top navigation bar, choose SQL Operations > SQL Window to open an SQL window.**

   Figure 10-3: Empty SQL window **shows the empty** `SQL window` **you opened.**

   Figure 10-3: Empty SQL window

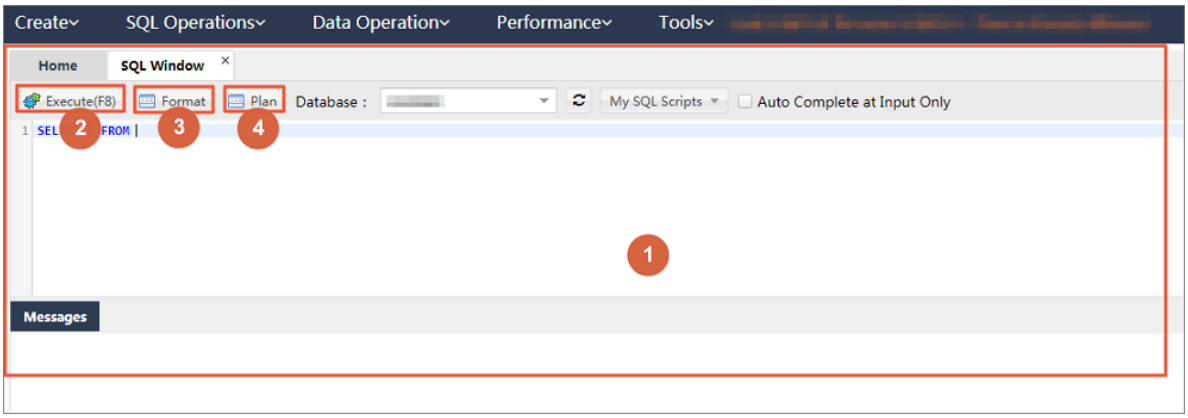


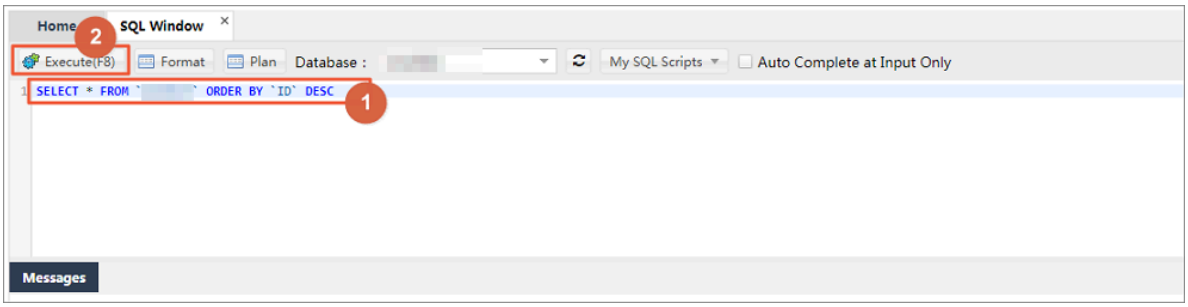
   Table 10-2: Numbered items in the SQL window **describes the numbered items in the SQL window.**

   Table 10-2: Numbered items in the SQL window

No.	Name	Description
1	SQL window	The green-framed area is the main body of the SQL window.
2	Run (F8) button	Click this button to run the entered SQL statements.
3	Format button	Click this button to format the entered SQL statements to make them more readable.
4	Execution Plan button	Click this button to display the execution plans of the selected SQL statements. You can optimize the SQL statements and improve SQL processing performance based on the execution plans.

3. **Enter the SQL statement you want to execute and click Run to complete the SQL query or update, as shown in *Figure 10-4: Run the SQL statement*.**

   Figure 10-4: Run the SQL statement



4. **You can view the result set, as shown in *Figure 10-5: View the result set*.**

   Figure 10-5: View the result set

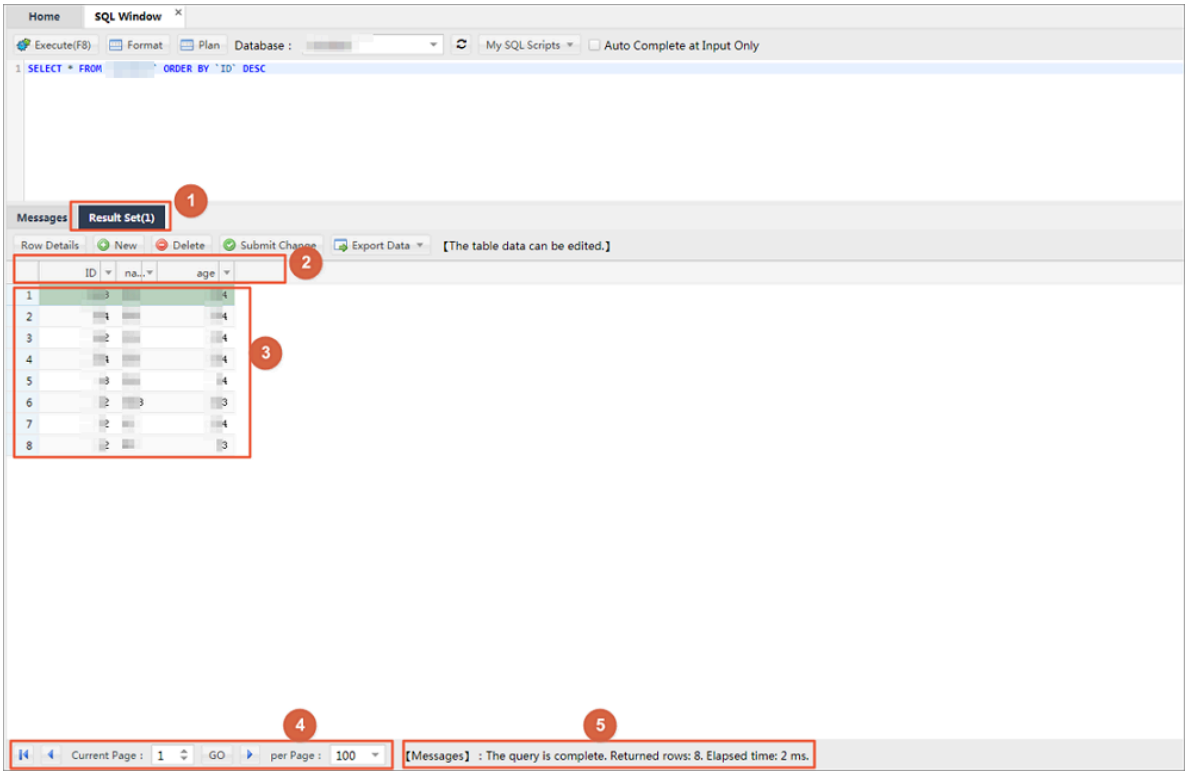


Table 10-3: Description of the numbered items in the result set

No.	Description
1	The Result Set tab shows the results returned by the SQL query statement.

No.	Description
2	The first row of the table shows the field names. If an alias has been specified for a field in the SQL statement, 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 not big enough 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	Progress of result acquisition and time elapsed.

5. **View the message about SQL execution.**

   Each time a data query (SELECT) or data correction (INSERT, UPDATE, or DELETE) statement is executed, DMS returns a message about the execution, including the status and impact.

   Figure 10-6: Data query **shows the message returned for data query.**

   Figure 10-6: Data query

   Figure 10-7: Data correction **shows the message returned for data correction.**

   Figure 10-7: Data correction

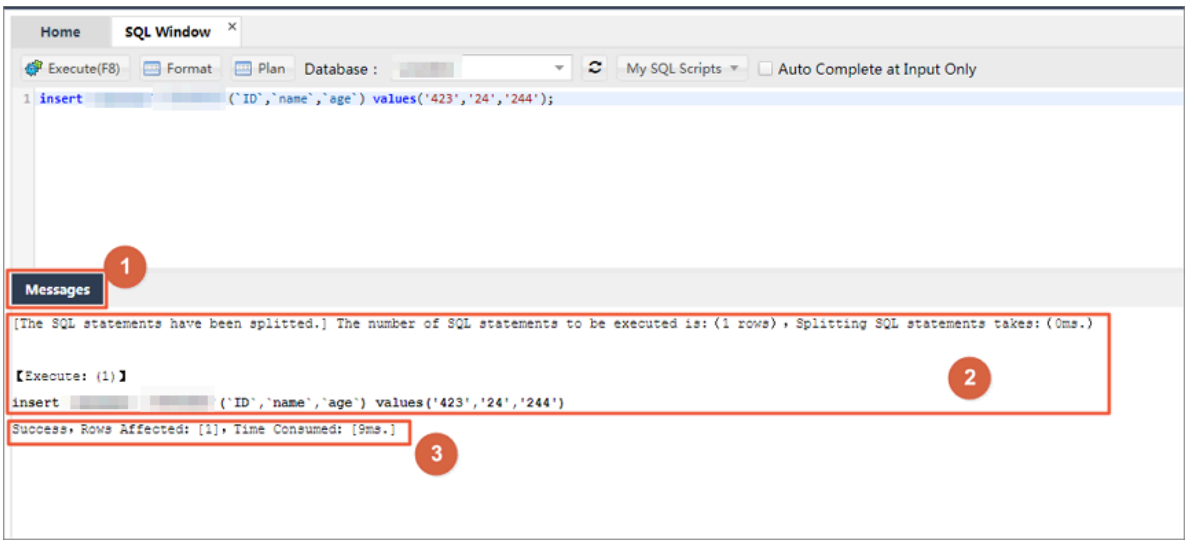


   Table 10-4: Description of the numbered items in the data correction window **describes the numbered items in the data correction window.**

Table 10-4: Description of the 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 statements step by step. · Analyzes the entered SQL statements. · Runs the SQL statements in the database. · Displays the queried data. · Displays 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 statements.

6. **Run SQL statements in batches.**

   DMS supports batch execution of SQL statements, as shown in *Figure 10-8: Batch execution*.

   Figure 10-8: Batch execution

   - **1: Shows the execution results of the first SQL statement.**
   - **2: Shows the execution results of the second SQL statement.**

   a) **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.**

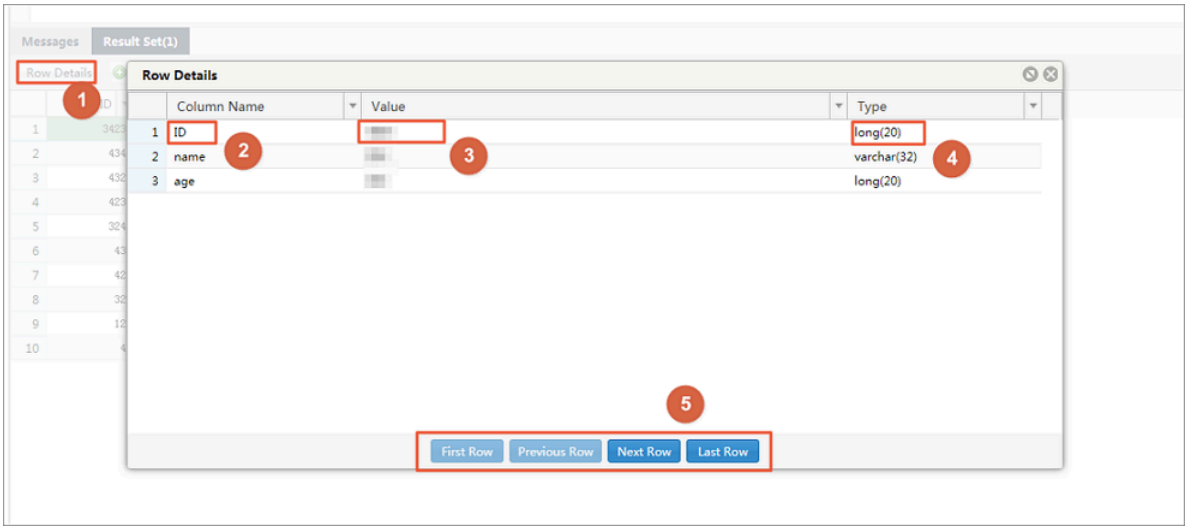
      Wait until all SQL statements are executed.

   c) **View the execution results.**

      If you run the 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** *Figure 10-9: Single row details*.

Figure 10-9: Single row details



The following table describes the numbered items in the Single Row Details window.

Table 10-5: Description of the 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 displays every `Field name`, `Field value`, and `Field type` of the record.
2	`Field name`: If you have specified aliases for fields, the aliases are displayed.
3	`Field value`: DMS automatically parses and displays the field values. Data such as time and binary code is formatted as a string for clear display.
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 previous and subsequent records.

8. **Optional: Edit the queried data in the result set.**

   - **Click Add to add a row of data to the currently queried table.**
   - **Click Delete to delete the selected row of data from the result-set table.**
   - **Select the 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 statements 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 Beautify to improve the readability and writability of the selected SQL statements.**

   - **Only the selected SQL statements are beautified. If you do not select any SQL statements, all the SQL statements that you entered are beautified.**
   - **The beautify function reformats your SQL statements to standard and readable statements, without changing the SQL execution logic and semantics or affecting the execution.**

   Examples:

   Figure 10-10: Original SQL statements **shows the original SQL statements.**

   Figure 10-10: Original SQL statements

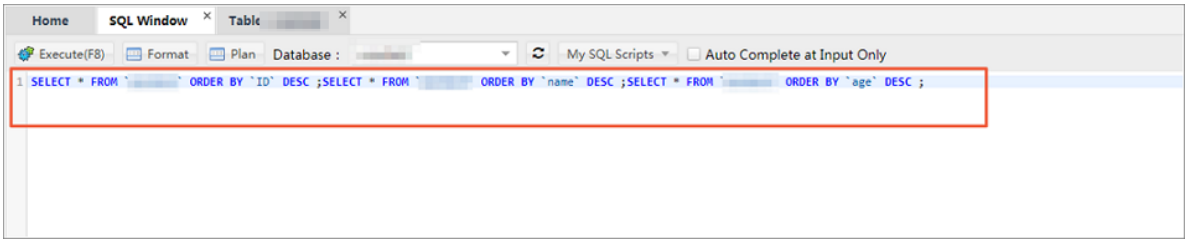
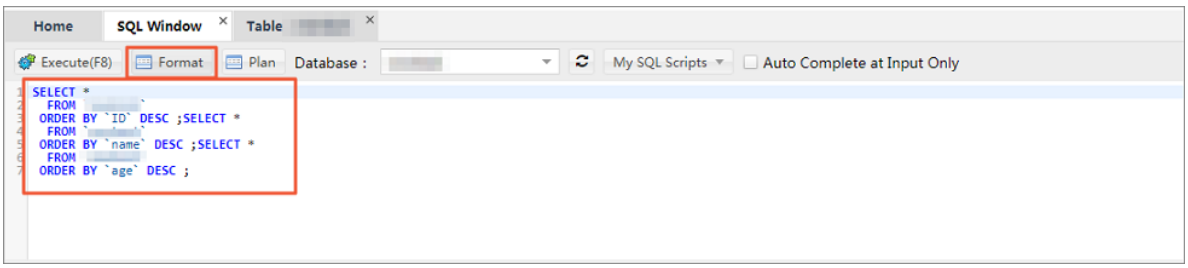


   Figure 10-11: Beautified SQL statements **shows the beautified SQL statements.**

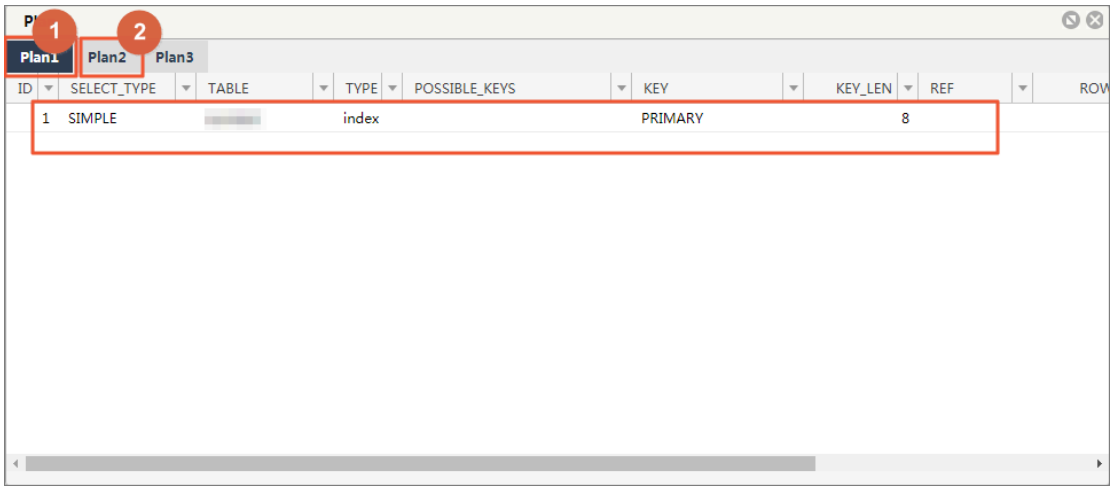
   Figure 10-11: Beautified SQL statements



10. Click Plan to view the execution plan if you want to troubleshoot SQL-related problems or optimize SQL performance.

- After you click 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 such as the type of the execution plan and possible keys.
  - Different databases display execution plans in different ways and to varying extents.
  - If you want to view the execution plans of several SQL statements, DMS displays the execution plan of each SQL statement in detail on different tabs, as shown in *Figure 10-12: Execution plan*.

Figure 10-12: 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.

10.3.2.2 Restore a saved SQL window

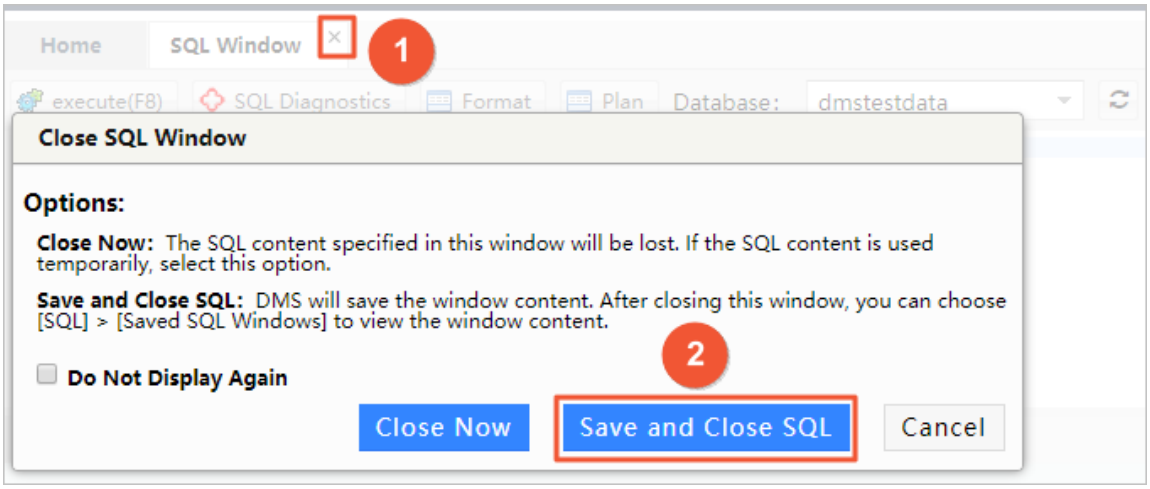
This topic describes how to restore a saved SQL window.

Context

- A MySQL example is used as an example.
- A maximum of 20 SQL windows (including the homepage) can be opened at the same time in DMS. We recommend that you open no more than five SQL windows .

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. In 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 windows you opened, and the SQL statements you entered in the SQL windows.
   - When you close a SQL window, DMS prompts you to confirm whether you want to save the content within the window.



   - 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 work content is saved.

     If you click Close, DMS does not save the present work in 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.

10.3.2.3 Manage frequently used SQL commands

This topic describes how to manage frequently used SQL commands in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. **In the top navigation bar, choose SQL Operations > SQL Window to open an SQL window.**
3. **Perform the following operations:**
   - **Add a frequently used SQL command.**

     Choose My SQL > Add My SQL to add a frequently used SQL command.

     - **Applicable scope: The custom SQL command is applicable to all 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 (with 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 frequently used SQL commands you saved.
   - **Manage your SQL commands.**

     Choose My SQL > Manage My SQL to manage frequently 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.**

10.3.2.4 Use the SQL template

This topic describes how to use the SQL template in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. **In the top navigation bar, choose SQL Operations > SQL Window. A SQL window appears.**

   The SQL template is displayed in the rightmost part of the SQL window.
3. **Double-click an SQL command or drag it into the SQL window. Then you can use or reference the command.**

   You can directly modify the commands referenced from the template even if you are not familiar with the commands.

10.3.3 Table operations (based on the Table directory tree)

10.3.3.1 Open a table-based SQL window

This topic describes how to open a table-based SQL window in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. **In the left-side directory tree, 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 records of the table.

10.3.3.2 Edit table data

This topic describes how to manage frequently used SQL commands in DMS.

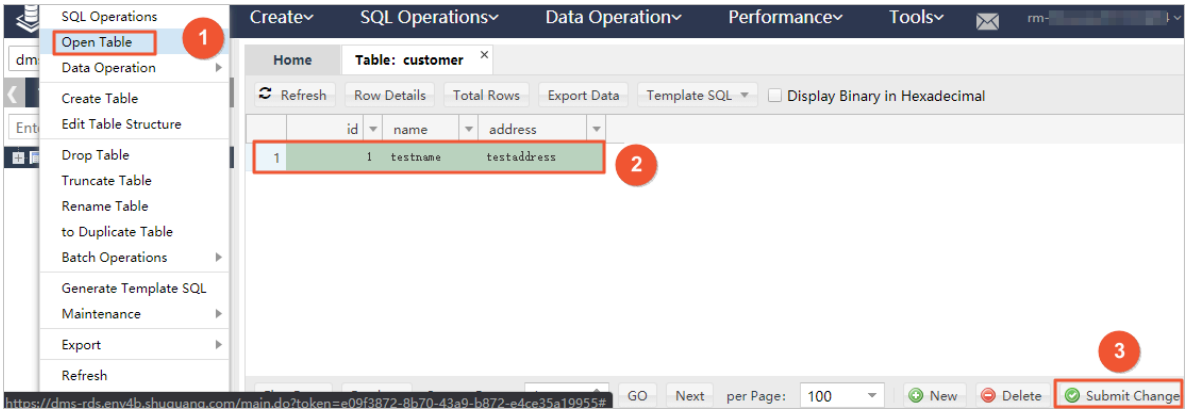
Context

- **A MySQL database is used as an example.**
- **This function applies to tables with average data volumes. For tables containing large volume of data, locate the data before editing. Data locating may take some time.**

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. **In the left-side directory tree, right-click a table and choose Open Table from the shortcut menu.**

   A window appears, indicating the data of the selected table.



- **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 the data, click Submit Changes to submit the modified data.**

10.4 Database development

10.4.1 Overview

This topic describes how to add, modify, delete, and manage objects such as indexes, foreign keys, and stored procedures.

10.4.2 Table

10.4.2.1 Create a table

This topic describes how to create a table in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.

2. You can create a table through any of the following methods:
   - In 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, enter the key name, column name, and information of the referenced databases, tables, and columns.
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 description.
   - You can click More to edit table parameters.
8. Click Save. DMS generates the SQL statements used to create a table.

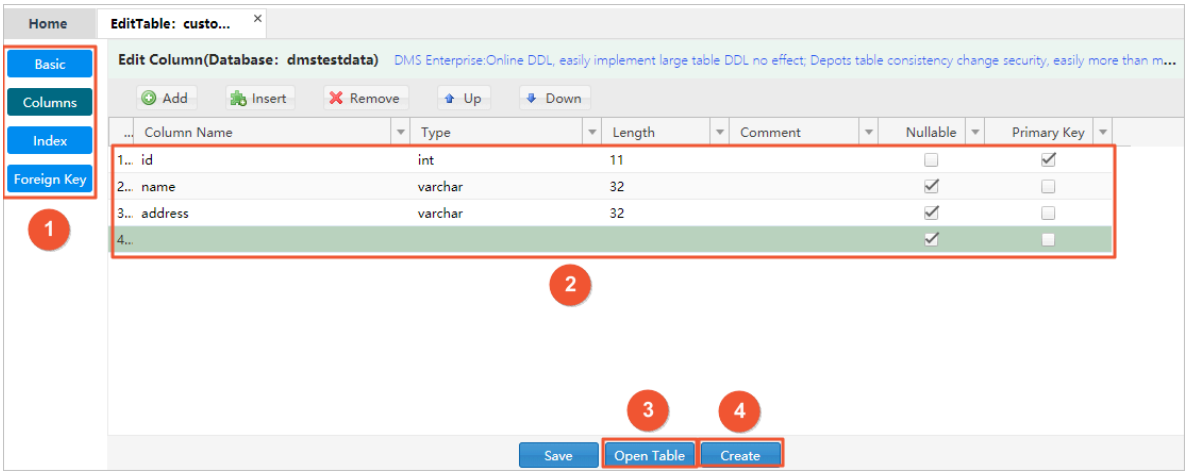
   Click OK after you confirm the SQL statements. DMS then adds the table to your database.

10.4.2.2 Edit a table

This topic describes how to edit a table in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS 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 into the window.



- 1: Select a table object type, such as `Columns` 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 to view and modify table data.
- 4: Click Create to view the statements used to create a table.

4. Click Save. DMS displays the SQL statements used to modify the table structure.

   Click OK after you confirm the SQL statements. DMS then saves the modified table structure to your database.

10.4.2.3 Delete a table

This topic describes how to delete a table in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS 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.

10.4.2.4 Create a similar table

This topic describes how to duplicate a table in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. In the left-side Table directory tree, right-click the table you want to copy and choose Create Table Like 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. The structure of the created table is the same as that of the source table.

   A similar table is created.

10.4.2.5 Generate SQL statement templates

This topic describes how to generate SQL templates in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. In the left-side Table directory tree, right-click the table you want to copy and choose Create SQL Template from the shortcut menu.
3. DMS generates SQL INSERT, UPDATE, SELECT and CREATE TABLE statement templates as a reference when you perform SQL operations.

10.4.2.6 Query table information

This topic describes how to query table information in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. In the left-side Table directory tree, right-click the table 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 of the table.
4. Click the Create Statement tab to view the table creation statements.

10.4.2.7 Clear data

This topic describes how to clear table data in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS 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.**

   Notice:
   Clearing table data is a high-risk operation and may affect your data usage. DMS prompts you to confirm whether to clear table data.
3. **Click Yes if you want to clear table data. DMS then clears data of the selected table.**
4. **Open the table to check whether its data is cleared.**

10.4.2.8 Perform operations on tables in batches

This topic describes how to perform operations on tables in batches in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS 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 to be deleted.**
   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 operations on tables in batches.**

   You can clear data, delete or 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 appears.

   b) **Select the tables to be operated 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.

10.4.2.9 Maintain a table

This topic describes how to maintain and optimize a table in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS 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.

10.4.3 Manage indexes

This topic describes how to add, modify, or delete indexes in DMS.

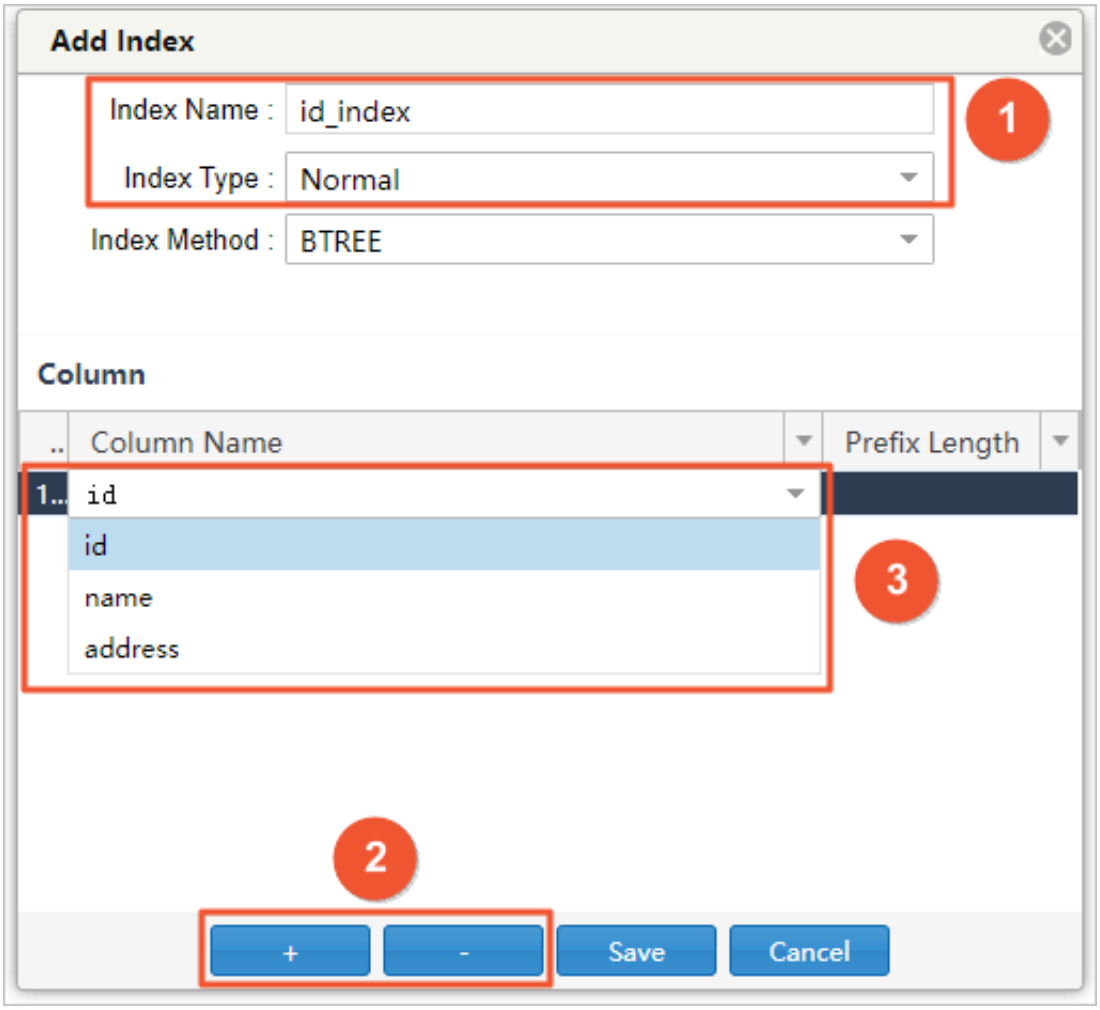
Procedure

1. *Log on to an ApsaraDB for RDS 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 appears.

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 enter or select values from the drop-down list. You can set a prefix length for a variable-length field (such as varchar) to save space occupied by the index.

4. Click Save.

   DMS generates SQL statements 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 appears.

     The method of modifying an index is similar to that of adding one, except that the SQL statements delete the old index before adding a new one.
   - In the left-side Table directory tree, right-click an index and choose Delete Index from the shortcut menu. The Delete Index window appears. Click OK to delete the index.

10.4.4 Manage foreign keys

This topic describes how to add foreign keys in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. In the left-side Table directory tree, right-click the table to be modified and choose Edit Table Structure from the shortcut menu.
3. On the Edit Table page that appears, click the Foreign Key tab to edit foreign keys.
4. Enter the foreign key information, and set the fields of foreign keys and referenced tables.
5. Click Save.

10.4.5 Create partitions

This topic describes how to create partitions in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. In the left-side Table directory tree, right-click a table and choose New 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 created table structure.

   A window is displayed for you to confirm the SQL statements used to create the table.

5. Click OK. DMS creates the partition table based on the partition fields and partitioning logic that you have configured.
6. After executing the SQL statements, check whether the partition table is created.

10.4.6 Create a stored procedure

This topic describes how to create and manage stored procedures in DMS.

Context

A MySQL database is used as an example.

Stored procedures, functions, triggers, and events are considered programmable objects in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. Click the left-side Programmable Object directory tree, and choose Stored Procedure > Create (Stored Procedure).

   The Create Stored Procedure tab is displayed.

3. Enter a name and a description for the stored procedure.
4. Click OK.
5. DMS provides a template for creating stored procedures. You only need to edit the stored procedure part.
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 set the input parameters for the stored procedure.

   Set the input parameters. In this example, set `cnt` to `80` to search for records that meet the `Value=80` condition.

8. Click Execute to execute the stored procedure.

   DMS displays output parameters or intermediate result set of the stored procedure, if any.

   - The Message tab displays messages about the execution, such as output variables and intermediate result sets.
   - The Intermediate Result Set 1 tab displays the result set generated during the execution. If multiple intermediate result sets are available, DMS will generate multiple tabs, such as Intermediate Result Set 1, Intermediate Result Set 2, and Intermediate Result Set 3.

9. Click the Intermediate Result Set 1 tab.

   DMS displays records with the value of 80.

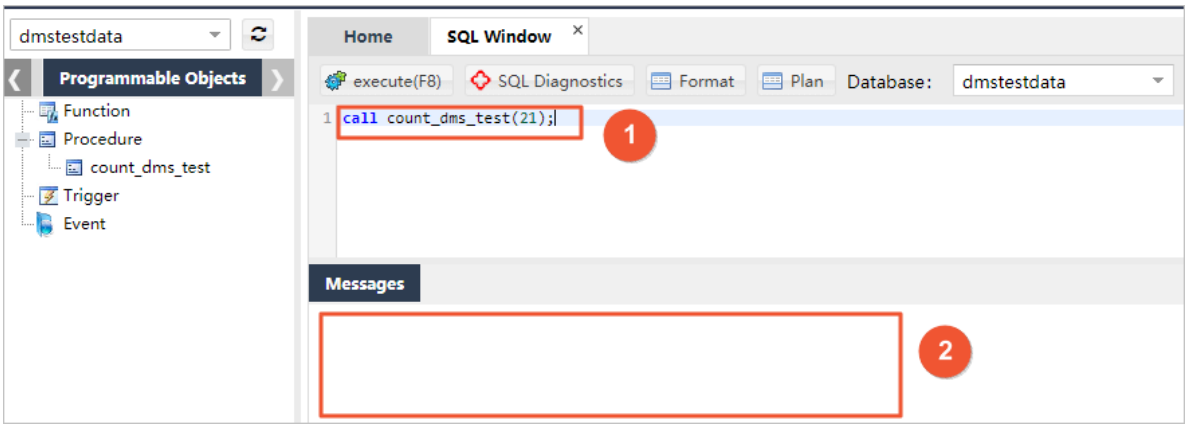
10. You can set the options when creating the stored procedure. Click Option Settings to set options for the stored procedure.

11. After a stored procedure is created, it is added to the Programmable Object directory tree.

    You can perform other operations related to the stored procedure through the following menu options:

    - Create
    - Edit
    - Delete
    - Execute

12. You can run the stored procedure in the SQL window.



- 1: Run the call stored_procedure_name command to call a stored procedure.
- 2: The SQL window shows the result set of the stored procedure, if any.

10.4.7 Create a function

This topic describes how to create a function in DMS.

Context

Functions, stored procedures, triggers, and events are considered programmable objects in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. **In the left-side Programmable Object directory tree, choose Function > Create (Function).**

   The Create Function page is displayed.
3. **Set basic information of the new function.**
4. **Click OK.**

   The Edit Function page appears. DMS generates a function creation template.
5. **Enter information in the function part.**
6. **Click Save. DMS then checks whether the function is correctly defined. If not, DMS returns an error message.**

   DMS runs the correct function definition in your database, and returns a message, indicating that the function is saved.
7. **Click Execute to execute the function.**
8. **Enter a parameter such as** `wednesday` **and click Execute to execute the function.**
9. **Click Option Settings to set different options for the function.**

   You can also run the function in the SQL window.

10.4.8 Create a view

This topic describes how to create and manage custom views in DMS.

Procedure

Create a view

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. **Click the View directory tree on the left side to check the views of the current database.**

3. Right-click the blank space and choose New View from the shortcut menu.

   The Create: View page is displayed.

4. Set basic information of the view.

   The following example shows how to filter records in the dmstest table whose values are even numbers, and output the id and name fields.

5. Click Save Changes. DMS generates SQL statements used to create the view based on your settings.
6. Click OK after you confirm the SQL statements. DMS saves the defined view to your database.
7. The saved view is added to the View directory tree on the left side. You can click the view to display its definition.

Check the view

8. Right-click View and choose Check View from the shortcut menu to query data through the newly created view.
9. You can perform view-related operations in DMS.

   The menu options include:

   - View Data
   - Create View
   - Edit View
   - Delete View
   - Refresh Views

10.4.9 Create a trigger

This topic describes how to create and manage a trigger in DMS.

Context

Triggers, functions, stored procedures, and events are considered programmable objects in DMS.

Procedure

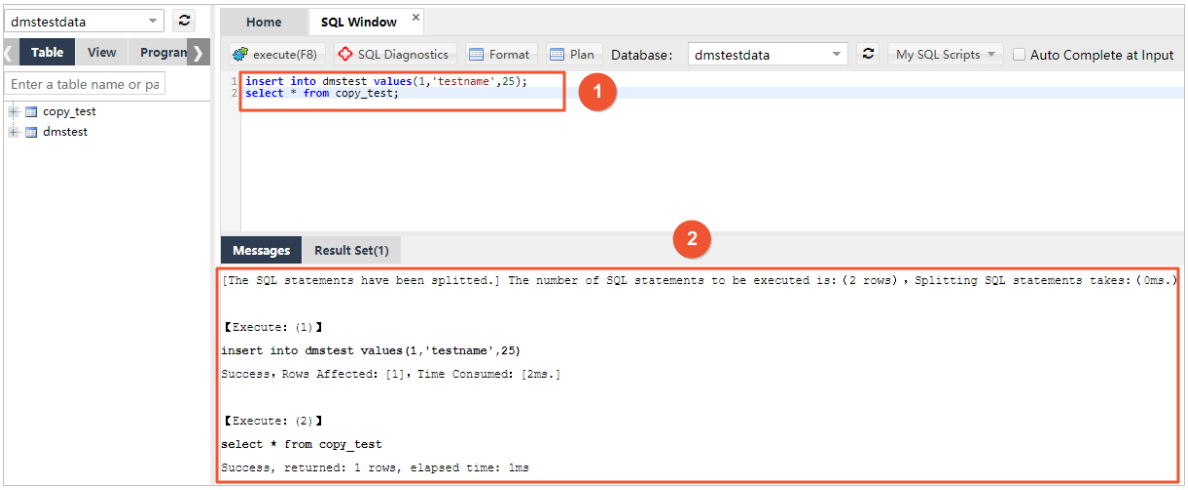
1. *Log on to an ApsaraDB for RDS instance through DMS*.

2. Click the Programmable Object directory tree on the left side, and choose Trigger > Create (Trigger).

   The Create: Trigger tab is displayed.

   - 1: Trigger table.
   - 2: Trigger settings.
     - Enter a name for the trigger.
     - Select dmstest from the drop-down list as the trigger table.
     - Select AFTER from the drop-down list as the trigger time.
     - Select INSERT from the drop-down list as the trigger event.
   - 3: Set a trigger statement.
     - Set the operations to be performed when the trigger event occurs.
     - When data is inserted into the dmstest table, the trigger in this example inserts data into the copy_test table and records the insertion time in copy_test.time.
3. Click Save after you finish the trigger settings. DMS then generates the SQL statement to be executed by 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 has been saved. In the left-side navigation pane, choose Programmable Object > Trigger to view the trigger you created.

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 messages about the execution of the SQL statements. The messages indicate that a row is inserted into the dmstest table and that this row is also added to the copy_test table.

6. Check the result set in the SQL window to verify whether the insert operation is correctly performed 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)

10.4.10 Create an event

This topic describes how to create and manage events in DMS.

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 disabled. You need to enable event support by modifying the configuration file or executing the `SET GLOBAL event_scheduler = ON;` statement.

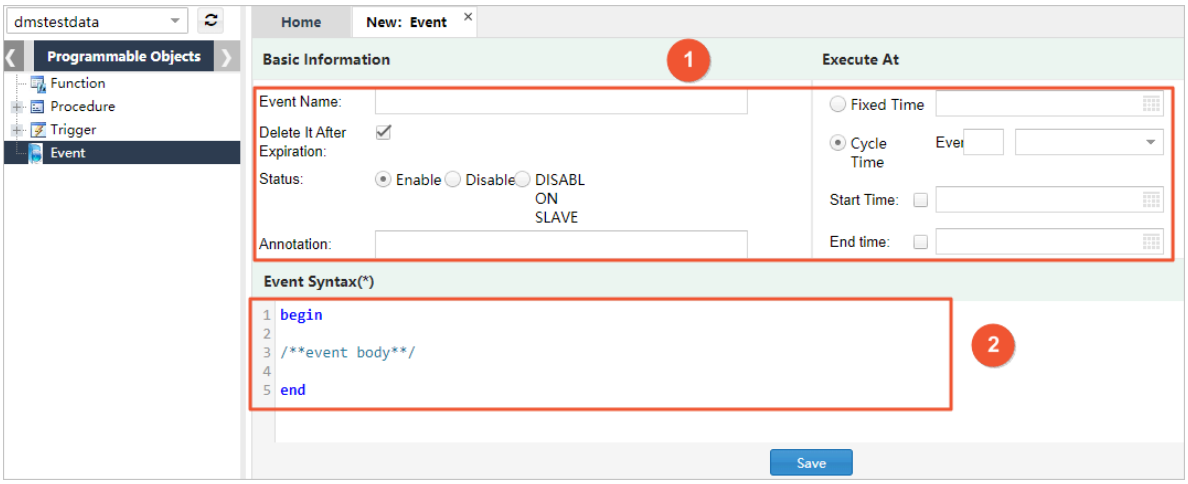
Context

Events, triggers, functions, and stored procedures are considered programmable objects in DMS.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. Click the Programmable Object directory tree on the left side, and choose Event > Create (Event).

   The Create Event page is displayed.



   1: In the event setup area, set the event name, cycle, start time, end time, status, and comment, and choose whether to enable cyclic execution.

   2: In the event execution Statement area, set the operations to be performed when a scheduled event is triggered.
3. Set an event trigger rule and the SQL statements for event execution.
4. Click Save. DMS generates the SQL statements used to create the event.

5. After you confirm that the SQL statements are correct, click OK. DMS then executes the edited event in your database.
   - If the event is created, DMS returns a message, indicating that the event is saved.
   - In the left-side navigation pane, choose Programmable Object > Event to view the event you created.
6. Check whether the event is properly executed in the SQL window.

   In this example, the event executes SQL statements to insert a piece of data into the `copy_test`table every minute. Check the copy_test table to see whether the data is inserted as programmed.
7. You can perform various event-related operations in DMS.

   The menu options include:
   - Create (Event)
   - Edit (Event)
   - Delete (Event)

10.5 Data processing

10.5.1 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 ApsaraDB for RDS instance through DMS*.
2. In the top navigation bar, choose Data Processing > Import.

   The Import tab is displayed.

   The Import tab contains the import toolbar and import history.

   If you have imported data, you can view previous operations in History.

3. Click New Task.

   The Import File dialog box is displayed.

   - Select the type of the file to be imported. Only SQL and CSV files are supported currently.
   - If the data file uses a character set, you can manually specify the character set . DMS automatically detects character sets of files.
   - DMS terminates the import task if an error occurs while executing an SQL statement. You can select `Ignore Errors` to proceed. However, this operation may affect subsequent operations.
   - You can enter a brief description of the import task for later review.

4. Click Start to start the import task.

   If the imported data has any error, DMS terminates the import process and return an error message. You can modify the data file to correct the error and import it again.

   If the imported data and SQL statements are correct, DMS displays the import progress, volume of the imported data, and time elapsed.

   After the import is completed, you can view the import task in History.

   Click Task Number to view the execution details of the task.

10.5.2 Export data

10.5.2.1 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 ApsaraDB for RDS instance through DMS*.
2. In the top navigation bar, choose Data Processing > Export to go to the Export page.
3. On the Export page, choose New Task > Export Database.

4. On the Export SQL Result Sets tab, select a database, file type (SQL or CSV), and content to be exported (structure and data, only data, or only 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 and review the export details, and download the exported data in the Export History List.

   After the export is complete, DMS automatically downloads the exported file to your local computer. You can also click Download File to download the exported file.

   You can view previously submitted export tasks in the Export History List. Click a task name to view the task details and download the exported data.

10.5.2.2 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 ApsaraDB for RDS instance through DMS*.
2. In 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 Export SQL Result Sets tab, 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 the SQL statements.

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.

10.6 Performance

10.6.1 Lock wait

10.6.1.1 View lock-waits

When an RDS for MySQL session is waiting for an exclusive InnoDB row lock held by another session, InnoDB lock wait will occur. This topic describes how to view lock wait in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. **In the top navigation bar, choose Performance > InnoDB Lock Wait.**

   If transactions of the current instance are waiting for locks, the lock hold and lock wait are displayed.
3. **Move the pointer over the `Lock` or `Lock-Wait` icon to view the locks or lock-waits, and related session IDs.**
4. **Click**  **to reload the data.**

10.6.1.2 Release lock wait

This topic describes how to release lock wait in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. In the top navigation bar, choose Performance > InnoDB Lock-Wait.

   If transactions of the current instance are waiting for locks, the Lock Hold and Lock Wait icons are displayed.
3. Move the pointer over the `Lock` or `Lock-Wait` icon to view the locks or lock-waits, and related session IDs.
4. Click the `Lock` or `Lock-Wait` icon.

   The Delete Session message appears.
5. Click Yes to terminate the current session.

10.7 Extended tools

10.7.1 Table data volume statistics

This topic describes how to view table data volume statistics in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. In the top navigation bar, choose Tools > Table data amount.
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 size (in bytes), data, 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 (in MB), number of table rows, global sorting, and storage engine. You can also perform the paging, refresh, and reset operations.

10.7.2 ER diagrams

This topic describes how to view entity-relationship (ER) diagrams in DMS.

Context

A MySQL database is used as an example.

Procedure

1. *Log on to an ApsaraDB for RDS instance through DMS*.
2. In the top navigation bar, choose Tools > ER Diagram.

   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. You can perform the following operations:
   - Select another database from the DB:mysql drop-down list.
   - Click the Sorting: Sorting Options drop-down list to sort tables in ascending or descending order of table name, field count, or relationship count.
   - Click Refresh to refresh the current ER diagram.
   - Click View SQL Scripts to list the SQL statements used to create all tables of the current database.
   - Click Download SQL Scripts to download the SQL statements used to create all tables of the current database.
   - Click Download XML Files to download the table-creating SQL statements of the current database in XML format.
   - 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.

11 Server Load Balancer (SLB)

11.1 What is Server Load Balancer?

Server Load Balancer (SLB) is a traffic distributing service that distributes inbound traffic to multiple ECS instances based on scheduling algorithms. SLB extends the service capability of applications and enhances their availability.

Overview

By setting a virtual service address, SLB virtualizes backend ECS instances into a high-performance and high-availability application service pool, and then distributes client requests to ECS instances in the pool based on scheduling algorithms.

SLB checks the health status of the ECS instances in the backend server pool and automatically isolates unhealthy ones to eliminate single points of failure (SPOFs), improving the overall service capability of applications.

Components

SLB consists of three components:

- **SLB instances**

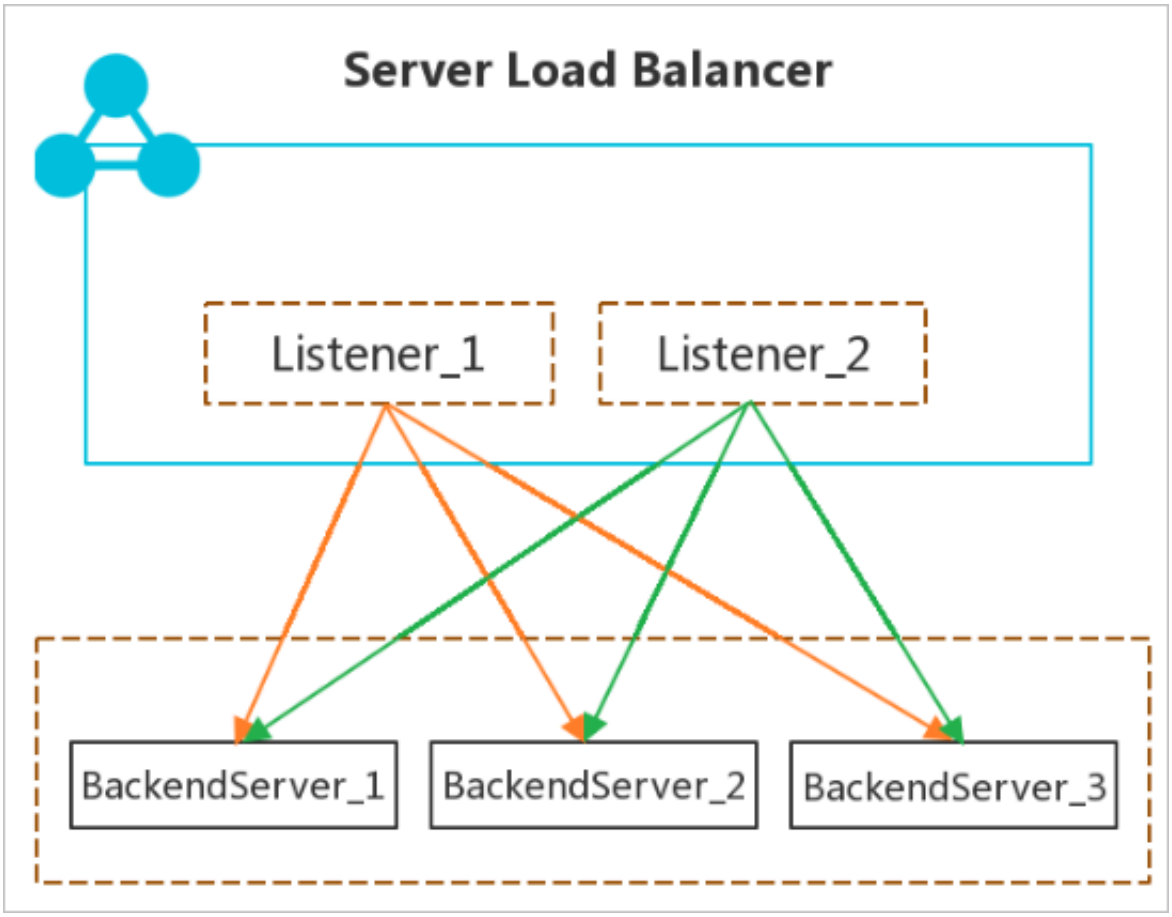
  An SLB instance is a running load balancing service that receives inbound traffic and distributes the traffic to the backend servers. To use the SLB service, you must create an SLB instance with at least one listener and two backend servers.

- **Listeners**

  A listener checks client requests and forwards them to backend servers based on the configured scheduling algorithms. Listeners also perform health checks on backend servers.

- **Backend servers**

  Backend servers are the ECS instances attached to SLB instances to receive the distributed client requests. You can add ECS instances to the default server group , a VServer group, or an active/standby server group for easy management.



Benefits

- **High availability**

  SLB is designed with full redundancy to avoid SPOFs and support zone-disaster recovery. SLB can be used with DNS for geo-disaster recovery to offer a service availability of 99.95%.

  SLB can be scaled based on application loads and can provide continuous service during traffic fluctuations.

- **Scalability**

  You can scale the service capability of applications as needed by adding or removing backend servers.

- **Cost-effectiveness**

  SLB is 60% more cost-efficient than traditional hardware load-balancing systems.
- **Security**

  SLB can be used with Apsara Stack Security to provide a DDoS protection capability of 5 Gbit/s.

11.2 Log on to the SLB console

This topic describes how to log on to the Apsara Stack Cloud Management (ASCM) console from Google Chrome and then access Server Load Balancer (SLB).

Prerequisites

- **Before logging on to the ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel. The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.**
- **We recommend that you use the Google Chrome browser.**

Procedure

1. **In the address bar, enter the URL used to access the ASCM console. Press Enter.**
2. **Enter your username and password.**

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

   Note:
   When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).
3. **Click Login to go to the ASCM console homepage.**
4. **In the top navigation bar, choose Products > Networking > Server Load Balancer.**

11.3 Quick start

11.3.1 Overview

This quick start tutotial describes how to create a public-facing Server Load Balancer (SLB) instance and how to forward requests to two backend servers.

> **Note:**
> **Before creating an SLB instance, you must determine the region, type, and billing method of the SLB instance. For more information, see *Before you begin*.**

This tutotial includes the following content:

1. *Create an SLB instance*

   Create an SLB instance. An SLB instance is a running entity of the SLB service.

2. *Add listeners and backend servers.*

   Configure listening rules and backend servers for the SLB instance.

3. *Delete an SLB instance*

   If you no longer need the SLB instance, delete it to avoid extra fees.

11.3.2 Before you begin

Before you create a Server Load Balancer (SLB) instance, you must determine the listener type and network type for the SLB instance based on your business needs.

Plan the region of the SLB instance

When you select a region, note that:

- **To reduce latency and increase the download speed, we recommend that you select a region closest to your customers.**
- **To provide more stable and reliable load balancing services, SLB allows you to deploy primary and secondary zones in most regions. This implements disaster recovery across data centers in the same region. We recommend that you select a region that supports deployment of primary and secondary zones.**
- **SLB does not support cross-region deployment. Therefore, you must select the same region as the backend ECS instances.**

Select the network type of the SLB instance (Public Network or Internal Network)

SLB provides load balancing services for the public network and internal network:

- **If you want to use SLB to distribute requests from the public network, create a public-facing SLB instance.**

  A public-facing SLB instance provides a public IP address to receive requests from the Internet.
- **If you want to use SLB to distribute requests from the internal network, create an internal SLB instance.**

  An internal SLB instance only has private IP addresses and is only accessible from the internal network and not from the Internet.

Select a listener protocol

SLB supports Layer 4 (TCP and UDP) and Layer 7 (HTTP and HTTPS) load balancing.

- **A Layer 4 listener distributes requests directly to backend servers without modifying packet headers. After a client request arrives at a Layer 4 listener, SLB uses the backend port number configured in the listener to establish a TCP connection with a backend server.**
- **A Layer 7 listener is an implementation of reverse proxy. After a client request arrives at a Layer 7 listener, SLB establishes a new TCP connection over HTTP to a backend server, instead of forwarding the request directly to the backend server.**

  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 further deteriorate due to factors such as having an insufficient number of client ports and having too many backend server connections. We recommend that you use Layer 4 listeners for high performance purposes.

Create backend servers

Before you use the SLB service, you must create ECS instances, deploy applications on them, and add them to the SLB instance to process client requests.

When you create ECS instances, take note of the following items:

- **Region and zone of each ECS instance**

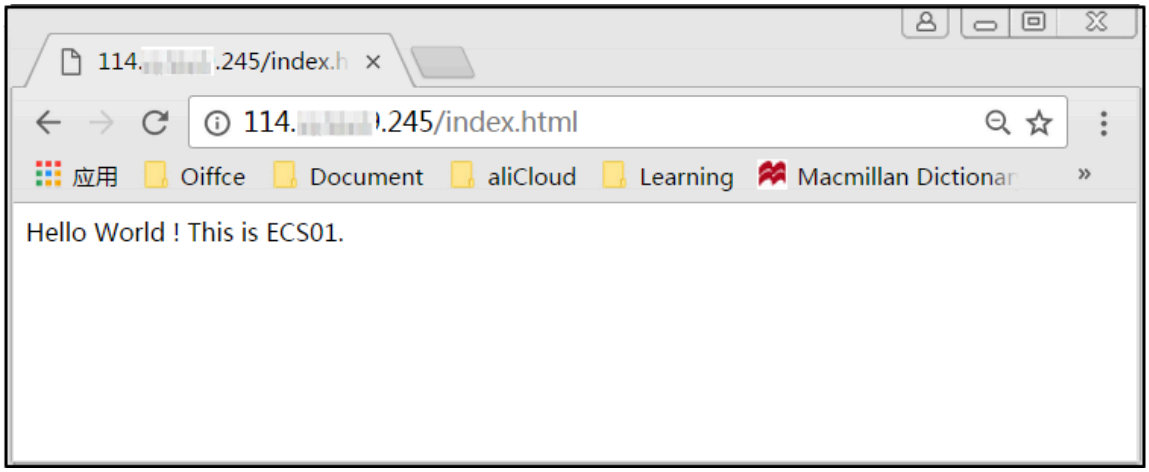
  Make sure that the region of each ECS instance is the same as that of the SLB instance.

  In this example, two ECS instances are created in the China (Hangzhou) region. They are named ECS01 and ECS02 as shown in the following figure. For more information about how to create an ECS instance, see the *Create an ECS instance* **topic in** *ECS User Guide*.

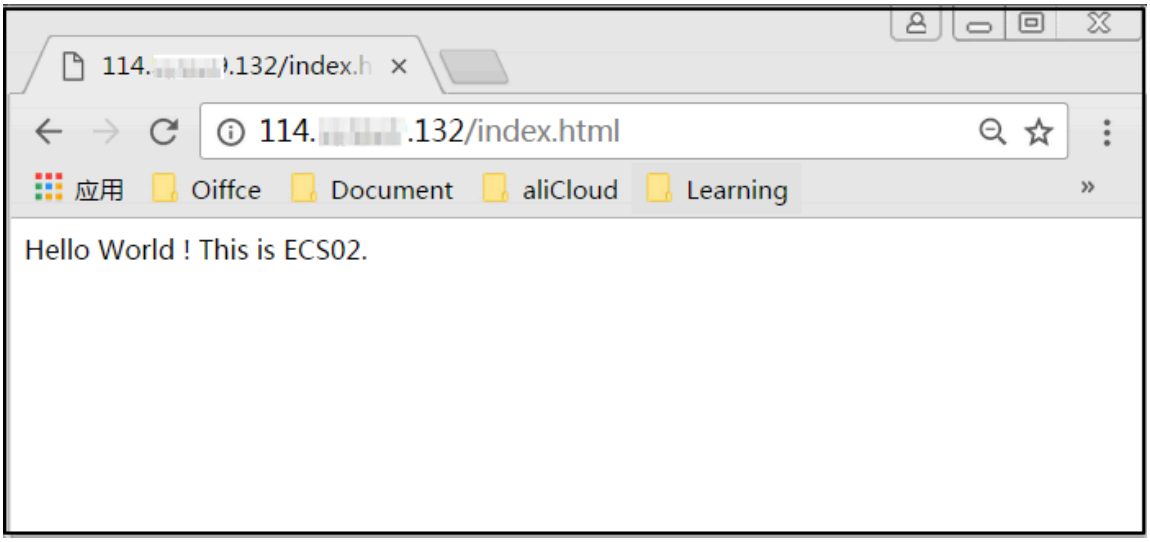
- **Application configuration**

  In this example, two static web pages are created on ECS01 and ECS02 by using Apache.

  - **Enter the EIP bound to ECS01 in the browser.**



  - **Enter the EIP bound to ECS02 in the browser.**



  No additional configuration is required after you deploy applications on the ECS instances. However, if you want to use a Layer 4 listener (TCP or UDP), and the ECS instances use a Linux operating system, ensure that the following parameters in the *net.ipv4.conf* file in */etc/sysctl.conf* are set to 0:

```
net.ipv4.conf.default.rp_filter = 0
net.ipv4.conf.all.rp_filter = 0
```

```
net.ipv4.conf.eth0.rp_filter = 0
```

11.3.3 Create an SLB instance

This topic describes how to create a Server Load Balancer (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

- **ECS instances are created and applications are deployed on them.**
- **The organization 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 SLB console*.
2. **On the Server Load Balancer page, click Create Instance.**
   - **Organization: Select an organization for the SLB instance to be created.**

Note:

Ensure that the organization of the SLB instance is the same as that of the backend ECS instances.

- **Resource Set: Select the resource set to which the SLB instance will belong from the drop-down list.**
- **Region: Enter the region of the SLB instance.**
- **Zone: Select a zone for the SLB instance from the drop-down list.**
- **Name: Enter the name of the SLB instance.**

  The name must be 2 to 128 characters in length and can contain letters, digits, hyphens (-), colons (:), commas (,), periods (.), and underscores (_). It must start with a letter and cannot start with `http://` or `https://`.
- **Network Access: Select the network access type, which can be Internal Network or Public Network. In this example, select Internal Network.**
- **Network Type: Select the network type, which can be Classic Network or VPC. In this example, select VPC.**
- **IP Version: Select the IP version.**
- **Service IP: Enter the service IP address. Ensure that the service IP address is valid. Otherwise, the SLB instance cannot be created. If this parameter is not set, the IP address that is automatically allocated by the system is used.**

3. **Click OK.**

What's next

Configure an SLB instance

11.3.4 Configure an SLB instance

This topic describes how to configure a Server Load Balancer (SLB) instance. After you create an SLB instance, you must configure the SLB instance so that it can forward traffic. You must add at least one listener and a group of backend servers. The following example sets a TCP listener and configures two ECS instances (ECS01 and ECS02) as backend servers. Static web pages are deployed on the ECS instances.

Procedure

1. *Log on to the SLB console*.
2. **On the Server Load Balancer page, click Configure Listener in the Actions column.**

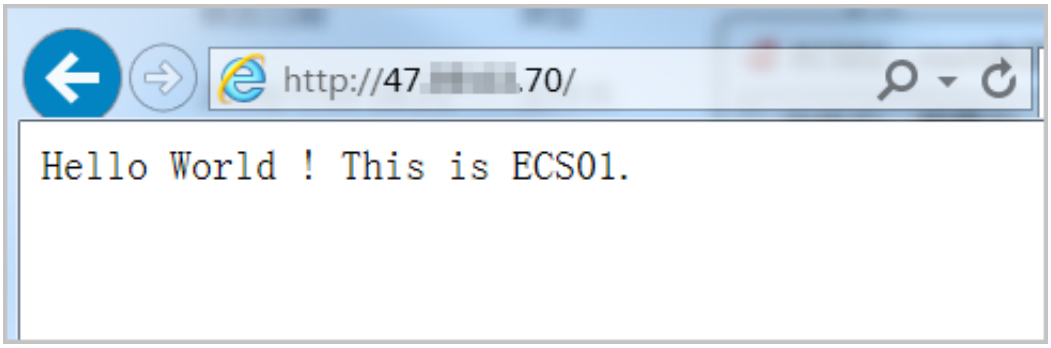
3. In the Protocol and Listener step, configure the listening rule based on the following information and use the default values for the remaining parameters.
   - Select Listener Protocol: In this example, select TCP.
   - Listening Port: the frontend port used to receive requests and forward the requests to backend servers.

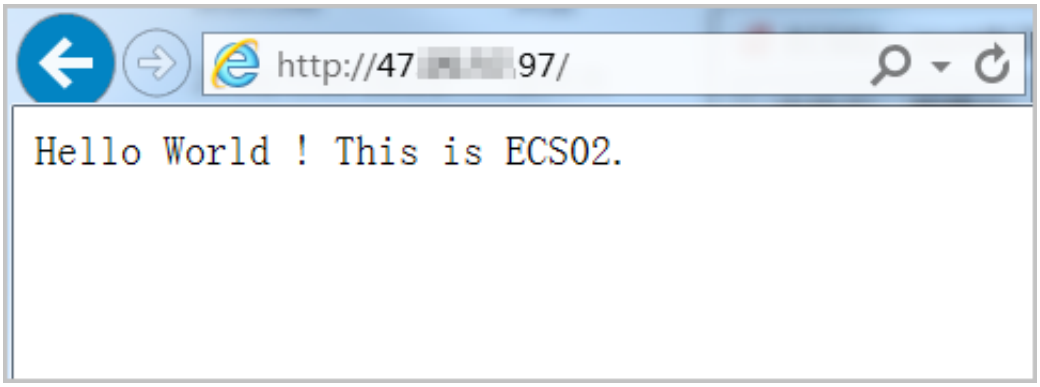
     In this example, set the port number to 80.
   - Enable Peak Bandwidth Limit: You can set a peak bandwidth to limit the external service capabilities that applications on the backend servers can provide.
   - Scheduling Algorithm: SLB supports the following scheduling algorithms. In this example, select Round-Robin (RR).
     - Weighted Round-Robin (WRR): Requests are distributed to backend servers in a specific scheduling sequence. A backend server with a higher weight receives more requests.
     - Weighted Least Connections (WLC): Requests are distributed to the backend server with the least number of connections. If the numbers of connection s of two backend servers are the same, the backend server with a higher weight will receive more requests.
     - Round-Robin (RR): Requests are distributed evenly and sequentially to backend servers.
4. Click Next. In the Backend Servers step, select Default Server Group and click Add More to add backend servers.
   a) In the dialog box, select previously created ECS01 and ECS02, and click Next: Set Weight and Port.
   b) Configure ports and weights for the added backend servers.
      - Port: the ports opened on backend servers to receive requests. They can be the same in an SLB instance. In this example, set the backend port numbers to `80`.
      - Weight: A backend server with a higher weight receives more requests. The default value is 100 and we recommend that you use the default value.

5. Click Next to configure health checks. In this example, use default health check configurations.

   With the health check feature enabled, when a backend server is declared unhealthy, SLB redistributes requests to healthy backend servers, and restores services to the original backend server when it is healthy again.
6. Click Next. In the Submit step, click Submit.
7. Click OK to go back to the Server Load Balancer page, and click  .

   If the health check status of a backend server is Normal, it indicates that the backend server is working properly and able to process requests.
8. In the web browser, enter the IP address of the SLB instance to test the service.





11.3.5 Delete an SLB instance

This topic describes how to delete a Server Load Balancer (SLB) instance. To avoid additional charges, you can delete an SLB instance when you no longer need the load balancing service. Deleting the SLB instance does not delete or affect backend ECS instances.

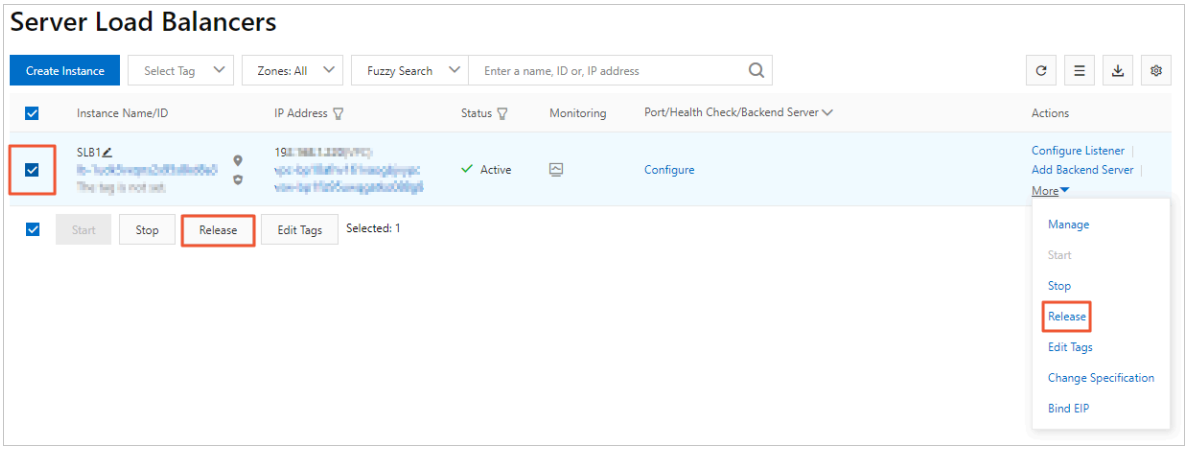
Context

Note:

- If you have resolved a domain name to the SLB endpoint, resolve it to another IP address first to avoid service interruptions.
- Only pay-as-you-go SLB instances can be released. Subscription SLB instances are automatically released if they are not renewed in a timely manner.
- The backend ECS instances are still running after the SLB instance is released. You can release the backend ECS instances if you do not need them anymore.

Procedure

1. *Log on to the SLB console*.
2. On the Instances page, select the region to which the target SLB instance belongs.
3. Find the target SLB instance, click Release at the bottom of the list or choose More > Release in the Actions column.



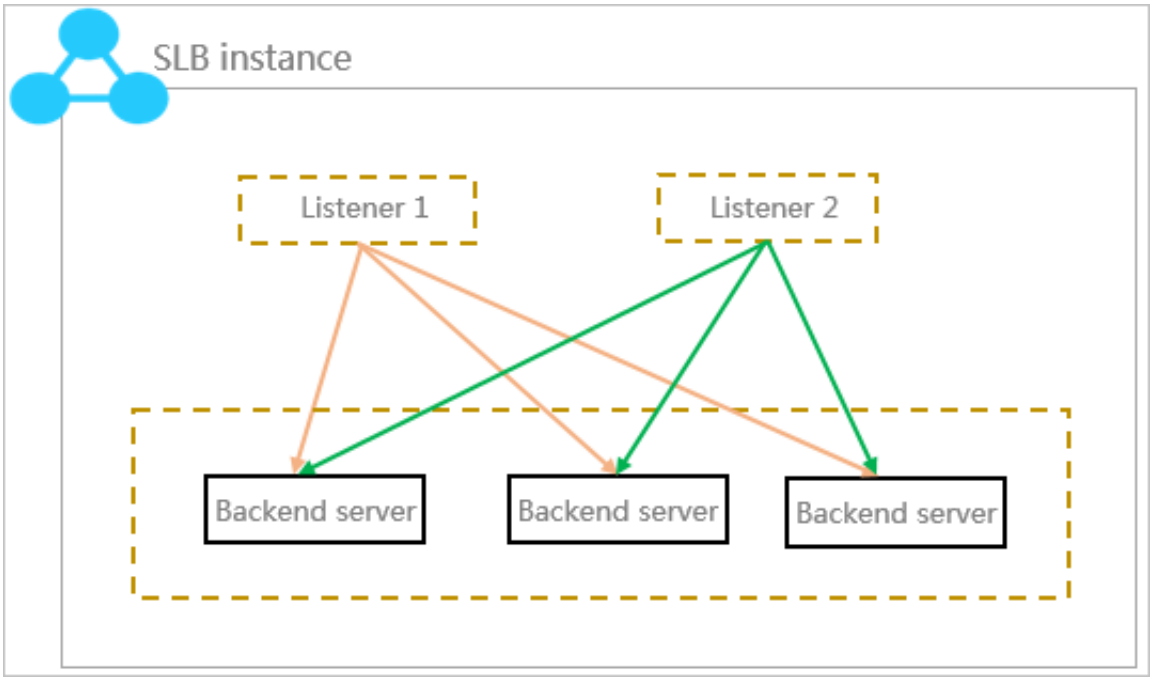
4. In the Release dialog box, select Release Now or Release on Schedule.

   If you select Release on Schedule, set a release time.
5. Click Next.
6. Click OK to release the SLB instance.

11.4 SLB instances

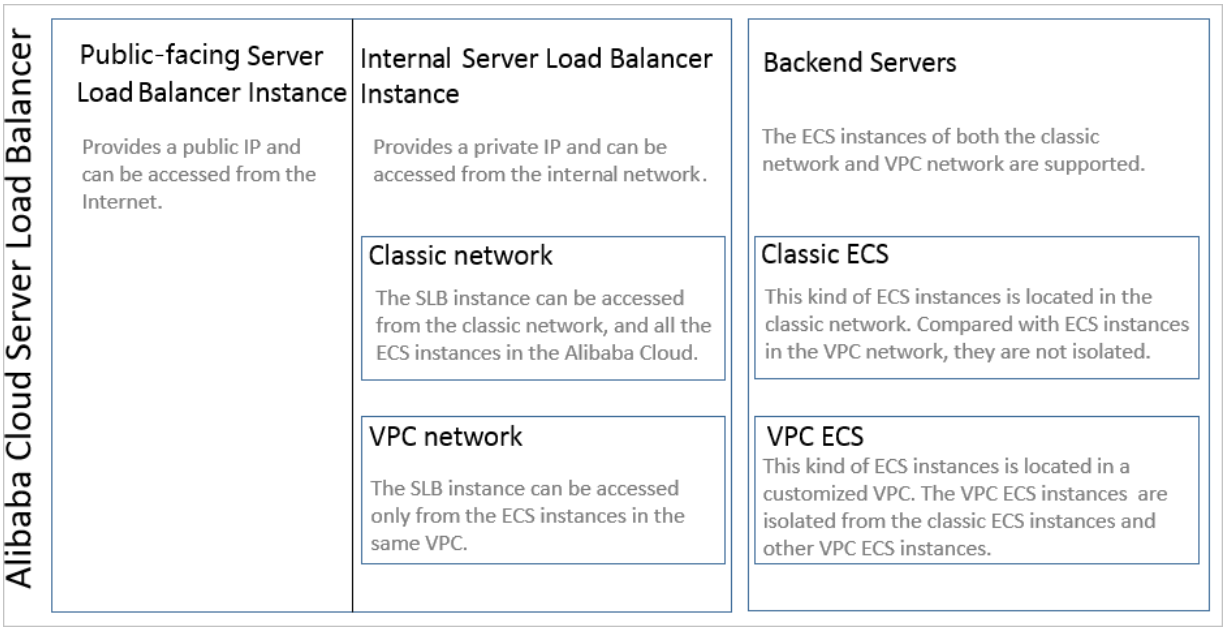
11.4.1 SLB instance overview

A Server Load Balancer (SLB) instance is a virtual machine in which the SLB service runs. To use the SLB service, you must create an SLB instance first, and then add listeners and backend servers to the SLB instance.



Instance network type

Alibaba Cloud provides Internet and intranet SLB services. If you create an Internet SLB instance, a public IP address is allocated to it. If you create an intranet SLB instance, a private IP address is allocated.



- **Internet SLB instances**

  An Internet SLB instance distributes client requests over the Internet to backend servers according to configured forwarding rules.

  When you create an Internet SLB instance, the system allocates a public IP address to the instance. You can resolve a domain name to the public IP address to provide public services.
- **Intranet SLB instances**

  Intranet SLB instances can only be used inside Alibaba Cloud and can only forward requests from clients that can access the intranet of SLB.

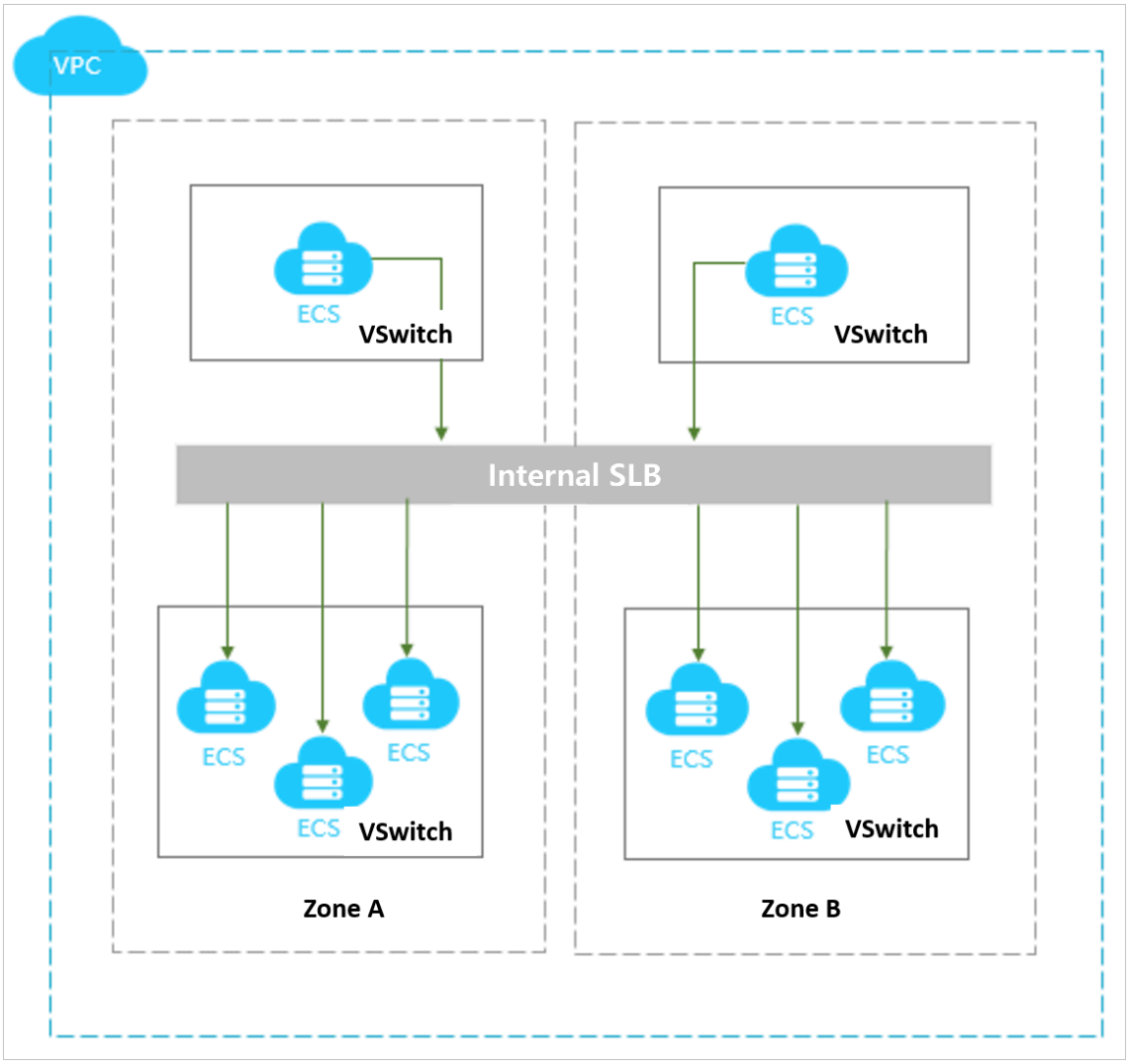
  For an intranet SLB instance, you can select the network type:

  - **Classic network**

    If you choose classic network for an intranet SLB instance, the IP address of the SLB instance is allocated and maintained by Alibaba Cloud. The instance can only be accessed by classic-network ECS instances.
  - **VPC network**

    If you choose VPC network for an intranet SLB instance, the IP address of the SLB instance is allocated from the CIDR block of the VSwitch that the instance

belongs to. An SLB instance of the VPC-type network can only be accessed by ECS instances in the same VPC.



Instance specifications

SLB provides shared-performance instances and guaranteed-performance instances.

- Shared-performance instances

  All shared-performance instances share SLB resources, which means their performance cannot be guaranteed.

- **Guaranteed-performance instances**

  The performance of guaranteed-performance instances are set according to their selected specifications. The following are three key performance indicators of guaranteed-performance instances:

  - **Max Connection**

    The maximum number of connections to an SLB instance. When the number of connections reaches the limit of the specification, new connection requests are dropped.

  - **Connection Per Second (CPS)**

    The number of new connections that are established per second. When the CPS reaches the limit of the specification, new connection requests are dropped.

  - **Queries Per Second (QPS)**

    The number of HTTP/HTTPS requests that can be processed per second. This metric is available only for layer-7 SLB listeners. When the QPS reaches the limit of the specification, new connection requests are dropped.

Currently, six specifications are available for guaranteed-performance instances.

Type	Specification	Max Connection	CPS	QPS	Purchase method
Specification 1	Small I (slb.s1.small)	5,000	3,000	1,000	Available for purchase from the official website of Alibaba Cloud.
Specification 2	Standard I (slb.s2.small)	50,000	5,000	5,000	Available for purchase from the official website of Alibaba Cloud.

Type	Specification	Max Connection	CPS	QPS	Purchase method
Specification 3	Standard II (slb.s2.medium)	100,000	10,000	10,000	Available for purchase from the official website of Alibaba Cloud.
Specification 4	Higher I (slb.s3.small)	200,000	20,000	20,000	Available for purchase from the official website of Alibaba Cloud.
Specification 5	Higher II (slb.s3.medium)	500,000	50,000	30,000	Available for purchase from the official website of Alibaba Cloud.
Specification 6	Super I (slb.s3.large)	1,000,000	100,000	50,000	Available for purchase from the official website of Alibaba Cloud.
Specification 7	Super II (slb.s3.xlarge)	2,000,000	200,000	100,000	Contact your account manager.

Type	Specification	Max Connection	CPS	QPS	Purchase method
Specification 8	Super III (slb.s3.xxlarge)	5,000,000	500,000	100,000	Contact your account manager.

11.4.2 Create an SLB instance

This topic describes how to create a Server Load Balancer (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

- **ECS instances are created and applications are deployed on them.**
- **You must make sure that the organization 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 SLB console*.
2. **On the Server Load Balancer page, click Create Instance.**
   - **Organization: Select an organization for the SLB instance to be created.**

     Note:

Make sure that the organization of the SLB instance is the same as that of backend ECS instances.

- **Resource Set: Select the resource set to which the SLB instance will belong from the drop-down list.**
- **Region: Enter the region of the SLB instance.**
- **Zone: Select a zone for the SLB instance from the drop-down list.**
- **Name: Enter the name of the SLB instance.**

  The name must be 2 to 128 characters in length and can contain letters, digits, hyphens (-), colons (:), commas (,), periods (.), and underscores (_). It must start with a letter and cannot start with `http://` or `https://`.
- **Network Access: Select the network access type, which can be Internal Network or Public Network.**
- **Network Type: Select the network type, which can be Classic Network or VPC.**
- **IP Version: Select the IP version.**
- **Service IP: Enter the service IP address. Ensure that the service IP address is valid. Otherwise, the SLB instance cannot be created. If this parameter is not set, the system will allocate an automatically generated IP address.**

3. **Click OK.**

What's next

Configure an SLB instance

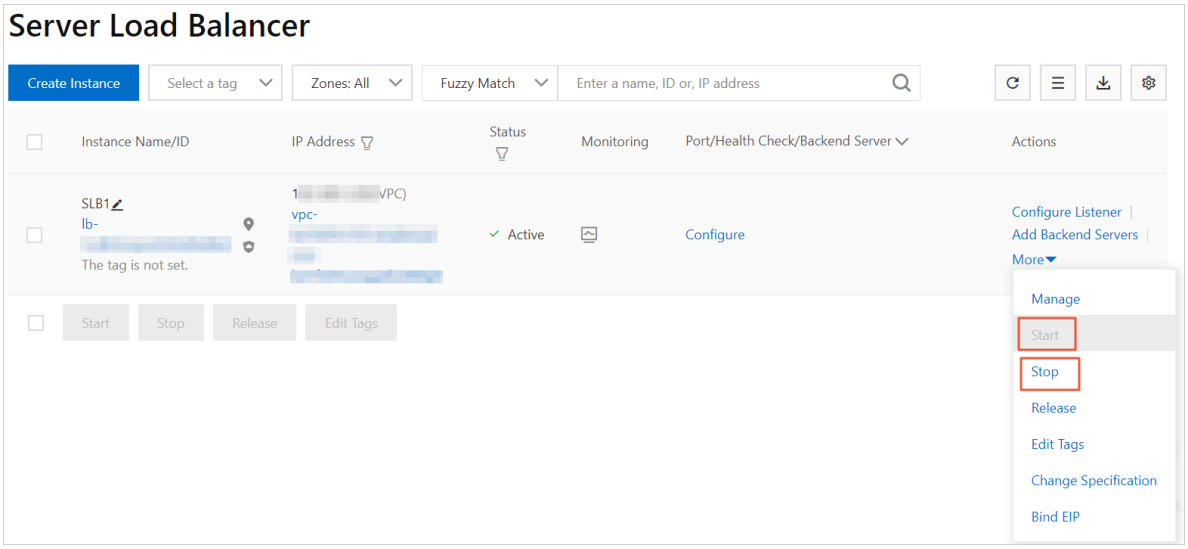
11.4.3 Start or stop an SLB instance

You can start or stop a Server Load Balancer (SLB) instance at any time. After being stopped, an SLB instance does not receive or forward requests any more.

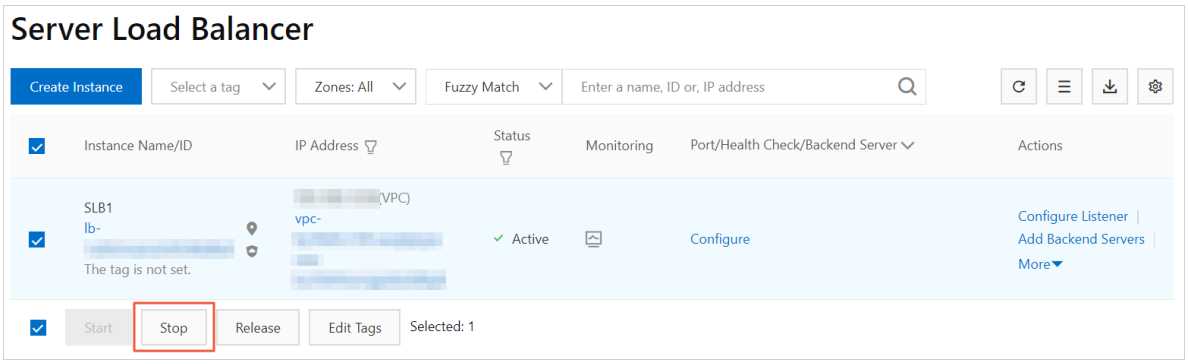
Procedure

1. *Log on to the SLB console*.
2. **In the left-side navigation pane, choose Instances > Server Load Balancer.**
3. **Select the region of the target SLB instance and find the target instance.**

4. In the Actions column, choose More > Start or More > Stop.



5. If you want to start or stop multiple instances at a time, select the target instances and click Start or Stop at the lower part of the page.



11.4.4 Tags

11.4.4.1 Overview

You can classify Server Load Balancer (SLB) instances by using tags.

Each tag consists of a key and a value. Before you use tags, note the following limits:

- **A tag cannot exist on its own and must be associated with an SLB instance.**
- **Up to 10 tags can be associated with an SLB instance.**
- **The key of each tag associated with an instance must be unique. Tags with the same key will be overwritten.**
- **Tags cannot be used across regions and are region-specific resources. For example, tags that belong to the China (Hangzhou) region are invisible to the China (Shanghai) region.**

11.4.4.2 Add a tag

This topic describes how to add a tag to a Server Load Balancer (SLB) instance.

Procedure

1. *Log on to the SLB console*
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. In the Actions column, choose More > Edit Tags.
4. Add a tag in the Edit Tags dialog box that appears.

   To add a tag, perform the following operations:

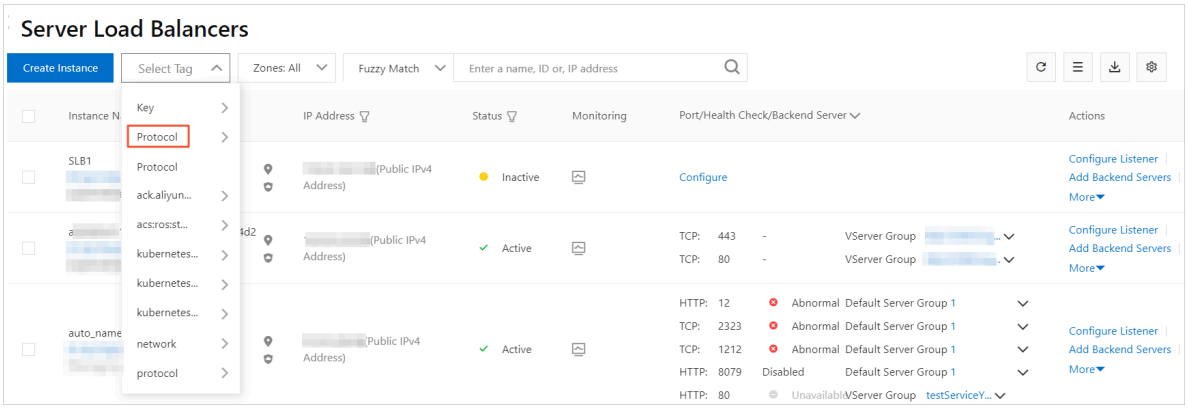
   - If existing tags are available, click Saved Tags and then select a tag to add.
   - To create a new tag, click New Tag in the Edit Tags dialog box, enter the key and value of the new tag, and then click OK next to the value.
5. Click OK.

11.4.4.3 Search for SLB instances by using a tag

This topic describes how to search for Server Load Balancer (SLB) instances by using a tag.

Procedure

1. *Log on to the SLB console*.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Select a region.
4. Click Select a tag and select the tag to be used as the search condition. The SLB instances associated with the selected tag are displayed.



5. To clear the search condition, rest the pointer over the selected tag and click the displayed delete icon.

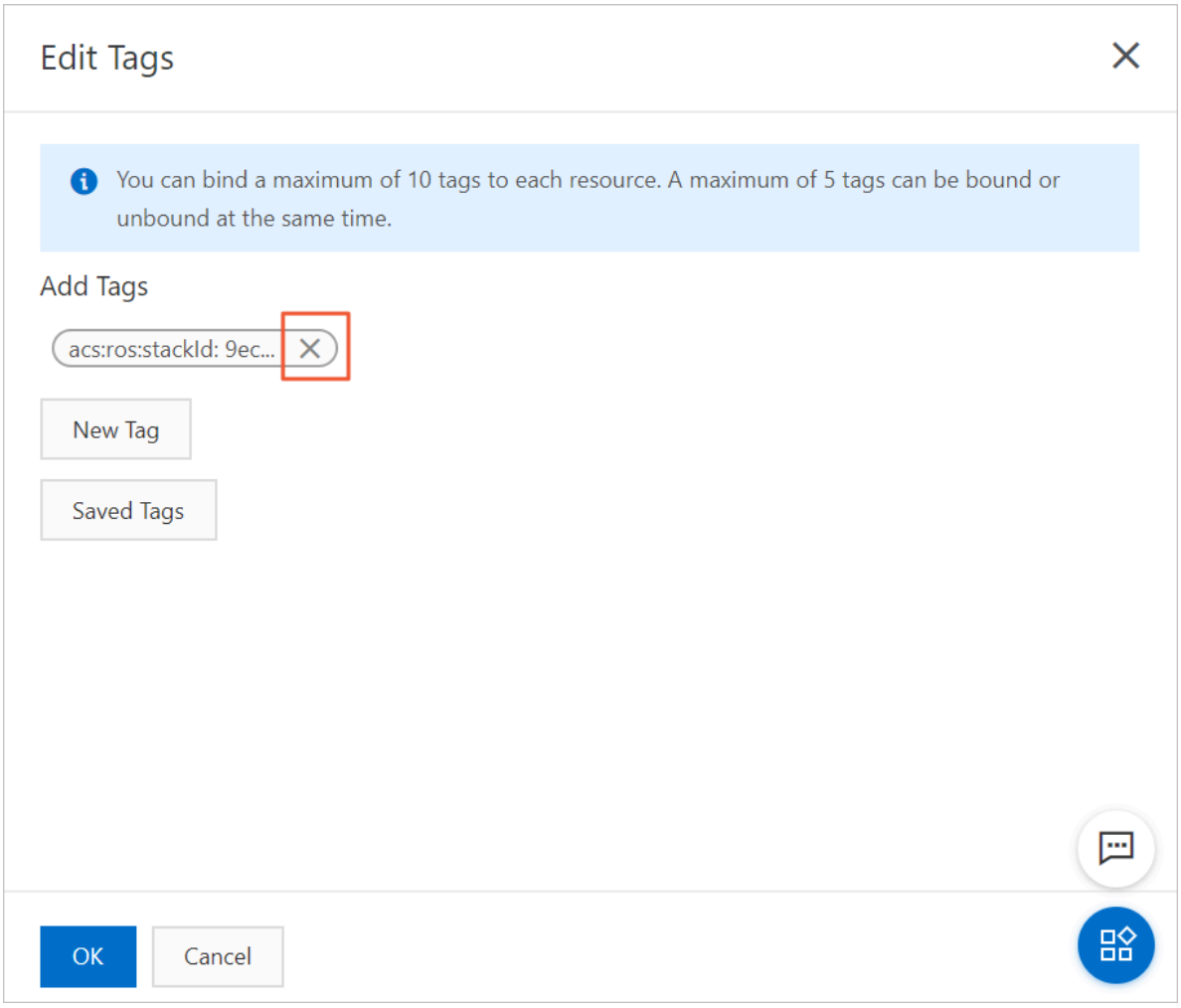
11.4.4.4 Delete a tag

Server Load Balancer (SLB) does not support deleting tags of multiple instances in batches. You can remove the tags of only one instance at a time.

Procedure

1. *Log on to the SLB console*.
2. **In the left-side navigation pane, choose Instances > Server Load Balancer.**
3. **Select the region of the target SLB instance and find the target SLB instance.**
4. **In the Actions column, choose More > Edit Tags.**
5. **On the Edit Tags page, click the delete icon next to the tag to be removed, and then click OK.**

> **Note:**
> **If a tag is removed from an SLB instance and is not associated with any other instances, the tag is deleted from the system.**



11.4.5 Release an SLB instance

This topic describes how to release a Server Load Balancer (SLB) instance. You can release an SLB instance immediately or at a specified time.

Procedure

1. *Log on to the SLB console*.
2. **Find the target instance and click More > Release.**

   You can select multiple SLB instances at a time and click Release at the bottom of the page to release SLB instances in batches.
3. **On the Release page, select Release Now or Release on Schedule.**

   Note:
   While the system executes the release operation every half hour or one hour cycle, the billing of the instance is stopped immediately at the release time you set.
4. **Click Next.**
5. **Confirm the displayed information and click OK to release the instance.**

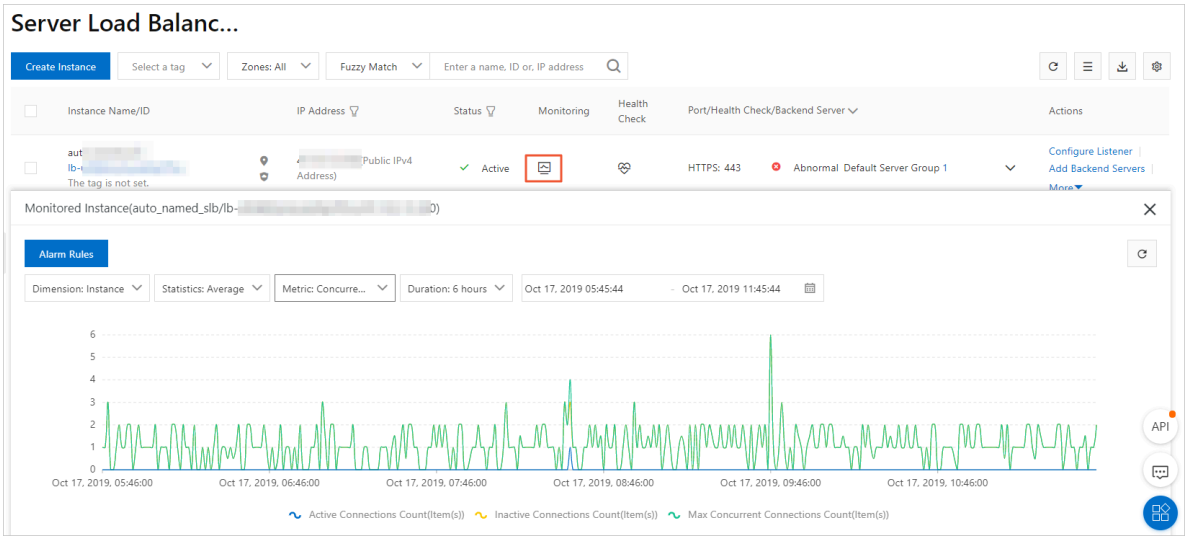
11.4.6 View monitoring data

This topic describes how to view the monitoring data of a Server Load Balancer (SLB) instance.

Procedure

1. *Log on to the SLB console*.
2. **Select the region of the target SLB instance.**
3. **Click the monitoring icon**  **next to the target SLB instance.**

4. Select the monitoring metrics that you want to view.



The following metrics are monitored for SLB instances.

Metric	Description
Traffic	· Inbound Traffic: the traffic sent from an external network to SLB · Outbound Traffic: the traffic sent from SLB
Packets	· RX Packets Count: the number of request packets received per second · TX Packets Count: the number of response packets sent per second
Concurrent Connections	· Active Connections Count: the number of established TCP connections. If persistent connections are used, a connection can transfer multiple file requests at one time. · Inactive Connections Count: the number of TCP connections not in the established state. You can use the `netstat -an` command to view the connections for both Windows and Linux servers. · Max Concurrent Connections Count: the total number of TCP connections.
Average Connection Requests Count	The average number of new TCP connections established between clients and SLB in a statistical period

Metric	Description
Dropped Traffic	· Dropped Inbound Traffic: the amount of inbound traffic dropped per second · Dropped Outbound Traffic: the amount of outbound traffic dropped per second
Dropped Packets	· Dropped RX Packets: the number of inbound packets dropped per second · Dropped TX Packets: the number of outbound packets dropped per second
Dropped Connections Count	The number of TCP connections dropped per second
The following metrics are specific to Layer-7 listeners.	
Layer-7 Protocol QPS	The number of HTTP/HTTPS requests that can be handled per second
Response Time (Listener)	The average response time of SLB
HTTP Status Code 2xx/3xx/4xx/5xx/Others (Listener)	The average number of HTTP response codes returned by listeners
Response Code 4xx/5xx (Server)	The average number of HTTP response codes returned by backend servers
Response Time (Server)	The average response time of backend servers

11.4.7 Configure alarm rules

After activating the CloudMonitor service, you can configure alarm rules for Server Load Balancer (SLB) instances in the CloudMonitor console.

Context

Note:
If a listener or an SLB instance is deleted, its alarm settings are deleted correspondingly.

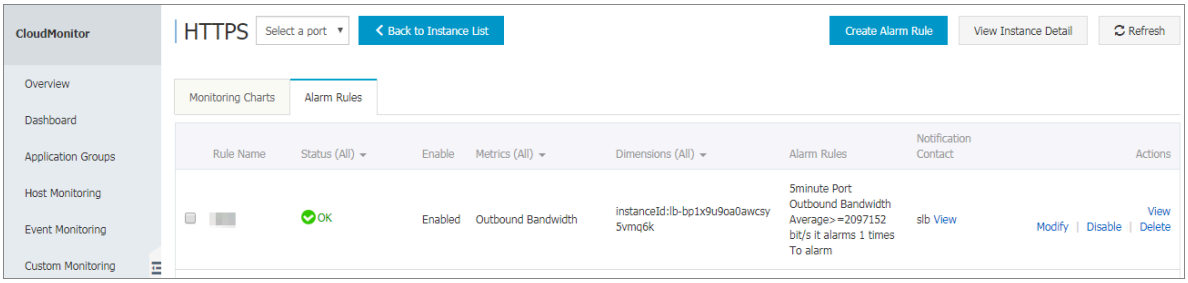
Procedure

1. *Log on to the SLB console*.
2. **Select the region to which the target SLB instance belongs.**

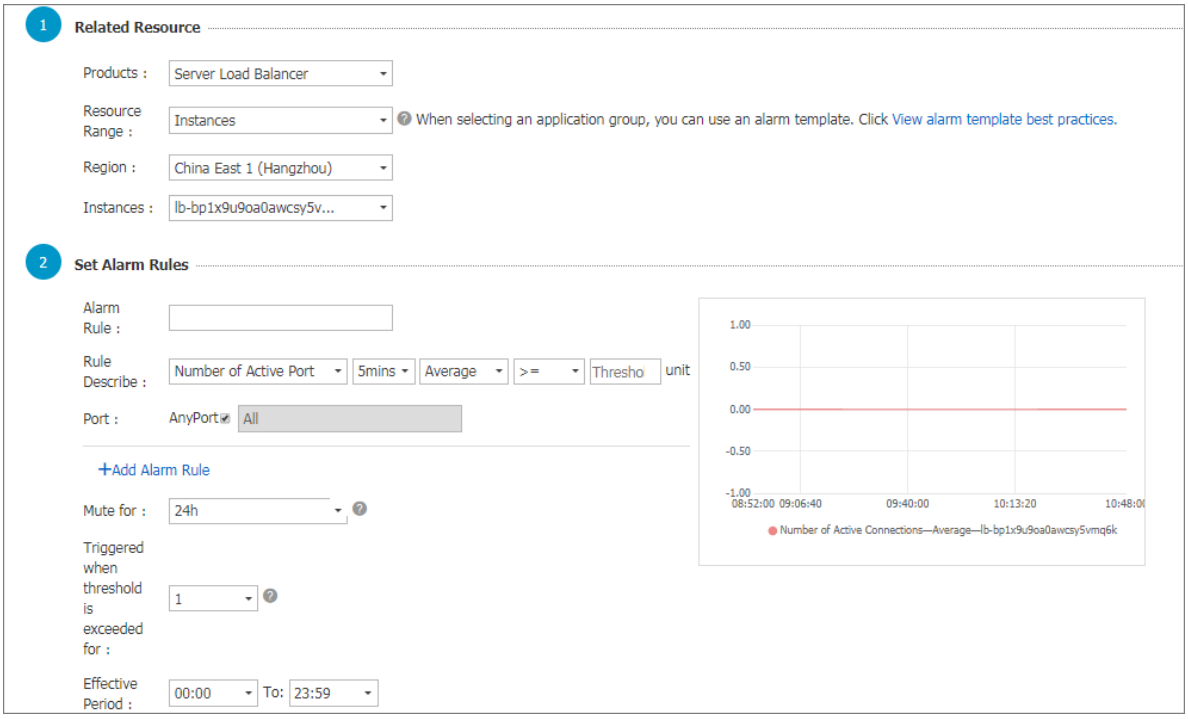
3. Find the target SLB instance and click  .

> **Notice:**
> **Make sure that you have configured a listener and enabled health checks for the SLB instance.**

4. Click Alarm Rules. You are then directed to the CloudMonitor console.



5. Click Create Alarm Rule.
6. Configure the alarm rule.



11.5 Listeners

11.5.1 Listener overview

After you create a Server Load Balancer (SLB) instance, you must configure a listener for it. The listener checks connection requests and then distributes the requests to backend servers according to configured forwarding rules.

Alibaba Cloud SLB provides Layer-4 (TCP and UDP protocols) and Layer-7 (HTTP and HTTPS protocols) listener services. Select the protocol based on your needs.

Protocol	Description	Scenario
TCP	· A connection-oriented protocol. A reliable connection must be established before data can be sent and received. · Source IP address-based session persistence. · Source IP addresses can be read at the network layer. · Fast data transmission.	· Applicable to scenarios where high transmission reliability and data accuracy are required, but some flexibility regarding network latency is permitted, such as file transmission, sending or receiving emails, and remote logons. · Web applications that have no special requirements. For more information, see *Add a TCP listener*.
UDP	· A non-connection-oriented protocol. UDP directly transmits data packets instead of making a three-way handshake with the other party before sending data. It does not provide error recovery and data re-transmission. · Fast data transmission, but the reliability is relatively low.	Applicable to scenarios with preference to real-time content over reliability, such as video chats and real-time financial quotations. For more information, see *Add a UDP listener*.

Protocol	Description	Scenario
HTTP	· An application layer protocol mainly used to package data. · Cookie-based session persistence. · Use X-Forward-For to obtain source IP addresses.	Applicable to applications that need to recognize data content, such as web applications and small-sized mobile games. For more information, see *Add an HTTP listener*.
HTTPS	· Encrypted data transmission that prevents unauthorized access. · Unified certificate management service. You can upload certificates to SLB and decryption operations are completed directly on SLB.	Applications that require encrypted transmission. For more information, see *Add an HTTPS listener*.

11.5.2 Add a TCP listener

This topic describes how to add a TCP listener to a Server Load Balancer (SLB) instance. The TCP protocol is applicable to scenarios with high requirements on reliability and data accuracy but with tolerance for low speed, such as file transmission, sending or receiving emails, and remote logons. You can add a TCP listener to forward requests from the TCP protocol.

Prerequisites

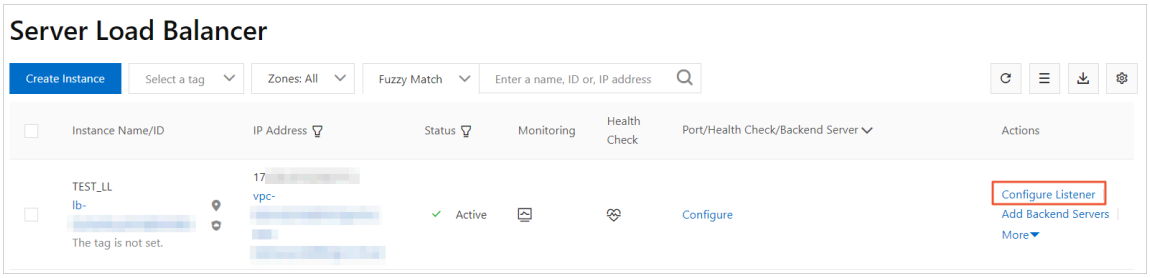
An SLB instance is created. For more information, see *Create an SLB instance*.

Step 1 Open the listener configuration wizard

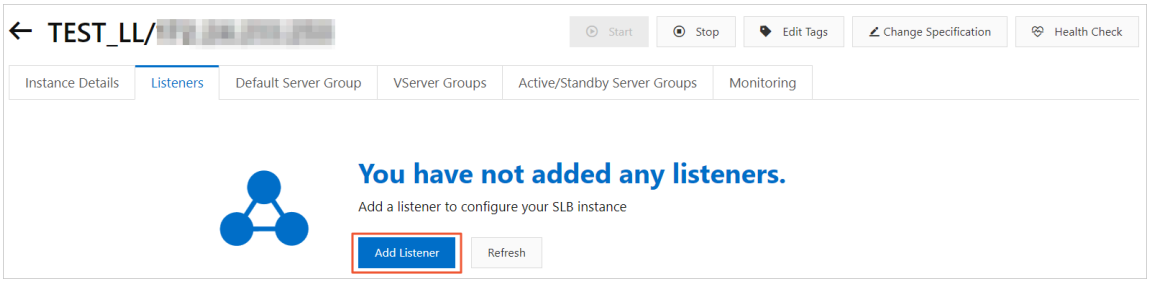
To open the listener configuration wizard, follow these steps:

1. *Log on to the SLB console*.
2. **In the left-side navigation pane, choose Instances > Server Load Balancer.**
3. **Select the region of the target SLB instance.**

4. Select one of the following methods to open the listener configuration wizard:
   - On the Server Load Balancer page, find the target SLB instance and then click Configure Listener in the Actions column.



   - On the Server Load Balancer page, click the ID of the target SLB instance. On the Listeners tab, click Add Listener.



Step 2 Configure the TCP listener

To configure the TCP listener, follow these steps:

1. Configure the TCP listener according to the following information:

Configuration	Description
Select Listener Protocol	Select the protocol type of the listener. In this topic, select TCP.

Configuration	Description
Listening Port	The listening port used to receive requests and forward the requests to backend servers. The port number is in the range of 1 to 65535. **Note:** In the same SLB instance, the UDP or TCP listener port numbers can be the same in the following regions. However, you must first apply for the privilege to use the beta function of configuring the same ports in TCP/UDP listeners on the *Quota Management* page of the SLB console. In other cases, the listener port numbers must be unique. · UAE (Dubai) · Australia (Sydney) · UAE (Dubai) · UK (London) · Germany (Frankfurt) · US (Silicon Valley) · US (Virginia) · Indonesia (Jakarta) · Japan (Tokyo) · India (Mumbai) · Singapore · Malaysia (Kuala Lumpur) · China (Hong Kong) · China (Shenzhen) · China (Hohhot) · China (Qingdao) · China (Chengdu) · China (Zhangjiakou) · China (Shanghai)
Advanced configurations	

Configuration	Description
Scheduling Algorithm	SLB supports four scheduling algorithms: round robin, weighted round robin (WRR), weighted least connections (WLC), and consistent hash. · Weighted Round-Robin (WRR): A backend server with a higher weight receives more requests. · Round-Robin (RR): Requests are evenly and sequentially distributed to backend servers. · Weighted Least Connections (WLC): A server with a higher weight receives more requests. When the weight values of two backend servers are the same, the backend server with a smaller number of connections is more likely to be polled. · Consistent Hash (CH): - Source IP: the consistent hash based on source IP addresses. Requests from the same source IP address are scheduled to the same backend server. - Tuple: the consistent hash based on four factors: source IP address + destination IP address + source port + destination port. The same streams are scheduled to the same backend server. Note: Currently, the Consistent Hash (CH) algorithm is only supported in the following regions: - Japan (Tokyo) - Australia (Sydney) - Malaysia (Kuala Lumpur) - Indonesia (Jakarta) - Germany (Frankfurt) - US (Silicon Valley) - US (Virginia) - UAE (Dubai) - China (Hohhot) - UK (London) - Zone B and Zone C of Singapore - China (Hong Kong) - China (Qingdao) - China (Zhangjiakou)

Configuration	Description
Enable Session Persistence	Select whether to enable session persistence. If you enable session persistence, all session requests from the same client are sent to the same backend server. For TCP listeners, session persistence is based on IP addresses. Requests from the same IP address are forwarded to the same backend server.
Enable Access Control	Select whether to enable the access control function.
Access Control Method	Select an access control method after you enable the access control function: · Whitelist: Only requests from IP addresses or CIDR blocks in the selected access control list are forwarded. It applies to scenarios where the application only allows access from some specific IP addresses. Enabling a whitelist poses some risks to your services. After a whitelist is configured, only the IP addresses in the list can access the SLB listener. If you enable the whitelist without adding any IP addresses in the corresponding access control list, all requests are forwarded. · Blacklist: Requests from IP addresses or CIDR blocks in the selected access control list are not forwarded. It applies to scenarios where the application only denies access from some specific IP addresses. If you enable a blacklist without adding any IP addresses in the corresponding access control list, all requests are forwarded.

Configuration	Description
Access Control List	Select an access control list as the whitelist or the blacklist. **Note:** An IPv6 instance can only be associated with IPv6 access control lists and an IPv4 instance can only be associated with IPv4 access control lists. For more information, see *Configure an access control list*.
Enable Peak Bandwidth Limit	Select whether to configure the listening bandwidth. If the SLB instance is charged based on bandwidth , you can set different peak bandwidth values for different listeners to limit the traffic passing through each listener. The sum of the peak bandwidth values of all listeners under an SLB instance cannot exceed the bandwidth value of that SLB instance. By default, all listeners share the bandwidth of the SLB instance. **Note:** SLB instances billed by traffic have no peak bandwidth limit by default.
Idle Timeout	Specify the idle connection timeout period. Value range: 10 to 900. Unit: seconds.
Listener Name	Enter a name for the TCP listener to be added.
Get Client Source IP Address	Backend servers of a Layer-4 listener can directly obtain the source IP addresses of clients.
Automatically Enable Listener after Creation	Choose whether to start the listener after the listener is configured. The listener is started by default.

2. Click Next.

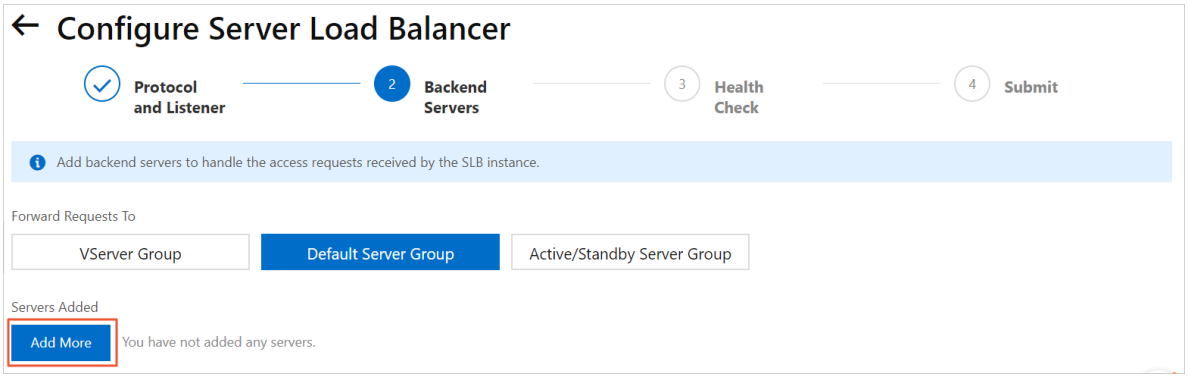
Step 3 Add backend servers

After configuring the listener, you need to add backend servers to process requests. You can use the default server group configured for the SLB instance, or configure

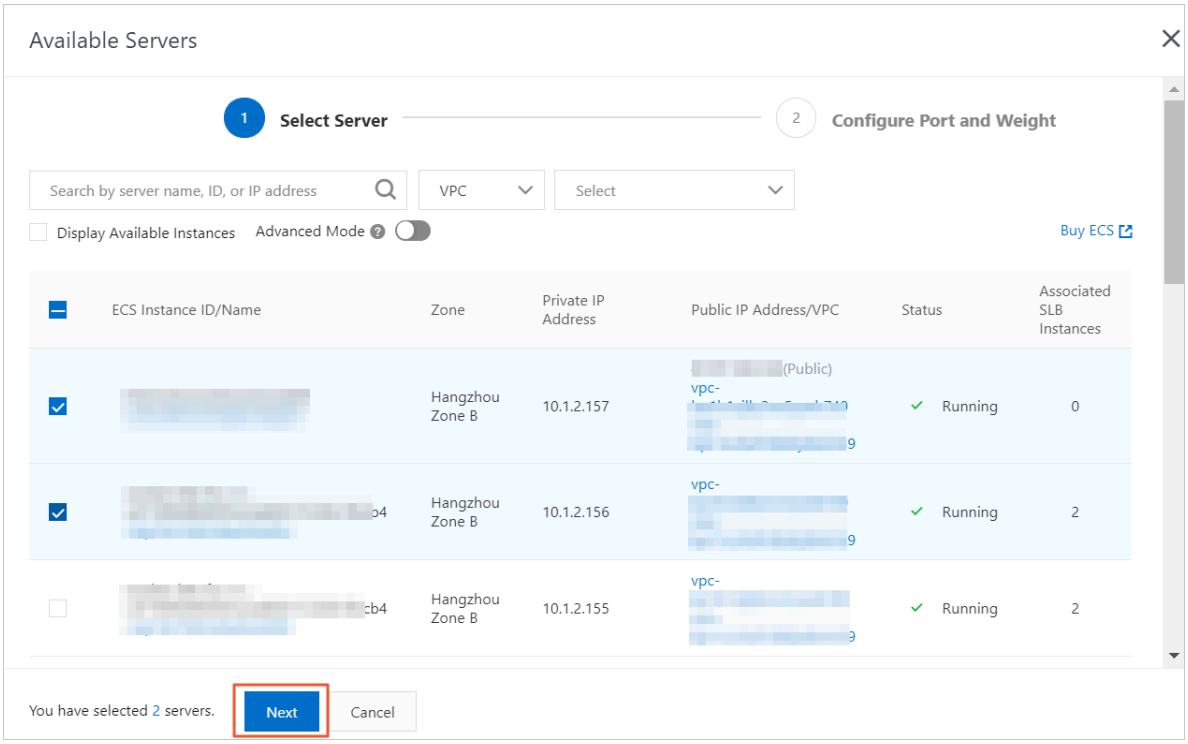
a VServer group or an active/standby server group for the listener. For more information, see *Backend server overview*.

In this example, use the default server group.

1. Select Default Server Group and then click Add More.



2. Select the ECS instances to add, and then click Next: Set Weight and Port.



3. Configure ports and weights for the added backend servers (ECS instances).

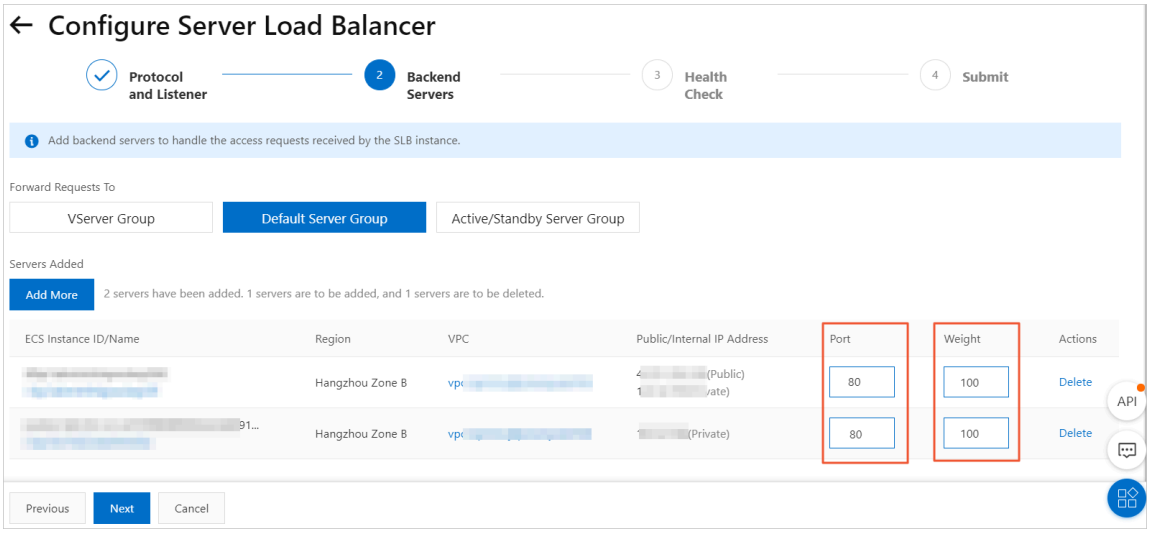
   - Port

     The port opened on the backend server to receive requests. The port number is in the range of 1 to 65535. Ports of backend servers can be the same in an SLB instance.

   - Weight

     The weight of the backend server. A backend server with a higher weight receives more requests.

> **Note:**
> If the weight is set to 0, no requests are sent to the backend server.



4. Click Next.

Step 4 Configure health checks

SLB checks the service availability of backend servers by performing health checks. The health check function improves the overall availability of your services and avoids the impact of backend server failures. Click Modify to change health check configurations. For more information, see *Health check overview*.

Step 5 Submit the configurations

To submit the listener configurations, follow these steps:

1. On the Submit page, check the listener configurations. You can click Modify to change the configurations.
2. Click Submit.

3. On the Submit page, click OK after the configurations are successful.

   After the configurations are successful, you can view the created listener on the Listeners page.

11.5.3 Add a UDP listener

This topic describes how to add a UDP listener to a Server Load Balancer (SLB) instance. You can add a UDP listener to forward UDP requests.

Context

Note the following before you add a UDP listener:

- Port 250, port 4789, and Port 4790 of the UDP listener are reserved by the system and are temporarily not open to the public.
- Currently, sharded packages are not supported.
- UDP listeners of an SLB instance of the classic network do not support viewing source IP addresses.
- The following operations require five minutes to take effect if they are performed in a UDP listener:
  - Removes backend servers.
  - Set the weight of a backend server to 0 after the backend server is declared as unhealthy.
- Because IPv6 has a longer IP header than IPv4, when you use a UDP listener on an IPv6 SLB instance, you must ensure that the MTU of the NIC on the backend server (most are ECS instances) communicating with the SLB instance is not greater than 1480 (some applications need to synchronize configuration files based on this MTU value). Otherwise, packets may be discarded because they are too large.

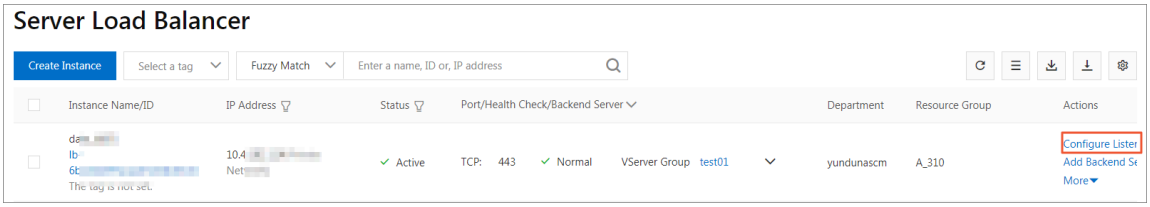
  If you use a TCP, HTTP, or HTTPS listener, no additional configurations are required because the TCP protocol supports MSS auto-negotiation.

Step 1 Open the listener configuration wizard

To open the listener configuration wizard, perform the following operations:

1. *Log on to the SLB console*.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.

3. Use one of the following methods to open the listener configuration wizard:
   - On the Server Load Balancer page, find the target instance and then click Configure Listener in the Actions column.



   - On the Server Load Balancer page, click the ID of the target SLB instance. On the Listeners tab, click Add Listener.

Step 2: Configure the UDP listener

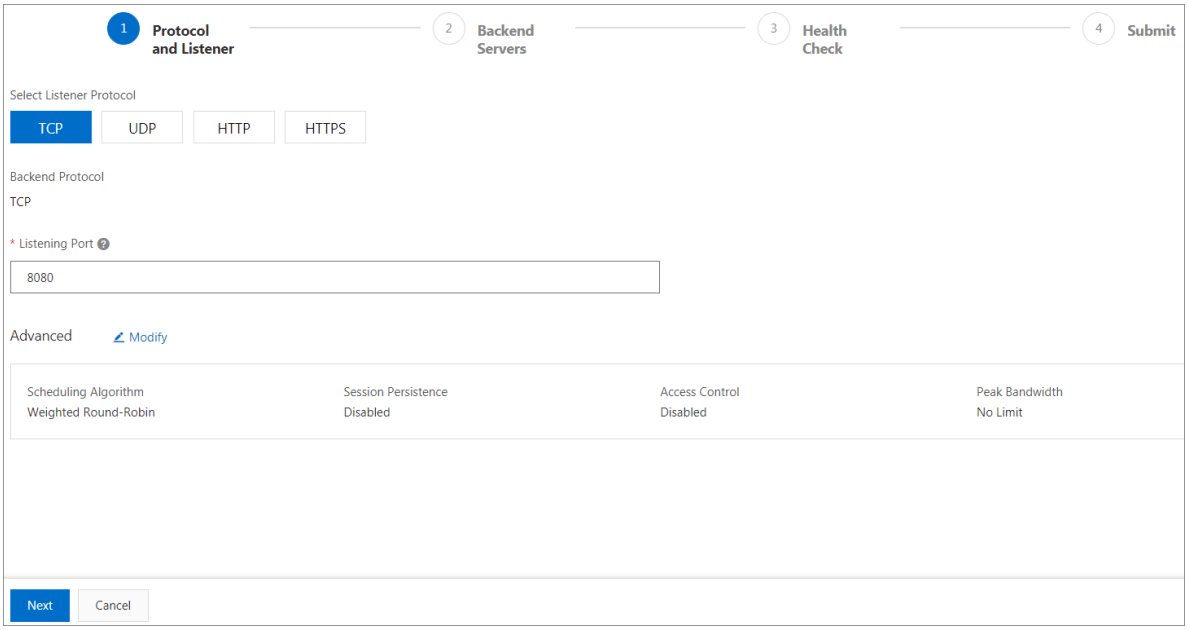
To configure the UDP listener, follow these steps:

1. On the Protocol and Listener page, configure the UDP listener according to the following information.

Configuration	Description
Select Listener Protocol	Select the protocol type of the listener. In this topic, select UDP.
Listening Port	The listening port used to receive requests and forward the requests to backend servers. The port number is in the range of 1 to 65535.
Advanced configurations	
Scheduling Algorithm	SLB supports four scheduling algorithms: round robin, weighted round robin (WRR), weighted least connections (WLC), and consistent hash. · Weighted Round-Robin (WRR): A backend server with a higher weight receives more requests. · Round-Robin (RR): Requests are evenly and sequentially distributed to backend servers. · Weighted Least Connections (WLC): A backend server with a higher weight receives more requests. If the weight values of two backend servers are the same, the backend server with a smaller number of connections is more likely to be polled.

Configuration	Description
Enable Peak Bandwidth Limit	Select whether to configure the listener bandwidth. If the SLB instance is charged based on bandwidth, you can set different peak bandwidth values for different listeners to limit the traffic passing through each listener. The sum of the peak bandwidth values of all listeners under an SLB instance cannot exceed the bandwidth value of the SLB instance. By default, this function is disabled and all listeners share the bandwidth of the SLB instance.
Get Client Source IP Address	Backend servers of a UDP listener can directly obtain source IP addresses of clients. **Note:** UDP listeners of an SLB instance of the classic network do not support viewing source IP addresses.
Automatically Enable Listener After Creation	Select whether to start the listener after the listener is configured. This function is enabled by default.

2. Click Next.



Step 3 Add backend servers

After configuring the listener, you must add backend servers to process frontend requests. You can use the default server group configured for the SLB instance, or configure a VServer group or an active/standby server group. For more information, see *Backend server overview*.

This example uses the default server group.

1. **Select Default Server Group, and click Add More.**
2. **Select the ECS instances to be added, and then click Next: Set Weight and Port.**
3. **Configure ports and weights for the added backend servers (ECS instances).**
   - **Port**

     The port on a backend server to receive requests. The port number is in the range of 1 to 65535. Port numbers of backend servers can be the same in an SLB instance.
   - **Weight**

     The weight of a backend server. A backend server with a higher weight receives more requests.

     Note:
     If the weight of a backend server is set to 0, the backend server will not receive new requests.
4. **Click Next.**

Step 4 Configure health checks

SLB checks the service availability of backend servers by performing health checks. The health check feature improves the overall availability of frontend business and avoids the impact of exceptions on backend servers. Click Modify to modify health check configurations. For more information, see *Health check overview*.

Step 5 Submit the configurations

Submit the configurations by performing the following steps:

1. **In the Submit step, check the configurations. You can click Modify to modify the configurations.**
2. **Click Submit.**

3. In the Configure Successful dialog box, click OK.

   After successful configuration, you can view the created listener on the Listeners tab.

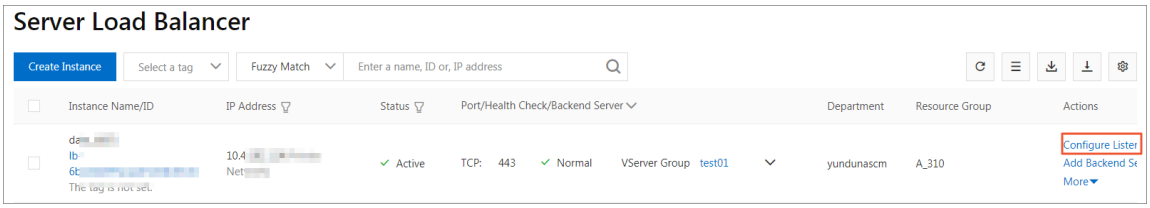
11.5.4 Add an HTTP listener

This topic describes how to add an HTTP listener to a Server Load Balancer (SLB) instance. The HTTP protocol is applicable for scenarios where data needs to be read, such as web applications and small mobile games. You can add an HTTP listener to forward HTTP requests.

Step 1 Open the listener configuration wizard

To open the listener configuration wizard, perform the following operations:

1. *Log on to the SLB console*.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Use one of the following methods to open the listener configuration wizard:
   - On the Server Load Balancer page, find the target instance and then click Configure Listener in the Actions column.



   - On the Server Load Balancer page, click the ID of the target SLB instance. On the Listeners tab, click Add Listener.

Step 2: Configure the HTTP listener

To configure the HTTP listener, follow these steps:

1. On the Protocol and Listener page, configure the HTTP listener according to the following information:

Configuration	Description
Select Listener Protocol	Select the protocol type of the listener. In this topic, select HTTP.

Configuration	Description
Listening Port	The listening port used to receive requests and forward the requests to backend servers. Value range: 1 to 65535. **Note:** The listening port must be unique in an SLB instance.
Advanced configurations	
Scheduling Algorithm	SLB supports three scheduling algorithms: round robin, weighted round robin (WRR), and weighted least connections (WLC). · Weighted Round-Robin (WRR): A backend server with a higher weight receives more requests. · Round-Robin (RR): Requests are evenly and sequentially distributed to backend servers. · Weighted Least Connections (WLC): A server with a higher weight receives more requests. When the weight values of two backend servers are the same, the backend server with a smaller number of connections is more likely to be polled.

Configuration	Description
Enable Session Persistence	Select whether to enable session persistence. After you enable session persistence, all session requests from the same client are sent to the same backend server. HTTP session persistence is based on cookies. The following two methods are supported: · Insert cookie: You only need to specify the cookie timeout period. SLB adds a cookie to the first response from the backend server (inserts SERVERID in the HTTP and HTTPS response packet). The next request will contain the cookie and the listener will distribute the request to the same backend server. · Rewrite cookie: You can set the cookie inserted to the HTTP or HTTPS response according to your needs. You must maintain the timeout period and lifespan of the cookie on the backend server. SLB will overwrite the original cookie when it discovers that a new cookie is set. The next time the client carries the new cookie to access SLB , the listener will distribute the request to the recorded backend server.

Configuration	Description
Enable Peak Bandwidth Limit	Select whether to configure the listener bandwidth. If the SLB instance incurs charges based on bandwidth, you can set different peak bandwidth values for different listeners to limit the traffic passing through each listener. The sum of the peak bandwidth values of all listeners under an SLB instance cannot exceed the bandwidth value of that SLB instance. This parameter is disabled by default, and all listeners share the total bandwidth of the SLB instance.
Enable Gzip Compression	Choose whether to enable Gzip compression to compress files of specific formats. Gzip supports the following file types: text/xml, text/plain, text/css, application/javascript, application/x-javascript, application/rss+xml, application/atom+xml, application/xml.
Add HTTP Header Fields	Select the custom HTTP headers that you want to add: · Use the `X-Forwarded-For` field to retrieve the source IP address of the client. · Use the `X-Forwarded-Proto` field to retrieve the listener protocol used by the SLB instance. · Use the `SLB-IP` field to retrieve the public IP address of the SLB instance. · Use the `SLB-ID` field to retrieve the ID of the SLB instance.
Get Client Source IP Address	HTTP listeners use X-Forwarded-For to obtain real IP addresses of clients.
Automatically Enable Listener After Creation	Choose whether to start the listener after the listener is configured. The listener is started by default.

2. Click Next.

Step 3 Add backend servers

After configuring the listener, you must add backend servers to process frontend requests. You can use the default server group configured for the SLB instance, or configure a VServer group or an active/standby server group. For more information, see *Backend server overview*.

This example uses the default server group.

1. **Select Default Server Group, and click Add More.**
2. **Select the ECS instances to be added, and then click Next: Set Weight and Port.**
3. **Configure ports and weights for the added backend servers (ECS instances).**
   - **Port**

     The port on a backend server to receive requests. The port number is in the range of 1 to 65535. Port numbers of backend servers can be the same in an SLB instance.
   - **Weight**

     The weight of a backend server. A backend server with a higher weight receives more requests.

     Note:
     If the weight of a backend server is set to 0, the backend server will not receive new requests.
4. **Click Next.**

Step 4 Configure health checks

SLB checks the service availability of backend servers by performing health checks. The health check feature improves the overall availability of frontend business and avoids the impact of exceptions on backend servers. Click Modify to modify health check configurations. For more information, see *Health check overview*.

Step 5 Submit the configurations

Submit the configurations by performing the following steps:

1. **In the Submit step, check the configurations. You can click Modify to modify the configurations.**
2. **Click Submit.**

3. In the Configure Successful dialog box, click OK.

   After successful configuration, you can view the created listener on the Listeners tab.

11.5.5 Add an HTTPS listener

This topic describes how to add an HTTPS listener to a Server Load Balancer (SLB) instance. You can add an HTTPS listener to forward requests from the HTTPS protocol.

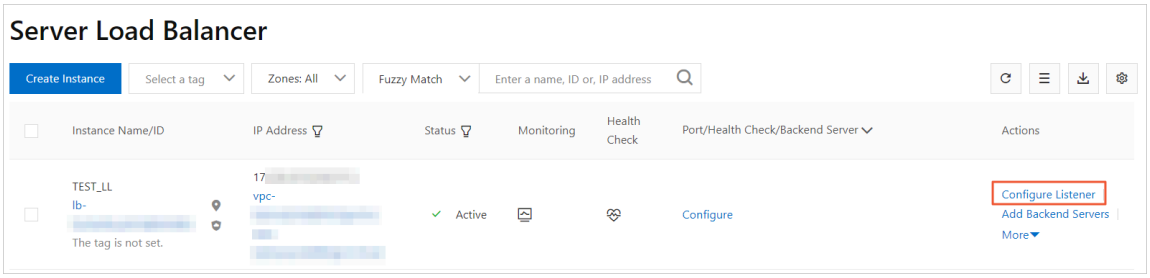
Prerequisites

An SLB instance is created. For more information, see *Create an SLB instance*.

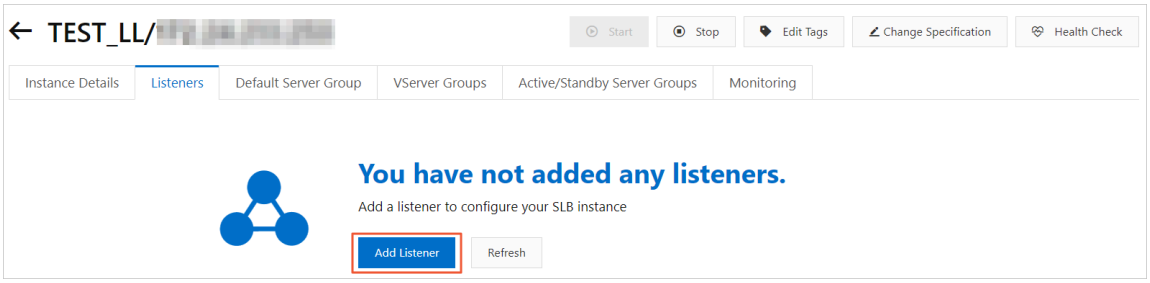
Step 1 Open the listener configuration wizard

To open the listener configuration wizard, follow these steps:

1. *Log on to the SLB console*.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Select the region of the target SLB instance.
4. Select one of the following methods to open the listener configuration wizard:
   - On the Server Load Balancer page, find the target SLB instance and then click Configure Listener in the Actions column.



   - On the Server Load Balancer page, click the ID of the target SLB instance. On the Listeners tab, click Add Listener.



Step 2 Configure the HTTPS listener

To configure the HTTPS listener, follow these steps:

1. **On the Protocol and Listener page, configure the HTTPS listener according to the following information:**

Configuration	Description
Select Listener Protocol	**Select the protocol type of the listener.** **In this topic, select HTTPS.**
Listening Port	**The listening port used to receive requests and forward the requests to backend servers.** **Value range: 1 to 65535** **Note:** **The listening port must be unique in an SLB instance.**
Advanced configurations	
Scheduling Algorithm	**SLB supports three scheduling algorithms: round robin, weighted round robin (WRR), and weighted least connections (WLC).** · **Weighted Round-Robin (WRR): A backend server with a higher weight receives more requests.** · **Round-Robin (RR): Requests are evenly and sequentially distributed to backend servers.** · **Weighted Least Connections (WLC): A server with a higher weight receives more requests. When the weight values of two backend servers are the same, the backend server with a smaller number of connections is more likely to be polled.**

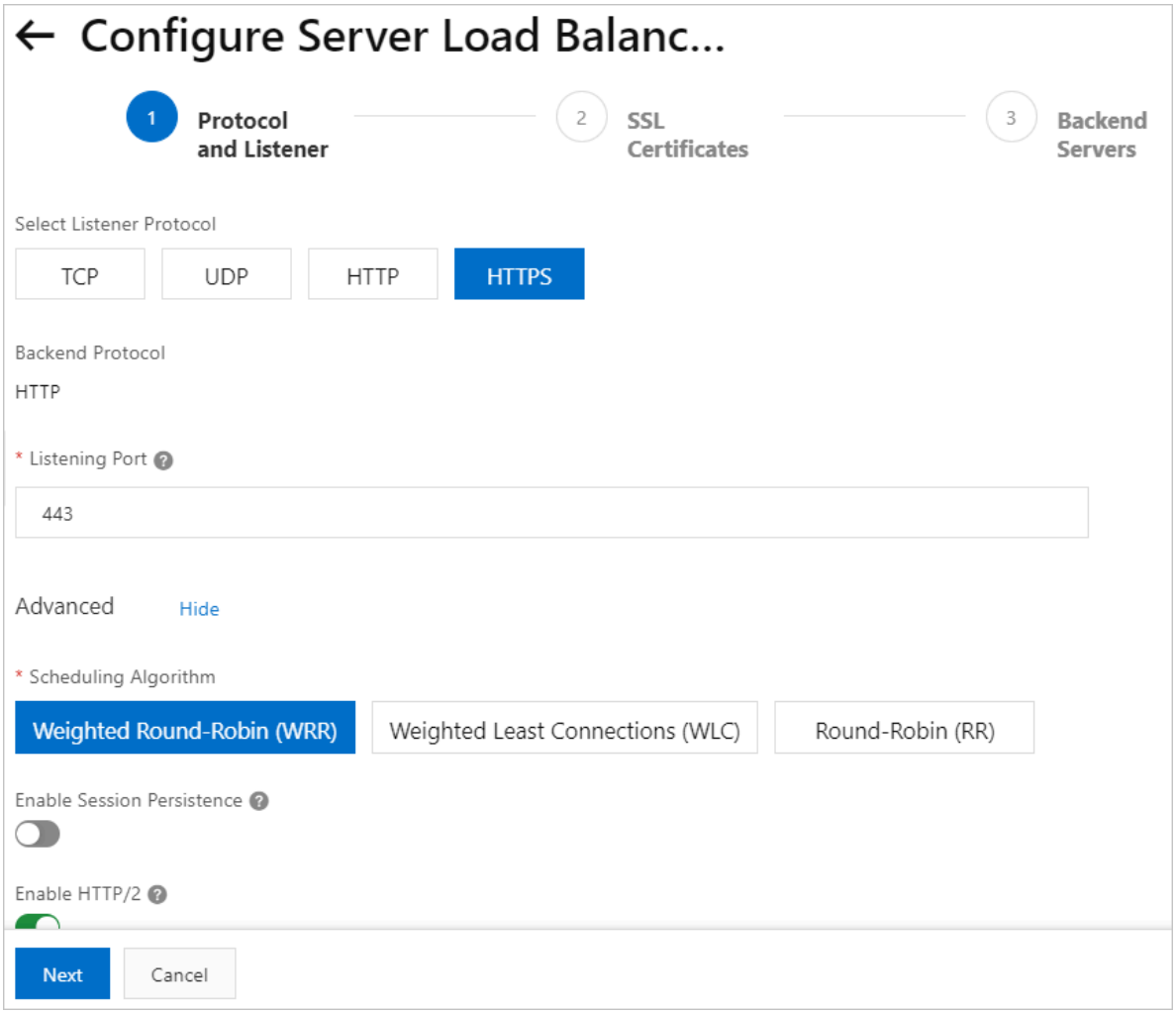
Configuration	Description
Enable Session Persistence	Select whether to enable session persistence. After you enable session persistence, all session requests from the same client are sent to the same backend server. HTTP session persistence is based on cookies. The following two methods are supported: · Insert cookie: You only need to specify the cookie timeout period. SLB adds a cookie to the first response from the backend server (inserts SERVERID in the HTTP and HTTPS response packet). The next request will contain the cookie and the listener will distribute the request to the same backend server. · Rewrite cookie: You can set the cookie to be inserted to the HTTP or HTTPS response according to your needs. You must maintain the timeout period and lifecycle of the cookie on the backend server. SLB will overwrite the original cookie when it discovers that a new cookie is set. The next time the client carries the new cookie to access SLB, the listener will distribute the request to the recorded backend server. For more information, see *Configure session persistence*.
Enable HTTP/2	Select whether to enable HTTP 2.0.
Enable Access Control	Select whether to enable the access control function.

Configuration	Description
Access Control Method	Select an access control method after you enable the access control function: · Whitelist: Only requests from IP addresses or CIDR blocks in the selected access control list are forwarded. It applies to scenarios where the application only allows access from some specific IP addresses. Enabling a whitelist poses some risks to your services. After a whitelist is configured, only the IP addresses in the list can access the listener. If you enable the whitelist without adding any IP addresses in the corresponding access control list, all requests are forwarded. · Blacklist: Requests from IP addresses or CIDR blocks in the selected access control list are not forwarded. It applies to scenarios where the application only denies access from some specific IP addresses. If you enable a blacklist without adding any IP addresses in the corresponding access control list, all requests are forwarded.
Access Control List	Select an access control list as the whitelist or the blacklist. Note: An IPv6 instance can only be associated with IPv6 access control lists and an IPv4 instance can only be associated with IPv4 access control lists. For more information, see *Configure an access control list*.

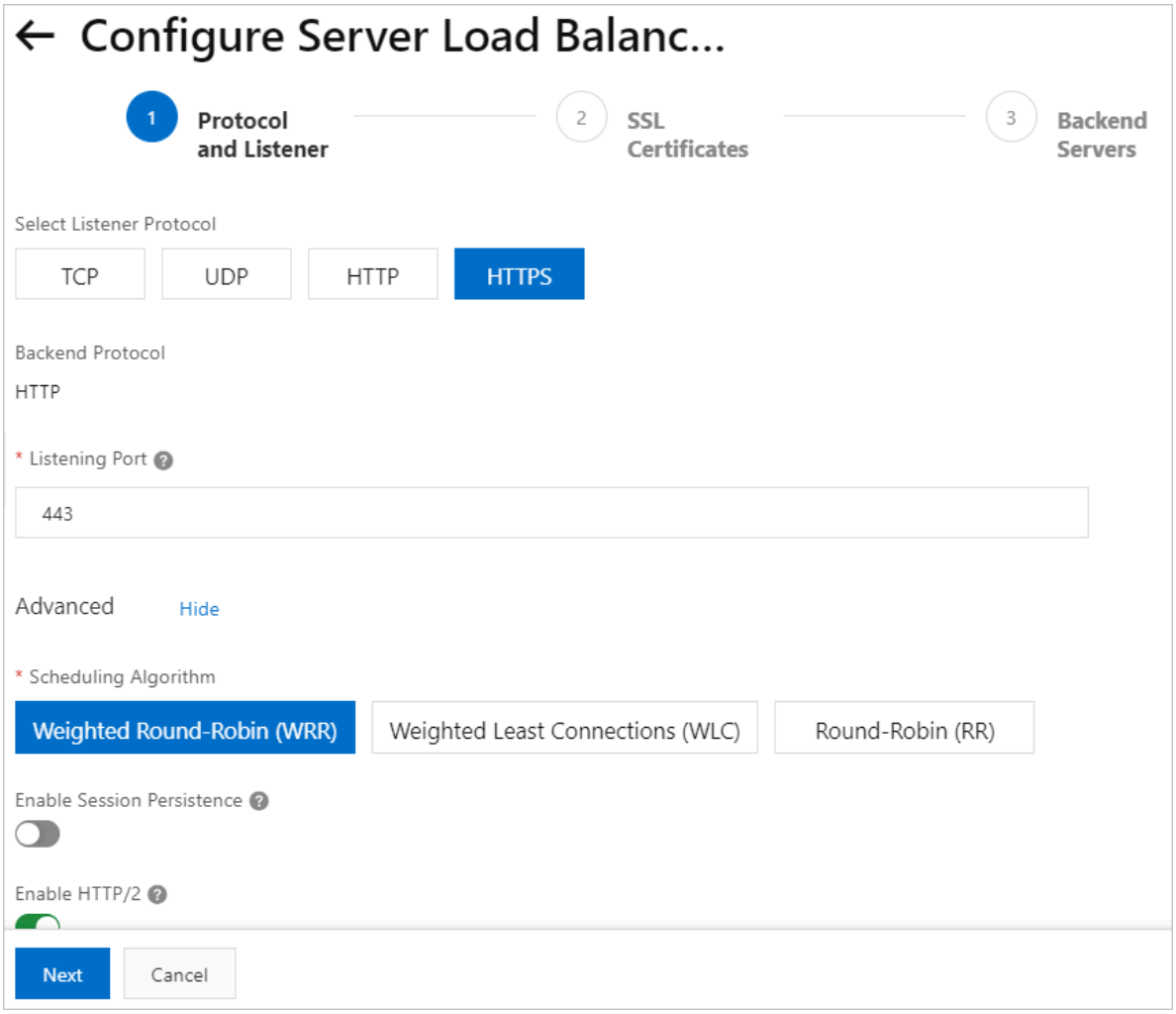
Configuration	Description
Enable Peak Bandwidth Limit	Select whether to configure the listening bandwidth. If the SLB instance is charged based on bandwidth, you can set different peak bandwidth values for different listeners to limit the traffic passing through each listener. The sum of the peak bandwidth values of all listeners under an SLB instance cannot exceed the bandwidth value of that SLB instance. By default, all listeners share the bandwidth of the SLB instance. Note: SLB instances billed by traffic have no peak bandwidth limit by default.
Idle Timeout	Specify the idle connection timeout period. Value range: 1 to 60. Unit: seconds. If no request is received during the specified timeout period, SLB temporarily terminates the connection and restarts the connection when the next request is received. This function is available in all regions.
Request Timeout	Specify the request timeout period. Value range: 1 to 180. Unit: seconds. If no response is received from the backend server during the specified timeout period, SLB stops waiting and sends an HTTP 504 error code to the client. Currently, this function is available in all regions.
TLS Security Policy	Only guaranteed-performance instances support selecting the TLS security policy.

Configuration	Description
Enable Gzip Compression	Choose whether to enable Gzip compression to compress files of specific formats. Now Gzip supports the following file types: text /xml, text/plain, text/css, application/javascript , application/x-javascript, application/rss+xml, application/atom+xml, and application/xml.
Add HTTP Header Fields	Select the custom HTTP headers that you want to add: · Use the `X-Forwarded-For` field to retrieve client source IP addresses. · Use the `X-Forwarded-Proto` field to retrieve the listener protocol used by the SLB instance. · Use the `SLB-IP` field to retrieve the public IP address of the SLB instance. · Use the `SLB-ID` field to retrieve the ID of the SLB instance.
Get Client Source IP Address	HTTP listeners use X-Forwarded-For to obtain real IP addresses of clients.

Configuration	Description
Automatically Enable Listener After Creation	Choose whether to start the listener after the listener is configured. The listener is started by default.



2. Click Next.



Step 3 Configure the SSL certificate

To add an HTTPS listener, you must upload a server certificate or CA certificate, as shown in the following table.

Certificate	Description	Required for one-way authentication?	Required for mutual authentication?
Server certificate	Used to identify a server. The client uses it to check whether the certificate sent by the server is issued by a trusted center.	Yes. You need to upload the server certificate to the certificate management system of SLB.	Yes. You need to upload the server certificate to the certificate management system of SLB.

Certificate	Description	Required for one-way authentication?	Required for mutual authentication?
Client certificate	Used to identify a client. The client user can prove its true identity when communicating with the server. You can sign a client certificate with a self-signed CA certificate.	No.	Yes. You need to install the client certificate on the client.
CA certificate	The server uses the CA certificate to authenticate the signature on the client certificate, as part of the authentication before launching a secure connection. If the authentication fails, the connection is rejected.	No.	Yes. You need to upload the CA certificate to the certificate management system of SLB.

Note the following before you upload a certificate:

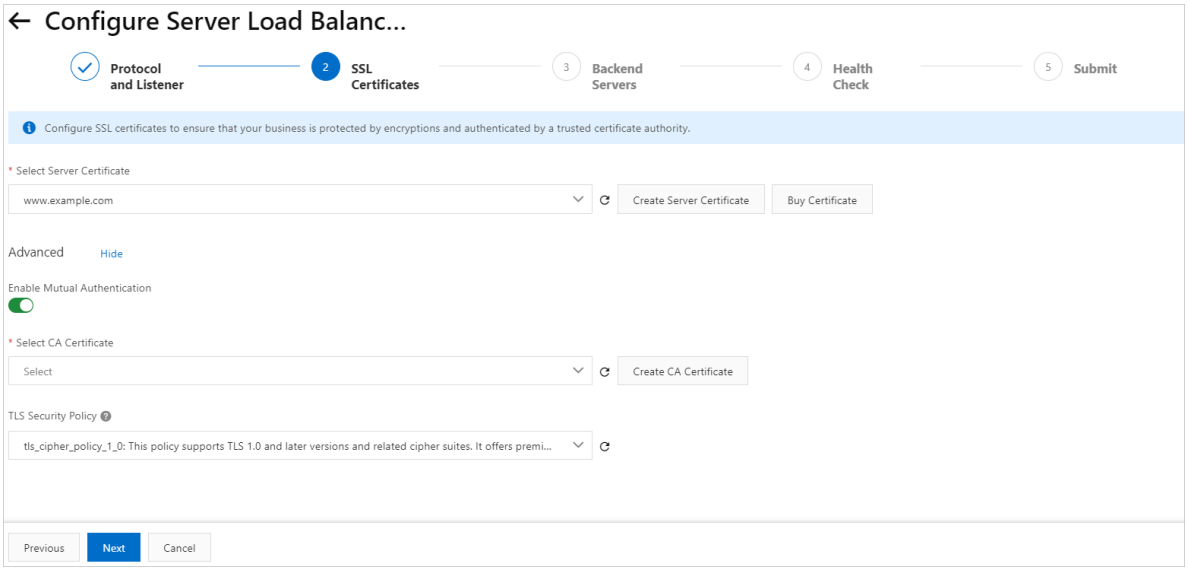
- The uploaded certificate must be in the PEM format. For more information, see *Certificate requirements*.
- After the certificate is uploaded to SLB, SLB can manage the certificate and you do not need to associate the certificate with backend ECS instances.
- It usually takes one to three minutes to activate the HTTPS listener because the uploading, loading, and validation of certificates take some time. Normally it takes effect in one minute and it will definitely take effect in three minutes.
- The ECDHE algorithm cluster used by HTTPS listeners supports forward secrecy, but does not support uploading security enhancement parameter files required by the DHE algorithm cluster, such as strings containing the `BEGIN DH PARAMETERS` field in the PEM certificate file. For more information, see *Certificate requirements*.
- Currently, SLB HTTPS listeners do not support SNI (Server Name Indication). You can use TCP listeners instead, and then configure SNI on backend ECS instances.

- The session ticket timeout period of HTTPS listeners is 300 seconds.
- The actual amount of traffic is larger than the billed traffic amount because some traffic is used for protocol handshaking.
- In the case of a large number of new connections, HTTPS listeners consume more traffic.

1. Select the server certificate that has been uploaded, or click Create Server Certificate to upload a server certificate.

   For more information, see *Overview*.

2. If you want to enable HTTPS mutual authentication or set a TLS security policy, click Modify.



3. Select an uploaded CA certificate, or click Create CA Certificate to upload a CA certificate.

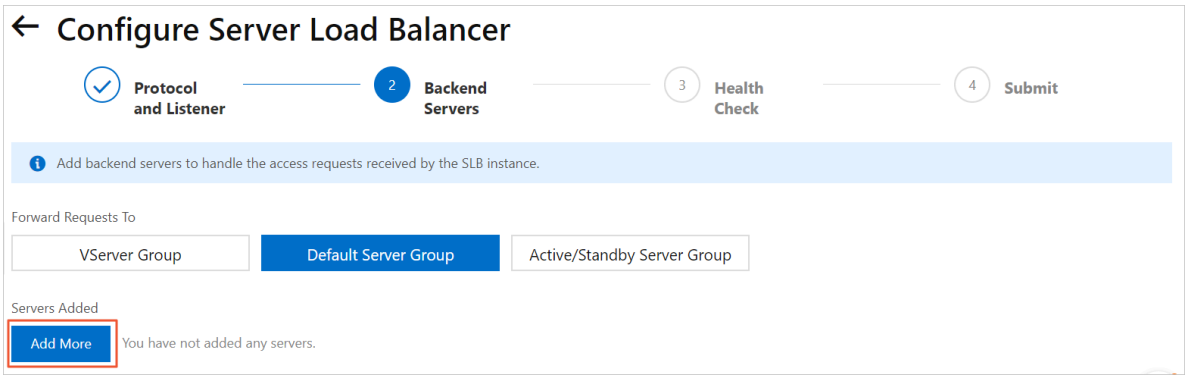
   You can use a self-signed CA certificate. For more information, see *Overview*.

Step 4 Add backend servers

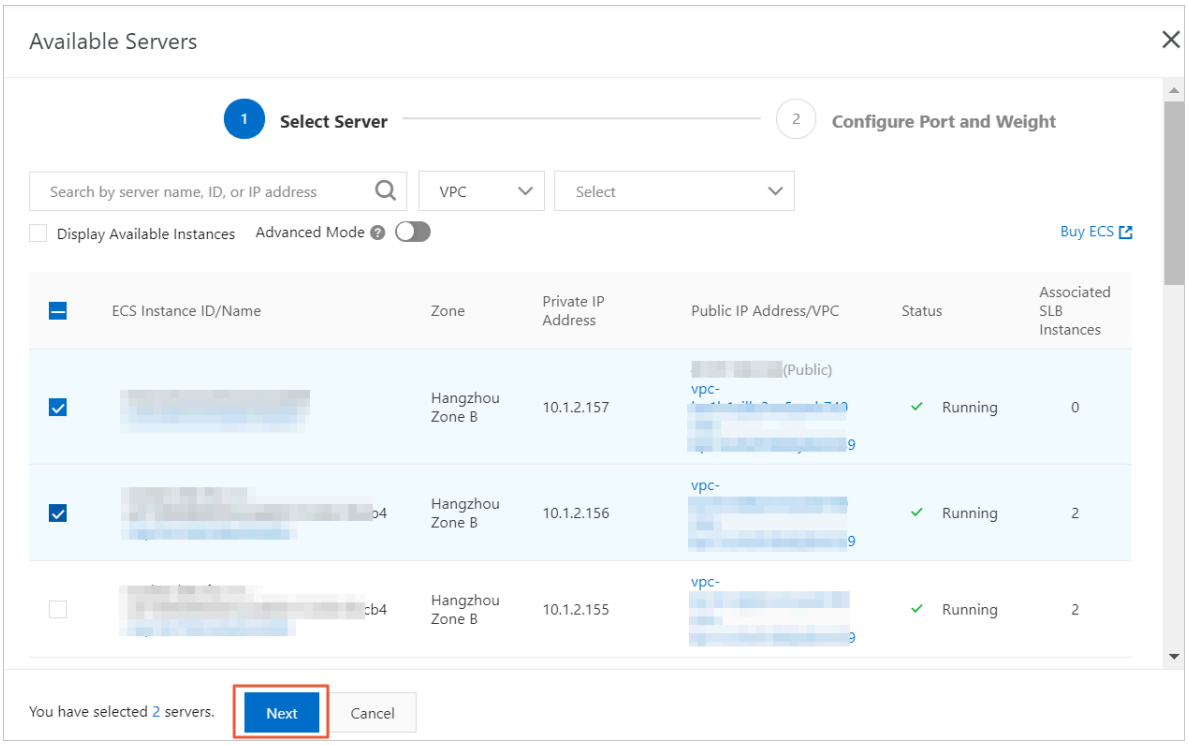
You need to add backend servers to process requests. You can use the default server group configured for the SLB instance, or configure a VServer group or an active/standby server group for the listener. For more information, see *Backend server overview*.

In this topic, use the default server group.

1. Select Default Server Group and then click Add More.



2. Select the ECS instances to add, and then click Next: Set Weight and Port.



3. Configure ports and weights for the added backend servers (ECS instances).

   - **Port**

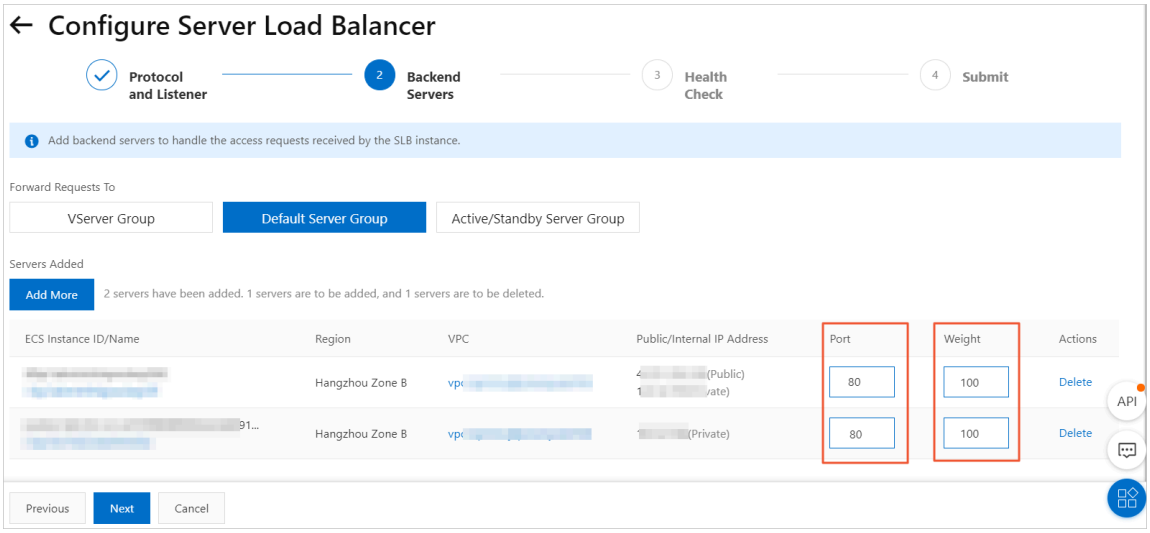
     The port opened on the backend server to receive requests. The port number is in the range of 1 to 65535. Ports of backend servers can be the same in an SLB instance.

   - **Weight**

     The weight of the backend server. A backend server with a higher weight receives more requests.

     Note:

If the weight is set to 0, no requests are sent to the backend server.



4. **Click Next.**

Step 5 Configure health checks

SLB checks the service availability of backend servers by performing health checks. The health check function improves the overall availability of your services and avoids the impact of backend server failures. Click Modify to change health check configurations. For more information, see *Health check overview*.

Step 6 Submit the configurations

To submit the listener configurations, follow these steps:

1. **On the Submit page, check the listener configurations. You can click Modify to change the configurations.**
2. **Click Submit.**
3. **On the Submit page, click OK after the configurations are successful.**

   After the configurations are successful, you can view the created listener on the Listeners page.

11.5.7 Enable access control

Server Load Balancer (SLB) provides an access control function for listeners. You can configure different whitelists or blacklists for different listeners.

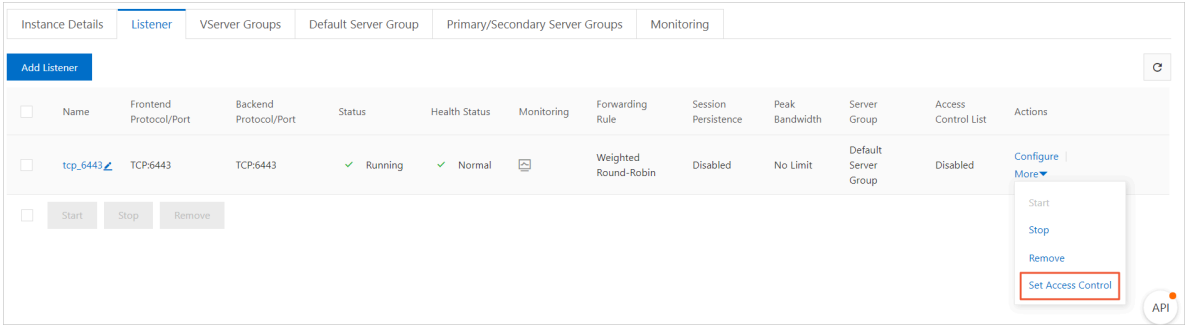
Prerequisites

Before you enable access control, make sure:

- An access control list is created. For more information.
- A listener is created.

Procedure

1. *Log on to the SLB console*.
2. Select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. On the Instance Details page, click the Listeners tab.
5. Find the target listener, choose More > Set Access Control.



6. On the Access Control Settings page, enable access control, select an access control method and an access control list, and click OK.
   - Whitelist: Only requests from IP addresses or CIDR blocks in the selected access control list are forwarded. It applies to scenarios where the application only allows access from some specific IP addresses.

     Enabling a whitelist poses some risks to your services. After a whitelist is configured, only the IP addresses in the list can access the listener. If you enable the whitelist without adding any IP addresses in the corresponding access control list, all requests are forwarded.
   - Blacklist: Requests from IP addresses or CIDR blocks in the selected access control list are not forwarded. It applies to scenarios where the application only denies access from some specific IP addresses.

     If you enable a blacklist without adding any IP addresses in the corresponding access control list, all requests are forwarded.

> **Note:**
> The access control function works only for new connection requests and does not affect existing connections.

11.5.8 Disable access control

If you do not need to set access restrictions, you can disable the access control function.

Procedure

1. *Log on to the SLB console*.
2. **Select the region of the target Server Load Balancer (SLB) instance.**
3. **Find the target SLB instance and click the instance ID.**
4. **On the Instance Details page, click the Listeners tab.**
5. **Find the target listener and choose More > Set Access Control.**
6. **On the Access Control Settings page, disable access control and click OK.**

11.6 Backend servers

11.6.1 Backend server overview

Before you use the Server Load Balancer (SLB) service, you must add one or more ECS instances as backend servers to an SLB instance to process distributed client requests.

SLB virtualizes the added ECS instances in the same region into an application pool featured with high performance and high availability. You can manage backend servers through a VServer group. Each listener can be associated with a specific server group. Each listener of an SLB instance can forward requests to the backend server that has a specific port.

Note:
After a VServer group is configured for a listener, the listener forward requests to the ECS instances in the associated VServer group instead of the ECS instances in the default server group.

You can increase or decrease the number of the backend ECS instances at any time and specify the ECS instances that receive requests. However, we recommend that you enable the health check function, and there must be at least one normal ECS instance to maintain service stability.

When you add ECS instances to an SLB instance, note the following:

- **SLB does not support cross-region deployment. Make sure that the ECS instances and the SLB instance are in the same region.**
- **You can use different operating systems for the backend ECS instances of an SLB instance, but the applications deployed in the ECS instances must be the same, and the data must be consistent. We recommend that you use the same operating system for better management and maintenance.**
- **Up to 50 listeners can be added to an SLB instance. Each listener corresponds to an application deployed on backend ECS instances. The listening port of an SLB instance corresponds to the application port opened on the ECS instance.**
- **You can specify a weight for each ECS instance in the backend server pool. An ECS instance with a higher weight receives a larger number of connection requests.**
- **If you enable session persistence, the requests distributed to backend ECS instances may be imbalanced. To solve this problem, we recommend that you disable session persistence and check if the problem persists.**

  When traffic is not distributed evenly, troubleshoot as follows:

  1. **Collect the access logs of the web service within a period of time.**
  2. **Check if the numbers of logs of backend ECS instances match SLB configurat ions. For example, if session persistence is enabled, check the access logs for the same IP address. If different weights are configured for backend ECS instances, calculate whether the percentage of access logs matches the weight.**
- **When an ECS instance is undergoing live migration, the persistent connections of SLB may be interrupted. You can solve this problem by reconnecting them.**

Default server group

A default server group contains ECS instances that receive requests. If a listener is not associated with a VServer group or an active/standby server group, requests are forwarded to ECS instances in the default server group by default.

For more information about how to create a default server group, see *Add a default backend server*.

Active/standby server groups

An active/standby server group only contains two ECS instances. One acts as the active server and the other acts as the standby server. No health check is performed on the standby server. When the active server is declared as unhealthy, the system

forwards traffic to the standby server. When the active server is declared as healthy and restores service, the traffic is forwarded to the active server again.

For more information about how to create an active/standby server group, see *Create an active/standby server group*.

> **Note:**
> **Only layer-4 listeners (TCP and UDP protocols) support configuring active/standby server groups.**

VServer groups

If you want to distribute different requests to different backend servers, or configure domain name-based or URL-based forwarding rules, you can use VServer groups.

For more information about how to create a VServer group, see *Create a VServer group*.

11.6.2 Default server groups

11.6.2.1 Add a default backend server

This topic describes how to add a default backend server. Before you use the Server Load Balancer (SLB) service, you must add at least one default backend server to receive client requests forwarded by SLB.

Prerequisites

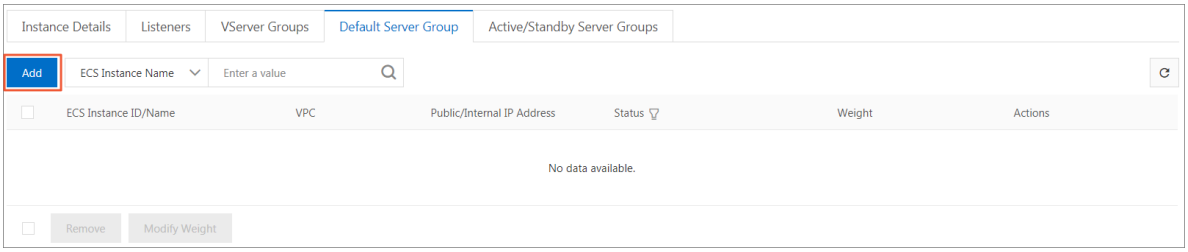
Before you add ECS instances to the default server group, make sure that the following conditions are met:

- **An SLB instance is created. For more information, see** *Create an SLB instance*.
- **ECS instances are created and applications are deployed on the ECS instances to process requests.**

Procedure

1. *Log on to the SLB console*.
2. **Find the target SLB instance and click its instance ID.**
3. **Click the Default Server Group tab.**

4. Click Add.



5. In the Available Servers dialog box that appears, select the ECS instances that you want to add to the default server group.
6. Click Next: Set Weight and Port.
7. In the Available Servers dialog box that appears, specify the weights of the ECS instances.

   An ECS instance with a higher weight receives more requests.

   To batch modify the weights of added servers, click the corresponding icon next to the weight value of the current server that you modified.

   -  : If you click this icon, the weights of all servers below the current server are also changed.
   -  : If you click this icon, the weights of all servers above the current server are also changed.
   -  : If you click this icon, the weights of all servers in the default server group are changed.
   -  : If you click this icon, the weights of all servers in the default server group are cleared.

   Notice:
   If the weight of a backend server is set to 0, no requests are sent to the backend server.

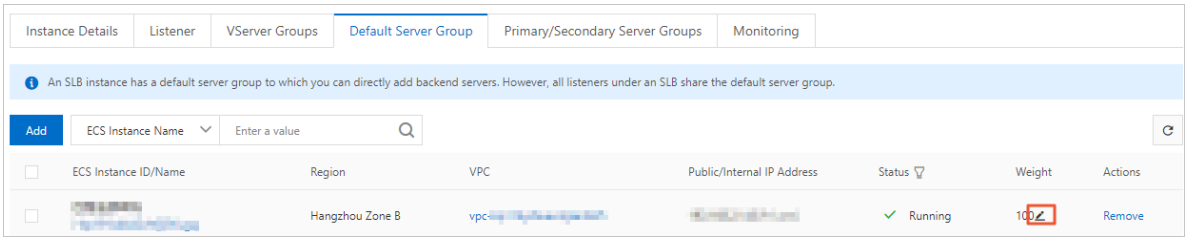
8. Click OK.

11.6.2.2 Modify the weight of a backend server

After you add a backend server to the default server group, you can modify the weight of the backend server.

Procedure

1. *Log on to the SLB console*.
2. On the Server Load Balancer page, select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. Click the Default Server Group tab.
5. Rest the pointer over the weight value of the target backend server, and then click the displayed pencil icon.



6. Modify the weight and then click OK.

   A backend server with a higher weight receives more requests.

> **Notice:**
> If the weight is set to 0, no request is sent to the backend server.

11.6.2.3 Remove a backend server

This topic describes how to remove a backend server that you no longer require to forward traffic.

Procedure

1. *Log on to the SLB console*.
2. Find the target SLB instance and click the instance ID.
3. Click the Default Server Group tab.
4. Find the target backend server and click Remove in the Actions column.

11.6.3 VServer groups

11.6.3.1 Create a VServer group

This topic describes how to create a VServer group. A VServer group consists of ECS instances which function as backend servers. If you associate a VServer group with a listener, the listener distributes requests to backend servers in the associated VServer group instead of other backend servers.

Prerequisites

Before you create a VServer group, make sure that the following conditions are met:

- **A Server Load Balancer (SLB) instance is created. For more information, see** *Create an SLB instance*.
- **ECS instances are created and applications are deployed on the ECS instances to process requests.**

Context

Take note of the following items before you create a VServer group:

- **The ECS instances added to a VServer group and the corresponding SLB instance must belong to the same region.**
- **A single ECS instance can be added to multiple VServer groups.**
- **A single VServer group can be associated with multiple listeners of an SLB instance.**
- **A VServer group consists of ECS instances and application ports.**

Procedure

1. *Log on to the SLB console*.
2. **Find the target SLB instance and click its instance ID.**
3. **Click the VServer Groups tab.**
4. **On the VServer Groups tab, click Create VServer Group.**

5. On the Create VServer Group page, perform the following operations:
   a) In the VServer Group Name field, enter a name for the VServer group to be created.
   b) Click Add. In the Available Servers dialog box that appears, select the ECS instances that you want to add.
   c) Click Next: Set Weight and Port.
   d) Enter the port number and weight of each ECS instance, and then click OK.

      Set the port numbers and weights based on the following information:

      - Port: The backend port opened on an ECS instance to receive requests.

        The backend ports in an SLB instance can be the same.
      - Weight: An ECS instance with a higher weight receives more requests.

        > **Notice:**
        > If the weight of an ECS instance is set to 0, no requests are sent to the ECS instance.

      To batch modify the port numbers and weights of added servers, click the corresponding icon next to the port or weight value of the current server that you modified.

      -  : If you click this icon, the ports or weights of all servers below the current server are also changed.
      -  : If you click this icon, the ports or weights of all servers above the current server are also changed.
      -  : If you click this icon, the ports or weights of all servers in the VServer group are changed.
      -  : If you click this icon, the ports or weights of all servers in the VServer group are cleared.

11.6.3.2 Edit a VServer group

After you create a VServer group, you can modify the configurations of the ECS instances in the VServer group.

Procedure

1. *Log on to the SLB console*.
2. On the Server Load Balancer page, select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. Click the VServer Groups tab.
5. Find the target VServer group, and then click Edit in the Actions column.
6. Modify the ports and weights of ECS instances or click Delete to remove ECS instances from the VServer group, and then click OK.

11.6.3.3 Delete a VServer group

This topic describes how to delete a VServer group. If a VServer group is no longer needed to forward traffic, you can delete the VServer group.

Procedure

1. *Log on to the SLB console*.
2. Find the target SLB instance and click its instance ID.
3. Click the VServer Groups tab.
4. Find the target VServer group, and then click Delete in the Actions column.

   Note:
   If the VServer group is associated with a listener or a forwarding rule, you must remove the listener or forwarding rule from the VServer group before the VServer group can be deleted.
5. In the dialog box that appears, click OK.

11.6.4 Active/standby server groups

11.6.4.1 Create an active/standby server group

If you need active/standby failover configurations, where one backend server is used as the active server and the other as the standby server, you can create an active/standby server group. When the active server works normally, requests are

distributed to the active server. If the active server fails, requests are distributed to the standby server.

Prerequisites

Before you create an active/standby server group, make sure the following conditions are met:

- A Server Load Balancer (SLB) instance is created. For more information, see *Create an SLB instance*.
- ECS instances are created and applications are deployed on the ECS instances to process distributed requests.

Procedure

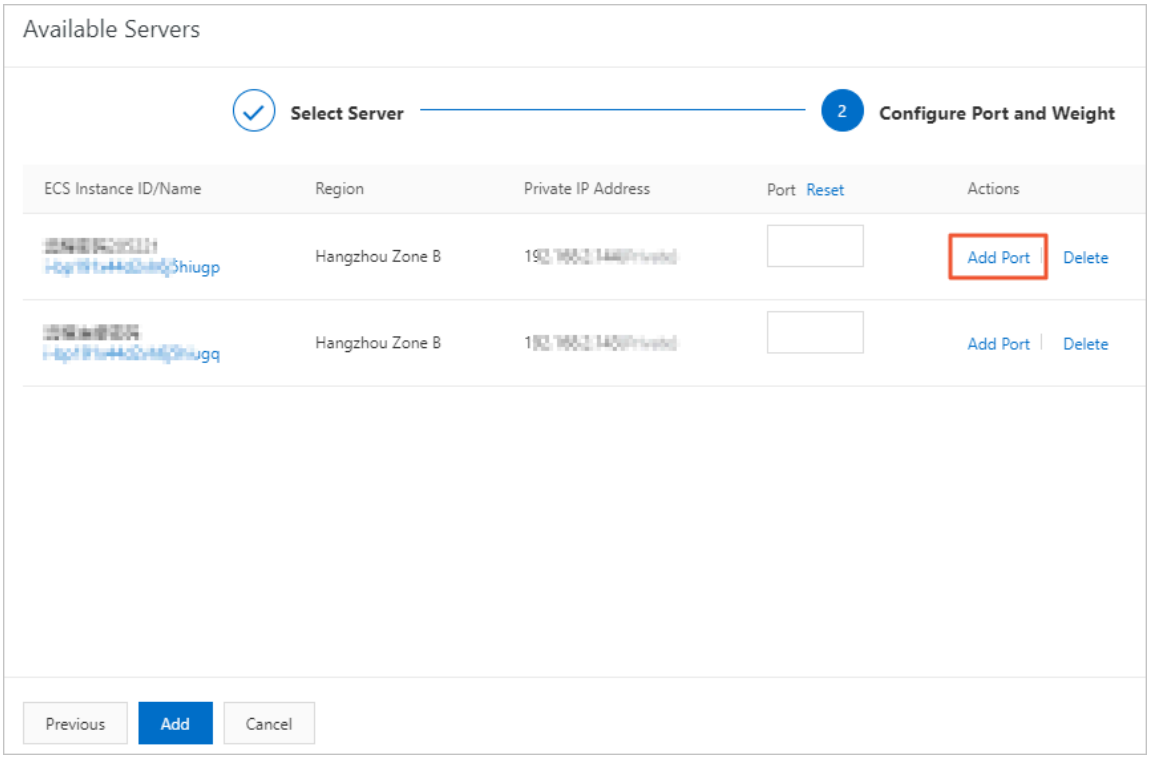
1. On the Server Load Balancer page, select the region of the target SLB instance.
2. Find the target SLB instance and click the instance ID.
3. Click the Active/Standby Server Groups tab.
4. On the Active/Standby Server Groups tab, click Create Active/Standby Server Group.

5. On the Create Active/Standby Server Group page, configure the active/standby server group.
   a) In the Name filed, enter a name for the active/standby server group to be created.
   b) Click Add and on the Available Servers page, select the servers you want to add to the active/standby server group.

      You can add only two ECS instances to an active/standby server group.
   c) Click Next: Set Weight and Port.
   d) In the Servers Added section, set the port, select an active server, and click OK.

      - Port: The backend port opened on the ECS instance to receive requests.

        The backend ports in an SLB instance can be the same.
      - Server: Select a server to act as the active server.



11.6.4.2 Delete an active/standby server group

If you no longer need an active/standby server group to forward traffic, you can delete the active/standby server group.

Procedure

1. *Log on to the SLB console*.
2. On the Server Load Balancer page, select the region of the target SLB instance.

3. **Find the target SLB instance and click the instance ID.**
4. **Click the Active/Standby Server Groups tab.**
5. **Find the target active/standby server group and click Delete in the Actions column.**
6. **In the displayed dialog box, click OK.**

11.7 Health check

11.7.1 Health check overview

This topic introduces the health check function of Server Load Balancer (SLB). SLB checks the service availability of backend servers (ECS instances) by performing health checks. Health checks improve the overall availability of your frontend service, and avoid impacts on the service caused by exceptions of backend ECS instances.

After you enable the health check function, SLB stops distributing requests to the instance that is unhealthy and restarts forwarding requests to the instance when it is declared healthy.

If your service is highly sensitive to traffic load, frequent health checks may impact your service. To reduce the impact on your service, you can reduce the health check frequency, increase the health check interval, or change a Layer-7 health check to a Layer-4 one based on the service conditions. To guarantee the service availability, we do not recommend disabling the health check function.

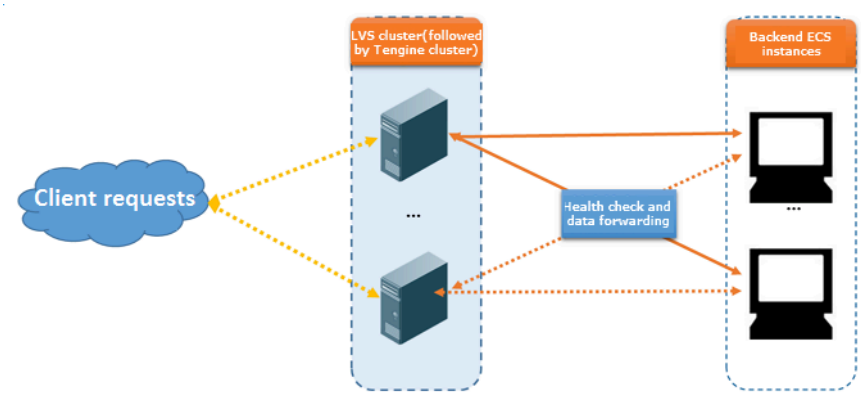
Health check process

SLB is deployed in clusters. Data forwarding and health checks are performed at the same time by node servers in the LVS cluster and Tengine cluster.

The node servers in the LVS cluster independently perform health checks in parallel, according to health check configurations. If an LVS node server detects that a backend ECS instance fails, the LVS node server no longer sends new client requests to this ECS instance. This operation is synchronized among all node servers.

The IP address range used for health checks is 100.64.0.0/10. Make sure that backend ECS instances do not block this CIDR block. You do not need to configure a security group rule to allow access from this CIDR block. However, if you have

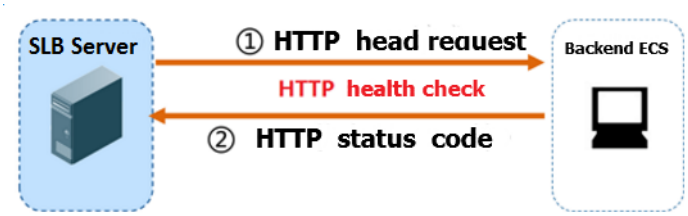
configured security rules such as iptables, you need to allow access from this CIDR block. (100.64.0.0/10 is reserved by Alibaba Cloud. Other users cannot use any IP address in this CIDR block and therefore no security risks exist.)



Health checks of HTTP/HTTPS listeners

For Layer-7 (HTTP or HTTPS) listeners, SLB checks the status of backend servers by sending HTTP HEAD requests, as shown in the following figure.

For HTTPS listeners, certificates are managed in SLB. HTTPS is not used for data exchange (including health check data and service interaction data) between SLB and backend ECS instances so that the system performance is improved.



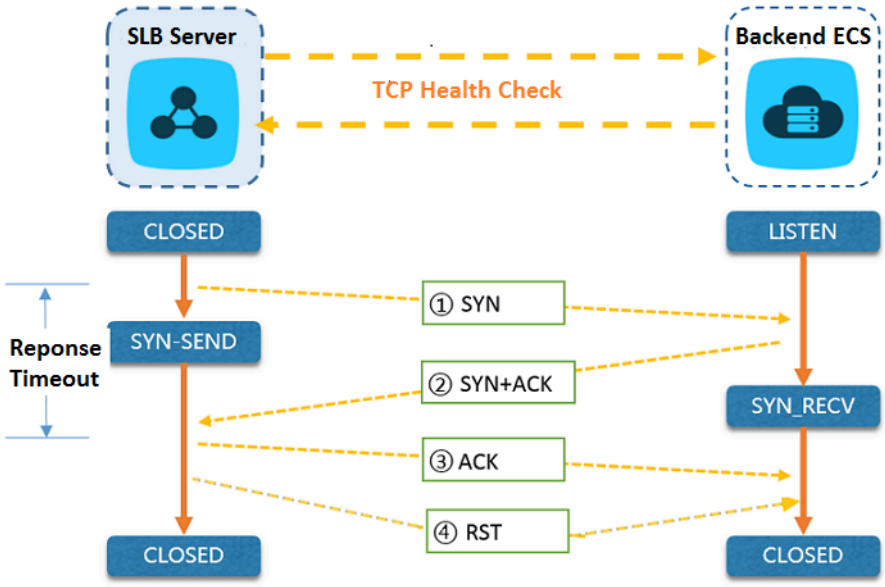
The health check process of a Layer-7 listener is as follows:

1. A Tengine node server sends an HTTP HEAD request to the intranet IP address, health check port, and health check path of a backend server according to the health check settings.
2. After receiving the request, the backend server returns an HTTP status code based on the running status.
3. If the Tengine node server does not receive the response from the backend server within the specified response timeout period, the backend server is declared as unhealthy.
4. If the Tengine node server receives a response from the backend ECS instance within the specified response timeout period, the node server compares the

response with the configured status code. If the response contains the status code that indicates a healthy server, the backend server is declared as healthy. Otherwise, the backend server is declared as unhealthy.

Health checks of TCP listeners

For TCP listeners, SLB checks the status of backend servers by establishing TCP connections, as the following figure shows.



The health check process of a TCP listener is as follows:

1. The LVS node server sends a TCP SYN packet to the intranet IP address and health check port of a backend ECS instance.
2. After receiving the request, the backend server returns a TCP SYN and ACK packet if the corresponding port is listening normally.
3. If the LVS node server does not receive the packet from the backend ECS instance within the specified response timeout period, the node server determines that the service does not respond and health check fails. Then, the node server sends an RST packet to the backend ECS instance to terminate the TCP connection.
4. If the LVS node server receives the packet from the backend ECS instance within the specified response timeout period, the node server determines that the service runs properly and the health check succeeds. Then, the server sends an RST packet to the backend ECS instance to terminate the TCP connection.

Note:

In general, TCP three-way handshakes are conducted to establish a TCP connection. After the LVS node server receives the SYN and ACK packet from the backend ECS instance, the LVS node server sends an ACK packet, and then immediately sends an RST packet to terminate the TCP connection.

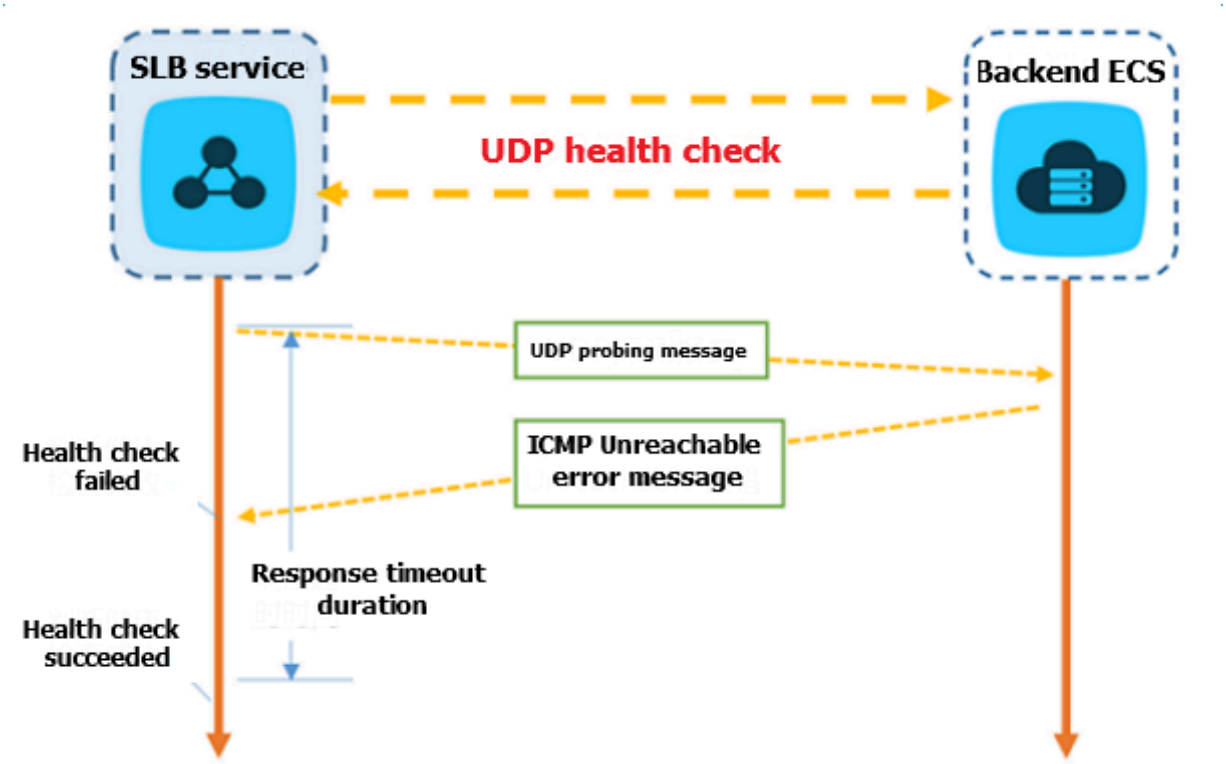
This process may cause backend servers to think an error occurred in the TCP connection, such as an abnormal exit, and then report a corresponding error message, such as `Connection reset by peer`.

Solution:

- **Use HTTP health checks.**
- **If you have enabled the function of obtaining real IP addresses, you can ignore the connection errors caused by accessing the preceding SLB CIDR block.**

Health checks of UDP listeners

For UDP listeners, SLB checks the status of backend servers by sending UDP packets , as shown in the following figure.



The health check process of a UDP listener is as follows:

1. **The LVS node server sends a UDP packet to the intranet IP address and health check port of the backend ECS instance according to health check configurations .**
2. **If the corresponding port of the ECS instance is not listening normally, the system returns an ICMP error message, such as `port XX unreachable`. Otherwise, no message is sent.**
3. **If the LVS node server receives the ICMP error message within the response timeout period, the ECS instance is declared as unhealthy.**
4. **If the LVS node server does not receive any message within the response timeout period, the ECS instance is declared as healthy.**

 **Note:**

For UDP health checks, the health check result may fail to reflect the real status of a backend server in the following situation:

If the ECS instance uses a Linux operating system, the speed of sending ICMP messages in high traffic hours is limited due to the anti-ICMP attack protection function of Linux. In this case, even if an exception occurs to the backend server, SLB may declare the backend server as healthy because the error message `port XX unreachable` is not returned. Then, the health check result deviates from the actual service status.

Solution:

Specify a pair of request and response for UDP health checks. If the specified response is returned, the ECS instance is considered healthy. Otherwise, the ECS instance is considered unhealthy. To achieve this, you must add corresponding configurations for the client.

Health check time window

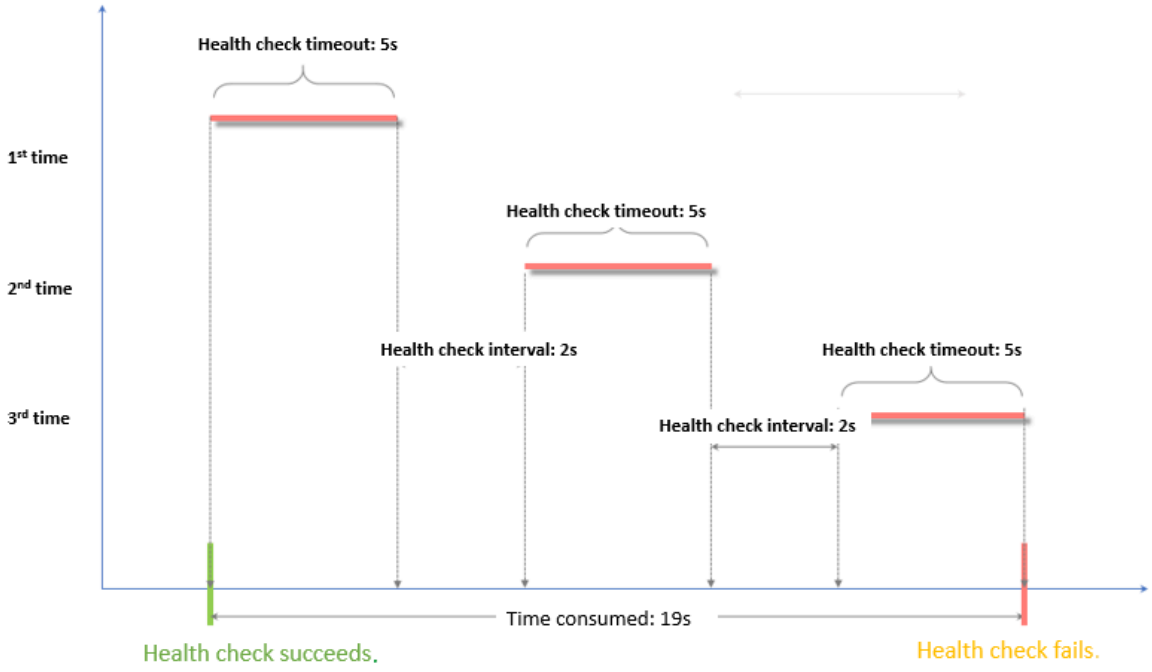
The health check function effectively improves the availability of your service. However, to avoid the impact of switching caused by frequent health check failures on system availability, status is switched (health check succeeded or failed) only when the health check succeeds or fails for a specified number of times in the time window. The health check time window is determined by the following three factors:

- **Health check interval: how often the health check is performed.**

- Response timeout: the length of time to wait for a response.
- Health check threshold: the number of consecutive successes or failures of health checks.

The health check time window is calculated as follows:

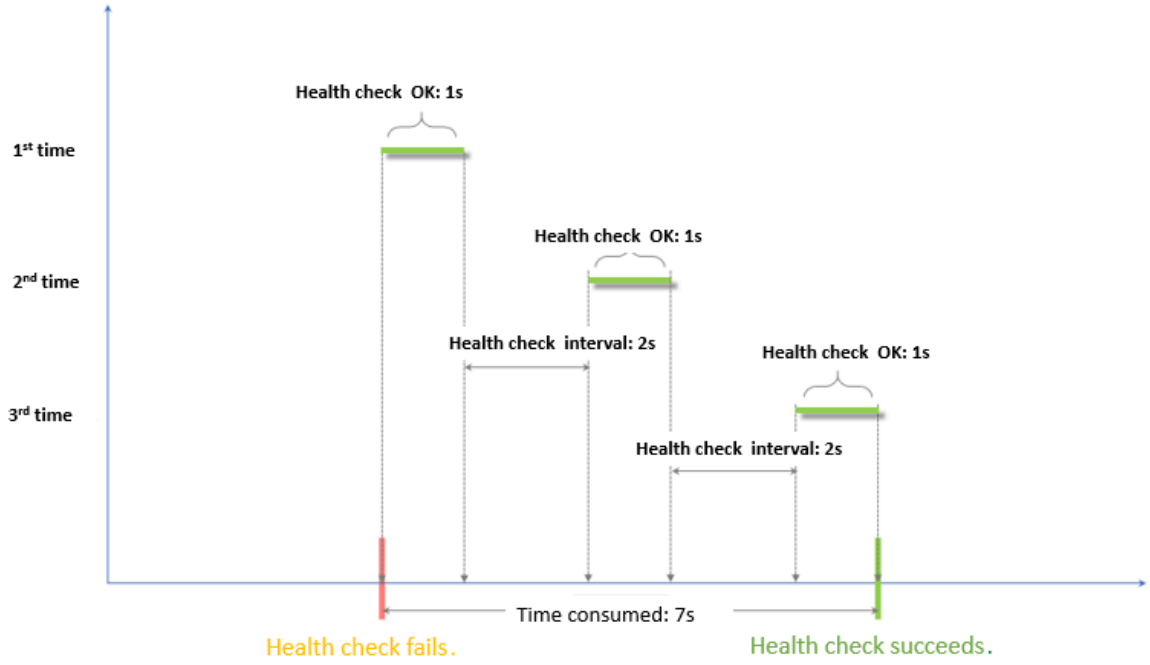
- Health check failure time window = response timeout × unhealthy threshold + health check interval × (unhealthy threshold -1)



- Health check success time window = response time of a successful health check × healthy threshold + health check interval × (healthy threshold - 1)

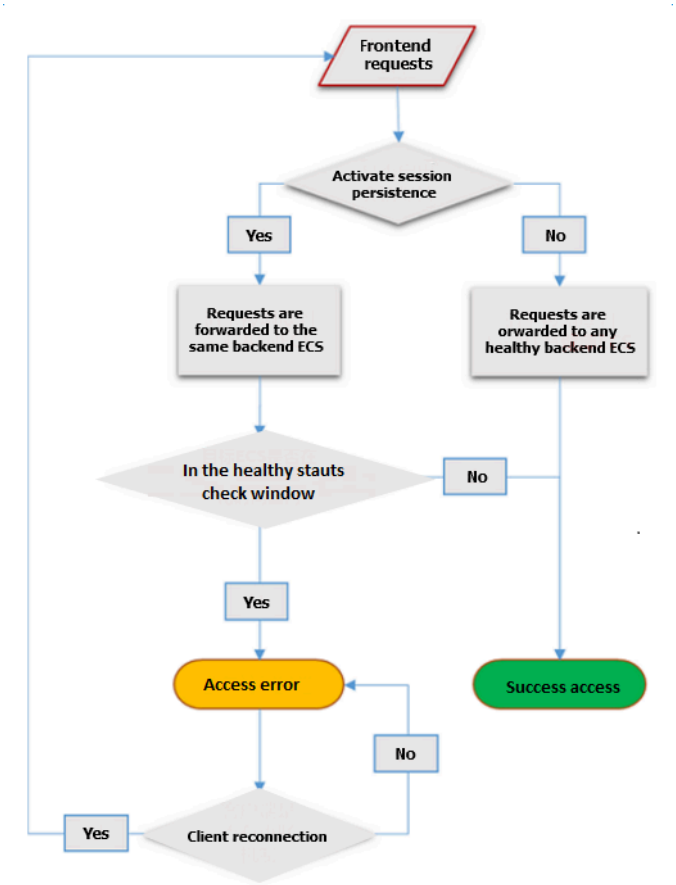
Note:
The response time of a successful health check is the duration from the time when the health check request is sent to the time when the response is received. When TCP health checks are used, the time is very short and almost negligible because the health check only checks whether the port is alive. For HTTP health

checks, the time depends on the performance and load of the application server and is generally within seconds.



The health check result has the following impact on request forwarding:

- **If the health check of the target ECS instance fails, new requests are distributed to other ECS instance. The client access is normal.**
- **If the health check of the target ECS instance succeeds, new requests are distributed to it. The client access is normal.**
- **If a request arrives during a health check failure window, the request is still sent to the ECS instance because the ECS instance is being checked and has not been declared unhealthy. Then, the client access fails.**



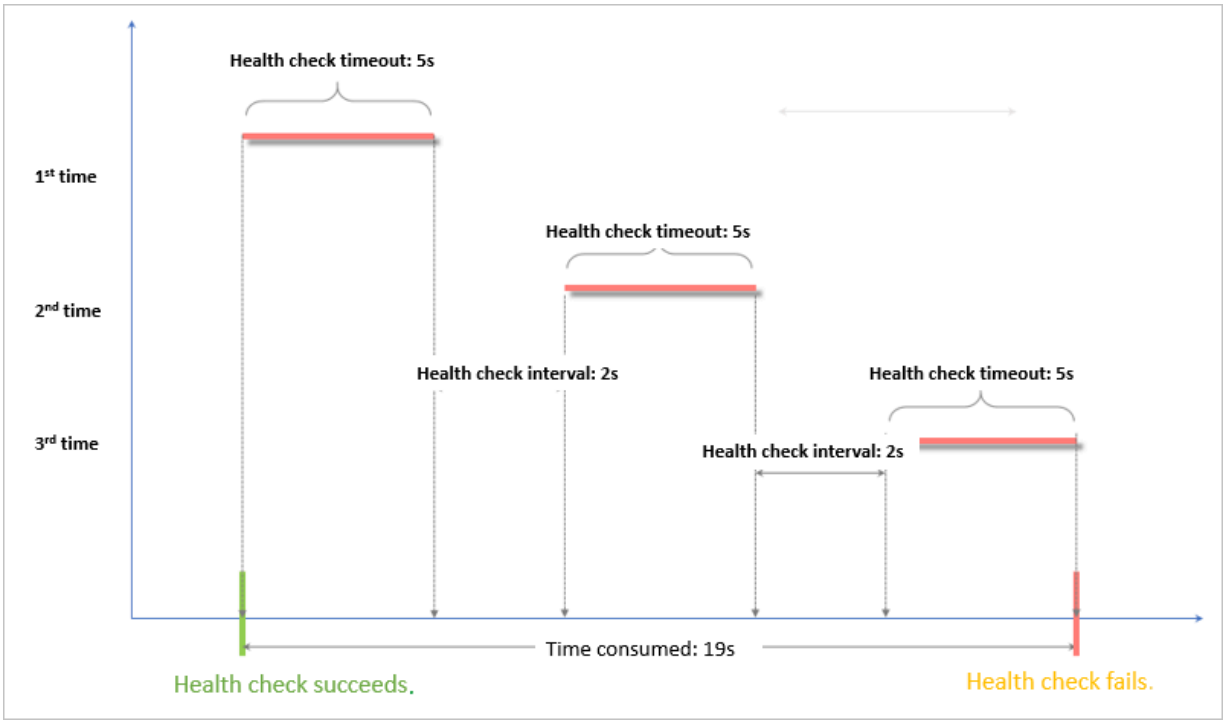
Health check examples

The following health check configurations are used as an example:

- **Response timeout: 5 seconds**
- **Health check interval: 2 seconds**
- **Healthy threshold: 3 times**
- **Unhealthy threshold: 3 times**

Health check failure time window = response timeout × unhealthy threshold + health check interval × (unhealthy threshold -1). In this example, the health check failure time window = 5 × 3 + 2 × (3-1) = 19s.

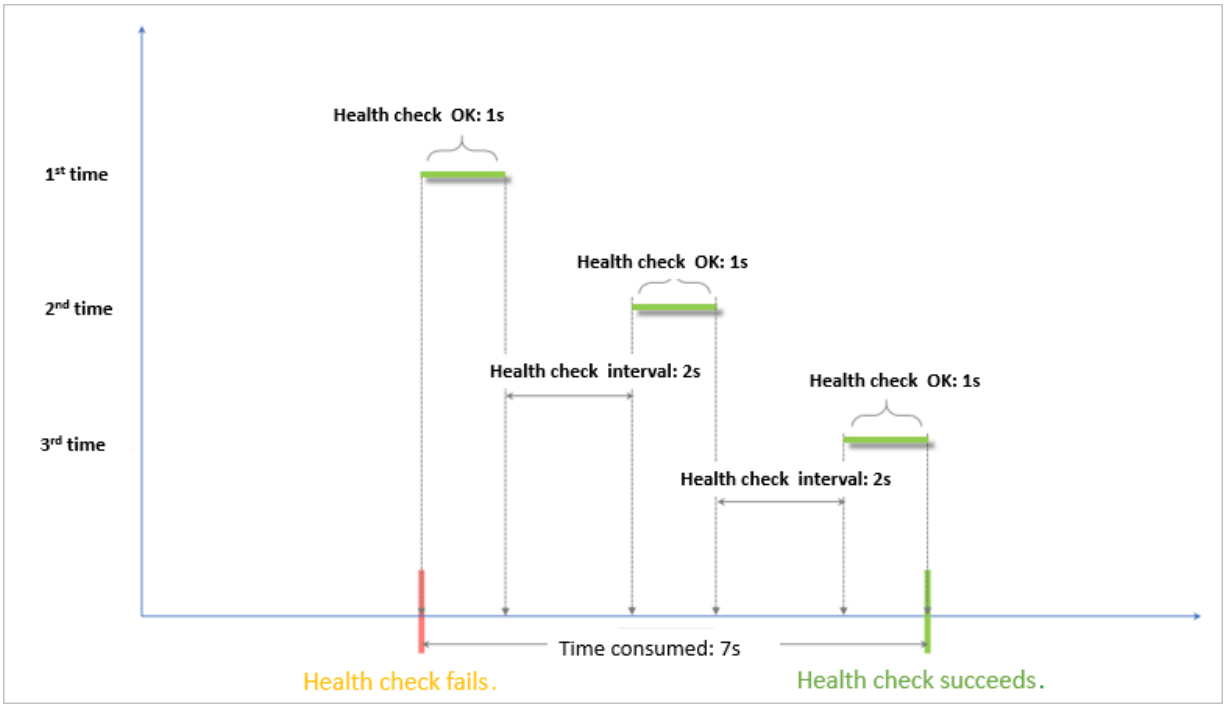
The following figure shows the time window from a healthy status to an unhealthy status.



Health check success time window = (health check success response time × healthy threshold) + health check interval × (healthy threshold-1). In this example, the health check success time window = (1 × 3) + 2 × (3-1) = 7s.

Note:
The health check success response time is the period from the time when a health check request is sent to the time when a response is received. When TCP health checks are used, the time is very short and almost negligible because the health check only checks whether the port is alive. For HTTP health checks, the time depends on the performance and load of the application server and is generally within seconds.

The following figure shows the time window from an unhealthy status to a healthy status (assume that it takes one second for the server to respond to a health check request).



Domain names in HTTP health checks

If you use HTTP health checks, you can set a domain name used for health checks. Some application servers verify the host field in the request. Therefore, the request header must contain the host field. If you set a domain name in the health check configurations, SLB adds the domain name to the host field. If no domain name is set, SLB does not include the host field in the request. Then, the health check request will be rejected by the preceding application servers and the health check may fail. Therefore, if your server requires the verification of the host field in a request, you must set a domain name in health check configurations to make sure that the health check function works normally.

11.7.2 Configure the health check feature

This topic describes how to configure the health check feature. You can configure this feature when you add a listener. The default settings can meet the requirements of most users.

Procedure

1. *Log on to the SLB console*.
2. **Find the target SLB instance and click its instance ID.**
3. **Click the Listeners tab.**

4. **Click Add Listener, or find the target listener and click Configure in the Actions column.**
5. **Click Next until the Health Check step appears.**

   We recommend that you use the default values when you configure the health check feature.

   Table 11-1: Health check configuration parameters

Parameter	Description
Health Check Protocol	**For TCP listeners, both TCP and HTTP health checks are supported.** **· TCP health checks are based on network layer detection.** **· HTTP health checks are based on HEAD requests.**
Health Check Method **(HTTP and HTTPS health checks only)**	**Health checks of Layer 7 listeners (HTTP and HTTPS listeners) support both the HEAD and the GET methods. The HEAD method is used by default.** **If your backend servers do not support the HEAD method or if the HEAD method is disabled, health checks may fail. To resolve this issue, you can use the GET method for health checks.** **If the GET method is used and the response length exceeds 8 KB, the response is truncated. However, the health check result is not affected.**

Parameter	Description
Health Check Path and Health Check Domain Name (Optional) (HTTP health checks only)	By default, SLB sends HTTP HEAD requests to the default homepage configured on the application server through the internal IP address of the backend ECS instance to perform health checks. If you do not use the default homepage of the application server for health checks, you must specify the URL for health checks. Because some application servers verify the host field in a request, the request header must contain the host field. If a domain name is configured in the health check feature, SLB adds the domain name to the host field when forwarding a request to the application server. If no domain name is configured, the health check request will be denied by the application server because it does not contain a host field and the health check may fail. If your application server verifies the host field in requests, you must configure a domain name to make sure that the health check feature works.
Normal Status Code (HTTP health checks only)	Select the HTTP status code that indicates successful health checks. Default values: http_2xx and http_3xx.
Health Check Port	The detection port used by the health check feature to access backend servers. By default, the backend port configured in the listener is used. **Note:** If a VServer group or an active/standby server group is configured for the listener, and the ECS instances in the group use different ports, leave this parameter empty. SLB uses the backend port of each ECS instance to perform health checks.

Parameter	Description
Response Timeout	The length of time to wait for a health check response. If the backend ECS instance does not send an expected response within the specified period of time, the health check will fail. Valid values: 1 to 300. Unit: seconds. Default value for UDP listeners: 10. Default value for HTTP, HTTPS, and TCP listeners: 5.
Health Check Interval	The interval between two consecutive health checks. All nodes in the LVS cluster perform health checks on backend ECS instances at the specified interval independently and in parallel. The health check statistics of a single ECS instance cannot reflect the health check interval because the nodes perform health checks at different times. Valid values: 1 to 50. Unit: seconds. Default value for UDP listeners: 5. Default value for HTTP, HTTPS, and TCP listeners: 2.
Unhealthy Threshold	The number of consecutive failed health checks that must occur on an ECS instance before the ECS instance is declared unhealthy. Valid values: 2 to 10. Default value: 3.
Healthy Threshold	The number of consecutive successful health checks that must occur on an ECS instance before the ECS instance is declared healthy. Valid values: 2 to 10. Default value: 3.

11.7.3 Disable the health check feature

This topic describes how to disable the health check feature. If you disable the health check feature, requests may be distributed to unhealthy ECS instances and cause impacts on your business. We recommend that you enable the health check feature.

Context

Note:

You can only disable the health check feature for HTTP and HTTPS listeners. The health check feature for UDP and TCP listeners cannot be disabled.

Procedure

1. *Log on to the SLB console*.
2. **On the Server Load Balancer page, find the target SLB instance and click its instance ID.**
3. **On the Listeners tab, find the target listener and click Configure in the Actions column.**
4. **On the Configure Listener page, click Next until the Health Check step appears.**
5. **Turn off Enable Health Check, click Next, click Submit, and then click OK.**

11.8 Certificate management

11.8.1 Overview

When you configure an HTTPS listener, you can directly use a certificate from Alibaba Cloud SSL Certificates Service or upload a third-party server certificate and CA certificate to Server Load Balancer (SLB). After you upload the certificate to SLB, you do not need to configure certificates on backend servers.

SLB supports the following two types of certificates:

- **Certificates issued or hosted by Alibaba Cloud SSL Certificates Service: You can select the required certificate from Alibaba Cloud SSL Certificates Service. When the certificate is about to expire, Alibaba Cloud will send alerts notifying you to renew the certificate to ensure its validity.**

  Currently, client CA certificates are not supported.
- **Third-party certificates: To upload a third-party certificate, you must have the public key and private key files of the certificate.**

  HTTPS server certificates and client CA certificates are supported.

Before you create a certificate, note the following:

- **If you need to use a certificate in multiple regions, you must select all the required regions when creating the certificate.**

- **Each Alibaba Cloud account can create up to 100 certificates.**

11.8.2 Certificate requirements

Server Load Balancer (SLB) only supports certificates in the PEM format. Before you upload a certificate, make sure that the certificate content, certificate chain, and private key conform to the corresponding format requirements.

Certificates issued by a root CA

If the certificate is issued by a root CA, the received certificate is the only one that needs to be uploaded to SLB. In this case, the website that is configured with the certificate is regarded as a trusted website and does not require additional certificates.

The certificate format must meet the following format requirements:

- **The certificate must start with `-----BEGIN CERTIFICATE-----`, and end with `-----END CERTIFICATE-----`, and both parts must be uploaded.**
- **Each line except the last line must contain 64 characters. The last line can contain 64 or fewer characters.**
- **Spaces are not allowed in the certificate content.**

The following is a sample certificate issued by a root CA.

Certificates issued by an intermediate CA

If the certificate is issued by an intermediate CA, you are able to obtain multiple intermediate certificates. You must upload both the server certificate and the required number of intermediate certificates to SLB.

The format of the certificate chain must meet the following requirements:

- **Paste the server certificate content first, and then paste the content of the one or more required intermediate certificates underneath without any blank lines in between the certificates.**
- **Spaces are not allowed in the content.**
- **Blank lines are not allowed in the content. Each line must contain 64 characters. For more information, see *RFC1421*.**
- **The certificate must conform to the corresponding format requirements. Generally, the intermediate CA provides instructions about the certificate format when issuing the certificate. The certificate chain must conform to the format requirements.**

The following is a sample certificate chain.

```
-----BEGIN CERTIFICATE-----
-----END CERTIFICATE-----
-----BEGIN CERTIFICATE-----
-----END CERTIFICATE-----
-----BEGIN CERTIFICATE-----
-----END CERTIFICATE-----
```

RSA private keys

When you upload the server certificate, you also need to upload the private key of the certificate.

The RSA private key format must meet the following requirements:

- **The private key must start with `-----BEGIN RSA PRIVATE KEY-----`, and end with `-----END RSA PRIVATE KEY-----`, and both parts must be uploaded together.**
- **Blank lines are not allowed in the content. Each line except the last line must contain 64 characters. The last line can contain 64 or fewer characters. For more information, see *RFC1421*.**

If your private key is encrypted (for example, the content at the beginning and end of the private key is `-----BEGIN PRIVATE KEY-----`, `-----END PRIVATE KEY-----` or `-----BEGIN ENCRYPTED PRIVATE KEY-----`, `-----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:

```
openssl rsa -in old_server_key.pem -out new_server_key.pem
```

The following is a sample RSA private key.

EC private keys

> **Note:**
> **Currently, EC private keys are supported only in the UK (London) region.**

When you upload the server certificate, you also need to upload the private key of the certificate.

The EC private key format must meet the following requirements:

- **The private key must start with `-----BEGIN EC PARAMETERS-----`, and end with `-----END EC PARAMETERS-----`, and both parts must be uploaded together.**
- **Blank lines are not allowed in the content. Each line except the last line must contain 64 characters. The last line can contain 64 or fewer characters. For more information, see *RFC1421*.**

If your private key is encrypted (for example, the content at the beginning and end of the private key is `-----BEGIN EC PRIVATE KEY-----`, `-----END EC PRIVATE KEY`

`-----`, or the private key contains `Proc-Type: 4,ENCRYPTED`), you must first run the following command to convert the private key:

```
openssl ec -in old_server_key.pem -out new_server_key.pem
```

The following is a sample EC private key.

```
-----BEGIN EC PARAMETERS-----
Bggq**********Bw==
-----END EC PARAMETERS-----
-----BEGIN EC PRIVATE KEY-----
MHcCAQEEICo9b+vQUhqFUWgWjE0YY4h0b3bE/udcubxVwcVY99MuoAoGCCqGSM49
AwEHoUQDQgAEgpla3Bj9rX**********4xz0SHsuQc/7XBmgmrMpAmE80c0DR
5HcMHFxRPtGLv22T62e5KqN1W3uN9Hplgg==
-----END EC PRIVATE KEY-----
```

11.8.3 Upload a certificate

Before you create an HTTPS listener, you must upload the required server certificate and CA certificate to SLB. You no longer need to configure certificates on backend servers after uploading the certificates to SLB.

Prerequisites

- A server certificate is purchased.
- A CA certificate and a client certificate are generated. For more information, see *Generate a CA certificate*.

Context

Before you upload a certificate, note the following:

- If you want to use a certificate in multiple regions, you must select all the required regions.
- Up to 100 certificates can be uploaded under one account.

Procedure

1. *Log on to the SLB console*.
2. In the left-side navigation pane, choose Certificates.
3. Click Create Certificate.

4. On the Create Certificate page, upload the certificate and then click OK.

Configuration	Description
Certificate Name	Enter a name for the certificate to be uploaded. The name must be 1 to 80 characters in length, and can only contain letters, numbers, and the following special characters: _ / . -
Regions	Select one or more regions to which the certificate to be uploaded belongs. A certificate cannot be used across regions. If you need to use a certificate in multiple regions, select all the required regions.
Certificate Type	Select the type of the certificate to be uploaded: · Server Certificate: For HTTPS one-way authentication, only the server certificate and the private key are required. · CA Certificate: For HTTPS mutual authentication, both the server certificate and the CA certificate are required.
Certificate Content	Paste the certificate content in the editor. Click View Sample Certificate to view the valid certificate formats. For more information, see *Certificate requirements*.
Private Key	Paste the private key of the server certificate in the editor. Click View Sample Certificate to view the valid certificate formats. For more information, see *Certificate requirements*. **Notice:** A private key is required only when you upload a server certificate.

To delete expired certificates in batches, click Remove All Expired Certificates.

11.8.4 Generate a CA certificate

When you configure an HTTPS listener, you can use a self-signed CA certificate. This topic describes how to generate a CA certificate and use the CA certificate to sign a client certificate.

Generate a CA certificate by using Open SSL

1. **Run the following commands to create a *ca* folder in the */root* directory and then create four subfolders under the *ca* folder.**

```
$ sudo mkdir ca
$ cd ca
$ sudo mkdir newcerts private conf server
```

   - ***newcerts* is used to store the digital certificate signed by the CA certificate.**
   - ***private* is used to store the private key of the CA certificate.**
   - ***conf* is used to store the configuration files used for simplifying parameters.**
   - ***server* is used to store the server certificate.**

2. **Create an *openssl.conf* file that contains the following information in the *conf* directory.**

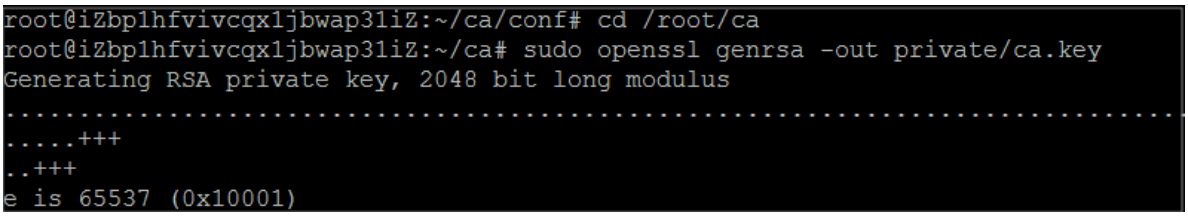
```
[ ca ]
default_ca = foo
[ foo ]
dir = /root/ca
database = /root/ca/index.txt
new_certs_dir = /root/ca/newcerts
certificate = /root/ca/private/ca.crt
serial = /root/ca/serial
private_key = /root/ca/private/ca.key
RANDFILE = /root/ca/private/.rand
default_days = 365
default_crl_days= 30
default_md = 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 command to generate a private key.**

```
$ cd /root/ca
```

```
$ sudo openssl genrsa -out private/ca.key
```

The following figure is an example of the key generation.

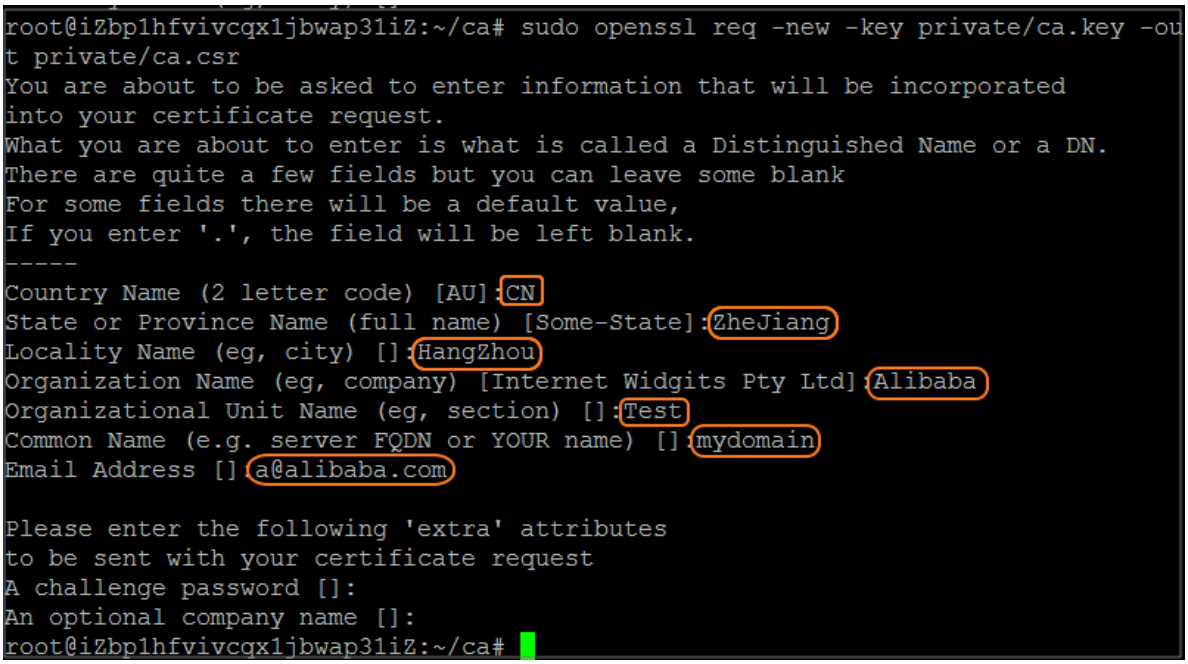


4. **Run the following command and input the required information according to the prompts. Press Enter to generate a `csr` file.**

```
$ sudo openssl req -new -key private/ca.key -out private/ca.csr
```

Note:

Common Name is the domain name of the SLB instance.



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 private 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 output is:

```
Using configuration from /root/ca/conf/openssl.conf
```

Sign the client certificate

1. **Run the following command to generate a `users` folder under the `ca` directory to store the client key.**

```
$ 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
```

Note:

Enter a pass phrase when creating the key. It is the password to protect the private key from unauthorized access. Enter the same password twice.

3. **Run the following command to create a `csr` file for the client key.**

```
$ sudo openssl req -new -key /root/ca/users/client.key -out /root/ca
/users/client.csr
```

Enter the pass phrase set in the previous step and other required information when prompted.

Note:

A challenge password is the password of the client certificate. Note that it is not the password of the client key.

4. **Run the following command 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.**

```
root@iZbp1hfvivcqx1jbwap31iZ:~/ca# sudo openssl ca -in /root/ca/users/client.csr
 -cert /root/ca/private/ca.crt -keyfile /root/ca/private/ca.key -out /root/ca/us
ers/client.crt -config "/root/ca/conf/openssl.conf"
Using configuration from /root/ca/conf/openssl.conf
Check that the request matches the signature
Signature ok
The Subject's Distinguished Name is as follows
countryName           :PRINTABLE:'CN'
stateOrProvinceName   :ASN.1 12:'ZheJiang'
localityName          :ASN.1 12:'HangZhou'
organizationName      :ASN.1 12:'Alibaba'
organizationalUnitName:ASN.1 12:'Test'
commonName            :ASN.1 12:'mydomain'
emailAddress          :IA5STRING:'a@alibaba.com'
Certificate is to be certified until Jun  4 15:28:55 2018 GMT (365 days)
Sign the certificate? [y/n]:y


1 out of 1 certificate requests certified, commit? [y/n]y
Write out database with 1 new entries
Data Base Updated
root@iZbp1hfvivcqx1jbwap31iZ:~/ca#
```

5. **Run the following command to convert the certificate to a** *PKCS12* **file.**

```
$ sudo openssl pkcs12 -export -clcerts -in /root/ca/users/client.crt
 -inkey /root/ca/users/client.key -out /root/ca/users/client.p12
```

Follow the prompts to enter the pass phrase of client key. Then enter the password used for exporting the client certificate. This password is used to protect the client certificate, which is required when you install the client certificate.

6. **Run the following commands to view the generated client certificate:**

```
 cd users
```

```
ls
```

11.8.5 Convert the certificate format

Server Load Balancer (SLB) supports PEM certificates only. Certificates in other formats must be converted to the PEM format before they can be uploaded to SLB. We recommend that you use Open SSL for conversion.

Convert DER to PEM

DER: This format is usually used on a Java platform. The certificate file suffix is generally *. der*, *. cer*, **or** *. crt*.

- **Run the following command to convert the certificate format:**

```
openssl x509 -inform der -in certificate.cer -out certificate.pem
```

- **Run the following command to convert the private key:**

```
openssl rsa -inform DER -outform PEM -in privatekey.der -out
privatekey.pem
```

Convert P7B to PEM

P7B: This format is usually used in a Windows server and Tomcat.

Run the following command to convert the certificate format:

```
openssl pkcs7 -print_certs -in incertificate.p7b -out outcertificate.
cer
```

Convert PFX to PEM

PFX: This format is usually used in a Windows server.

- **Run the following command to extract the certificate:**

```
openssl pkcs12 -in certname.pfx -nokeys -out cert.pem
```

- **Run the following command to extract the private key:**

```
openssl pkcs12 -in certname.pfx -nocerts -out key.pem -nodes
```

11.8.6 Replace a certificate

This topic describes how to replace a certificate. To avoid the impact of certificate expiration on your service, we recommend that you replace the certificate before the certificate expires.

Procedure

1. **Create and upload a new certificate.**

   For more information, see *Overview*.

2. **Configure the new certificate in HTTPS listener configuration.**

   For more information, see *Add an HTTPS listener*.

3. **On the Certificates page, find the target certificate, and click Delete.**
4. **In the displayed dialog box, click OK.**

12 Virtual Private Cloud (VPC)

12.1 Quick start

12.1.1 Overview

This tutorial provides guidance for you to create a VPC and deploy an ECS instance in the VPC.

Prerequisites

Before you create a VPC that contains an ECS instance, you must plan a CIDR block first. For more information, see *Network planning*.

Configuration process

This tutorial includes the following topics:

- *Log on to the VPC console*

  This topic describes how to log on to the VPC console.

- *Create a VPC*

  This topic describes how to create a VPC. Before you can use cloud resources in a VPC, you must first create a VPC.

- *Create a VSwitch*

  This topic describes how to create a VSwitch after you create a VPC.

- *Create a security group*

  This topic describes how to create a security group. Before you create an ECS instance in a VPC, you must create a security group first. Security groups can be used to control access to or from ECS instances.

- *Create an ECS instance*

  An ECS instance is a virtual computing environment that consists of basic server components, such as CPUs, memory, operating system, disks, and bandwidth.

12.1.2 Network planning

When you create a VPC and a VSwitch, you must specify the private IP address segment for them in the form of a CIDR block.

CIDR is a bitwise, prefix-based standard for the interpretation of IP addresses. It allows address segments to be grouped into a route entry. You can flexibly allocate IP address segments with subnet masks such as /25, /26, and /27. These IP address segments are called CIDR blocks.

Plan the CIDR block for a VPC

When planning the CIDR block for a VPC, note the following limits:

- **You can use the standard private network segments (192.168.0.0/16, 10.0.0.0/8, 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 available CIDR blocks in the global configuration during the delivery planning phase.**
- **When creating a VPC using the API, the allowed block size is between a /8 netmask and /24 netmask.**
- **After you create a VPC, you cannot modify its CIDR block.**

Plan the CIDR block for a VSwitch

When you plan the CIDR block for a VSwitch, note the following limits:

- **The allowed block size of the VSwitch is between a /16 netmask and /29 netmask. This means the VSwitch can provide 8 to 65,536 IP addresses.**
- **The CIDR block of a VSwitch must be within the range of the CIDR block of the VPC to which the VSwitch belongs.**

  Note:
  If the CIDR blocks of your VSwitch and VPC are the same, only this single VSwitch can be created.

- **The first and last three IP addresses of a VSwitch are reserved. For example, if the CIDR block of a VSwitch is 192.168.1.0/24, IP addresses 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved.**
- **The CIDR block of a VSwitch cannot be the same as the destination CIDR block in the route entry of the VPC where the VSwitch resides. However, it can be a subset of the destination CIDR block in the current route entry.**

- After you create a VSwitch, you cannot modify its CIDR block.

12.1.3 Log on to the VPC console

This topic describes how to log on to the VPC console by using Google Chrome.

Prerequisites

- Before logging on to the ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel. The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.
- We recommend that you use the Google Chrome browser.

Procedure

1. In the address bar, enter the URL used to access the ASCM console. Press Enter.
2. Enter your username and password.

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

   > **Note:**
   > When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).

3. Click Login to go to the ASCM console homepage.
4. In the top navigation bar, choose Products > Networking > Virtual Private Cloud.

12.1.4 Create a VPC

This topic describes how to create a VPC. Before you can use cloud resources in a VPC, you must first create a VPC.

Context

When you create a VPC, the following limits apply:

- You can specify only one CIDR block for each VPC.
- After you create a VPC, the system automatically creates a VRouter and a route table for the VPC. Each VPC can have only one VRouter and one route table.

Procedure

1. *Log on to the VPC console*.
2. On the VPCs page, click Create VPC.
3. In the Create VPC dialog box that appears, set the parameters and click Submit. The following table describes the parameters.

Parameter	Description
Organization	The organization to which the VPC belongs.
Resource Set	The resource set to which the VPC belongs.
Region	The region where your VPC is deployed.
Shared with Subdepartments	Indicates whether to share the VPC. If you select Yes, administrators of sub-organizations can create resources in the VPC.
VPC Name	The name of the VPC. The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.
IPv4 CIDR Block	The IPv4 CIDR block of the VPC. You can specify an IPv4 CIDR block in either of the following ways: - Recommended CIDR Block: You can use one of the three standard IPv4 CIDR blocks: 192.168.0.0/16, 172.16.0.0/12, and 10.0.0.0/8. - Custom CIDR Block: You can use 192.168.0.0/16, 172.16.0.0/12, 10.0.0.0/8, and their subnets as the CIDR blocks. The allowed block size is between a /24 netmask and /8 netmask. Example: 192.168.0.0/16.  **Note:** After you create a VPC, you cannot change its IPv4 CIDR block.

Parameter	Description
IPv6 CIDR Block	Indicates whether to assign an IPv6 CIDR block. · Do Not Assign: The system does not assign an IPv6 CIDR block to the VPC. · Assign: The system assigns an IPv6 CIDR block to the VPC. If you set this parameter to Assign, the system automatically creates a free IPv6 gateway for your VPC, and assigns an IPv6 CIDR block with netmask /56, such as 2xx1: db8::/56. IPv6 addresses only support communication in private networks by default. If you want to enable the VPC to access the Internet or be accessed by an IPv6 client on the Internet, you must enable IPv6 Internet bandwidth. For more information, see *Enable Internet bandwidth for an IPv6 address*.
Description	The description of the VPC. The description must be 2 to 256 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.

12.1.5 Create a VSwitch

This topic describes how to create a VSwitch after you create a VPC.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click VSwitches.
3. Select the region of the VPC to which the VSwitch belongs.
4. On the VSwitches page, click Create VSwitch.
5. On the VSwitch page that appears, configure the parameters and click Submit. The following table describes the parameters.

Parameter	Description
Organization	The organization to which the VSwitch belongs.
Resource Set	The resource set to which the VSwitch belongs.
Region	The region to which the VSwitch belongs.

Parameter	Description
Zone	The zone to which the VSwitch belongs. In a VPC, a VSwitch can belong to only one zone. However, you can deploy cloud resources on VSwitches of different zones to achieve zone-disaster recovery. **Note:** You can add a cloud resource instance to only one VSwitch.
VPC	The VPC for which you want to create the VSwitch.
VSwitch Name	The name of the VSwitch. The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.
IPv4 CIDR Block	The IPv4 CIDR block of the VSwitch. · You must specify the IP address segment for the VSwitch in the form of a CIDR block. The allowed block size is between a /29 netmask and /16 netmask. This means the VSwitch can provide 8 to 65,536 IP addresses. · The CIDR block of the 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 VSwitch can be created. · The first and last three IP addresses of a VSwitch are reserved. For example, If the CIDR block of a VSwitch is 192.168.1.0/24, the IP addresses 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved. · The CIDR block of the VSwitch cannot be the same as the destination CIDR block of a route entry of the VPC where the VSwitch resides. However, the CIDR block of the VSwitch can be a subset of the destination CIDR block of the route entry. · After you create a VSwitch, you cannot modify its CIDR block.

Parameter	Description
IPv6 CIDR Block	The IPv6 CIDR block of the VSwitch. · If IPv6 is not enabled for the VPC for which you want to create the VSwitch, you cannot assign an IPv6 CIDR block to the VSwitch. · If IPv6 is enabled for the VPC, you can enter a digit that ranges from 0 to 255 to specify the last 8 bits of the IPv6 CIDR block. For example, if the IPv6 CIDR block of the VPC is 2xx1: db8::/64, you can enter 255 (ff in hexadecimal notation) for the IPv6 CIDR block of the VSwitch. Then, the IPv6 CIDR block of the VSwitch is 2xx1: db8: ff::/64.
Description	The description of the VSwitch. The description must be 2 to 256 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.

12.1.6 Create a security group

This topic describes how to create a security group. Before you create an ECS instance in a VPC, you must create a security group first. Security groups can be used to control access to or from ECS instances.

Procedure

1. **Log on to the ASCM console.**
2. **In the top navigation bar, choose Products > Elastic Computing > Elastic Compute Service.**
3. **Choose Networks and Security > Security Groups.**
4. **On the Security Groups page, click New Security Group.**
5. **In the Create Security Group dialog box that appears, set the parameters and click Submit. The following table describes the parameters.**

Parameter	Description
Organization	The organization to which the security group belongs.
Resource Set	The resource set to which the security group belongs.

Parameter	Description
Region	The region to which the security group belongs. The security group and the VPC must belong to the same region.
Zone	The zone to which the security group belongs.
VPC	The VPC to which the security group belongs.
Security Group Name	The name of the security group. The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.
Description	The description of the security group. The description must be 2 to 256 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.

12.1.7 Create an ECS instance

This topic describes how to create an ECS instance. An ECS instance is a virtual computing environment that consists of basic server components, such as CPUs, memory, operating system, disks, and bandwidth.

Procedure

1. **Log on to the ASCM console.**
2. **In the top navigation bar, choose Products > Elastic Computing > Elastic Compute Service.**
3. **Click Instances.**
4. **On the Instances page, click Create Instance.**
5. **In the Create ECS instance dialog box that appears, set the parameters and click Submit.**

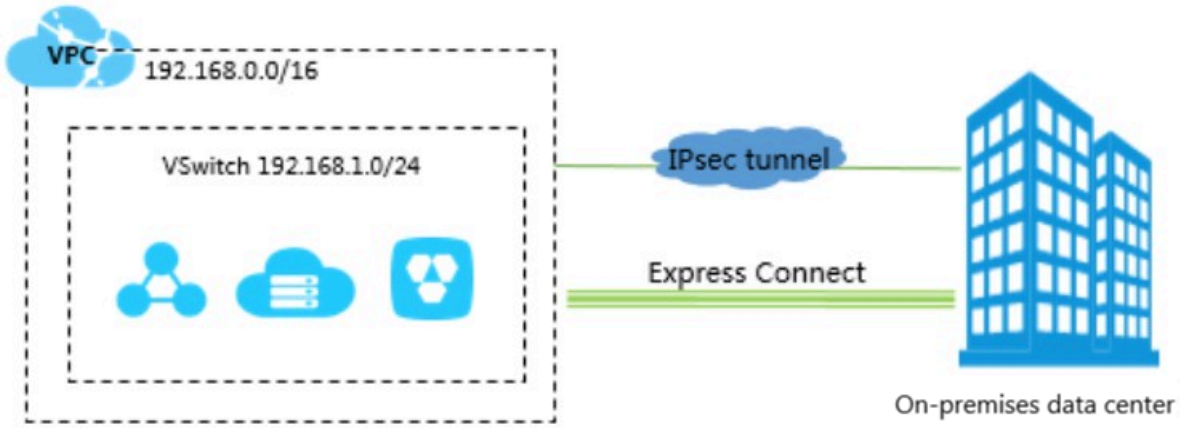
   For more information about how to create an ECS instance, *see* **Create an instance in Quick start in the ECS user guide.**

12.2 VPCs

12.2.1 VPC overview

You can configure the IP address ranges, route tables, and gateways of your VPCs. You can also use Apsara Stack resources, such as ECS, ApsaraDB for RDS, and Server Load Balancer (SLB) instances, in your VPCs.

Additionally, you can connect a VPC with another VPC or an on-premises data center to establish a custom network. In this way, you can smoothly migrate applications to the cloud and expand the on-premises data center.



12.2.2 Create a VPC

This topic describes how to create a VPC. Before you can use cloud resources in a VPC, you must first create a VPC.

Context

When you create a VPC, the following limits apply:

- **You can specify only one CIDR block for each VPC.**
- **After you create a VPC, the system automatically creates a VRouter and a route table for the VPC. Each VPC can have only one VRouter and one route table.**

Procedure

1. *Log on to the VPC console*.
2. **On the VPCs page, click Create VPC.**

3. In the Create VPC dialog box that appears, set the parameters and click Submit.

   The following table describes the parameters.

Parameter	Description
Organization	The organization to which the VPC belongs.
Resource Set	The resource set to which the VPC belongs.
Region	The region where your VPC is deployed.
Shared with Subdepartments	Indicates whether to share the VPC. If you select Yes, administrators of sub-organizations can create resources in the VPC.
VPC Name	The name of the VPC. The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.
IPv4 CIDR Block	The IPv4 CIDR block of the VPC. You can specify an IPv4 CIDR block in either of the following ways: · Recommended CIDR Block: You can use one of the three standard IPv4 CIDR blocks: 192.168.0.0/16, 172.16.0.0/12, and 10.0.0.0/8. · Custom CIDR Block: You can use 192.168.0.0/16, 172.16.0.0/12, 10.0.0.0/8, and their subnets as the CIDR blocks. The allowed block size is between a /24 netmask and /8 netmask. Example: 192.168.0.0/16.  Note: After you create a VPC, you cannot change its IPv4 CIDR block.

Parameter	Description
IPv6 CIDR Block	Indicates whether to assign an IPv6 CIDR block. · Do Not Assign: The system does not assign an IPv6 CIDR block to the VPC. · Assign: The system assigns an IPv6 CIDR block to the VPC. If you set this parameter to Assign, the system automatically creates a free IPv6 gateway for your VPC, and assigns an IPv6 CIDR block with netmask /56, such as 2xx1: db8::/56. IPv6 addresses only support communication in private networks by default. If you want to enable the VPC to access the Internet or be accessed by an IPv6 client on the Internet, you must enable IPv6 Internet bandwidth. For more information, see *Enable Internet bandwidth for an IPv6 address*.
Description	The description of the VPC. The description must be 2 to 256 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.

12.2.3 Modify a VPC

This topic describes how to modify the name and description of a VPC.

Prerequisites

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

Procedure

1. *Log on to the VPC console*.
2. Select the region where your VPC is deployed.
3. On the VPCs page, find the VPC, and click Manage in the Actions column.
4. In the VPC Details section, click Edit next to Name. In the dialog box that appears, enter a new VPC name and click OK.

   The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), and hyphens (-). It must start with a letter.

5. Click Edit next to Description. In the dialog box that appears, enter a new description of the VPC and click OK.

   The description must be 2 to 256 characters in length and cannot start with `http://` or `https://`.

12.2.4 Delete a VPC

This topic describes how to delete a VPC. After you delete a VPC, the associated VRouter and route table are also deleted.

Prerequisites

Before you delete a VPC, ensure that the following conditions are met:

- The VSwitch of the VPC is deleted. For more information, see *Delete a VSwitch*.
- The NAT gateway of the VPC is deleted. For more information, see *Delete a NAT gateway*.

Procedure

1. *Log on to the VPC console*.
2. Select the region where your VPC is deployed.
3. On the VPCs page, find the VPC, and click Delete in the Actions column.
4. In the Delete VPC dialog box that appears, click OK.

12.3 VSwitches

12.3.1 VSwitch overview

A VSwitch is a basic network device in a VPC and is used to connect cloud resources. After you create a VPC, you can create VSwitches to define subnets in the VPC. The VSwitches within a VPC are interconnected. You can deploy applications in VSwitches of different zones to improve service availability.

A cloud resource cannot be directly deployed in a VPC, but can be deployed in a VSwitch (subnet) of the VPC.

12.3.2 Create a VSwitch

This topic describes how to create a VSwitch after you create a VPC.

Procedure

1. *Log on to the VPC console*.

2. In the left-side navigation pane, click VSwitches.
3. Select the region of the VPC to which the VSwitch belongs.
4. On the VSwitches page, click Create VSwitch.
5. On the VSwitch page that appears, configure the parameters and click Submit. The following table describes the parameters.

Parameter	Description
Organization	The organization to which the VSwitch belongs.
Resource Set	The resource set to which the VSwitch belongs.
Region	The region to which the VSwitch belongs.
Zone	The zone to which the VSwitch belongs. In a VPC, a VSwitch can belong to only one zone. However, you can deploy cloud resources on VSwitches of different zones to achieve zone-disaster recovery. Note: You can add a cloud resource instance to only one VSwitch.
VPC	The VPC for which you want to create the VSwitch.
VSwitch Name	The name of the VSwitch. The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.

Parameter	Description
IPv4 CIDR Block	The IPv4 CIDR block of the VSwitch. · You must specify the IP address segment for the VSwitch in the form of a CIDR block. The allowed block size is between a /29 netmask and /16 netmask. This means the VSwitch can provide 8 to 65,536 IP addresses. · The CIDR block of the 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 VSwitch can be created. · The first and last three IP addresses of a VSwitch are reserved . For example, If the CIDR block of a VSwitch is 192.168.1.0/24 , the IP addresses 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved. · The CIDR block of the VSwitch cannot be the same as the destination CIDR block of a route entry of the VPC where the VSwitch resides. However, the CIDR block of the VSwitch can be a subset of the destination CIDR block of the route entry. · After you create a VSwitch, you cannot modify its CIDR block.
IPv6 CIDR Block	The IPv6 CIDR block of the VSwitch. · If IPv6 is not enabled for the VPC for which you want to create the VSwitch, you cannot assign an IPv6 CIDR block to the VSwitch. · If IPv6 is enabled for the VPC, you can enter a digit that ranges from 0 to 255 to specify the last 8 bits of the IPv6 CIDR block. For example, if the IPv6 CIDR block of the VPC is 2xx1: db8::/64, you can enter 255 (ff in hexadecimal notation) for the IPv6 CIDR block of the VSwitch. Then, the IPv6 CIDR block of the VSwitch is 2xx1: db8: ff::/64.
Description	The description of the VSwitch. The description must be 2 to 256 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.

12.3.3 Create a cloud resource

This topic describes how to create a cloud resource. A cloud resource cannot be directly deployed in a VPC, but can be deployed in a VSwitch (subnet) of the VPC.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click VSwitches.
3. Select the region of the VPC to which the VSwitch belongs.
4. On the VSwitches page, find the VSwitch, move the pointer over Purchase in the Actions column, and select the cloud resource you want to create.

   Cloud resources that can be created in VSwitches include ECS instances, SLB instances, and RDS instances.
5. On the page that appears, set the parameters and click Submit.
   - For information about how to create an ECS instance, see Create an instance under Quick start in *ECS User Guide* .
   - For information about how to create an SLB instance, see Create an SLB instance under Quick start in *SLB User Guide* .
   - For information about how to create an RDS instance, see Create an instance under Quick start in *RDS User Guide* .

12.3.4 Modify a VSwitch

This topic describes how to modify the name and description of a VSwitch.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click VSwitches.
3. Select the region of the VPC to which the VSwitch belongs.
4. On the VSwitches page, find the target VSwitch, and click Manage in the Actions column.
5. In the VSwitch Basic Information section, click Edit next to Name. In the dialog box that appears, enter a new name for the VSwitch, and click OK.

   The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), and hyphens (-). It must start with a letter.

6. Click Edit next to Description. In the dialog box that appears, enter a new description of the VSwitch and click OK.

   The description must be 2 to 256 characters in length and cannot start with `http://` or `https://`.

12.3.5 Delete a VSwitch

This topic describes how to delete a VSwitch. After you create a VSwitch, cloud resources cannot be deployed in it.

Prerequisites

Before you delete a VSwitch, ensure that the following conditions are met:

- The cloud resources created under the VSwitch, such as ECS instances, SLB instances, and RDS instances, are deleted.
- SNAT entries associated with the VSwitch are deleted. For more information, see *Delete a SNAT entry*.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click VSwitches.
3. Select the region of the VPC to which the VSwitch belongs.
4. On the VSwitches page, find the target VSwitch, and click Delete in the Actions column.
5. In the Delete VSwitch dialog box that appears, click Ok.

12.4 Route tables

12.4.1 Route table overview

After you create a VPC, the system creates a default route table for the VPC and adds system routes to the VPC for traffic management.

A route table is a list of route entries on a router. The system creates a route table for a VPC after the VPC is created. When the VPC is deleted, the route table is also deleted. You cannot create or delete a route table.

Each entry in the route table is a route entry. A route entry, which specifies a destination for network traffic, consists of a destination CIDR block, next hop type,

and next hop. There are two types of route entries, system route entries and custom route entries.

- System route entries

  After you create a VPC, the system automatically creates a system route. This route is used for cloud resources in the VPC to communicate with each other. After you create a VSwitch, the system also creates a system route. You cannot create or delete a system route.

- Custom route entries

  You can add a custom route to forward traffic to a specific next hop. For more information, see *Add a custom route entry*.

12.4.2 Add a custom route entry

This topic describes how to add a custom route entry. After you create a VPC, the system creates a default route table for the VPC and adds system routes to the VPC for traffic management. You cannot create or delete system routes. However, you can create custom routes to route traffic from specific CIDR blocks to specific destinations.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click Route Tables.
3. Select the region of the route table.
4. On the Route Tables page, find the route table to which you want to add a custom route entry, and click Manage in the Actions column.
5. In the Route Entry List section, click Add Route Entry.
6. In the Add Route Entry dialog box that appears, set the parameters, and click OK. The following table describes the parameters.

Parameter	Description
Name	The name of the route entry. The name must be 2 to 128 characters and can contain letters, digits, underscores (_), and hyphens (-). The name must start with a letter.

Parameter	Description
Destination CIDR Block	CIDR block to which you want to forward traffic.

Parameter	Description
Next Hop Type	The type of the next hop. Valid values: · ECS Instance: Traffic from the destination CIDR block is routed to the selected ECS instance. Select this type if you want to route traffic to an ECS instance for centralized forwarding and management. For example, configure an ECS instance as an Internet gateway to manage the access of other ECS instances to the Internet. · NAT Gateway: Traffic from the destination CIDR block is routed to the selected NAT Gateway. · Secondary NetworkInterface: Traffic from the destination CIDR block is routed to the selected secondary ENI. · Router Interface (To VPC): Traffic from the destination CIDR block is routed to the selected VPC. Select this type if you want to connect to a VPC by using Express Connect. · Router Interface (To VBR): Traffic from the destination CIDR block is routed to the router interface associated with a virtual border router (VBR). Select this type if you want to connect to a local IDC by using Express Connect. In this mode, you also need to select a routing mode: - General Routing: Select an associated router interface. - Active/Standby Routing: You can use only two instances as the next hop. The active router is weighted 100 and the standby router is weighted 0. The standby router interface takes over services only if the active router interface fails the health check. - Load Balancing: Select 2 to 4 router interfaces as the next hop. The peer router of the router interface must be a VBR. The valid instance weight is an integer ranging from 1 to 255. The default value is 100. The weight of the selected instances must be the same so that traffic can be evenly distributed to the next-hop instances.

Parameter	Description
ECS Instance/ NAT Gateway/ Secondary NetworkInterface/ VPC	Select the next-hop instance.

12.4.3 Modify a route table

After you create a route table, you can modify its name and description.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click Route Tables.
3. Select the region of the route table.
4. On the Route Tables page, find the route table, and click Manage in the Actions column.
5. In the Route Table Details section, click Edit next to Name. In the dialog box that appears, enter a new name for the route table, and click OK.

   The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), and hyphens (-). It must start with a letter.
6. Click Edit next to Description. In the dialog box that appears, enter a new description of the route table, and click OK.

   The description must be 2 to 256 characters in length and cannot start with `http://` or `https://`.

12.4.4 Export route entries

This topic describes how to export route entries from a route table for backup.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click Route Tables.
3. Select the region of the route table.
4. On the Route Tables page, find the route table, and click Manage in the Actions column.

5. In the Route Entry List section, click Export.

   The exported route entries are saved in a `.csv` file. You can view the exported route entries on a local PC.

12.4.5 Delete a custom route entry

This topic describes how to delete a custom route entry. You can delete custom route entries, but you cannot delete system route entries.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click Route Tables.
3. Select the region where the route table resides.
4. On the Route Tables page, find the route table, and click Manage in the Actions column.
5. In the Route Entry List section, find the route entry, and click Delete in the Actions column.
6. In the Delete Route Entry dialog box that appears, click Ok.

12.5 Elastic IP Addresses (EIPs)

12.5.1 EIP overview

An Elastic IP Address (EIP) is a public IP address that you can purchase and possess without purchasing an instance. You can associate an EIP with a VPC ECS instance, a VPC private SLB instance, a VPC ENI, or a NAT gateway.

An EIP is a NAT IP address. It is allocated to an Internet gateway of Apsara Stack and is mapped to the associated resource through NAT. After an EIP is associated with a cloud resource, the cloud resource can access the Internet by using the EIP.

EIP has the following benefits:

- Independent purchase and possession

  You can purchase an EIP independently as a resource of your account.
- Flexible association

  You can associate an EIP to a resource based on your business needs and release the EIP when the resource is no longer required.

- **Configurable network capabilities**

  You can change the bandwidth of an EIP at any time. After the change, the new bandwidth takes effect immediately.

12.5.2 Create an EIP

This topic describes how to create an EIP. An EIP is a public IP address that you can purchase and possess without purchasing an instance.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, choose Elastic IP Addresses > Elastic IP Addresses.**
3. **On the EIP page, click Create EIP.**
4. **On the Create Elastic IP Address page, set the parameters, and click Submit. The following table describes the parameters.**

Parameter	Description
Organization	The organization to which the EIP belongs.
Resource Set	The resource set to which the EIP belongs.
Region	The region to which the EIP belongs. The EIP must be in the same region as the cloud resources (ECS instances , NAT Gateways, SLB instances, and secondary ENIs) with which you want to associate the EIP.
Zone	The zone to which the EIP belongs.
Line Type	The line type of the EIP you want to create. You can only select BGP.
Peak Bandwidth	The peak bandwidth of the EIP.

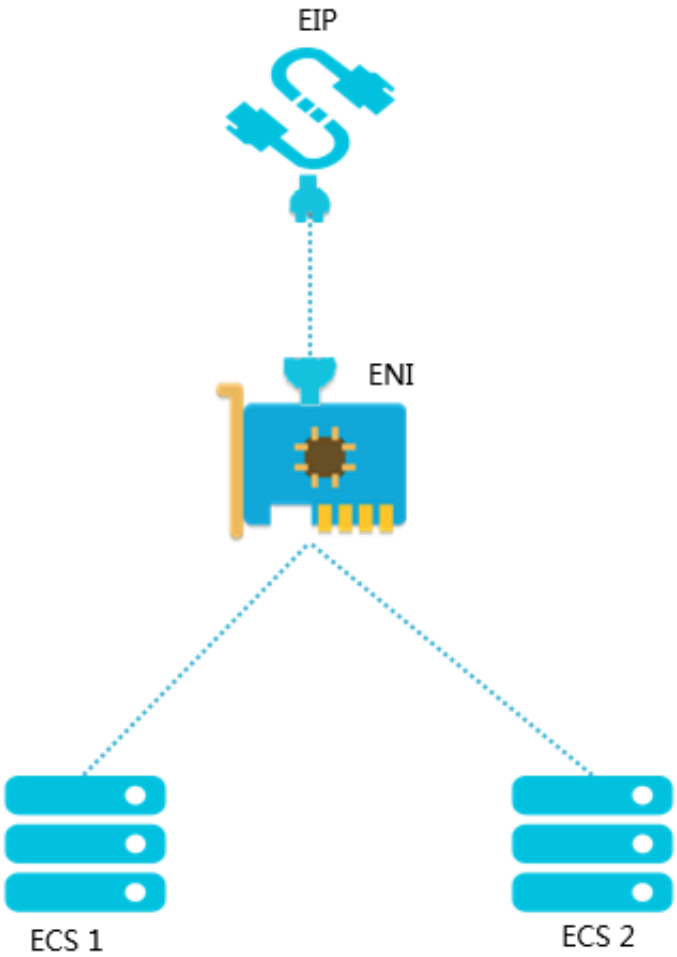
12.5.3 Associate an EIP with a cloud resource

12.5.3.1 Associate an EIP with a secondary ENI

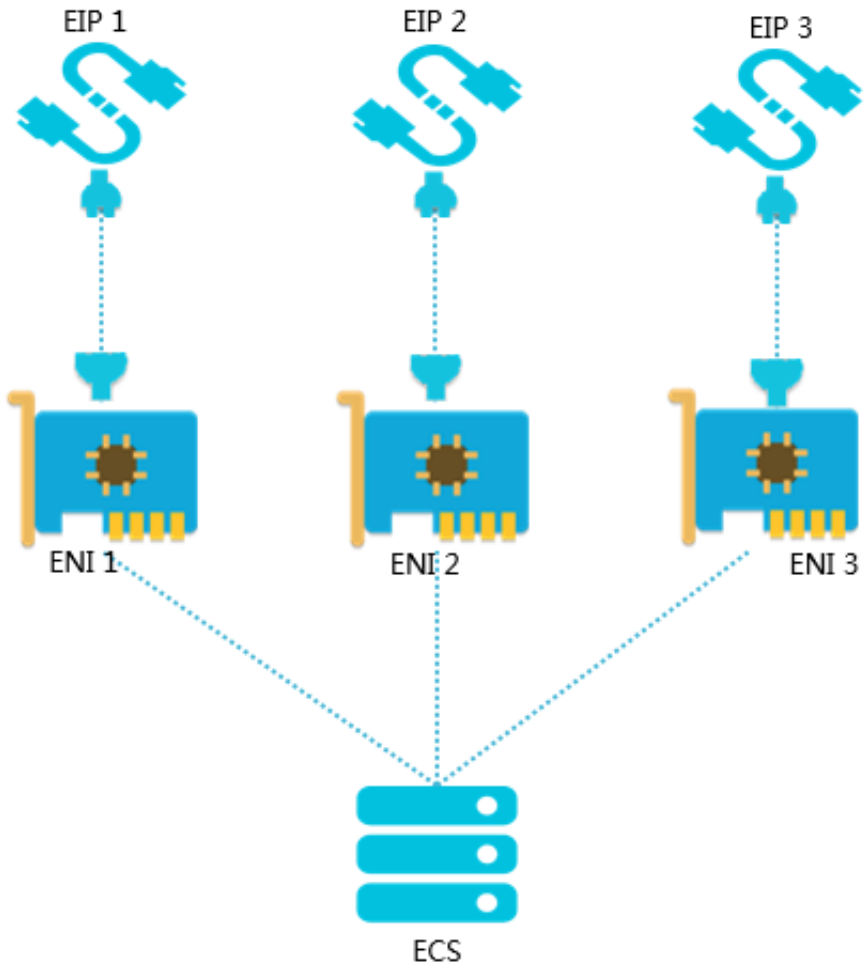
12.5.3.1.1 Overview

You can associate an EIP with an ENI. This allows you to build a more robust, flexible, and scalable IT solution. It also enables a single server to use multiple public IP addresses.

An ENI has a private IP address. After an ENI is associated with an Elastic IP address , the ENI has both a private IP address and a public IP address. When you move an ENI that is associated with an EIP from an Elastic Compute Service (ECS) instance to another ECS instance, the public IP address and private IP address are also migrated.



You can also associate multiple ENIs with an ECS instance and associate an Elastic IP address with each ENI so that the ECS instance has multiple public IP addresses. You can use these public IP addresses to provide external services with correspond ing security group rules.



Association modes

You can associate an EIP with an ENI in either of the following modes:

- **NAT mode**
- **Cut-through mode**

The following table lists differences between the two association modes.

Item	NAT mode	Cut-through mode
Whether the EIP is visible on the ENI in the operating system	**No**	**Yes** **Note:** You can run the `ifconfig` or `ipconfig` to obtain the public IP address of the ENI.
Type of ENIs that can be associated with EIPs	**Primary and secondary ENIs**	**Secondary ENIs**

Item	NAT mode	Cut-through mode
The maximum number of EIPs that a primary ENI can be associated with	1	The primary ENI cannot be associated.
The maximum number of EIPs that a secondary ENI can be associated with	Depends on the number of private IP addresses of the secondary ENI. Note: EIPs are in one-to-one mapping with the private IP addresses of the secondary ENI. If a secondary ENI has 10 private IP addresses, a maximum of 10 EIPs can be associated with the ENI.	1 Note: In the cut-through mode, an EIP can only be associated with the primary private IP address of the secondary ENI.
Availability of the private network function of the secondary ENI associated with an EIP	Yes	No
Supported protocols	When an EIP is deployed as a NAT ALG, protocols such as H.323, SIP, DNS, and RTSP are not supported.	EIPs support all IP protocols such as FTP, H.323, SIP, DNS, RTSP, and TFTP.

12.5.3.1.2 Configure the NAT mode

In the NAT mode, both the private and public IP addresses of the ENI are available, and the EIP of the ENI is invisible.

Prerequisites

Before you configure the NAT mode, ensure that the following conditions are met:

- **The network type of the secondary ENI is VPC.**
- **The secondary ENI and the EIP reside in the same region.**

- The secondary ENI is not associated with any ECS instance.

  If the secondary ENI has been associated with an ECS instance, disassociate it from the ECS instance. After the NAT mode is configured, associate it to the ECS instance again.

Context

The NAT mode has the following characteristics:

- The number of EIPs that a secondary ENI can be associated is determined by the number of private IP addresses of the secondary ENI.
- The EIP is associated with the ENI in NAT mode. The private and public IP addresses of the ENI are available at the same time.
- The EIP is invisible to the operating system. You need to call the DescribeEipAddresses operation to query the public IP address that is associated with the ENI. For more information, see "DescribeEipAddresses" in *Development Guide* .
- When an EIP is deployed as a NAT ALG, protocols such as H.323, SIP, DNS, and RTSP are not supported.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, choose Elastic IP Addresses > Elastic IP Addresses.
3. Select the region where the EIP resides.
4. On the Elastic IP Addresses page, find the EIP and click Bind in the Actions column.
5. In the Bind Elastic IP Address dialog box that appears, set the parameters and click OK. The following table describes the parameters.

Parameter	Description
Instance Type	The type of the instance. Select Secondary ENI.
Mode	The association mode. Select NAT mode.
Secondary ENI	The secondary ENI with which you want to associate the EIP.

12.5.3.1.3 Configure the cut-through mode

In the cut-through mode, the EIP replaces the private IP address of the secondary Elastic Network Interface (ENI) and the secondary ENI becomes a pure Internet NIC. You can view the EIP in the NIC information of the operating system.

Prerequisites

Before you configure the cut-through mode, ensure that the following conditions are met:

- **The network type of the secondary ENI is VPC.**
- **The secondary ENI and the EIP reside in the same region.**
- **The secondary ENI is not associated with any ECS instance.**

  If the secondary ENI has been associated with an ECS instance, disassociate it from the ECS instance. After the cut-through mode is configured, associate it to the ECS instance again.
- **Each secondary ENI can be associated with only one EIP.**
- **DHCP is enabled in the ECS system image.**

Context

EIPs are NAT IP addresses. In the NAT mode, the public IP address is assigned to the gateway device rather than the NIC of the ECS instance. In the operating system of the ECS instance, the private IP address of the NIC is visible, but the public IP address is invisible. You must record the mapping between the network interfaces or servers and the public IP addresses. This is inconvenient for operation and maintenance. Additionally, when an EIP is deployed as a NAT application layer gateway (NAT ALG), protocols such as H.323, SIP, DNS, RTSP, and TFTP are not supported.

Cut-through mode makes the EIP visible on the NIC and solves these problems. In the cut-through mode:

- **The EIP replaces the private IP address of the ENI. The ENI becomes a pure Internet NIC and its private network functions are not available.**
- **You can obtain the EIP on the ENI in the operating system and run the `ifconfig` or `ipconfig` command to obtain the public IP address of the ENI.**
- **EIPs support all IP protocols such as FTP, H.323, SIP, DNS, RTSP, and TFTP.**

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, choose Elastic IP Addresses > Elastic IP Addresses.
3. Select the region where the EIP resides.
4. On the Elastic IP Addresses page, find the EIP and click Bind in the Actions column.
5. In the Bind Elastic IP Address dialog box that appears, set the parameters and click OK. The following table describes the parameters.

Parameter	Description
Instance Type	The type of the instance. Select Secondary ENI.
Mode	The association mode. Select Cut-Through Mode.
Secondary ENI	The secondary ENI with which you want to associate the EIP.

12.5.3.2 Associate an EIP with a NAT gateway

This topic describes how to associate an EIP with a NAT gateway. After you associate an EIP with a NAT gateway, you can use the EIP to configure DNAT and SNAT entries.

Prerequisites

You have created a NAT gateway. For more information, see *Create a NAT gateway*.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, choose Elastic IP Addresses > Elastic IP Addresses.
3. Select the region where the EIP resides.
4. On the Elastic IP Addresses page, find the EIP and click Bind in the Actions column.
5. In the Bind Elastic IP Address dialog box that appears, set the parameters and click OK. The following table describes the parameters.

Parameter	Description
Instance Type	The type of the instance. Select NAT Gateway Instance.

Parameter	Description
NAT Gateway Instance	The NAT gateway instance with which you want to associate the EIP. The NAT gateway and the EIP must be in the same region.

12.5.3.3 Associate an EIP with an ECS instance

This topic describes how to associate an EIP with a VPC ECS instance. After an ECS instance is associated with an EIP, the ECS instance can communicate with the Internet.

Prerequisites

An ECS instance is created. For more information, see Create an instance under Quick start in *ECS User Guide* **.**

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, choose Elastic IP Addresses > Elastic IP Addresses.**
3. **Select the region where the EIP resides.**
4. **On the Elastic IP Addresses page, find the EIP and click Bind in the Actions column.**
5. **In the Bind Elastic IP Address dialog box that appears, set the parameters and click OK. The following table describes the parameters.**

Parameter	Description
Instance Type	The type of the instance. Select ECS Instance.

Parameter	Description
ECS Instance	The ECS instance with which you want to associate the EIP. The ECS instance to be associated must meet the following requirements: · The network type of ECS instance is VPC. · The ECS instance and the EIP reside in the same region. · The ECS instance is in the Running or Stopped state. · The ECS instance is not associated with a public IP address or other EIPs. · An ECS instance can be associated with only one EIP.

12.5.3.4 Associate an EIP with an SLB instance

This topic describes how to associate an Elastic IP address (EIP) to a Server Load Balancer (SLB) instance. After you associate an Elastic IP address with an SLB instance, the SLB instance can forward requests from the Internet.

Prerequisites

An SLB instance is created. For more information, see Create an SLB instance under Quick start in *SLB User Guide* .

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, choose Elastic IP Addresses > Elastic IP Addresses.
3. Select the region where the EIP resides.
4. On the Elastic IP Addresses page, find the EIP and click Bind in the Actions column.
5. In the Bind Elastic IP Address dialog box that appears, set the parameters and click OK. The following table describes the parameters.

Parameter	Description
Instance Type	The type of the instance. Select SLB Instance.

Parameter	Description
SLB Instance	The SLB instance with which you want to associate the EIP. The SLB instance must meet the following requirements: · The network type of the SLB instance is VPC. · The SLB instance and the EIP reside in the same region. · Each SLB instance can be associated with only one EIP.

12.5.4 Upgrade an EIP

This topic describes how to upgrade the peak bandwidth of an EIP. After an upgrade, the new peak bandwidth takes effect immediately.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, choose Elastic IP Addresses > Elastic IP Addresses.
3. Select the region where the EIP resides.
4. On the Elastic IP Addresses page, find the target EIP and choose More > Upgrade in the Actions column.
5. In the Change Specifications dialog box that appears, enter a new peak bandwidth and click Submit.

   The peak bandwidth cannot exceed 1,000 Mbit/s.

12.5.5 Disassociate an EIP

You can unbind an EIP from a cloud resource when the cloud resource does not need to communicate with the Internet.

Prerequisites

The DNAT or SNAT entries that have been added to the NAT gateway associated with the EIP are deleted. For more information, see *Delete a DNAT entry* and *Delete a SNAT entry*.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, choose Elastic IP Addresses > Elastic IP Addresses.

3. Select the region where the EIP resides.
4. On the Elastic IP Addresses page, find the EIP and click Unbind in the Actions column.
5. In the Unbind Elastic IP Address dialog box that appears, click OK.

12.5.6 Release an EIP

This topic describes how to release an Elastic IP address (EIP).

Prerequisites

The EIP is disassociated from the corresponding cloud resource. For more information, see *Disassociate an EIP*.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, choose Elastic IP Addresses > Elastic IP Addresses.
3. Select the region where the EIP resides.
4. On the Elastic IP Addresses page, find the EIP and choose More > Release in the Actions column.
5. In the Release Elastic IP dialog box that appears, click OK.

12.6 NAT gateways

12.6.1 NAT gateway overview

NAT Gateway is an enterprise gateway for communication between VPCs and the Internet. It provides NAT proxy services (SNAT and DNAT), up to 10 Gbit/s forwarding capacity, and cross-zone disaster tolerance.

Features

Each NAT gateway must be associated with a public IP address. After you create a NAT gateway, you can associate it with an Elastic IP address (EIP).

NAT Gateway supports SNAT and DNAT.

- SNAT allows an ECS instance without a public IP address to access the Internet.
- DNAT maps the public IP address of a NAT Gateway to an ECS instance so that the instance can be assessed from the Internet.

Benefits

NAT Gateway has the following benefits:

- **High performance: NAT Gateway provides a forwarding capacity of up to 10 Gbit/s for each instance.**
- **High availability: Based on the SDN technology, NAT Gateway uses a distributed architecture that allows you to deploy multiple instances across zones. Each instance can take over services upon failures in a certain zone.**
- **Flexible configurations: You can change the instance specifications, bandwidth, and number of public IP addresses of NAT Gateway at any time to meet your business needs.**

12.6.2 Manage a NAT gateway

12.6.2.1 Create a NAT gateway

A NAT gateway is an enterprise-class gateway that provides NAT proxy services. Before you configure SNAT and DNAT rules, you must create a NAT gateway.

Prerequisites

A VPC and a VSwitch are created. For more information, see *Create a VPC* and *Create a VSwitch*.

Procedure

1. ***Log on to the VPC console*.**
2. **In the left-side navigation pane, click NAT Gateways.**
3. **On the NAT Gateways page, click Create NAT Gateway.**
4. **Set the parameters and click Submit. The following table describes the parameters.**

Parameter	Description
Organization	The organization to which the NAT gateway belongs.
Resource Set	The resource set to which the NAT gateway belongs.
Region	The region to which the NAT gateway belongs.

Parameter	Description
VPC	The VPC to which the NAT gateway belongs. If you cannot find the VPC in the VPC list, perform the following operations: · Check whether the VPC is configured with a NAT gateway. Only one NAT gateway can be configured for each VPC. · Check whether the VPC has a custom route entry with destination CIDR block set to 0.0.0.0/0. If there is such a route entry, delete the custom route. · Check whether the RAM user is authorized to read and access the VPC. If the RAM user is not authorized, contact your Alibaba Cloud account owner to grant permissions.
Type	The specifications of the NAT gateway. Values: · Small: The maximum number of SNAT connections is 10,000. · Medium: The maximum number of SNAT connections is 50,000. · Large: The maximum number of SNAT connections is 200,000. · Super Large: The maximum number of SNAT connections is 1 million. Note: The specifications of a NAT gateway affects the maximum number of SNAT connections, but do not affect the number of DNAT connections.
Name	The name of the NAT gateway. The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.

12.6.2.2 Modify a NAT gateway

This topic describes how to modify the name and description of a NAT gateway.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click NAT Gateways.
3. Select the region where the NAT gateway resides.
4. On the NAT Gateways page, find the NAT gateway, and click Manage in the Actions column.
5. On the NAT Gateway Details page, click Edit next to Name. In the dialog box that appears, enter a new name for the NAT gateway, and click OK.

   The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), and hyphens (-). It must start with a letter.
6. Click Edit next to Description. In the dialog box that appears, enter a new description of the NAT gateway, and click OK.

   The description must be 2 to 256 characters in length and cannot start with `http://` or `https://`.

12.6.2.3 Delete a NAT gateway

This topic describes how to delete a NAT gateway.

Prerequisites

Before you delete a NAT gateway, the following conditions must be met:

- The EIP associated with the NAT gateway is disassociated. For more information, see *Disassociate an EIP from a NAT gateway*.
- The DNAT entry created in the DNAT table is deleted. For more information, see *Delete a DNAT entry*.
- The SNAT entry created in the SNAT table is deleted. For more information, see *Delete a SNAT entry*.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click NAT Gateways.
3. Select the region where the NAT gateway resides.

4. On the NAT Gateways page, find the NAT gateway, and choose More > Delete in the Actions column.
5. In the Delete Gateway dialog box that appears, click OK.

   After you select Delete (Delete NAT gateway and resources), the NAT gateway and its DNAT and SNAT entries are deleted, and the EIP associated with the NAT gateway is disassociated.

12.6.3 Manage a DNAT table

12.6.3.1 DNAT table overview

You can use the DNAT function to map a public IP address on the NAT Gateway to a private IP address in the CIDR block of the target ECS instance, so that the ECS instance can provide Internet services.

DNAT entries

You can create DNAT entries in a DNAT table to implement port-based forwarding. After a DNAT entry is configured, data packets received by the specified public IP address are forwarded to the ECS instance according to the custom mapping rule.

Each DNAT entry consists of the following parts:

- **Public IP: The Elastic IP Address (EIP) associated with the NAT Gateway.**
- **Private IP Address: The private IP address of the ECS instance in the VPC.**
- **Public Port: The public port used for forwarding traffic.**
- **Private Port: The private port used for forwarding traffic.**
- **IP Protocol: The protocol type of the port used for forwarding traffic.**

Port mapping and IP mapping

The DNAT function includes port mapping and IP mapping:

- **Configure port mapping**

  If port mapping is enabled, the NAT Gateway forwards requests from the specified protocol and port to the specified port of the target ECS instance in the VPC. For example,

DNAT entry	Public IP address	Public port	Private IP address	Private port	Protocol
Entry 1	139.224.xx.xx	80	192.168.x.x	80	TCP

DNAT entry	Public IP address	Public port	Private IP address	Private port	Protocol
Entry 2	139.224.xx.xx	8080	192.168.x.x	8000	UDP

Entry 1: The NAT Gateway forwards traffic that is destined for 139.224.xx.xx through port 80 to port 80 of the ECS instance whose private IP address is 192.168.x.x.

Entry 2: The NAT Gateway forwards traffic that is destined for 39.224.xx.xx through port 8080 to port 8000 of the ECS instance whose private IP address is 192.168.x.x.

- IP mapping

  If IP mapping is enabled, the NAT Gateway forwards all traffic destined for the public IP address to the target ECS instance. The following is an example:

DNAT entry	Public IP address	Public port	Private IP address	Private port	Protocol
Entry 3	139.224.xx.xx	Any	192.168.x.x	Any	Any

Entry 3: The NAT Gateway forwards all requests destined for 139.224.xx.xx to the ECS instance whose private IP address is 192.168.x.x.

12.6.3.2 Create a DNAT entry

This topic describes how to create a destination network address translation (DNAT) entry. NAT Gateway supports DNAT. DNAT maps a public IP address to an ECS instance in a VPC so that the ECS instance can provide services over the Internet. DNAT supports port mapping and IP mapping.

Prerequisites

A NAT gateway is created and associated with an Elastic IP address (EIP). For more information, see *Create a NAT gateway* and *Associate an EIP with a NAT gateway*.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click NAT Gateways.
3. Select the region where the NAT gateway resides.

4. On the NAT Gateways page, find the NAT gateway, and click Configure DNAT in the Actions column.
5. On the DNAT Table page, click Create DNAT Entry.
6. In the Create DNAT Entry dialog box that appears, set the parameters, and click OK. The following table describes the parameters.

Configuration	Description
Public IP	Select a public IP address.  Note: An IP address that is already being used in an SNAT entry cannot be selected.
Private IP	Select the private IP address of the ECS instance to access the Internet. You can specify the private IP address in the following ways: · Auto Fill: Select an ECS instance from the ECS instance or ENI list. · Manually Input: Enter the private IP address that you want to map.  Note: It must be within the private CIDR block of the VPC. You can also enter an existing private IP address of the ECS instance.
Port Settings	DNAT supports IP mapping and port mapping. Select a mapping method: · All Ports: Select this option to configure IP mapping. This is the same as associating an EIP with the ECS instance. If this method is used, all requests destined for the public IP address are directed to the ECS instance. · Specific Port: Select this option to configure port mapping. After this method is used, NAT Gateway forwards the requests from the specified protocol and port to the specified port of the ECS instance. After you select Specific Port, enter Public Port (the external port used for traffic forwarding), Private Port (the internal port used for traffic forwarding) and the IP Protocol (the protocol type of the port).

12.6.3.3 Modify a DNAT entry

This topic describes how to modify a DNAT entry. You can modify the public IP address, private IP address, port, and name of a DNAT entry.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, click NAT Gateways.**
3. **Select the region where the NAT gateway resides.**
4. **On the NAT Gateways page, find the NAT Gateway, and click Configure DNAT in the Actions column.**
5. **On the DNAT table page, find the DNAT entry, and click Edit in the Actions column.**
6. **In the Edit DNAT Entry dialog box that appears, change the public IP address, private IP address, port settings, and name of the DNAT entry, and click OK.**

12.6.3.4 Delete a DNAT entry

This topic describes how to delete a DNAT entry.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, select NAT Gateways.**
3. **Select the region where the NAT gateway resides.**
4. **On the NAT Gateways page, find the NAT Gateway, and click Configure DNAT in the Actions column.**
5. **On the DNAT table page, find the DNAT entry, and click Remove in the Actions column.**
6. **In the Delete DNAT Entry dialog box that appears, click OK.**

12.6.4 Manage a SNAT table

12.6.4.1 SNAT table overview

NAT Gateways support the SNAT function. This function allows ECS instances that are not associated with public IP addresses in a VPC to access the Internet.

SNAT entries

You can create SNAT entries in a source network address translation (SNAT) table to allow ECS instances to access the Internet.

A SNAT entry consists of the following parts:

- VSwitch or ECS instance: the VSwitch or ECS instance that requires the SNAT proxy service.
- Public IP address: the public IP address used to access the Internet.

> **Note:**
> You can select multiple public IP addresses to build a SNAT IP address pool. When an ECS instance in a VPC initiates a request to access the Internet, the ECS instance randomly uses a public IP address in the SNAT address pool.

VSwitch granularity and ECS granularity

The SNAT feature provides the following two granularities to enable ECS instances in a VPC to access the Internet.

- VSwitch granularity

  If you select VSwitch granularity to create a SNAT entry, the NAT Gateway provides the Internet proxy service for an ECS instance in the specified VSwitch when the ECS instance initiates an Internet access request. In this way, the ECS instance can use the specified public IP address to access the Internet. By default, all ECS instances in the VSwitch can use the specified public IP address to access the Internet.

  > **Note:**
  > If an ECS instance has a public IP address (for example, a fixed public IP address is assigned, an EIP is associated, or DNAT IP mapping is configured) and initiates an Internet access request, the instance preferentially accesses the Internet by using the public IP address instead of the SNAT feature of NAT Gateway.

- ECS granularity

  If you select VSwitch granularity to create a SNAT entry, the specified ECS instance uses the specified public IP address to access the Internet. When the ECS instance initiates an Internet access request, the NAT Gateway provides the Internet proxy service for the ECS instance.

12.6.4.2 Create a SNAT entry

The SNAT function of NAT Gateway allows ECS instances without public IP addresses in a VPC to access the Internet.

Prerequisites

Ensure that the following conditions are met when you create a SNAT entry:

- **A NAT gateway is created and associated with an Elastic IP address (EIP). For more information, see *Create a NAT gateway* and *Associate an EIP with a NAT gateway*.**
- **A VSwitch has been created in the VPC associated with the NAT gateway if you want to create a SNAT entry with VSwitch granularity. For more information, see *Create a VSwitch*.**
- **If you want to create a SNAT entry with ECS granularity, ensure that an ECS instance has been created in the VPC associated with the NAT gateway. For more information, *see* "Create an instance" in ECS User Guide.**

Procedure

1. ***Log on to the VPC console*.**
2. **In the left-side navigation pane, click NAT Gateways.**
3. **Select the region where the NAT gateway resides.**
4. **On the NAT Gateways page, find the NAT gateway, and click Configure SNAT in the Actions column.**
5. **On the SNAT Table page, click Create SNAT Entry.**
6. **In the Create SNAT Entry dialog box that appears, set the parameters, and clickOK. The following table describes the parameters.**

Parameter	Description
VSwitch Granularity	
VSwitch	**Select a VSwitch in the VPC. All ECS instances under this VSwitch can access the Internet by using the SNAT feature.** **Note:** **If an ECS instance has a public IP address (for example, a fixed public IP address is allocated, an EIP is associated, or DNAT IP mapping is configured) and initiates Internet access, the instance preferentially accesses the Internet by using the public IP address instead of the SNAT function of NAT Gateway.**

Parameter	Description
VSwitch CIDR Block	Displays the CIDR block of the selected VSwitch.
Public IP	Select the public IP address that is used to access the Internet. You can select multiple public IP addresses to build a SNAT IP address pool. Note: A public IP address that is used in a DNAT entry cannot be used to create a SNAT entry.
Entry Name	Enter a name for the SNAT entry. The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), and hyphens (-). It must start with a letter.
ECS Granularity	
Available ECS Instances	Select an ECS instance in the VPC. The ECS instance accesses the Internet by using the configured public IP address. Ensure that the following conditions are met: · The ECS instance is in the running state. · The ECS instance is not associated with other EIPs or with any fixed public IP address.
ECS CIDR Block	Displays the CIDR block of the ECS instance.
Public IP	Select the public IP address that is used to access the Internet. You can select multiple public IP addresses to build a SNAT IP address pool. Note: A public IP address that is used in a DNAT entry cannot be used to create a SNAT entry.
Entry Name	Enter a name for the SNAT entry. The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), and hyphens (-). It must start with a letter.

12.6.4.3 Modify a SNAT entry

This topic describes how to change the public IP address and name of a SNAT entry.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, click NAT Gateways.**
3. **Select the region where the NAT gateway resides.**
4. **On the NAT Gateways page, find the NAT gateway, and click Configure SNAT in the Actions column.**
5. **On the SNAT table page, find the SNAT entry, and click Edit in the Actions column.**
6. **In the Edit SNAT Entry dialog box that appears, change the public IP address and name of the SNAT entry, and click OK.**

12.6.4.4 Delete a SNAT entry

This topic describes how to delete a SNAT entry.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, click NAT Gateways.**
3. **Select the region where the NAT gateway resides.**
4. **On the NAT Gateways page, find the NAT gateway, and click Configure SNAT in the Actions column.**
5. **On the SNAT table page, find the SNAT entry, and click Remove in the Actions column.**
6. **In the Delete SNAT Entry dialog box that appears, click OK.**

12.6.5 Manage an EIP

12.6.5.1 Associate an EIP with a NAT gateway

This topic describes how to associate an Elastic IP address (EIP) with a NAT gateway. Each NAT gateway must be associated with a public IP address. After you create a NAT gateway, you can associate it with an EIP.

Prerequisites

A NAT gateway and an EIP are created. For more information, see *Create a NAT gateway* **and** *Create an EIP*.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, select NAT Gateways.
3. Select the region where the NAT gateway resides.
4. On the NAT Gateways page, find the NAT gateway, and choose More > Bind Elastic IP Address in the Actions column.
5. In the Bind Elastic IP Address dialog box that appears, set the parameters and click OK. The following table describes the parameters.

Parameter	Description
Select from EIP list	
Usable EIP list	Select the EIP that is used to access the Internet.
Allocate one EIP and associate it to a NAT gateway	
Buy EIP	Indicates the number of EIPs to be purchased. The default value is 1, which cannot be changed. The system creates a pay-as-you-go EIP and associate it with the NAT gateway.

12.6.5.2 Disassociate an EIP from a NAT gateway

This topic describes how to disassociate an EIP from a NAT gateway.

Prerequisites

The EIP that you want to disassociate is not used by any SNAT or DNAT entry.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, select NAT Gateways.
3. Select the region where the NAT gateway resides.
4. On the NAT Gateways page, find the NAT gateway, and choose More > Unbind Elastic IP Address in the Actions column.
5. In the Unbind Elastic IP Address dialog box that appears, select the EIP that you want to disassociate, and click OK.

12.6.6 View monitoring data

CloudMonitor allows you to view monitoring data of NAT gateways, such as the number of SNAT connections, cumulative dropped connections per second due

to maximum limit, and cumulative dropped connections per second due to new connection limit.

Procedure

1. *Log on to the VPC console*
2. In the left-side navigation pane, click NAT Gateways.
3. Select the region where the NAT gateway resides.
4. On the NAT Gateways page, find the NAT gateway, and click  in the Monitoring column.

   The following table describes the monitoring metrics of NAT gateways.

Item	Description	Dimension	Unit	Minimum monitoring granularity
SNAT connections	The number of SNAT connections of a NAT Gateway instance.	Instance	Count/Min	30s

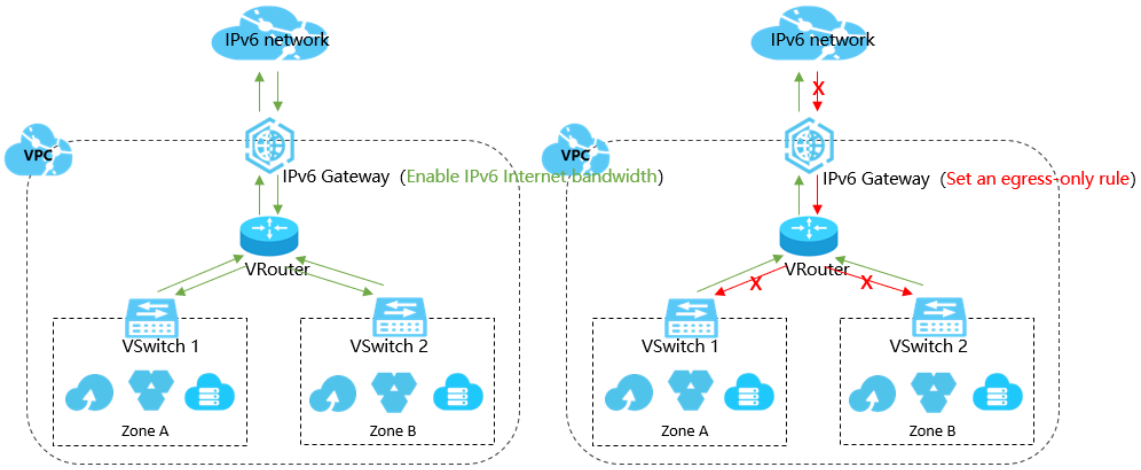
Item	Description	Dimension	Unit	Minimum monitoring granularity
Capacity Limit discarded connections	The maximum number of SNAT connections vary according to the NAT Gateway specification. Capacity limit discarded connections indicate the SNAT connections that are dropped when the number of connections to the instance exceeds the maximum number of SNAT connections corresponding to the specification of the instance. Note: This metric is an accumulated value and will not be reset. · If the number of capacity limit discarded connections increase continuously during a certain period of time, we recommend that you upgrade the specification of NAT Gateway. · If a horizontal line is displayed during a certain period of time, it indicates that no packets were dropped during this time period.	Instance	Count/Min	30s

Item	Description	Dimension	Unit	Minimum monitoring granularity
Speed limit discarded connections	The maximum number of SNAT connections per second vary according to the NAT Gateway specification. Speed limit discarded connections indicate the number of SNAT connections that are dropped when the number of SNAT connections to the instance per second exceeds the maximum number of SNAT connections per second corresponding to the specification of the instance. Note: This metric is an accumulated value and will not be reset. · If the number of speed limit discarded connections increase continuously during a certain period of time, we recommend that you upgrade the specification of the NAT Gateway. · If a horizontal line is displayed during a certain period of time, it indicates that no packets were dropped during this time period.	Instance	Count/Min	30s

12.7 IPv6 gateways

12.7.1 IPv6 gateway overview

This topic provides an overview of the IPv6 Gateways of Virtual Private Cloud (VPC). An IPv6 Gateway functions as an IPv6 traffic gateway for a VPC. You can configure the IPv6 Internet bandwidth and egress-only rules to manage the inbound and outbound IPv6 traffic.



Functions

The functions of an IPv6 gateway are as follows:

- **IPv6 internal network communication**

  By default, an IPv6 address in a VPC is allocated with an Internet bandwidth of 0 Mbit/s and only supports communication over the internal network. Specifically, the cloud instances in a VPC can only access other IPv6 addresses in the same VPC through the IPv6 address. The resources cannot access the Internet with these IPv6 addresses or be accessed by IPv6 clients over the Internet.

- **IPv6 public network communication**

  You can purchase an Internet bandwidth for the IPv6 address for which you have applied. In this way, the resources in the VPC can access the Internet through the IPv6 address and be accessed by IPv6 clients over the Internet.

  You can set the Internet bandwidth to 0 Mbit/s at any time to deny the IPv6 address Internet access. After this configuration, the IPv6 address can only communicate over the internal network.

- **IPv6 public network communication with an egress-only rule**

  You can set an egress-only rule for an IPv6 Gateway. In this way, the IPv6 address can access the Internet, but IPv6 clients are denied access to your cloud resources in the VPC over the Internet.

  You can delete the egress-only rule at any time. After the rule is deleted, your resources in the VPC can access the Internet through the IPv6 address for which you have purchased Internet bandwidth, and IPv6 clients can access the resources in the VPC over the Internet.

The network access capability of IPv6 addresses is dependent on the settings of the network type, Internet bandwidth, and egress-only rule, as shown in the following table.

IPv6 network type	Enable IPv6 Internet bandwidth?	Set an egress-only rule?	IPv6 network access capability
Internal network	No	No	Internal network communication
Public network	Yes	No	Internal network communication Public network communication
		Yes	Internal network communication Public network communication when access is initiated by VPCs

Benefits

IPv6 Gateway provides the following benefits:

- **High availability**

  IPv6 Gateways provide cross-zone high availability and stable IPv6 Internet gateway services.
- **High performance**

  A single IPv6 Gateway provides a 10-gigabit level throughput.
- **Flexible management of public network communication**

  You can manage the Internet communication capability of an IPv6 Gateway by adjusting its Internet bandwidth and setting an egress-only rule.

12.7.2 Enable an IPv6 CIDR block for a VPC

12.7.2.1 Create an IPv4 and IPv6 dual-stack VPC

This topic describes how to create a dual-stack VPC that supports both IPv4 and IPv6 addresses. When you create a VPC, you can configure both IPv4 and IPv6 CIDR blocks. By default, IPv4 CIDR blocks are enabled and cannot be removed. However, you can select whether to enable IPv6 CIDR blocks. After IPv6 CIDR blocks are enabled, the system creates an IPv6 Gateway free of charge for the VPC so that you can manage the Internet bandwidth and traffic of IPv6 CIDR blocks.

Procedure

1. *Log on to the VPC console*.
2. **On the VPCs page, click Create VPC.**
3. **In the Create VPC dialog box that appears, set the parameters and click Submit. The following table describes the parameters.**

Parameter	Description
Organization	The organization to which the VPC belongs.
Resource Set	The resource set to which the VPC belongs.
Region	The region where your VPC network is deployed.
Share with Sub-organizations	Indicates whether to share the VPC. If you select Yes, administrators of sub-organizations can create resources in the VPC.

Parameter	Description
VPC Name	The name of the VPC. The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.
IPv4 CIDR Block	The IPv4 CIDR block of the VPC. You can specify an IPv4 CIDR block in either of the following ways: · Recommended CIDR Block: You can use one of the three standard IPv4 CIDR blocks: 192.168.0.0/16, 172.16.0.0/12, and 10.0.0.0/8. · Custom CIDR Block: You can use 192.168.0.0/16, 172.16.0.0/12, 10.0.0.0/8, and their subnets as the CIDR blocks. The allowed block size is between a /24 netmask and /8 netmask.. Example: 192.168.0.0/16. Note: After you create a VPC, you cannot change its IPv4 CIDR block.
IPv6 CIDR Block	Indicates whether to assign an IPv6 CIDR block. In this example, select Assign. If you set this parameter to Assign, the system automatically creates a free IPv6 gateway for your VPC, and assigns an IPv6 CIDR block with netmask /56, such as 2xx1: db8::/56. IPv6 addresses only support communication in private networks by default. If you want to enable the VPC to access the Internet by using an IPv6 address or be accessed by an IPv6 client on the Internet, you must enable IPv6 Internet bandwidth. For more information, see *Enable Internet bandwidth for an IPv6 address*. Note: After you create a VPC, you cannot change its IPv6 CIDR block.

Parameter	Description
Description	Enter a description of the VPC. The description must be 2 to 256 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.

12.7.2.2 Enable an IPv6 CIDR block for a VPC

This topic describes how to configure an IPv6 CIDR block for an existing VPC. After you enable IPv6 for a VPC, the system automatically creates an IPv6 Gateway of the Free Version specification for the VPC. You can use the IPv6 Gateway to manage IPv6 Internet bandwidth and set egress-only rules.

Procedure

1. *Log on to the VPC console*.
2. Select the region where your VPC resides.
3. On the VPCs page, find the target VPC, and click Enable IPv6 CIDR Block in the IPv6 CIDR Block column.
4. In the dialog box that appears, select Enable IPv6 CIDR Block of all VSwitches in VPC and click OK.

   If you do not select Enable IPv6 CIDR Block of all VSwitches in VPC, you must enable an IPv6 CIDR block for each VSwitch. For more information, see *Enable an IPv6 CIDR block for a VSwitch*.

12.7.3 Enable an IPv6 CIDR block for a VSwitch

12.7.3.1 Create an IPv4 and IPv6 dual-stack VSwitch

This topic describes how to create a dual-stack VSwitch that supports both IPv4 and IPv6 addressing. When you create a VSwitch, you can enable an IPv6 CIDR block for it.

Prerequisites

An IPv6 CIDR block is enabled for the VPC to which the VSwitch belongs. For more information, see *Enable an IPv6 CIDR block for a VPC*.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click VSwitches.
3. Select the region of the VPC to which the VSwitch belongs.
4. On the VSwitches page, click Create VSwitch.
5. On the VSwitch page that appears, set the parameters and click Submit. The following table describes the parameters.

Parameter	Description
Organization	The organization to which the VSwitch belongs.
Resource Set	The resource set to which the VSwitch belongs.
Region	The region to which the VSwitch belongs.
Zone	The zone to which the VSwitch belongs. In a VPC, a VSwitch can belong to only one zone. However, you can deploy cloud resources on VSwitches of different zones to achieve cross-zone disaster recovery. **Note:** You can add a cloud resource instance to only one VSwitch.
VPC	The VPC for which you want to create the VSwitch.
VSwitch Name	The name of the VSwitch. The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.

Parameter	Description
IPv4 CIDR Block	The IPv4 CIDR block of the VSwitch. · You must specify the IP address segment for the VSwitch in the form of a CIDR block. The allowed block size is between a /29 netmask and /16 netmask. This means the VSwitch can provide 8 to 65,536 IP addresses. · The CIDR block of the 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 VSwitch can be created. · The first and last three IP addresses of a VSwitch are reserved . For example, If the CIDR block of a VSwitch is 192.168.1.0/24 , the IP addresses 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved. · The CIDR block of the VSwitch cannot be the same as the destination CIDR block of a route entry of the VPC where the VSwitch resides. However, the CIDR block of the VSwitch can be a subset of the destination CIDR block of the route entry. · After you create a VSwitch, you cannot modify its CIDR block.
IPv6 CIDR Block	The IPv6 CIDR block of the VSwitch. The mask for the IPv6 CIDR block of a VSwitch is /64 by default. You can enter a number ranging from 0 to 255 to define the last 8 bits of the IPv6 CIDR block. For example, if the IPv6 CIDR block of the VPC is 2xx1: db8::/64, you can enter 255 (ff in hexadecimal notation) for the IPv6 CIDR block of the VSwitch. Then, the IPv6 CIDR block of the VSwitch is 2xx1: db8: ff::/64.
Description	The description of the VSwitch. The description must be 2 to 256 characters in length and can contain letters, digits, underscores (_), hyphens (-), periods (.), colons (:), and commas (,). It must start with a letter but cannot start with `http://` or `https://`.

12.7.3.2 Enable an IPv6 CIDR block for a VSwitch

This topic describes how to enable an IPv6 CIDR block for a VSwitch.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, click VSwitches.**
3. **Select the region where your VSwitch resides.**
4. **On the VSwitches page, find the target VSwitch, and click Enable IPv6 CIDR Block in the IPv6 CIDR Block column.**
5. **If no IPv6 CIDR block is enabled for the VPC to which the VSwitch belongs, click OK.**
6. **Specify the IPv6 CIDR block, and click OK.**

   The mask for the IPv6 CIDR block of a VSwitch is /64 by default. You can enter a number ranging from 0 to 255 to define the last 8 bits of the IPv6 CIDR block. For example, if the IPv6 CIDR block of a VPC is 2xx1: db8::/64, you can enter 255 (ff in hexadecimal notation) for the IPv6 CIDR block of the VSwitch. Then, the IPv6 CIDR block of the VSwitch is 2xx1: db8: ff::/64.

12.7.4 Manage an IPv6 gateway

12.7.4.1 IPv6 gateway specifications

IPv6 gateways are available in different specifications. The specifications of an IPv6 gateway affect the IPv6 Internet forwarding capacity, the maximum bandwidth quota of an IPv6 address, and the number of egress-only rules.

Specification	Maximum Internet traffic forwarding capacity	Maximum Internet bandwidth of a single IPv6 address		Egress-only rule number quota
		Bandwidth-based billing	Traffic-based billing	
Free Version	10 Gbit/s	2 Gbit/s	200 Mbit/s	0
Medium	20 Gbit/s	2 Gbit/s	500 Mbit/s	50
Large	50 Gbit/s	2 Gbit/s	1 Gbit/s	200

12.7.4.2 Create an IPv6 gateway

This topic describes how to create an IPv6 Gateway instance. An IPv6 Gateway is an IPv6 Internet traffic gateway of VPC. After you purchase an IPv6 Gateway instance, you can purchase IPv6 Internet bandwidth and set egress-only rules for the instance.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, click IPv6 Gateway.**
3. **Select the region where the IPv6 gateway resides.**
4. **On the IPv6 Gateway page, click Create IPv6 Gateway.**
5. **Set the parameters and click Submit. The following table describes the parameters.**

Parameter	Description
Organization	The organization to which the IPv6 gateway belongs.
Resource Set	The resource set to which the IPv6 gateway belongs.
Region	The region where the IPv6 gateway resides. The IPv6 gateway must belong to the same region as the VPC for which you want to enable the IPv6 CIDR block.
VPC	The VPC for which you want to enable the IPv6 gateway. If you cannot find the VPC in the list, perform the following operations: · Check whether the VPC has enabled an IPv6 gateway. · Check whether the VPC has a custom route with the destination CIDR block set to ::/0. If yes, delete the custom route. **Note:** After an IPv6 gateway is created, you cannot change the VPC.
Edition	The specifications of the IPv6 gateway. The specifications of an IPv6 gateway affect the IPv6 Internet forwarding capacity, the maximum bandwidth quota of an IPv6 address, and the number of egress-only rules. For more information, see *IPv6 gateway specifications*.

12.7.4.3 Modify an IPv6 gateway

This topic describes how to modify the name and description of an IPv6 gateway.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, click IPv6 Gateway.**
3. **Select the region where the IPv6 gateway resides.**
4. **On the IPv6 Gateway page, find the target IPv6 gateway, and click Manage in the Actions column.**
5. **On the IPv6 Gateway Details page, click Edit next to Name to change the name of the IPv6 gateway.**

   The name must be 2 to 128 characters in length and can contain letters, digits, underscores (_), and hyphens (-). It must start with a letter.
6. **Click Edit next to Description to modify the description of the IPv6 gateway.**

   The description must be 2 to 256 characters in length and cannot start with `http://` **or** `https://`**.**

12.7.4.4 Delete an IPv6 gateway

This topic describes how to delete an IPv6 gateway.

Prerequisites

Before you can delete an IPv6 gateway of the medium or large specifications, you must delete the egress-only rule. For more information, see *Delete an egress-only rule*.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, click IPv6 Gateway.**
3. **Select the region where the IPv6 gateway resides.**
4. **On the IPv6 Gateway page, find the IPv6 gateway, and click Delete in the Actions column.**
5. **In the dialog box that appears, click OK.**

12.7.5 Manage Internet bandwidth for an IPv6 address

12.7.5.1 Enable Internet bandwidth for an IPv6 address

This topic describes how to enable Internet bandwidth for an IPv6 address. You can enable Internet bandwidth for an IPv6 address so that the IPv6 address can be used to access the Internet.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click IPv6 Gateway.
3. Select the region where the IPv6 gateway resides.
4. On the IPv6 Gateway page, find the target IPv6 gateway, and click Manage in the Actions column.
5. In the left-side navigation pane, click IPv6 Internet Bandwidth.
6. In the IPv6 Internet Bandwidth page, find the IPv6 address, and click Create IPv6 Internet Bandwidth in the Actions column.
7. Specify a bandwidth, and click Submit.

   The specifications of an IPv6 gateway affect the maximum bandwidth quota of the relevant IPv6 address. For more information, see *IPv6 gateway specifications*.

12.7.5.2 Change the peak bandwidth for an IPv6 gateway

This section describes how to change the peak bandwidth of an IPv6 gateway.

Procedure

1. *Log on to the VPC console*.
2. In the left-side navigation pane, click IPv6 Gateway.
3. Select the region where the IPv6 gateway resides.
4. On the IPv6 Gateway page, find the IPv6 gateway, and click Manage in the Actions column.
5. In the left-side navigation pane, click IPv6 Internet Bandwidth.
6. In the IPv6 AddressList section, find the IPv6 address, and choose More > Modify IPv6 Internet Bandwidth in the Actions column.
7. Specify a bandwidth, and click Submit.

12.7.5.3 Delete the Internet bandwidth for an IPv6 address

This topic describes how to delete the Internet bandwidth for an IPv6 address if you do not need to use the IPv6 address to communicate with the Internet.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, click IPv6 Gateway.**
3. **Select the region where the IPv6 gateway resides.**
4. **On the IPv6 Gateway page, find the target IPv6 gateway, and click Manage in the Actions column.**
5. **In the left-side navigation pane, click IPv6 Internet Bandwidth.**
6. **In the IPv6 AddressList section, find the IPv6 address, and click Delete Internet Bandwidth.**
7. **In the dialog box that appears, click OK.**

12.7.6 Manage an egress-only rule

12.7.6.1 Create an egress-only rule

This topic describes how to create an egress-only rule for your IPv6 gateway. After you create an egress-only rule, the instances in your VPC can use IPv6 addresses to access the Internet but cannot be accessed by IPV6-compliant devices from the Internet.

Prerequisites

IPv6 Internet bandwidth is enabled for IPv6 addresses. For more information, see *Enable Internet bandwidth for an IPv6 address*.

Context

You cannot create an egress-only rule for the free version of IPv6 gateways. However, you can create an egress-only rule for IPv6 gateways of medium and large specifications.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, click IPv6 Gateway.**
3. **Select the region where the IPv6 gateway resides.**

4. **On the IPv6 Gateway page, find the IPv6 gateway, and click Manage in the Actions column.**
5. **In the left-side navigation pane, click Egress-only Rule.**
6. **Click Create Egress-only Rule. On the page that appears, select the ECS instance that uses the IPv6 address to communicate with the Internet, and click OK.**

12.7.6.2 Delete an egress-only rule

This topic describes how to delete an egress-only rule. You can delete an egress-only rule from an IPv6 gateway if cloud instances do not need to use IPv6 addresses to communicate with the Internet.

Procedure

1. *Log on to the VPC console*.
2. **In the left-side navigation pane, click IPv6 Gateway.**
3. **Select the region where the IPv6 gateway resides.**
4. **On the IPv6 Gateway page, find the IPv6 gateway, and click Manage in the Actions column.**
5. **In the left-side navigation pane, click Egress-only Rule.**
6. **On the Egress-only Rule page, find the egress-only rule that you want to delete, and click Delete in the Actions column.**
7. **In the dialog box that appears, click OK.**

13 Apsara Stack Security

13.1 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.

Background

Traditional security solutions for IT services detect attacks on network perimeters . They use hardware products such as firewalls and intrusion prevention systems (IPSs) to protect networks against attacks.

However, with the development of cloud computing, an increasing number of enterprises and organizations now use cloud computing services instead of traditional IT services. Cloud computing features low costs, on-demand flexible configuration, and high resource utilization. Cloud computing environments do not have definite network perimeters. As a result, traditional security solutions cannot effectively secure cloud assets.

With the powerful data analysis capabilities and professional security operations team of Alibaba Cloud, Apsara Stack Security provides integrated security protection services at the network layer, application layer, and server layer.

Complete security solution

Apsara Stack Security consists of Apsara Stack Security Standard Edition and optional security services, and provides users with a comprehensive security solution.

Security domain	Service	Description
Security management	Threat Detection Service	Monitors traffic and overall security status to audit and centrally manage security.
Server security	Server Guard	Protects Elastic Compute Service (ECS) instances against intrusions and malicious code.

Security domain	Service	Description
Application security	Web Application Firewall	Protects web applications against attacks and ensures that mobile and PC users can securely access web applications over the Internet.
Network security	Anti-DDoS	Ensures the availability of network links and improves business continuity.
Data security	Sensitive Data Discovery and Protection	Prevents data leaks and helps your business systems meet compliance requirements.
Security O&M service	On-premises security services	Help you establish and optimize your cloud security system to protect your business system against attacks by making full use of the security features of Apsara Stack Security and other Apsara Stack services.

13.2 Restrictions

Before logging on to Apsara Stack Security Center, make sure that your local PC meets the requirements.

Table 13-1: 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

13.3 Quick start

13.3.1 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 13-2: Default roles in Apsara Stack Security*.

Table 13-2: 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.  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*.

13.3.2 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 ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel. The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.**
- **We recommend that you use the Google Chrome browser.**

Procedure

1. **In the address bar, enter the URL used to access the ASCM console. Press Enter.**
2. **Enter your username and password.**

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

   Note:
   When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).
3. **Click Login to go to the ASCM console homepage.**
4. **In the top navigation bar, choose Security > Apsara Stack.**
5. **On the Apsara Stack Security Center page, set Region.**
6. **Click YD to go to Apsara Stack Security Center.**

13.3.3 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 selected a region. To manage the servers 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 is switched to the selected one.

13.4 Threat Detection Service

13.4.1 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, attacker intrusions, web attacks, DDoS attacks, threat intelligence, enterprise security reputation, and other security threats. The service uses modeling and analysis to obtain key information from traffic features, server behavior, and server operations logs and detects intrusions that cannot be identified by only detecting traffic or scanning files. TDS identifies threat sources and attack behavior, and assesses threat levels based on the outputs from cloud-based analysis models and intelligence data.

TDS provides the following features:

- Overview: provides a security situation overview and information about security screens.
- Security alerts: displays security alerts that have occurred in the business system.
- Attack analysis: displays application attacks and brute-force attacks that have occurred in the system.
- Asset management: allows you to manage the server assets and NAT assets in Apsara Stack.
- Security reports: allows you to configure security report tasks for Apsara Stack.

13.4.2 Security overview

13.4.2.1 View security overview information

This topic describes how to view the security statistics, attack trends, and network traffic information of the Apsara Stack platform.

Context

The Security Overview tab presents an overview of the detected security events, latest threats, and inherent vulnerabilities in the current system. The security administrator can view the information on the Security Overview tab to have a comprehensive understanding of the system security situation.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Threat Detection > Overview. Then click the Security Overview tab.**

3. **View the security situation of the Apsara Stack platform.**

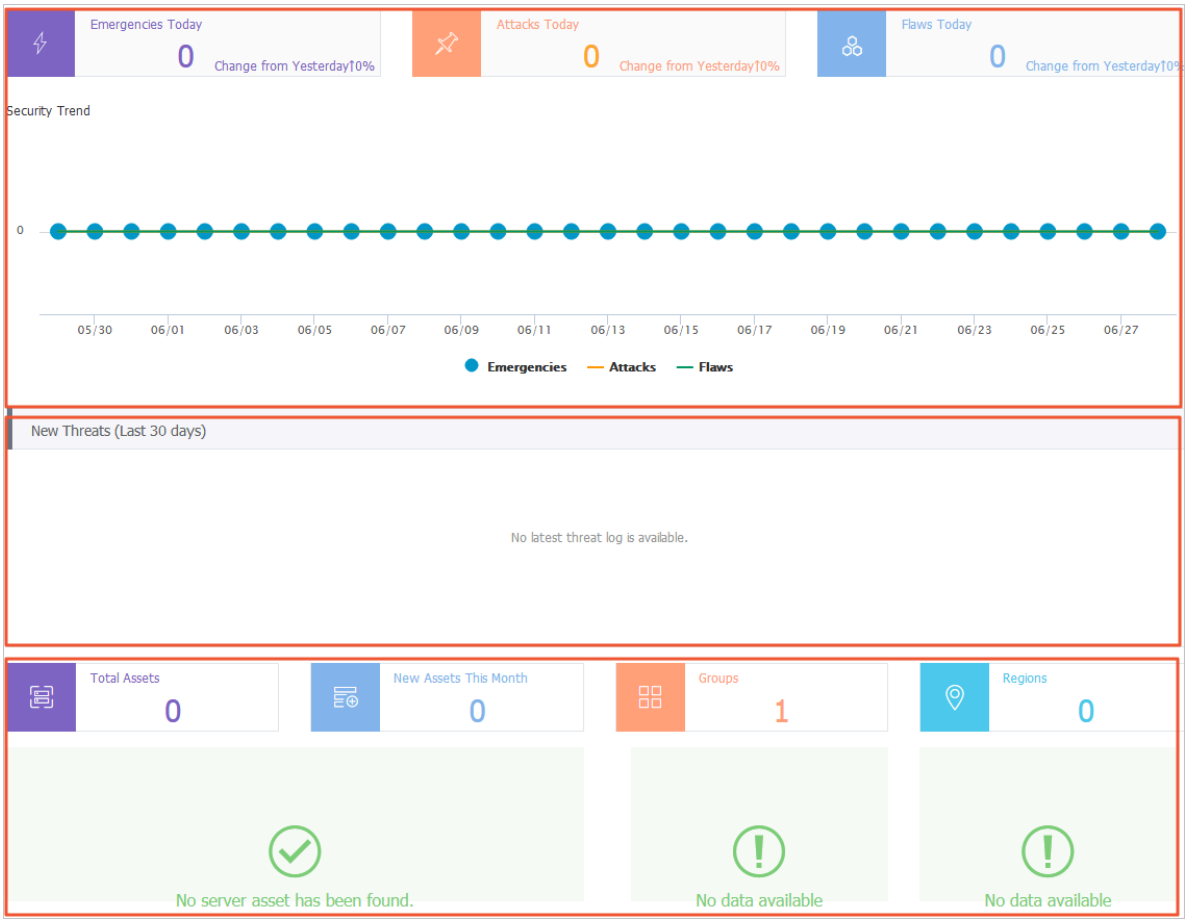


Table 13-3: Sections on the Security Overview tab

Section	Function	Description
1	Asset details	Displays the number of Apsara Stack assets, the number of regions, and other information.
2	Attack trends	Displays the security events, security attacks, and vulnerabilities that have been detected in the current system, as well as how the number of attacks changes with time.
3	Network traffic	Displays inbound and outbound traffic and QPS .

13.4.2.2 View information on visual screens

This topic describes how to view information on visual screens.

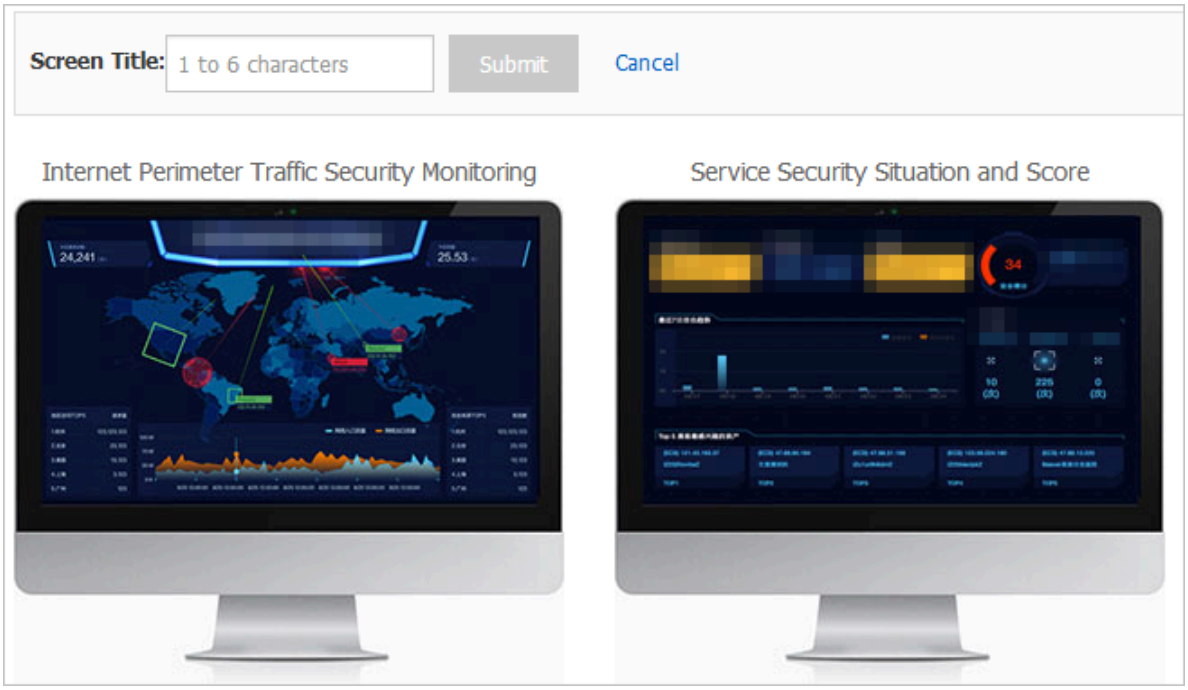
Context

The visual screens use dynamic graphs to present key security event metrics. This allows the security administrator to get a good grasp of the security situation and make informed security decisions.

The visual screens include the screen to monitor the security of the Internet perimeter traffic and the screen to display the service security situation and score.

Procedure

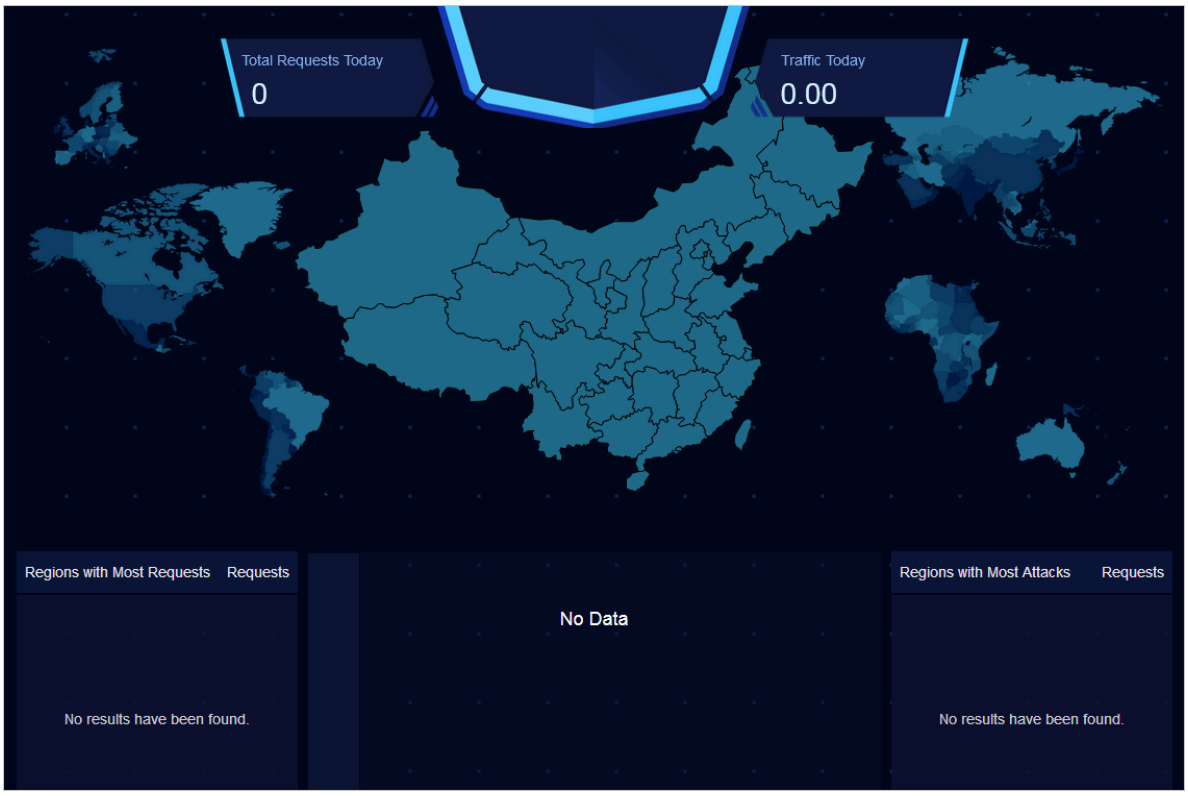
1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Threat Detection > Overview. Then click the Screens tab.



Click Modify on the Screens tab to modify the displayed title for Internet Perimeter Traffic Security Monitoring.

3. **Click the screen under Internet Perimeter Traffic Security Monitoring.**

   You can view information displayed on the screen.



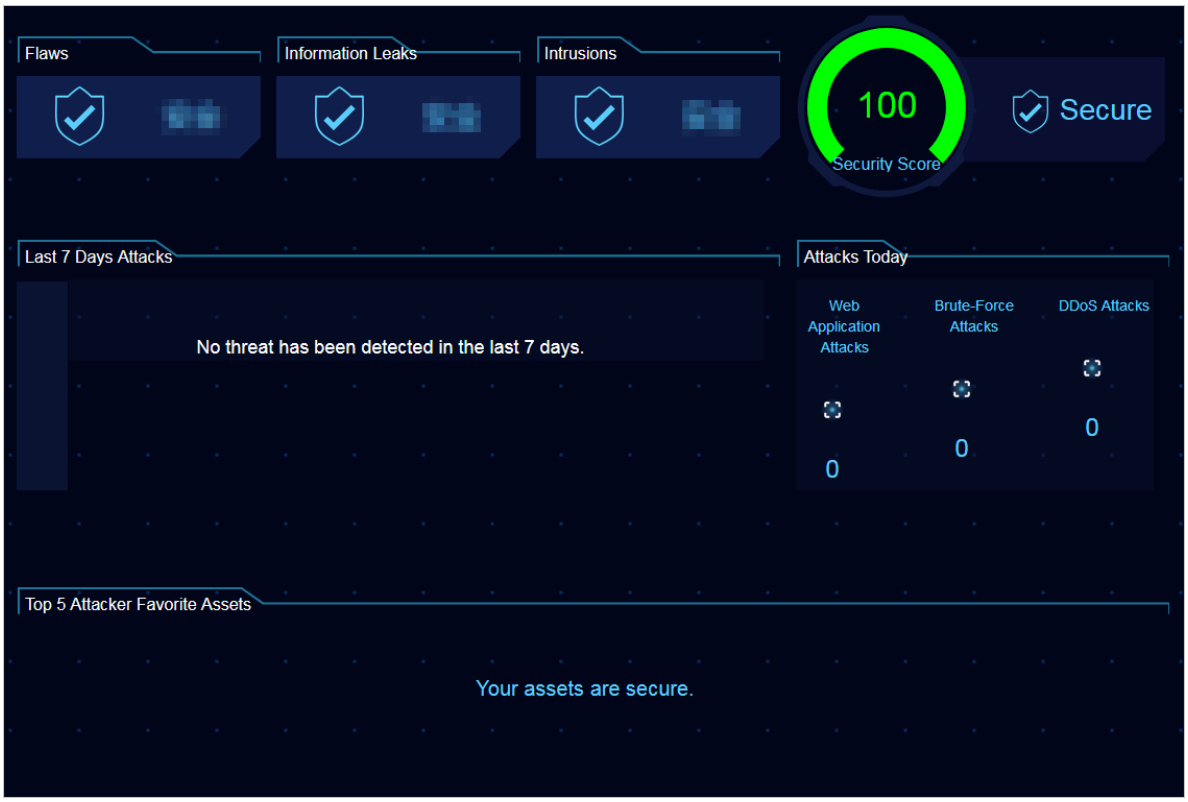
The Internet Perimeter Traffic Security Monitoring screen provides a system traffic overview and statistics on regions with requests and those with attacks. The screen lists the top five regions with requests and top five regions with attacks. This helps the security administrator obtain up-to-date information about the regional distribution of requests and attacks.

Table 13-4: Traffic sources

Type	Implementation mechanism
Request analysis	The assets that interest users are pushed to Network Traffic Monitoring System, which then reports access information for these assets.
Attack analysis	Network Traffic Monitoring System of Apsara Stack Security detects, reports, and displays events such as web attacks.
Traffic display	Network Traffic Monitoring System collects traffic and reports it to the console for recording.

4. **Click the screen under Service Security Situation and Score.**

   You can view the information displayed on the screen.



   The Service Security Situation and Score screen displays detailed information about the current security events in the system. After analyzing the system vulnerabilities and the assets that have been attacked or that interest attackers, this screen evaluates the security situation of the system and displays the security level.

   The data on this screen is collected and reported by Apsara Stack Security modules such as Network Traffic Monitoring System, Server Guard, and flaw analysis. The top five assets that interest attackers are obtained by a big data engine through data modeling.

13.4.3 Security alerts

13.4.3.1 View and handle security alerts

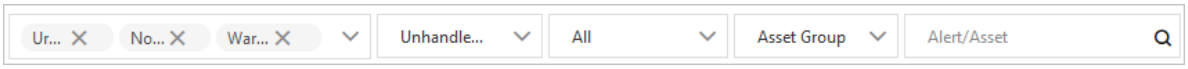
This topic describes how to view and handle security alerts on the Security Alerts page.

Procedure

1. *Log on to Apsara Stack Security Center*.

2. In the left-side navigation pane, choose Threat Detection > Security Alerts.
3. Optional: Set conditions for filtering security alerts.

> **Note:**
> If you want to view all alerts, the conditions are not required.



Filter condition	Description
Severity level	The severity level of an alert. You can select multiple levels. Valid values: • `Urgent` • `Warning` • `Notice`
Alert status	The status of an alert. Valid values: • `Unhandled Alerts` • `Handled`
Alert type	The type of an alert. Select `All` or a specific type.
Affected asset group	The affected asset groups. Select `Asset Group` or a specific group.
Alert name or asset keyword	The name of an alert or the keyword of affected assets. You can specify a name or keyword to search for related alerts.

4. View security alerts and their details in the alert list.
5. Click Processing in the Actions column corresponding to an alert to handle the alert.

13.4.3.2 Manage a quarantine

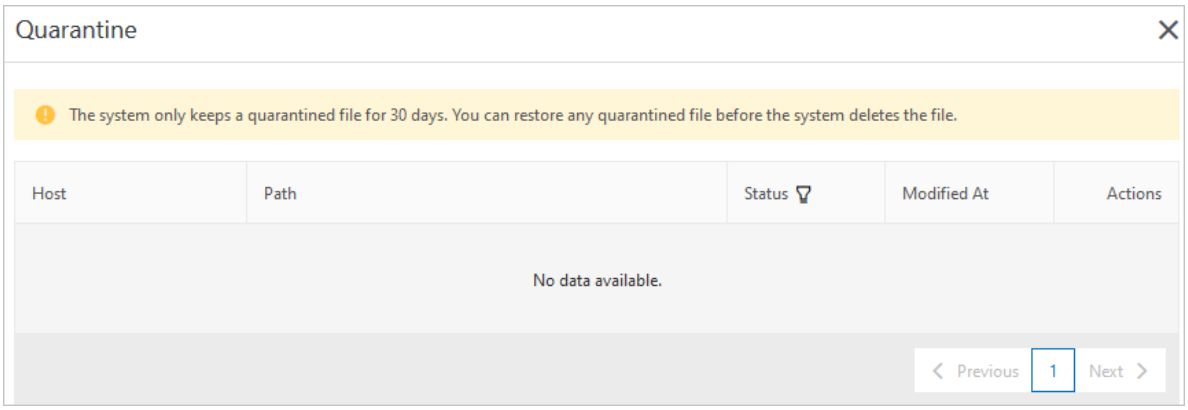
This topic describes how to manage a quarantine. The system quarantines detected threat files. You can restore quarantined files within 30 days after they have been quarantined, and the system automatically deletes these files after they expire.

Context

> **Note:**
> Before you restore a quarantined threat file, make sure that the file is a false positive.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Threat Detection > Security Alerts.
3. In the upper-right corner of the Alerts page, click Quarantine.
4. On the Quarantine page that appears, view information about quarantined files, such as server IP addresses, directories for storing the files, status, and operation time.



5. Optional: Click Restore in the Actions column corresponding to a quarantined file.

   You can perform this operation to restore a specified file from the quarantine. The restored file is displayed in the security alert list again.

13.4.4 Attack analysis

13.4.4.1 View application attacks

This topic describes how to view application attacks. The Network Traffic Monitoring System module of Apsara Stack Security monitors all traffic to web servers and extracts attack information.

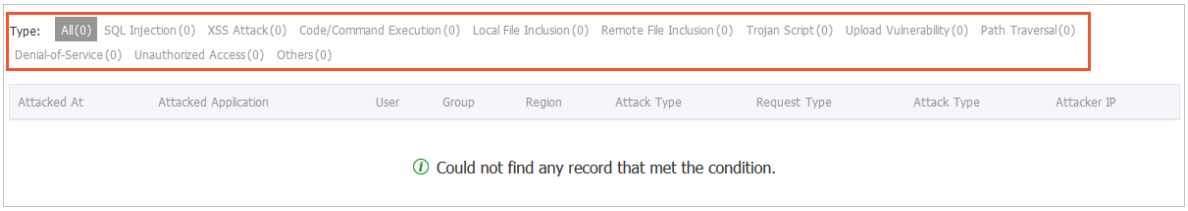
Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Threat Detection > Attack Analysis.
3. Click the Application Attacks tab.
4. Select a region from the Region drop-down list.

5. View attacks launched in the last seven days and their types.

Item	Description
Last 7 Days Attacks	Displays recent attacks targeting applications on the Apsara Stack platform.
Attack Types of Last 7 Days	Displays the types of recent attacks targeting applications on the Apsara Stack platform.

6. Click an application attack type in the Type section.



Application attack type	Description
All	All types of attacks.
SQL Injection	A web application does not check the validity of the data provided by users. An attacker creates SQL statements to submit special characters and commands from an input area on the web page, such as a form or the address box for entering a URL. This allows the attacker to interact with the database to obtain private information or tamper with data in the database.
XSS Attack	A web application does not filter or restrict the statements and variables provided by users. 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, the malicious code automatically runs on the browser.
Code/Command Execution	An attacker uses URLs to issue requests and runs unauthorized code or commands on the web server.
Local File Inclusion/Remote File Inclusion	An attacker adds invalid parameters to URLs and issues requests to a 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 this vulnerability.

Application attack type	Description
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 into a website , an attacker usually mixes ASP or PHP webshell files with normal web page files in the web directory of the web server. Then, the attacker can access the webshell files from a browser and obtain the command execution environment to control the web server.
Upload Vulnerability	When a web application processes a file uploaded by a user , it does not check the validity of the file name extension or that of the content in the file before it stores the file in a web server. 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 unauthorized directories or run commands in directories except for the root directory of the web server.
Denial-of-Service	An attacker uses DoS to exhaust the network or system resources on a server and interrupt or stop services on this server. As a result, authorized users cannot access the server.
Unauthorized Access	If an application has vulnerabilities in the authentication process, an attacker exploits the vulnerabilities to bypass authentication and access or edit unauthorized code.
Others	Other types of application attacks.

7. View the details of application attacks in the list.

13.4.4.2 View brute-force attacks

This topic describes how to view brute-force attacks. The Server Guard agent installed on a server detects brute-force attacks launched by attackers at this server and reports these attacks to the console.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Threat Detection > Attack Analysis.
3. Click the Brute-Force Attack tab.

4. Select a region from the Region drop-down list.
5. View brute-force attacks in the event list.

13.4.5 Manage assets

13.4.5.1 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 13-1: Asset Overview page

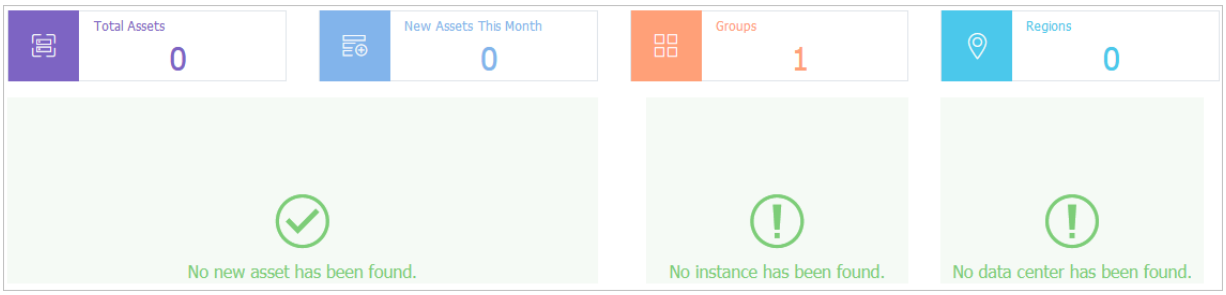


Table 13-5: Parameters on the Asset Overview page

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.

Parameter	Description
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.

13.4.5.2 Manage groups

13.4.5.2.1 Add a group

This topic describes how to add a group.

Context

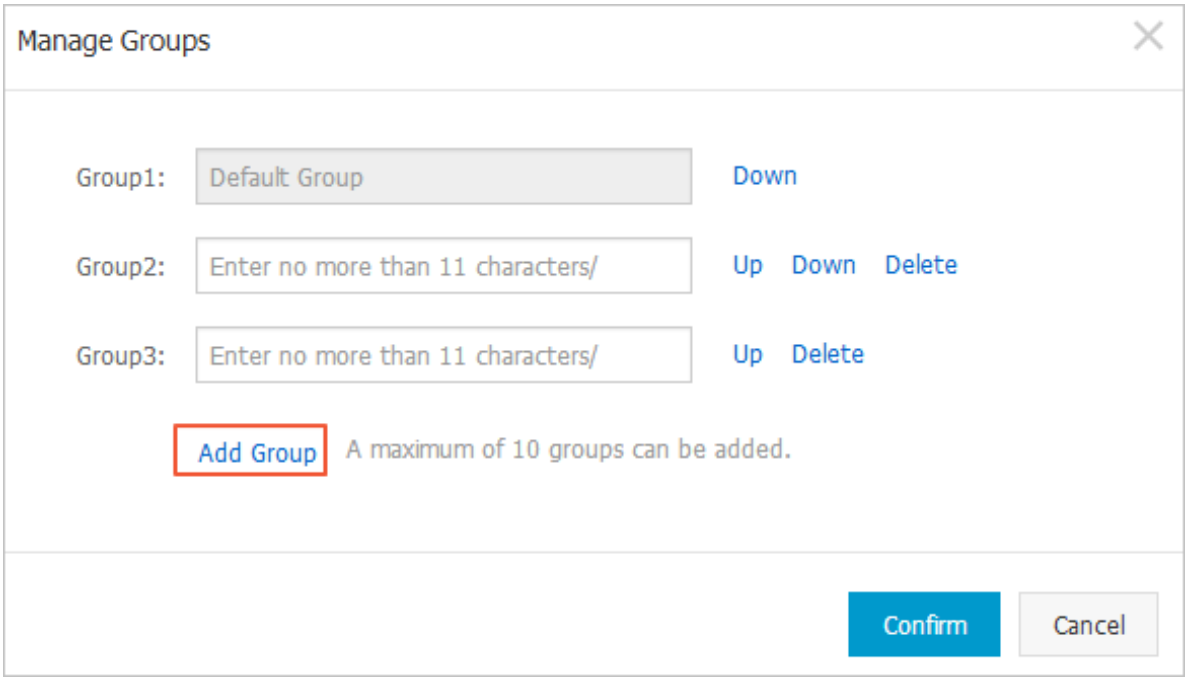
Asset groups are used to differentiate assets, making it easier for you to query and modify the asset information.

> **Note:**
> **A maximum of 10 asset groups are supported.**

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Threat Detection > Assets.**
3. **In the Group area, click Manage Groups.**

4. In the Manage Groups dialog box, click Add Group.



5. Enter a group name, and click Confirm.

13.4.5.2.2 Delete a group

This topic describes how to delete unnecessary groups to facilitate asset information query and modification.

Context

- You cannot delete or rename the default group.
- You cannot delete groups that contain assets.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Threat Detection > Assets.
3. In the Group area, click Manage Groups.
4. In the Manage Groups dialog box, click Delete next to a group.
5. Click Confirm.

13.4.5.2.3 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 Threat Detection > Assets.
3. In the Group area, click Manage Groups.
4. In the Manage Groups dialog box, click Up or Down to sort the groups.
5. Click Confirm to save the new group order.

13.4.5.3 Asset information

13.4.5.3.1 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 view the server asset 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.

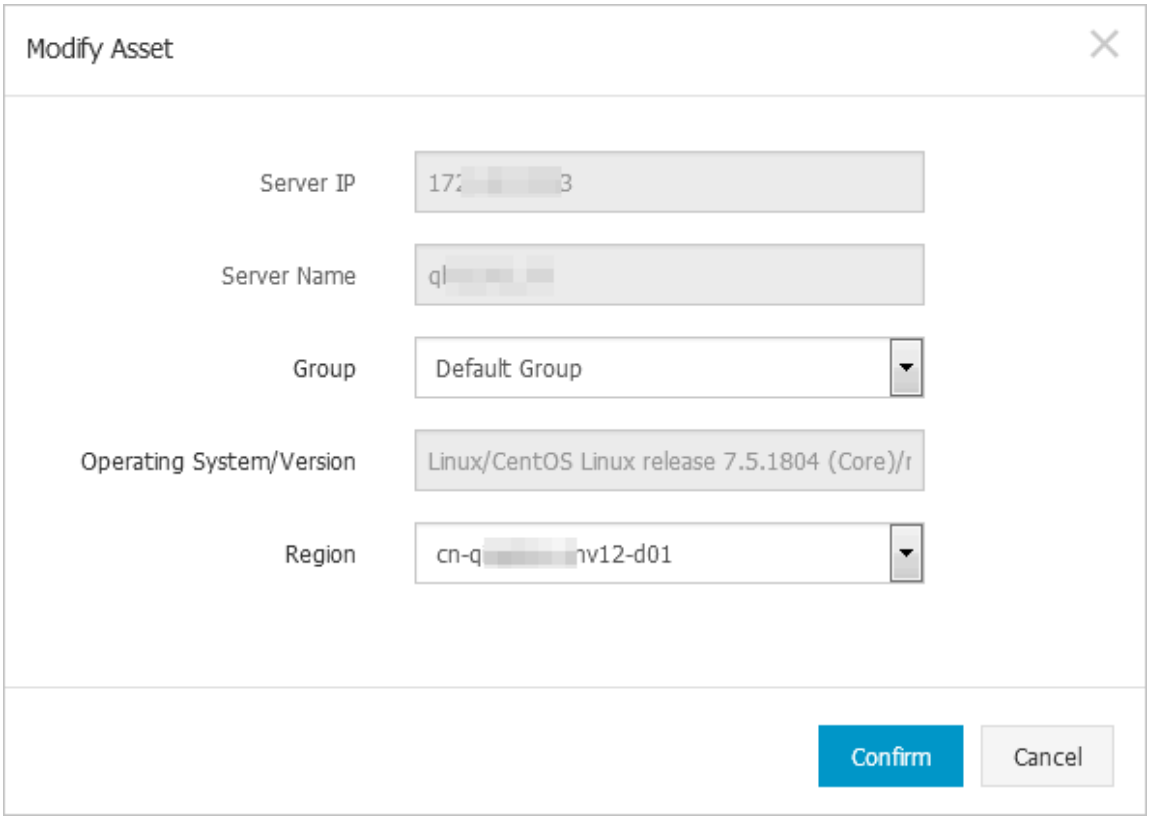
Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Threat Detection > Assets.
3. In the Server Asset/NAT Asset area, click the Server Asset tab.
4. Set the search criteria and click Search to view a server asset.

> **Note:**
> **You can filter server assets by operating system, region, or group. You can also enter a server IP address or server name for a fuzzy search. By default, the server assets in all regions are displayed and sorted by IP address.**

5. Maintain the server asset information.

   - 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.

13.4.5.3.2 Manage NAT assets

This topic describes how to add, view, and delete NAT assets.

Context

NAT assets are external IP addresses that are converted from internal IP addresses through NAT, namely, IP addresses that are exposed to the Internet. Multiple servers can share one external IP address but use different ports to receive Internet requests. After an IP address is set as a NAT asset, Threat Detection Service analyzes this asset to detect attack events.

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 or region. The security administrator can also add one NAT asset or add multiple NAT assets by specifying a CIDR block.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Threat Detection > Assets.
3. In the Server Asset/NAT Asset area, click the NAT Asset tab.
4. Set the search criteria and click Search to view a NAT asset.

   Note:

   You can filter NAT assets by region or group. You can also enter a NAT IP address for a fuzzy search. By default, the NAT assets in all regions are displayed and sorted by IP address.

5. Add a NAT asset.

   Note:

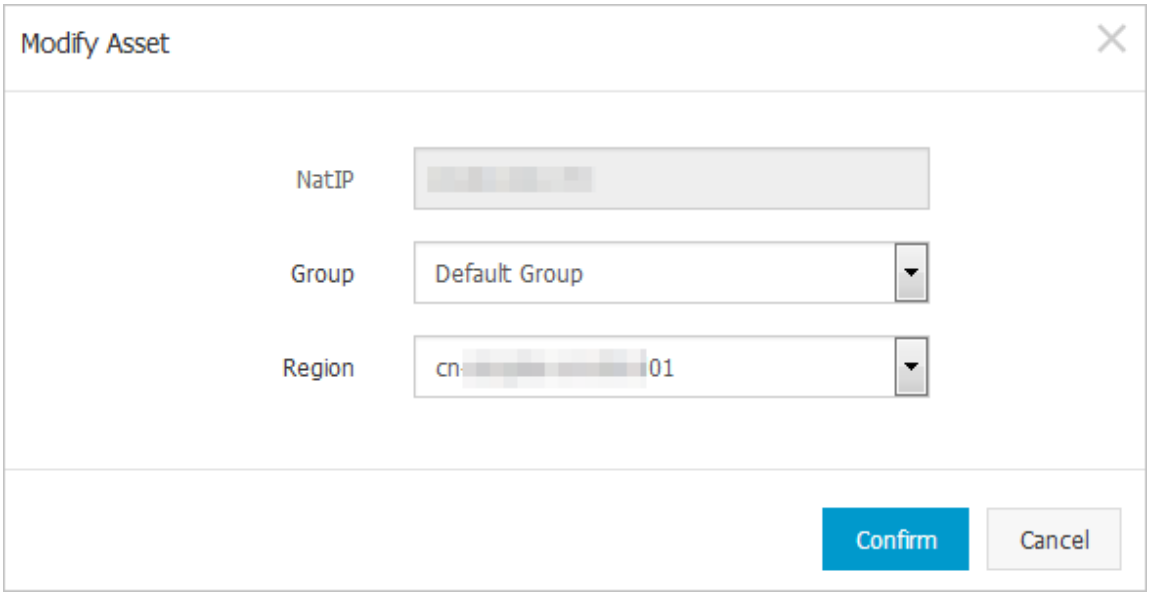
   The NAT IP address to be added cannot be the same as an existing IP address. Specify a valid IP address or CIDR block as the NAT IP address.

   a) Click Add in the upper-right corner of the NAT Asset tab page.

   b) In the Add Asset dialog box, enter an IP address or a CIDR block and select a group and a region.

   c) Click Confirm.

6. Maintain NAT asset information.
   - Click Modify. In the Modify Asset dialog box, change the business group and region of the asset.



   - Click Delete. In the Delete Asset message, click Confirm to delete the NAT asset.

13.4.5.3.3 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 one or more assets at one time.

- **Modify an attribute for one asset.**

  This method applies when you modify only one asset or when the assets to be modified are not in the same CIDR block and use server names without similarities. For more information about how to modify one asset, see *Manage server assets* **and** *Manage NAT assets*.

- **Modify an attribute for multiple assets.**

  This method applies when you modify multiple assets that are in the same CIDR block or have similar server names.

Note:
You cannot modify the server IP address, server name, operating system, or operating system version.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Threat Detection > Assets.**
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 that contains the server assets or NAT assets to be modified.**

     Note:
     If you specify a CIDR block, all server assets and NAT assets in the specified CIDR block are modified.
   - **Select `Server Name` from the Type drop-down list, and enter the common part of the names of servers to be modified.**

     Note:
     If you specify the common part of server names, all servers whose names contain the specified common part are modified.

13.4.6 Security reports

13.4.6.1 Create a report task

This topic describes how to create a report task. A report task regularly sends security reports of the Apsara Stack platform to a 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.

Item	Sub-item	Description
Threat Detection Service	Security statistics	The security overview information on the overview page of Threat Detection Service (TDS).

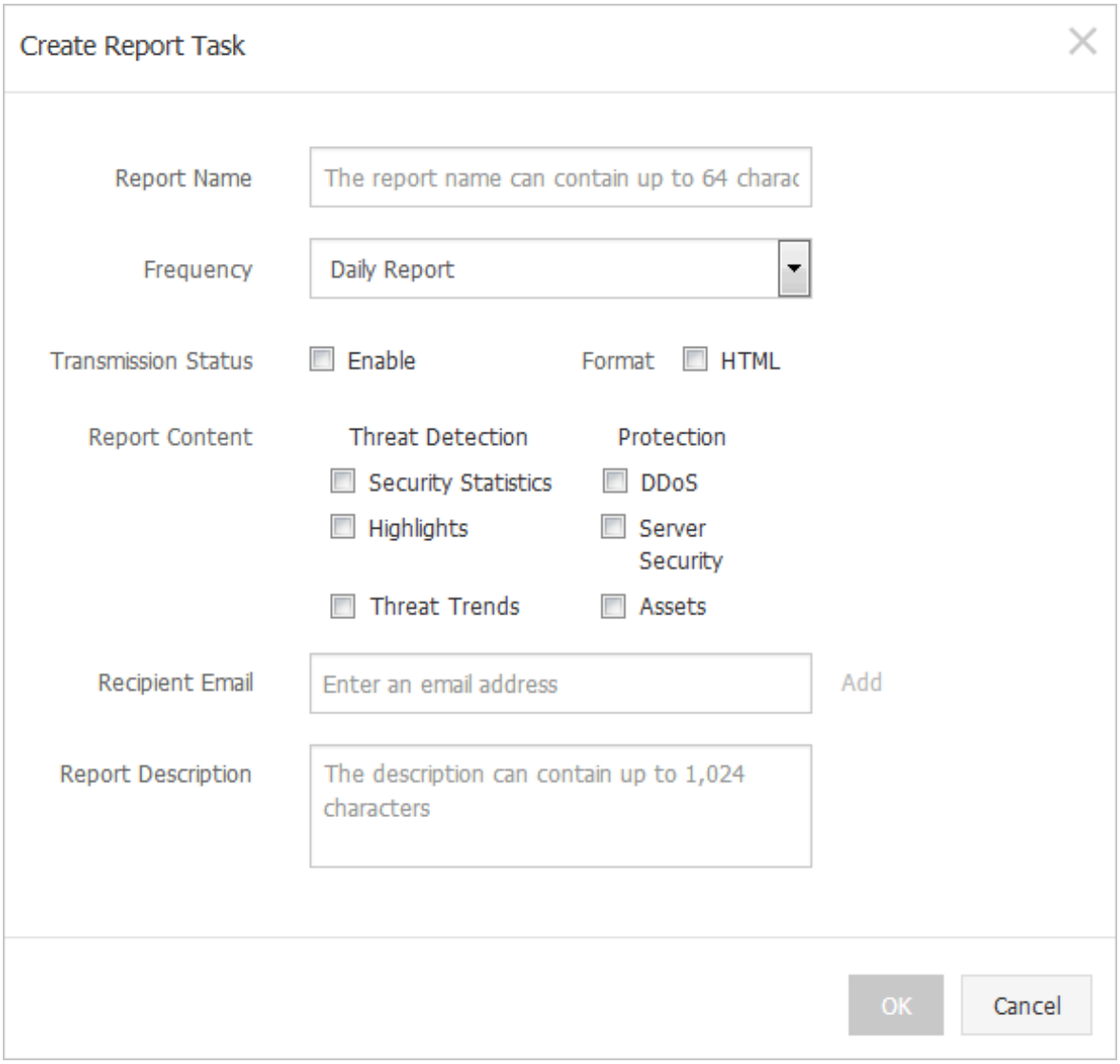
Item	Sub-item	Description
	Highlights	The important emergency information on the event analysis page of TDS.
	Threat trend	The attack trends and analysis information on the threat analysis page of TDS.
Security protection	Distributed denial of service (DDoS)	The DDoS attack events detected by Apsara Stack Security Center.
	Server security	The server security vulnerabilities, unusual logons, brute-force attacks, and configuration risks detected by Apsara Stack Security Center.
	Protected assets	The 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 Service > Security Reports.
3. On the Security Reports page, click Create Task.

4. **In the Create Report Task dialog box, set relevant parameters.**

Figure 13-2: Create a report task



Parameter	Description
Report Name	**The name of the report task.**
Frequency	**The interval at which security reports are sent.** **Value values:** • `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.**

Parameter	Description
Transmission Status	Specifies whether to enable this report task.
Format	Specifies whether to generate reports in HTML format.
Report Content	The items to be contained in a security report.
Recipient Email	The email address that receives security reports. **Note:** Click Add next to the field to add an email address. You can add up to 10 email addresses.
Report Description	The report description.

5. Click OK.

Result

After the report task is created, the specified email addresses receive security reports at the specified intervals, as shown in *Figure 13-3: Security report*.

Figure 13-3: Security report

13.4.6.2 Manage report tasks

This topic describes how to view, modify, or delete report tasks.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Threat Detection Service > 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.

13.5 Network Traffic Monitoring System

13.5.1 View traffic trends

This topic describes how to view the trends of network traffic, inbound traffic statistics, and outbound traffic statistics.

Context

By analyzing traffic trends, the security administrator can obtain the traffic rates and the peaks and troughs of traffic periods. In addition, the security administrator can block access from malicious IP addresses by viewing the five IP addresses with the most traffic.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Network Security > Traffic Analysis > Traffic Trend.
3. In the upper-right corner of the Traffic Trends page, select the time range, which can be Last 1 Hour, Last 24 Hours, or Last 7 Days.
4. View network traffic information.
   - Network traffic trends

     View the network traffic trends from the selected time period, including inbound and outbound traffic measured in bytes per second (BPS) and packets per second (PPS).
   - Inbound Traffic

     View the following statistics: Inbound Sessions, Inbound Applications, and Destination IPs with Most Requests.
   - Outbound Traffic

     View the following statistics: Outbound Sessions, Outbound Applications, and Source IPs with Most Requests.

5. Optional: Click to export traffic trends as a PDF file.

13.5.2 View traffic at the Internet border

This topic describes how to view the traffic at the Internet border. You can obtain up-to-date information about network security.

Prerequisites

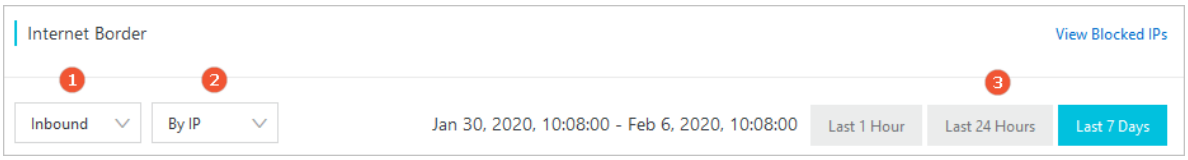
The Network Traffic Monitoring System module is purchased and deployed at the egress (ISW) of Apsara Stack. This module is used to audit, analyze, and manage both inbound and outbound traffic at Internet borders.

Context

You can use traffic information to identify abnormal Internet traffic and block malicious requests.

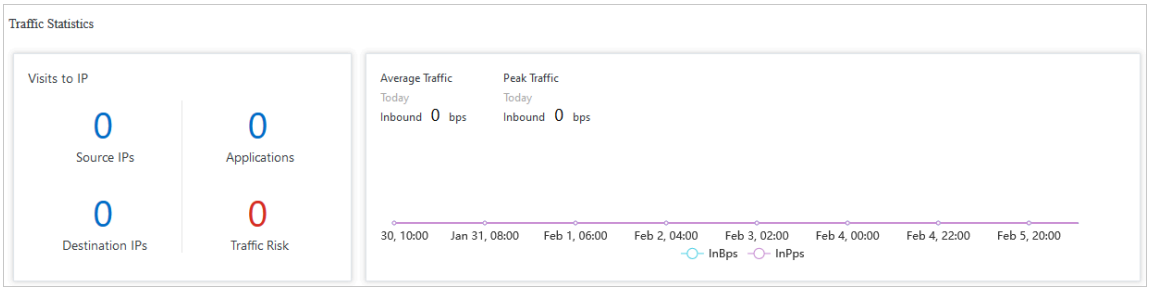
Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Network Security > Traffic Analysis > Internet Border.
3. Set traffic filter conditions.

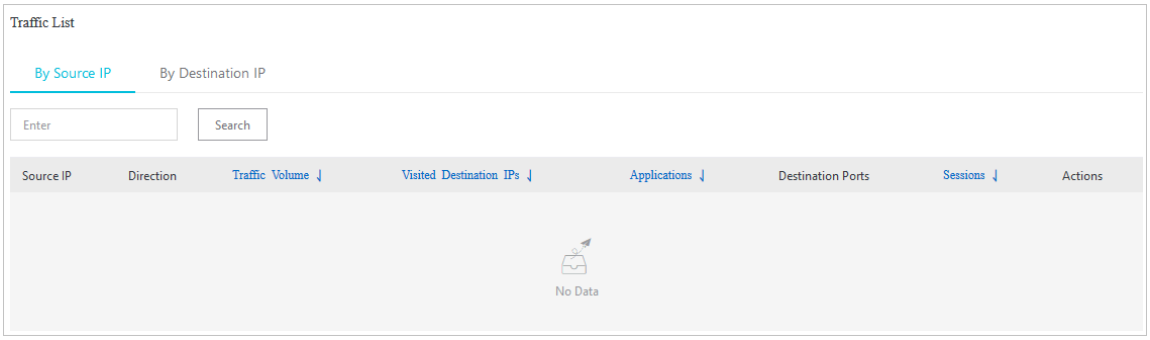


Item	Description
1	Set the direction of traffic. Valid values: • `Inbound`: The traffic flows from the Internet to the internal network. • `Outbound`: The traffic flows from the internal network to the Internet.
2	Specify whether you want to view the traffic by IP address or application. Valid values: `By IP` and `By Application`.
3	Set the time range. Valid values: `Last 1 Hour`, `Last 24 Hours`, and `Last 7 Days`.

4. View details about the traffic at the Internet border.
   - Traffic Statistics



     - The Visits to IP section includes Source IPs, Destination IPs, Applications, and Traffic Risk.
     - In the traffic chart on the right, you can view the average traffic, peak traffic, and traffic trends.
   - Traffic List



     In the Traffic List section, you can view traffic details.
5. In the Traffic List section, view abnormal traffic of the specified IP address.
   - If `Inbound` is specified, you can view abnormal traffic on the By Destination IP tab in the Traffic List section.
   - If `Outbound` is specified, you can view abnormal traffic in the Traffic List section.

13.5.3 View traffic at the internal network border

This topic describes how to view the traffic at the internal network border. You can obtain up-to-date information about network security based on the traffic.

Prerequisites

The Network Traffic Monitoring System module is purchased and deployed at the ingress (CSW) of Apsara Stack. This module is used to audit, analyze, and manage

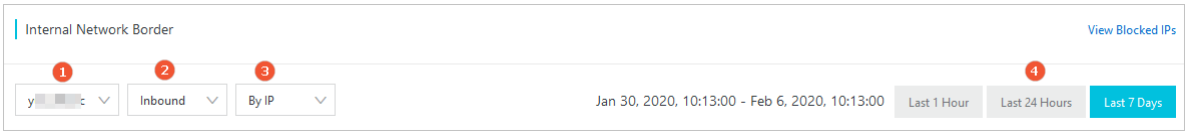
both inbound and outbound traffic routed over physical connections between on-premises data centers and VPCs.

Context

You can use traffic information to identify abnormal internal network traffic and block malicious requests.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Network Security > Traffic Analysis > Internal Network Border.
3. Set traffic filter conditions.



Item	Description
1	Select a VPC instance name from the drop-down list.
2	Set the traffic direction. Valid values: • `Inbound`: The traffic flows from the Internet to the internal network. • `Outbound`: The traffic flows from the internal network to the Internet.
3	Specify whether you want to view the traffic by IP address or application. Valid values: `By IP` and `By Application`.
4	Set the time range. Valid values: `Last 1 Hour`, `Last 24 Hours`, and `Last 7 Days`.

4. **View details about the traffic at the internal network border.**
   - **Traffic Statistics**
     - **The Visits to IP section includes Source IPs, Destination IPs, Applications, and Traffic Risk.**
     - **In the traffic chart on the right, you can view the average traffic, peak traffic, and traffic trends.**
   - **Traffic List**

     In the Traffic List section, you can view traffic details.

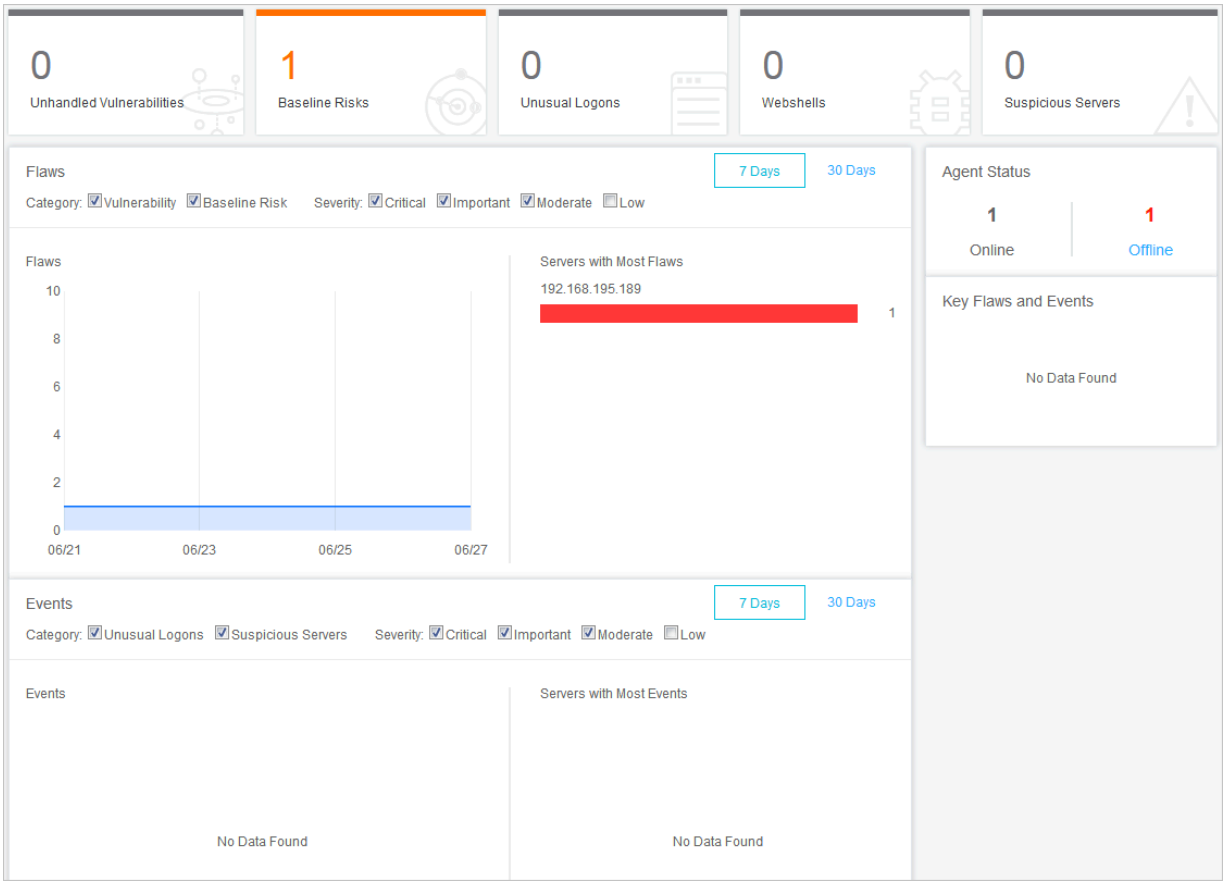
     If By IP is specified, you can view the abnormal traffic of the specified IP address in the Traffic List section.

13.6 Server security

13.6.1 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: 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, webshells, and suspicious servers.**
- **Flaws: This area displays the trend of security flaws on your servers. Your server security is threatened if you do not handle the security flaws.**
- **Events: This area displays the trend of security events on your servers. A security event is an intrusion event that has been detected on your server.**
- **Agent Status: This area displays the number of servers being protected and the number of servers with an offline agent.**
- **Key Flaws and Events: This area displays the recent key flaws and events on your servers. You can click a flaw or event to view the details.**

13.6.2 Server list

13.6.2.1 Manage servers

This topic describes how to manage servers. On the Servers page, you can view the status of servers protected by Server Guard.

Context

The protection status of a server includes:

- Online: Server Guard provides comprehensive security protection for this server.
- Offline: Server Guard cannot provide security protection for this server because the Server Guard agent on this server is offline.
- Disable Protection: Security protection is temporarily disabled for this server.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. In the left-side navigation pane, choose Server Security > Servers.
3. Optional: Search for a server.

   To view the agent status of a server, enter the server IP address in the search bar, and click Search. Detailed server information, such as security information, is displayed.
4. View the agent status and detailed security information of the server.

   Click  in the upper-right corner of the page to select the information columns you want to display. The following table lists the information categories.

Category	Information
Basic information	· Server IP/Name · Tag · OS · Region
Agent status	Agent Status
Threat prevention	· Vulnerability · Baseline Risk
Intrusion detection	· Unusual Logons · Webshells · Suspicious Servers
Server fingerprints	· Processes · Ports · Root Accounts/Total Accounts

5. Manage servers.

Action	Description
Change Group	Select servers and click Change Group to add the selected servers to a new group.
Modify Tag	Select servers and click Modify Tag to modify tags for the servers.
Security Inspection	Select servers and click Security Inspection to select the items to be checked.
Delete External Servers	Select external servers, and choose More > Delete External Servers.
Disable Protection	Select the servers whose agent status is Online, and choose More > Disable Protection. This temporarily disables protection for these servers to reduce server resource consumption.
Enable Protection	Select the servers whose agent status is Disable Protection, and choose More > Enable Protection. This enables protection for these servers.

13.6.2.2 Manage server groups

This topic describes how to manage server groups. To facilitate the security management of specific servers, you can add them to groups and view their security events by group.

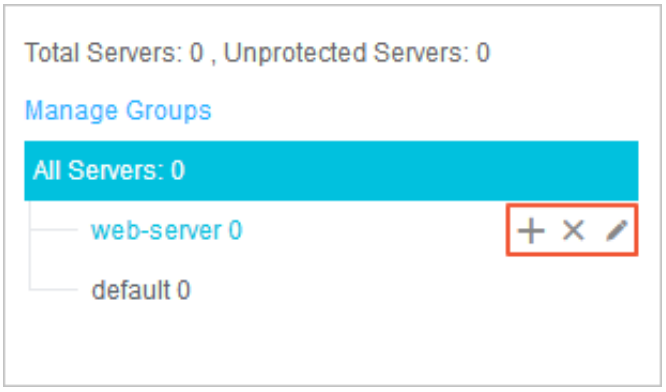
Context

Servers that are not added to any group are assigned to the default group. If you delete a group, all servers in the group are automatically moved to the default group.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. In the left-side navigation pane, choose Server Security > Servers.

3. Manage server groups.



- Create a group.

  Click the Add Subgroup icon next to All Servers or a specific group, enter a group name, and click OK.

  > **Note:**
  > The system supports a maximum of 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 a group, all servers in the group are automatically moved to the default group.

- Sort groups.

  Click Manage Groups to sort groups in descending order by priority.

4. Change the groups to which servers belong.
   a) Select servers from the list on the right.
   b) Click Change Group.
   c) In the Change Group dialog box that appears, select a group from the drop-down list.
   d) Click OK.

13.6.3 Threat protection

13.6.3.1 Vulnerability management

13.6.3.1.1 Manage Linux software vulnerabilities

This topic describes how to manage Linux software vulnerabilities.

Context

Apsara Stack Security automatically scans the software that has been installed on your servers for vulnerabilities on the Common Vulnerabilities and Exposures (CVE) list and sends you alerts about the detected vulnerabilities. Apsara Stack Security also provides commands to fix vulnerabilities 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 the detected Linux vulnerabilities.**

   Note:
   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.**

   Note:
   You can quickly locate specific affected assets by using the search and filter functions.

   - **Basic Information: The basic information of the vulnerability, including the name, CVSS score, description, and resolution.**
   - **Affected Assets: The servers that are affected by the vulnerability.**

5. **Select an action based on the impact of the vulnerability.**

   Table 13-6: 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 takes effect only after a server reboot, you should 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 does not alert you about ignored vulnerabilities.
Verify	Click Verify to verify the vulnerability fix. If you do not manually verify the fix, the system automatically verifies the fix within 48 hours.

   You can manage a vulnerability on one or more affected assets at one time.

   - **To manage a vulnerability on one asset, select an action from the Actions column of this asset.**
   - **To manage a vulnerability on one or more assets, select one or more affected servers, and select an action in the lower-left corner.**

13.6.3.1.2 Manage Windows vulnerabilities

This topic describes how to manage Window vulnerabilities.

Context

Apsara Stack Security automatically checks if your servers have the latest Microsoft updates installed, and notifies you of the detected vulnerabilities. Apsara Stack Security also automatically detects and fixes major vulnerabilities on your servers.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Threat Prevention > Vulnerabilities, and click the Windows Vulnerabilities tab.**

3. Check the detected Windows vulnerabilities.

   Note:
   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.

   Note:
   You can quickly locate specific affected assets by using the search and filter functions.

   - Basic Information: The basic information of the vulnerability, including the name, CVSS score, description, and resolution.
   - Affected Assets: The servers that are affected by the vulnerability.

5. Select an action based on the impact of the vulnerability. *Table 13-7: Actions on vulnerabilities* describes the actions.

   Table 13-7: Actions on vulnerabilities

Action	Description
Fix Now	Select this option to fix the vulnerability directly. The system caches an official Windows patch in the cloud for your server to download and update.
Ignore	Select this option to ignore the vulnerability. The system does not alert you about ignored vulnerabilities.
Verify	Click Verify to verify the vulnerability fix.
Restarted and Verified	If a vulnerability fix takes effect only after a server reboot, you should 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 manage a vulnerability on one or more affected assets at one time.

   - To manage a vulnerability on one asset, select an action from the Actions column of this asset.
   - To manage a vulnerability on one or more assets, select one or more affected servers, and select an action in the lower-left corner.

13.6.3.1.3 Manage Web CMS vulnerabilities

This topic describes how to manage Web CMS vulnerabilities.

Context

The Web CMS vulnerability detection feature obtains the information of the latest vulnerabilities and provides patches in the cloud. This helps you quickly detect and fix vulnerabilities.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Threat Prevention > Vulnerabilities, and click the Web CMS Vulnerabilities tab.**
3. **View all vulnerabilities.**

   Note:
   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.**

   Note:
   You can quickly locate specific affected assets by using the search and filter functions.
5. **Select an action based on the impact of the vulnerability.** *Table 13-8: Actions on vulnerabilities* **describes the actions.**

   Table 13-8: Actions on vulnerabilities

Action	Description
Fix Now	Select this option to fix the Web CMS 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 this vulnerability. For more information about the paths of the Web files, see the paths specified in the vulnerability remarks.

Action	Description
Ignore	Select this option to ignore the vulnerability. The system does not alert you about ignored vulnerabilities.
Verify	Click Verify to verify the vulnerability fix. If you do not manually verify the fix, the system automatically verifies the fix within 48 hours.
Undo Fix	For vulnerabilities that have been fixed, click Undo Fix to restore the Web files that have been replaced.

You can manage a vulnerability on one or more affected assets at one time.

- To manage a vulnerability on one asset, select an action from the Actions column of this asset.
- To manage a vulnerability on one or more assets, select one or more affected servers, and select an action in the lower-left corner.

13.6.3.1.4 Manage other vulnerabilities

This topic describes how to manage other vulnerabilities.

Context

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 whether the fix is successful.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Server Security > Intrusion Prevention > Vulnerabilities, and click the Others 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 based on the impact of the vulnerability.** *Table 13-9: Actions on vulnerabilities* **describes the actions.**

   Follow the instructions to manually fix the vulnerabilities on the Others tab page.

   Table 13-9: Actions on vulnerabilities

Action	Description
Ignore	Select this option to ignore a vulnerability. The system does not alert you about an ignored vulnerability.
Verify	Click Verify to verify the fix after you have manually fixed a vulnerability. If you do not manually verify a fix, the system automatically verifies the fix within 48 hours after the vulnerability fix is complete.

   You can fix a vulnerability for one or multiple affected assets at one time.

   - **To fix a vulnerability for one affected asset, select an action from the Actions column of the asset.**
   - **To fix a vulnerability for one or multiple affected assets, select the target servers, and select an action in the lower-left corner.**

13.6.3.1.5 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 configure a vulnerability whitelist.

Context

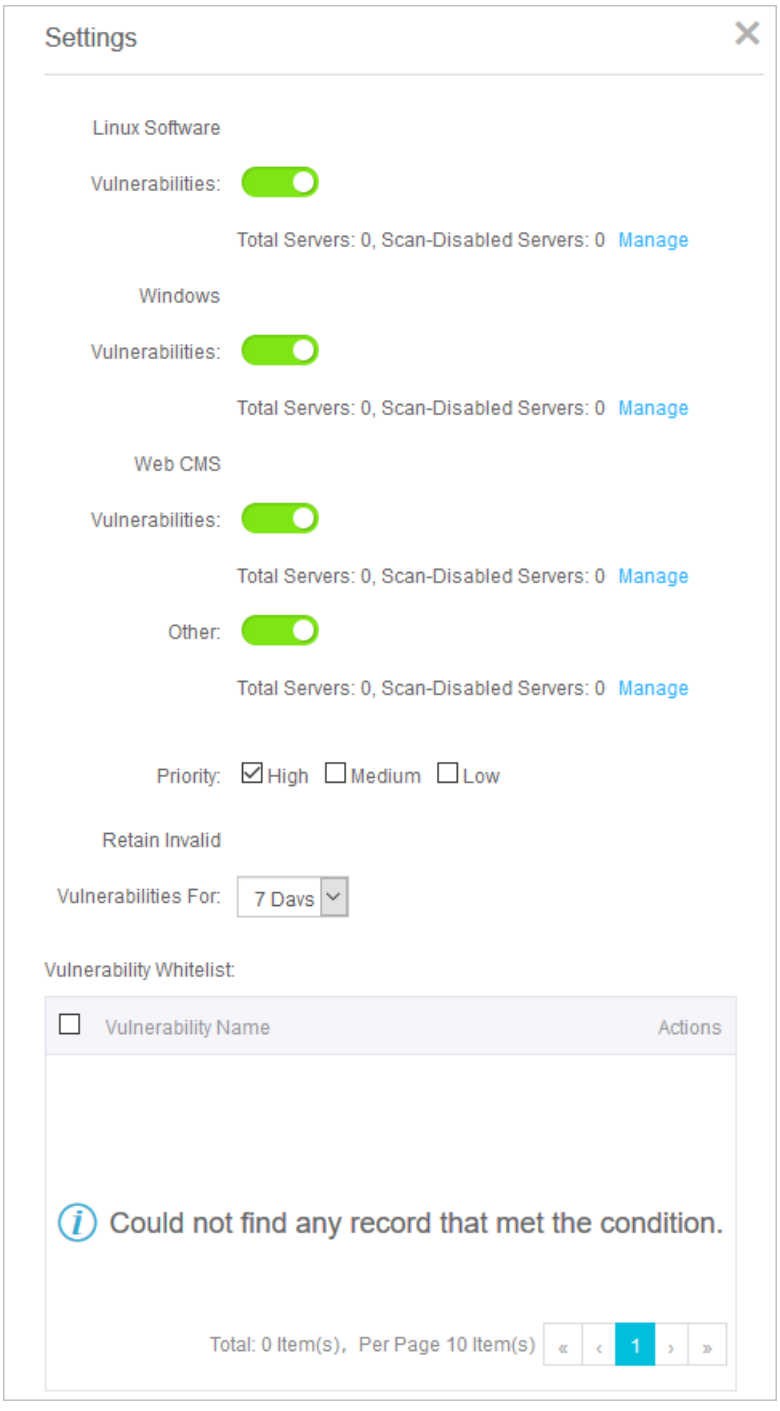
A vulnerability whitelist allows you to exclude vulnerabilities from the detection list. You can add multiple vulnerabilities in the vulnerability list to the whitelist . The system does not detect whitelisted vulnerabilities. You can manage the vulnerability whitelist on the vulnerability management settings page.

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 policies.**

   Figure 13-4: 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 the priorities of vulnerabilities to be detected. The priorities include `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 is invalid. The system deletes 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.

13.6.3.2 Baseline check

13.6.3.2.1 Overview

The baseline check feature automatically checks the security configurations on servers, and provides the detailed check results and suggestions for baseline reinforcement.

Features

After you enable the baseline check feature, Apsara Stack Security automatically checks for risks related to the operating systems, accounts, databases, passwords, and security compliance configurations of your servers, and provides reinforcement suggestions. For more information, see *Baseline check items*.

By default, a full baseline check is performed automatically from 0:00 to 6:00 every day. You can create and manage the scan policies. When you create or modify a policy, you can customize the check items, interval, and time period of a baseline check, and select the servers to which you want to apply this policy. For more information, see *Add a custom baseline check policy*.

Notes

The following check items are disabled by default. To check these items, make sure that these items do not affect your business and select them when you customize a scan policy.

- **Weak password check for specific applications such as MySQL, PostgreSQL, and SQL Server**

  > **Note:**
  > **When you run these check items, you attempt to log on to servers with weak passwords. The logon attempts consume server resources and generate many logon failure records.**

- **Check items related to classified protection**
- **Check items related to the Center for the Internet Security (CIS) standard**

Baseline check items

Category	Check item
Database	Memcached security baseline check
	Redis security baseline check
Operating system	Security baseline check based on the Alibaba Cloud standard · Windows Server 2008 R2 · Windows Server 2012 R2 · Windows Server 2016 R2 · Ubuntu · Debian Linux 8 · CentOS Linux 6 · CentOS Linux 7
	Security baseline check based on the CIS standard · Windows Server 2008 R2 · Windows Server 2012 R2 · Windows Server 2016 R2 · Ubuntu 14 · Ubuntu 16 · Debian Linux 8 · CentOS Linux 6 · CentOS Linux 7

Category	Check item
	Compliance baseline check based on Grade II Protection of Information Security · Windows Server 2008 R2 · Windows Server 2012 R2 · Windows Server 2016 R2 · Ubuntu · Debian Linux 8 · CentOS Linux 6 · CentOS Linux 7
	Compliance baseline check based on Grade III Protection of Information Security · Windows Server 2008 R2 · Windows Server 2012 R2 · Windows Server 2016 R2 · Ubuntu · Debian Linux 8 · CentOS Linux 6 · CentOS Linux 7
Weak password	Weak password check for Linux
	Anonymous FTP logon check
	Weak password check for Microsoft SQL Server
	Weak password check for MySQL
	Weak password check for PostgreSQL
	Weak password check for Windows
	Weak password check for FTP
Middleware	Apache Tomcat security baseline check

13.6.3.2.2 Add a custom baseline check policy

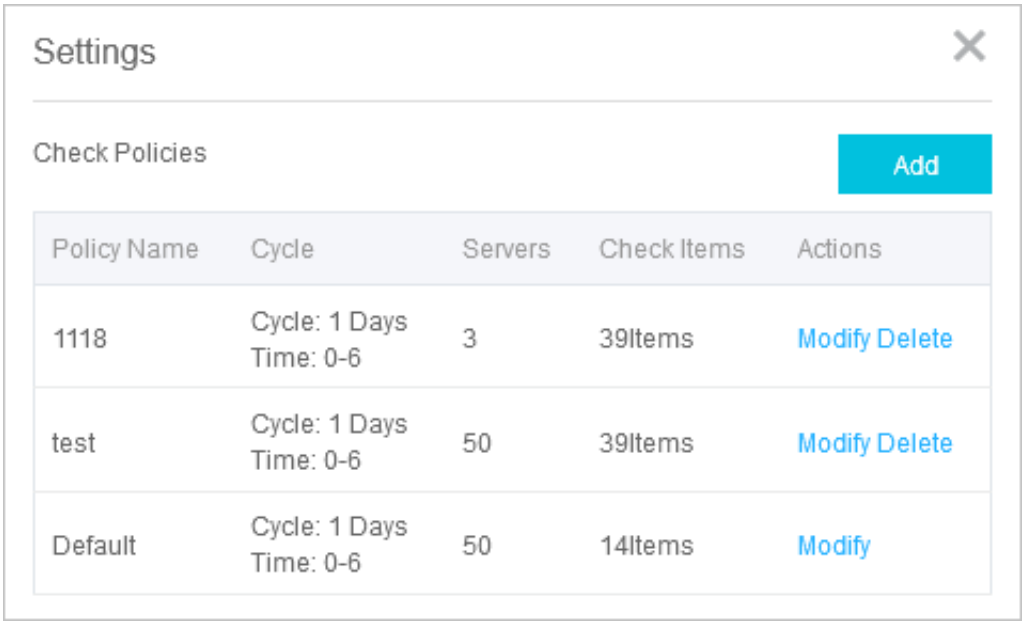
This topic describes how to add a custom baseline check policy.

Context

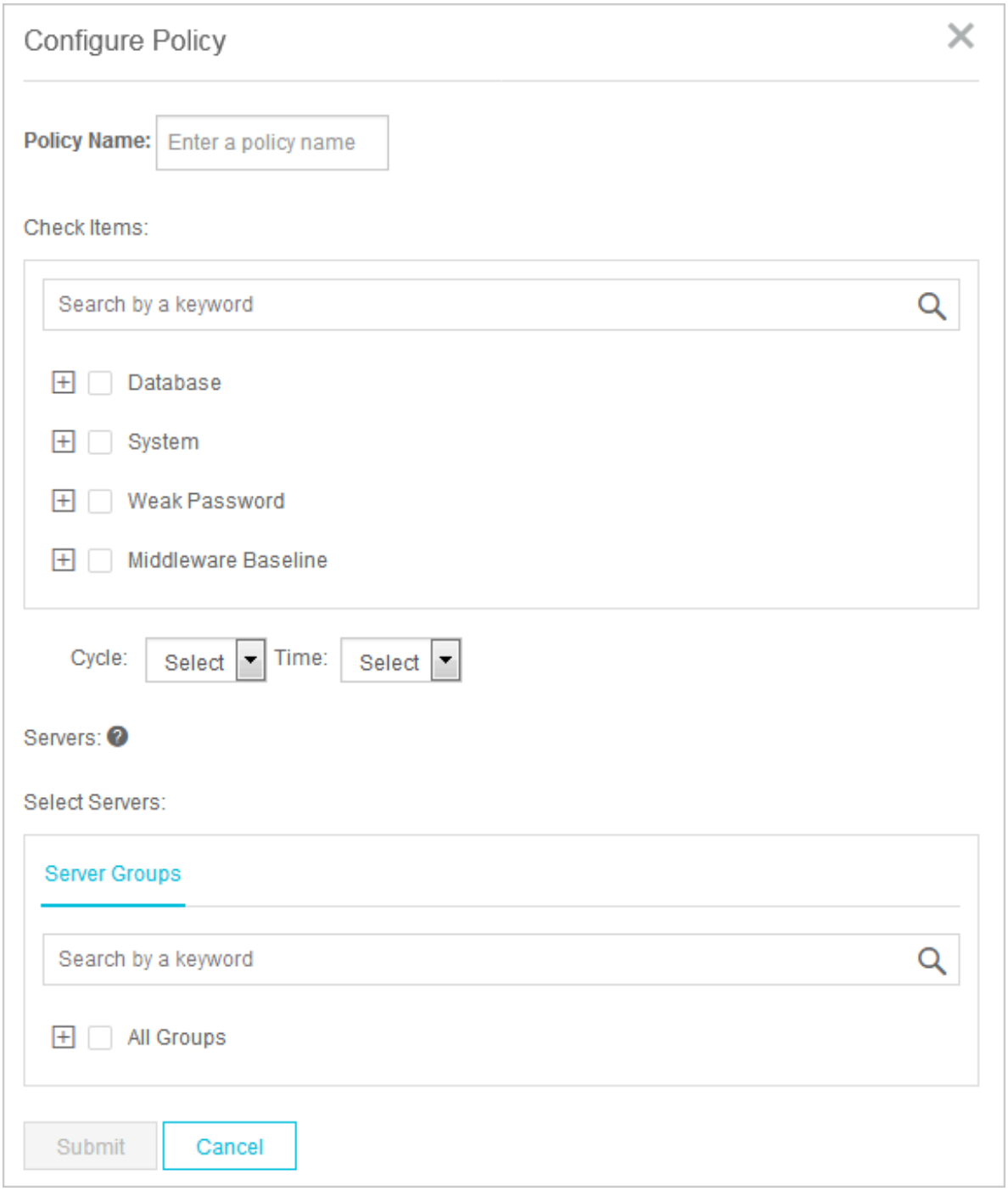
By default, the baseline check feature uses the default policy to check the baseline security of assets. You can also customize baseline check policies based on your business needs, for example, to check the compliance with Grade II Protection of Information Security.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Server Security > Threat Prevention > Baseline Check.**
3. **On the Baseline Check page, click Settings in the upper-right corner.**
4. **In the Settings window, click Add.**



5. In the Configure Policy window, set the policy parameters.



Parameter	Description
Policy Name	The name of the custom policy.
Check Items	The items that the custom policy checks. For more information, see *Baseline check items*.

Parameter	Description
Cycle	The check interval and period. · Check interval: the frequency that a baseline check is performed. Valid values: `1 Day`, `3 Days`, `7 Days`, and `30 Days`. · Check period: the period during which a baseline check is performed. Valid values: `00:00-06:00`, `06:00-12:00`, `12:00-18:00`, and `18:00-24:00`.
Servers	The group of servers on which the baseline check is performed. · By default, newly created servers are in the Default group. To apply the policy to newly created servers, you must select the Default group. · For more information about how to manage server groups, see *Manage server groups*.

6. Click Submit.

13.6.3.2.3 Manage baseline check settings

This topic describes how to manage baseline check settings.

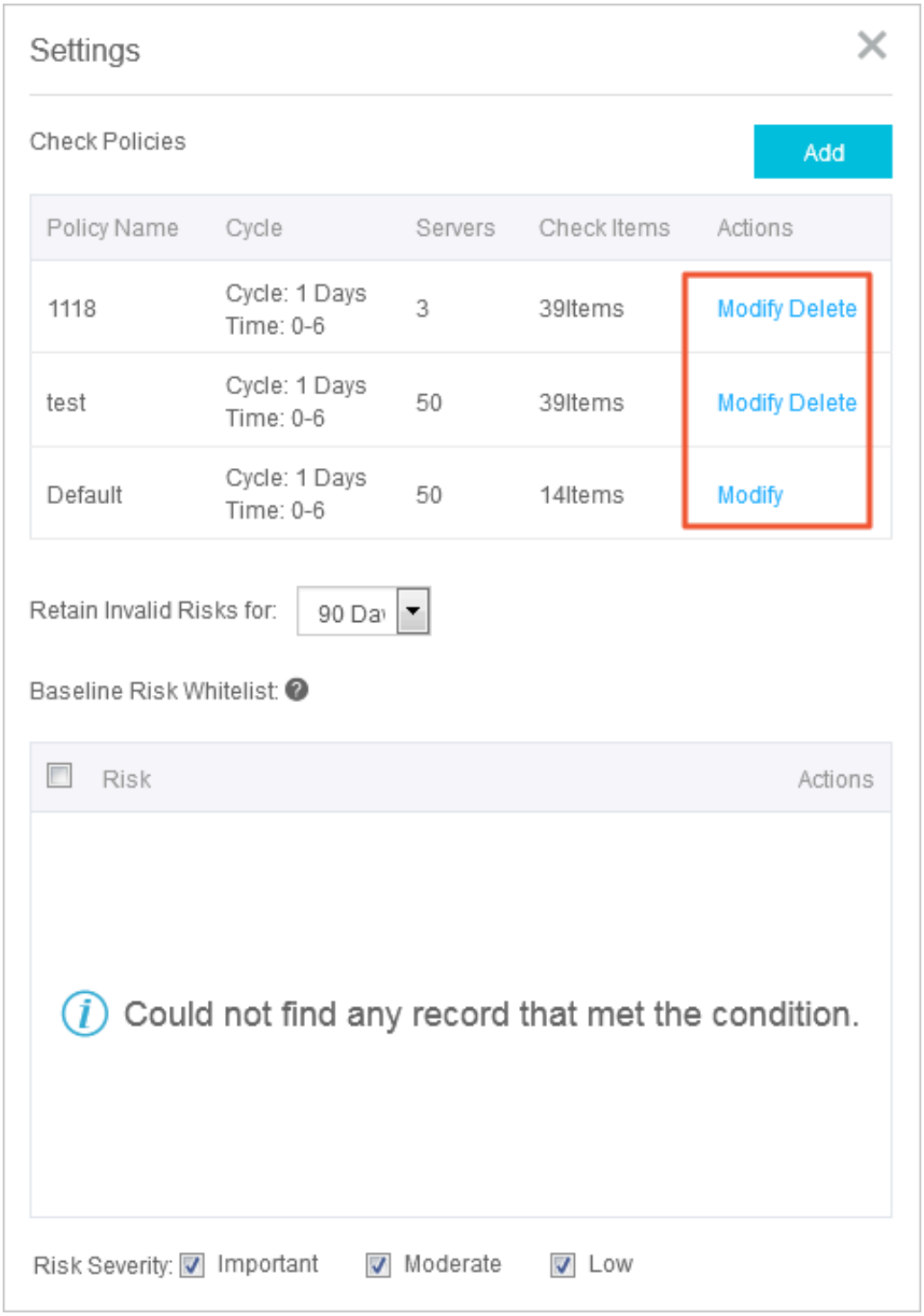
Context

You can manage baseline check settings, such as the scan policies, baseline whitelist, and baseline risk levels.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Server Security > Threat Prevention > Baseline Check.
3. On the Baseline Check page, click Settings in the upper-right corner.

4. In the Settings window, manage the scan policies.



- Delete a scan policy.

  In the Actions column of the target policy, click Delete.

- Edit a scan policy.

  In the Actions column of the target policy, click Edit. For more information about how to set the parameters, see *Add a custom baseline check policy*.

Note:

You cannot delete the default policy or modify the check items of the default policy. However, you can modify the server group to which the default policy applies.

5. **Set Retain Invalid Risks for.**

   Select a time period from the drop-down list. Valid values: `90 Days`, `30 Days`, **and** `7 Days`.

6. **Manage the check items in the Baseline Risk Whitelist section.**

   The baseline check feature does not check the items added to the whitelist. To check an item that has been added to the whitelist, click Remove in the Actions column of the item to remove it from the whitelist.

7. **Set Risk Severity.**

   You can set Risk Severity to `Important`, `Moderate`, **or** `Low`. **After you set Risk Severity, the baseline check feature only reports the baseline risks of the corresponding risk level.**

   For example, if you set Risk Severity to `Important`, **only the baseline risks of the Important risk level appear on the Baseline Check tab.**

13.6.3.2.4 View baseline check results and resolve baseline risks

This topic describes how to view baseline check results and resolve baseline risks.

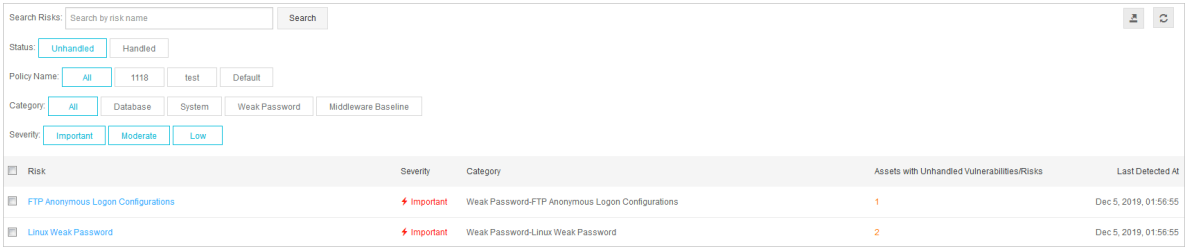
Context

A security baseline is a series of security configuration standards used to identify the basic protection capabilities of devices and systems in a network environment.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Server Security > Threat Prevention > Baseline Check.**

3. View baseline check results.



- Search for risks.

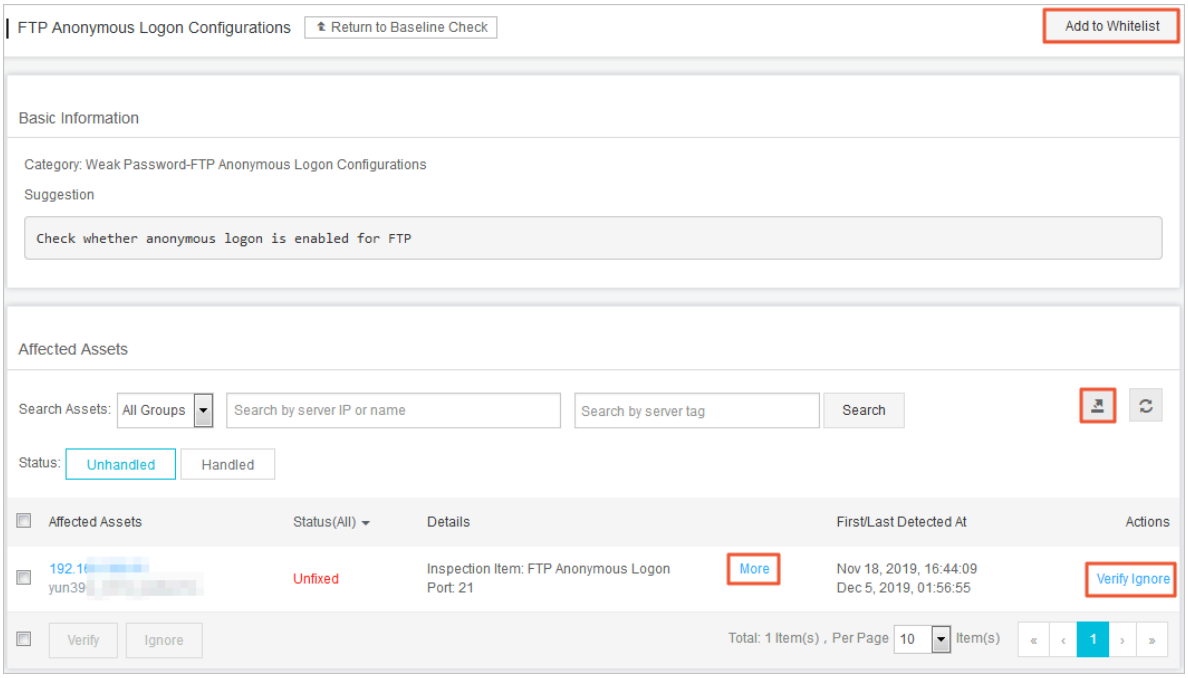
  Enter a keyword in the Search Risks field to search for risks.

- Filter risks.

  Set Status, Policy Name, Category, and Severity to filter risks.

> **Note:**
> For more information about how to set Severity, see the method for setting the Risk Severity parameter in *Manage baseline check settings*.

4. Manage a single baseline risk.



Determine the impact of the risk on your servers.

- To export the list of affected assets, click .
- If you are certain that this risk does not affect the security of your servers and you do not want the system to report the risk in the future, click Add to Whitelist in the upper-right corner.

> **Note:**
> If you want the system to report the risk again, you can remove the risk from the whitelist. For more information, see the method for setting the Baseline Risk Whitelist parameter in *Manage baseline check settings*.

- If the risk does not affect the security of some servers, click Ignore.
- If the risk affects the security of your servers, follow these steps to reinforce the servers:

a) Click the name of the baseline risk.

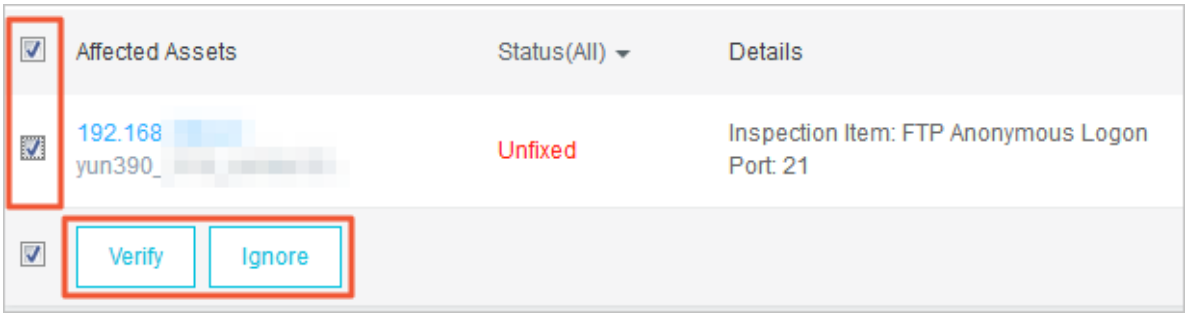
b) View the basic information about the baseline risk and the affected assets.

c) In the Details column, click More to view the check items and reinforcement suggestions.

d) Manually reinforce the relevant servers based on the reinforcement suggestions.

e) After reinforcing the servers, click Verify to check the reinforcement result.

5. Verify or ignore a risk on multiple servers at the same time.



- If a risk does not affect the security of some servers, select these servers and click Ignore to ignore the risk on these servers.
- If the affected servers have been reinforced, select these servers and click Verify to check the security reinforcement results on these servers.

13.6.4 Intrusion detection

13.6.4.1 Unusual logons

13.6.4.1.1 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.

13.6.4.1.2 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 the 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 baseline risks, or specify a blacklist or a whitelist. Then, click Label as Handled.**

13.6.4.1.3 Configure logon security

This topic describes how to configure logon security. You can set approved locations, IP addresses, time periods, and accounts for logons.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **In the left-side navigation pane, choose Server Security > Intrusion Detection.**
3. **Click the Unusual Logons tab.**
4. **Click Logon Security in the upper-right corner of the page.**

5. Set approved logon locations.

   To add an approved logon location, follow these steps:

   a) Click Add.

   b) Select a logon location from the drop-down list.

   c) Specify the servers on which the selected logon location takes effect.

      - Click All Servers to select specific servers.
      - Click Server Groups to select servers by group.

   d) Click OK.

   > **Note:**
   > Click Modify or Delete to modify or delete an approved logon location.

6. Set approved logon IP addresses.

   Turn on the Disapproved IP Alert switch. The switch is turned on if it turns green.

   To add an approved logon IP address, follow these steps:

   a) Click Add.

   b) In the Specify an Approved Logon IP section, enter an IP address.

   c) Specify the servers on which the specified IP address takes effect.

      - Click All Servers to select specific servers.
      - Click Server Groups to select servers by group.

   d) Click OK.

   Click Modify or Delete to modify or delete an approved logon IP address.

7. **Set approved logon time periods.**

   Turn on the Disapproved Time Alert switch. The switch is turned on if it turns green.

   To add an approved logon time period, follow these steps:

   a) Click Add.

   b) In the Specify an Approved Logon Duration section, specify a time period.

   c) Specify the servers on which the specified logon time period takes effect.

   - **Click All Servers to select specific servers.**
   - **Click Server Groups to select servers by group.**

   d) Click OK.

   Click Modify or Delete to modify or delete an approved logon time period.

8. **Set approved accounts.**

   Turn on the Disapproved Account Alert switch. The switch is turned on if it turns green.

   To add an approved account, follow these steps:

   a) Click Add.

   b) In the Specify an Approved Account section, enter an account.

   c) Select the servers on which the specified account takes effect.

   - **Click All Servers to select specific servers.**
   - **Click Server Groups to select servers by group.**

   d) Click OK.

   Click Modify or Delete to modify or delete an approved account.

13.6.4.2 Webshells

13.6.4.2.1 Manage webshells

This topic describes how to view and quarantine webshell files.

Context

Server Guard scans the Web directory on your server to check whether any webshell file exists. If a webshell file is detected, an alert is triggered.

Server Guard detects webshell files in PHP, JSP, or other common formats in real time or at scheduled time locally or in the cloud. Server Guard also allows you to quickly quarantine the detected webshell files.

Server Guard detects webshell files through the following methods:

- Dynamic detection: When any file in the Web directory is modified, Server Guard scans the modified content.
- Scheduled detection: Server Guard scans the entire Web directory every early morning.

> **Note:**
> By default, scheduled detection is enabled for all servers protected by Server Guard. To enable scheduled detection for a specific server, choose Settings > Security Settings. In the Trojan Scan area, click Manage on the right of Web Directory Periodic Scan, and specify the server for which you want to enable scheduled detection.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Server Security > Intrusion Detection > Webshells.
3. Select a server to check the detected webshell files.
4. Handle the webshell files.
   - Click Quarantine to quarantine a file. You can select and quarantine multiple files at one time.
   - Click Restore to restore a file that has been quarantined by mistake.
   - Click Ignore to ignore a file. Server Guard no longer generates alerts for an ignored file.

> **Note:**
> Server Guard does not delete webshell files on your server. Instead, it quarantines the files. You can restore a quarantined file if you determine that the file is trusted. After a webshell file is restored, Server Guard no longer generates alerts for this file.

13.6.4.3 Suspicious servers

13.6.4.3.1 Manage server exceptions

This topic describes how to view the 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.**

Action	Description
Handle	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.

13.6.5 Server fingerprints

13.6.5.1 Manage listening ports

Security Center regularly collects information about listening ports on a server.

Context

This task is applicable to the following scenarios:

- **Check for servers that listen to the specified port.**
- **Check for the listening ports of a server.**

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Server Security > Server Security, and click the Listening Ports tab.**
3. **View the listening ports, network protocol, and the number of servers.**

   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.**

13.6.5.2 Manage processes

Security Center regularly collects the process information on a server.

Context

This task is applicable to the following scenarios:

- **Check for servers that run the specified process.**
- **Check for 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.**

13.6.5.3 Manage account information

Security Center regularly collects the account information on a server.

Context

This task is applicable to the following scenarios:

- **Check for 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.**

13.6.5.4 Manage software versions

Security Center regularly collects software version information of a server.

Context

This task is applicable to the following scenarios:

- **Check for software that has been installed without authorization.**
- **Check for software of outdated versions.**
- **Quickly locate the affected assets when vulnerabilities are detected.**

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.**

13.6.5.5 Set the server fingerprint refresh frequency

You can set the frequency at which the data of running processes, system accounts, listening 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 in the upper-right corner.**
3. **Select the refresh frequency from each drop-down list.**
4. **Click OK.**

13.6.6 Log retrieval

13.6.6.1 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 13-10: 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

13.6.6.2 Log retrieval

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 13-11: Search condition parameters

Parameter	Description
Log source	**Select a supported log source. For more information, see *Table 13-12: Log sources*.**
Field	**Select a field that is supported by the specified log source. For more information, see *Table 13-12: Log sources*.**
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 13-13: 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 use.**
   - **Saved Searches: Click Saved Searches to select and apply a search condition configuration that has been saved.**

13.6.6.3 Supported log sources and fields

This topic describes log source types and fields that are supported by the Log Retrieval feature.

Log Retrieval allows you to query the following types of logs. You can click a log source to view the fields that can be retrieved.

Table 13-12: Log sources

Log source	Description
Account	The log entries of successful system logons.
Brute Force	The log entries of system logon failures generated during brute-force attacks.
Process Snapshot	The log entries of server processes at a specific time.
Network Snapshot	The log entries of listening ports on a server at a specific time.
Account Snapshot	The log entries of account logon information on a server at a specific time.
Process	The log entries of process startup on a server.
Network	The log entries of active connections from a server to external networks.

Account

The following table lists fields that are supported in queries.

Field	Date type	Description
uuid	String	The agent ID.
ip	String	The server IP address.
warn_ip	String	The source IP address.
warn_port	String	The logon port.
warn_user	String	The username for the logon.
warn_type	String	The logon type.
warn_count	String	The number of logon attempts.

Brute Force

The following table lists fields that are supported in queries.

Field	Date type	Description
uuid	String	The agent ID.
ip	String	The server IP address.

Field	Date type	Description
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.

Process

The following table lists fields that are supported in queries.

Field	Date 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 ID of the parent process.
uid	String	The ID of the user.
username	String	The name of the user.
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.

Network Snapshot

The following table lists fields that are supported in queries.

Field	Date type	Description
uuid	String	The agent ID.
ip	String	The server IP address.

Field	Date type	Description
src_port	String	The listening port.
src_ip	String	The listening IP address.
proc_path	String	The process path.
pid	String	The process ID.
proc_name	String	The name of the process.
proto	String	The protocol.

Account Snapshot

The following table lists fields that are supported in queries.

Field	Date type	Description
uuid	String	The agent ID.
ip	String	The server IP address.
perm	String	Indicates whether the agent has root permissions.
home_dir	String	The home directory.
warn_time	String	The time when the password expiration notification was sent.
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 time when the account expired.
passwd_expire	String	The time when the password expired.
last_logon	String	The last logon time.

Field	Date type	Description
user	String	The user.
status	String	The user status. Valid values: · 0: disabled · 1: normal

Process Snapshot

The following table lists fields that are supported in queries.

Field	Date type	Description
uuid	String	The agent ID.
ip	String	The server IP address.
path	String	The process path.
start_time	String	The time when the process was started.
uid	String	The ID of the user.
cmdline	String	The command line.
pname	String	The name of the parent process.
name	String	The name of the process.
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.

Network

The following table lists fields that are supported in queries.

Field	Date type	Description
uuid	String	The agent ID.
ip	String	The server IP address.
src_ip	String	The source IP address.

Field	Date type	Description
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.

13.6.6.4 Logical operators

The Log Retrieval feature supports multiple search conditions. You can add multiple logical operators in one search condition for one log source, or combine multiple search conditions for several log sources by using different logical operators. This topic describes the logical operators that are supported in log queries. Examples are provided to help you understand these operators.

The following table describes the logical operators that are supported in log queries.

Table 13-13: Logical operators

Logical operator	Description
and	**Binary operator.** **This operator is in the format of `query1 and query2`, which indicates the intersection of the query results of `query1` and `query2`.** **Note:** **If no logical operators are used for multiple keywords, the default operator is "and".**
or	**Binary operator.** **This operator is in the format of `query1 or query2`, which indicates the union of the query results of `query1` and `query2`.**

Logical operator	Description
not	**Binary operator.** **This operator is in the format of `query1 not query2`, which indicates a result that matches `query1` but does not match `query2`. This format is equivalent to `query1 - query2`.** **Note:** **`not query1` indicates that the log data that does not contain the query results of `query1` is returned.**

13.6.7 Settings

13.6.7.1 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. **In the left-side navigation pane, choose Server Security > Settings.**
3. **Enable periodic Trojan scan for servers.**
   a) **Click Manage in the Trojan Scan section.**
   b) **Select servers that require periodic Trojan scan from the All Servers section, and click the rightwards arrow.**
   c) **Click OK.**
4. **Specify the resource usage mode of the Server Guard agent for servers.**
   - **Business First Mode: The peak CPU utilization is less than 10% and the peak memory usage is less than 50 MB.**
   - **Protection First Mode: The peak CPU utilization is less than 20% and the peak memory usage is less than 80 MB.**

   a) **Click Manage in the Agent section.**
   b) **Specify the working mode of the Server Guard agent for servers.**
   c) **Click OK.**

13.6.7.2 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, such as Safedog and Yunsuo, on your server, the system may fail to install the Server Guard agent. We recommend that you disable or uninstall the security software before you install the 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 commands 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 use the preceding methods to manually reinstall the agent:

- **An image that includes the Server Guard agent is used to install the agent on multiple external servers at one time.**
- **You have copied the Server Guard agent files from a server that has been installed with the agent to your external servers.**

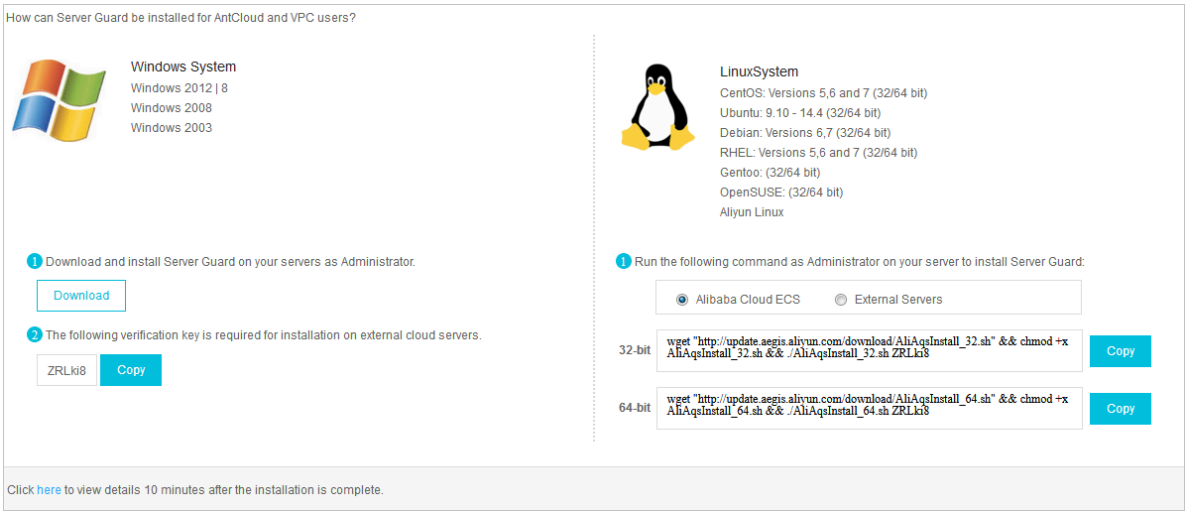
Procedure

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 13-5: Install the agent*.

   Figure 13-5: Install the agent



3. **Obtain and install the Server Guard agent based on the operating system type of your server.**

   - **Windows OS**

     a. **In the lower-left area 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 OS

  a. In the lower-right area of the page, select `Alibaba Cloud ECS` or `External Servers`.

  b. Select the installation command for your 32-bit or 64-bit 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 five minutes after you install 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.

13.6.7.3 Uninstall the Server Guard agent from a server

If you decide not to use any of the Server Guard features on your server, you can use the following procedure to uninstall the Server Guard agent.

Context

Before you uninstall the Server Guard agent from a 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 run the install command 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.

13.7 Application security

13.7.1 Quick start

This topic helps you get started with Web Application Firewall (WAF).

WAF uses intelligent semantic analysis algorithms to identify Web attacks. WAF also integrates a learning model to enhance its analysis capabilities and meet your daily security protection requirements without relying on traditional rule libraries.

The procedure for using WAF is as follows:

1. Customize WAF protection rules.

   WAF provides a default protection policy. You can also customize policies that suit your business needs.

   - For more information about how to configure protection policies, see *Configure protection policies*.
   - For more information about how to configure custom rules, see *Create a custom rule*.
   - For more information about how to configure HTTP flood protection rules, see *Configure an HTTP flood detection rule*.

2. Add protected websites.

   WAF can protect Internet websites and Virtual Private Cloud (VPC) websites.

   - For more information about how to add an Internet website to WAF for protection, see *Add an Internet website for protection*.
   - For more information about how to add a VPC website to WAF for protection, see *Add a VPC website for protection*.

3. Configure Domain Name System (DNS) resolution.

   For more information about how to change the DNS-resolved source IP address of a website to a virtual IP address of WAF, see *Modify DNS resolution settings*.

4. **View WAF protection results.**
   - **For more information about how to view the overall protection information, see** *View protection overview*.
   - **For more information about how to view the access status, see** *View Web service access information*.
   - **For more information about how to view the detection logs for Web attacks, see** *View attack detection logs*.
   - **For more information about how to view the detection logs for HTTP flood attacks, see** *View HTTP flood protection logs*.

13.7.2 Protection configuration

13.7.2.1 Configure protection policies

This topic describes how to configure Web Application Firewall (WAF) protection policies.

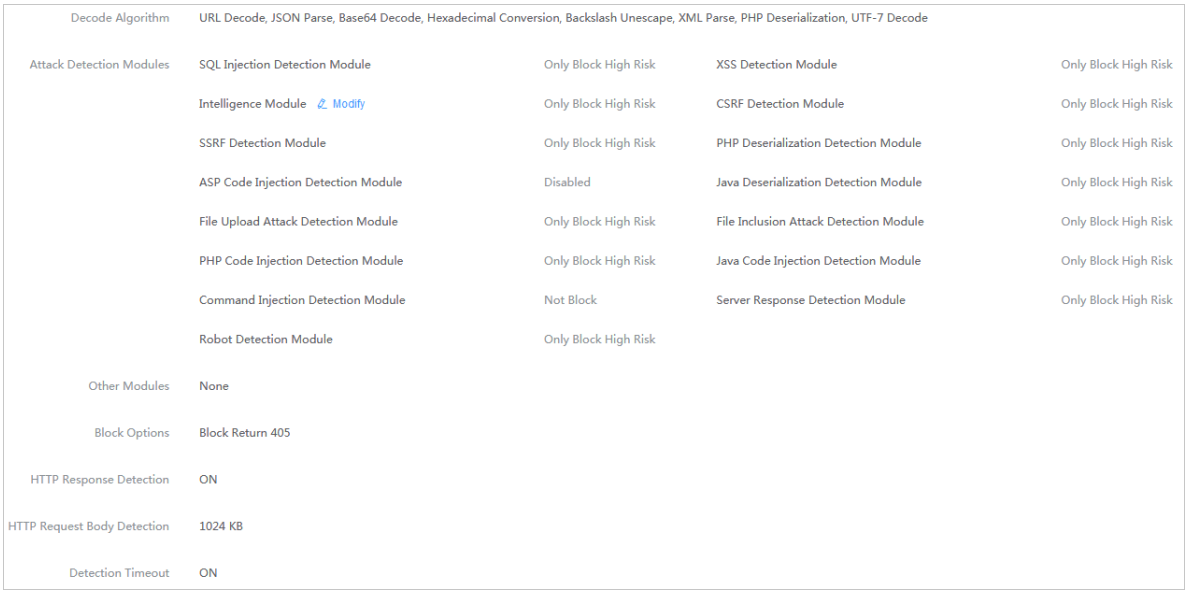
Context

WAF provides a default protection policy. You can also customize policies that suit your business needs.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Application Security > WAF. Choose Protection Configuration > Website Protection Policies.**
3. **Click Add Protection Policy. In the dialog box that appears, specify Policy Name, and click Confirm.**

4. Click the name of the new policy to modify the policy.



Parameter	Description
Decode Algorithm	Select algorithms for decoding the requests.
Attack Detection Modules	Specify the types of attacks to be detected and the risk levels of attacks to be blocked.
Block Options	Specify the status code and image to be returned when an attack is blocked.
HTTP Response Detection	Set the status of the Enable HTTP Response Detection toggle and specify Response Detection Max Body Size.
HTTP Request Body Detection	Specify Response Detection Max Body Size.

Parameter	Description
Detection Timeout	Set the status of the Enable Detection Timeout toggle, and specify Timeout Threshold.

For example, take the following steps to configure Attack Detection Modules:

a) Place the pointer over a specific module in the Attack Detection Modules area, for example, SQL Injection Detection Module, and click Modify.

b) In the SQL Injection Detection Module dialog box, set the detection parameters.

Parameter	Description
Enabled	Indicates whether to enable the detection module.
Blocking Threshold	You can select `Not Forbid`, `Only Forbid High Risk`, `Forbid Medium or High Risk`, or `Forbid All`.
Record Threshold	You can select `Not Record`, `Only Record High Risk`, `Record Medium or High Risk`, or `Record All`.
Detect Non-Injected SQL	Indicates whether to enable detection for non-injected SQL attacks.

c) Click Confirm.

5. **Manage the protection policies.**

   If you want to delete a protection policy, select this policy, and click Delete Selected Protection Policies.

   > **Note:**
   > **You cannot delete the default policy.**

13.7.2.2 Create a custom rule

This topic describes how to create a custom rule for Web Application Firewall (WAF).

Context

Security administrators can customize rules to meet various requirements for attack detection. The administrators can add, edit, or delete the rules in the console . You can use rules to filter requests that meet specific conditions.

Multiple custom rules have the OR logical relation. If two custom rules have the same conditions but differentiate in the operating mode such as blocking and allowing, the system runs the first rule.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Application Security > WAF. On the page that appears, choose Protection Configuration > Customized Rules.**

3. Click Add New Customized Rule in the upper-right corner of the page. In the Add Customized Rule dialog box that appears, set parameters.

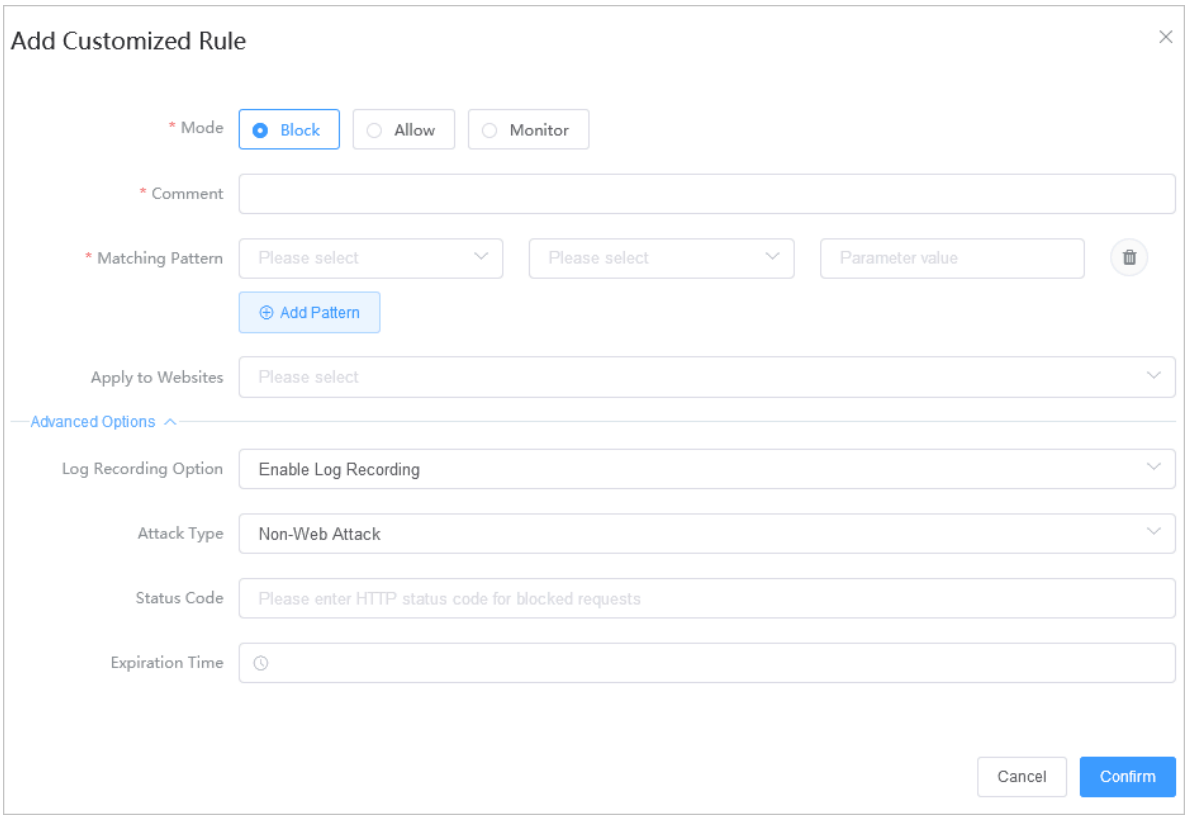


Table 13-14: Parameters for creating a custom rule

Parameter	Description
Mode	The operating mode of the rule. Valid values: · Block: An HTTP request is blocked if it meets the conditions of the rule. · Allow: An HTTP request is allowed if it meets the conditions of the rule. · Monitor: An HTTP request is allowed and recorded if it meets the conditions of the rule.
Comment	The remarks of the rule, such as the purpose of the rule.
Matching Pattern	The conditions that trigger the rule. Click Add Pattern to set multiple conditions. Multiple conditions have the AND logical relation. The custom rule takes effect only when all conditions are met.
Apply to Websites	The websites to be protected by the rule.

Parameter	Description
Log Recording Option	Indicates whether to record a protection event in the intrusion detection logs if the rule is triggered. The default value is Enable Log Recording. After Log Recording Option is set to Enable Log Recording, all interception records are recorded in the intrusion detection logs.
Attack Type	The type of attacks to be blocked by the rule.
Status Code	The status code to be displayed after an attack is blocked by the rule.
Expiration Time	The time when the rule expires.

4. Click Confirm.
5. Manage custom rules.
   - Edit a rule.

     In the Actions column corresponding to a rule, click the Edit icon.
   - Enable a rule.

     Select a rule and click Enable Selected Rules.
   - Disable a rule.

     Select a rule and click Disable Selected Rules.
   - Delete a rule.

     Select a rule and click Delete Selected Rules.

13.7.2.3 Configure an HTTP flood detection rule

This topic describes how to configure an HTTP flood detection rule.

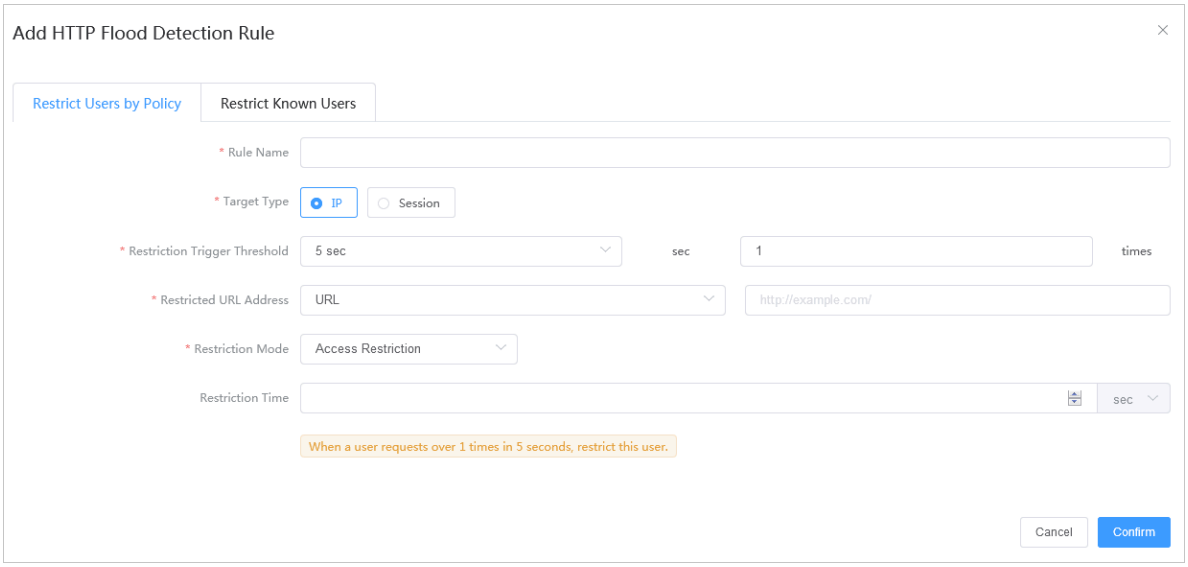
Context

An HTTP flood is a type of distributed denial of service (DDoS) attack that targets at web applications. Attackers use proxy servers or zombies to overwhelm targeted web servers by sending HTTP requests.

Create an HTTP flood detection rule for all users

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Application Security > WAF. On the page that appears, choose Protection Configuration > HTTP Flood Detection Rules.

3. **Click Add Flood Detection Rule in the upper-right corner. The Add HTTP Flood Detection Rule dialog box appears.**
4. **Click the Restrict Users by Policy tab, set parameters, and click Confirm.**

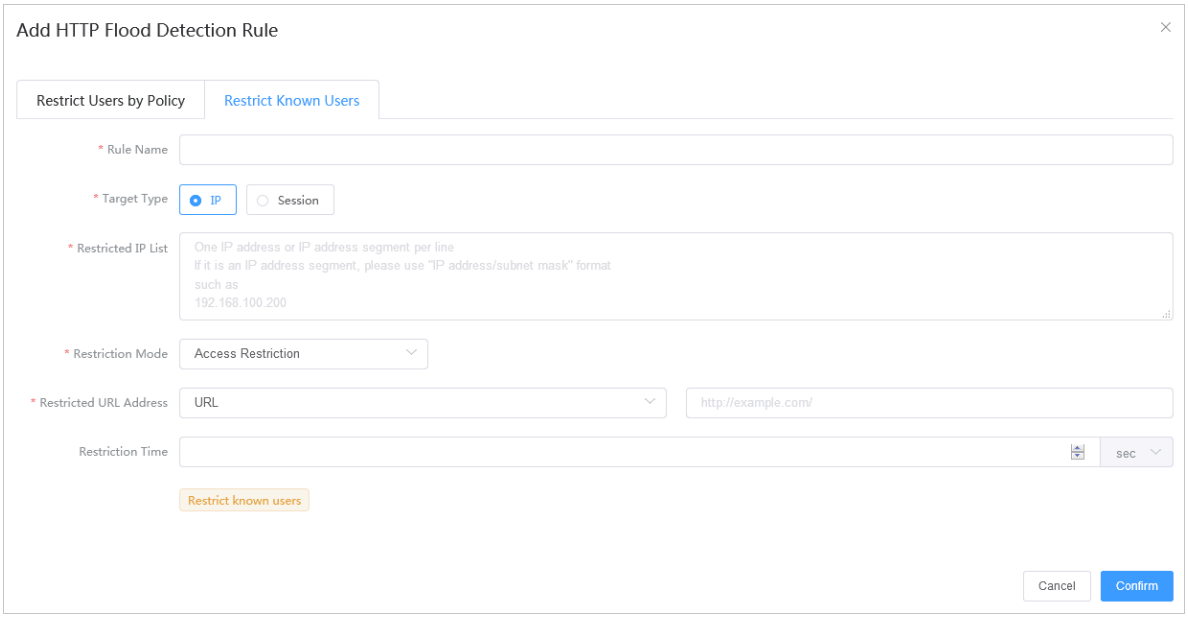


Parameter	Description
Rule Name	The name of the HTTP flood detection rule.
Target Type	The type of the restricted target. Valid values: `IP` and `Session`.
Restriction Trigger Threshold	The condition that triggers the restriction.
Restricted URL Address	The URL to be protected by the rule. Valid values: • `URL` • `URL Prefix`
Restriction Mode	The mode in which the requests from a user are restricted. Valid values: • `Access Restriction`: All requests from the restricted user to the specified URL are blocked. • `Access Frequency Restriction`: The frequency of access from the restricted user to the specified URL is limited.
Restriction Time	The time when the restriction takes effect.

Create an HTTP flood detection rule for known users

1. *Log on to Apsara Stack Security Center*.

2. In the left-side navigation pane, choose Application Security > WAF. On the page that appears, choose Protection Configuration > HTTP Flood Detection Rules.
3. Click Add Flood Detection Rule in the upper-right corner. The Add HTTP Flood Detection Rule dialog box appears.
4. Click the Restrict Known Users tab, set parameters, and click Confirm.



Parameter	Description
Rule Name	The name of the HTTP flood detection rule.
Target Type	The type of the restricted target. Valid values: `IP` and `Session`.
Restricted IP List/Restricted Session List	Enter the IP addresses or sessions to be restricted based on the setting of Target Type. Specify only one IP address or session in each line.
Restriction Mode	The mode in which the specified request source is restricted. Valid values: • `Access Restriction`: All requests from the restricted user to the specified URL are blocked. • `Access Frequency Restriction`: The frequency of access from the restricted user to the specified URL is limited.

Parameter	Description
Restricted URL Address	The URL to be protected by the rule. Valid values: • URL • URL Prefix
Restriction Time	The time when the restriction takes effect.

Manage HTTP flood detection rules

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Application Security > WAF. On the page that appears, choose Protection Configuration > HTTP Flood Detection Rules.
3. In the rule list, manage the added HTTP flood detection rules.
   - Search for a rule.

     Click Add Filter and add filter conditions to quickly locate an HTTP flood detection rule.
   - Enable a rule.

     Select a rule and click Enable Selected Rules.
   - Disable a rule.

     Select a rule and click Disable Selected Rules.
   - Delete a rule.

     Select a rule and click Delete Selected Rules.

13.7.2.4 Configure an HTTP flood protection whitelist

This topic describes how to configure an HTTP flood protection whitelist.

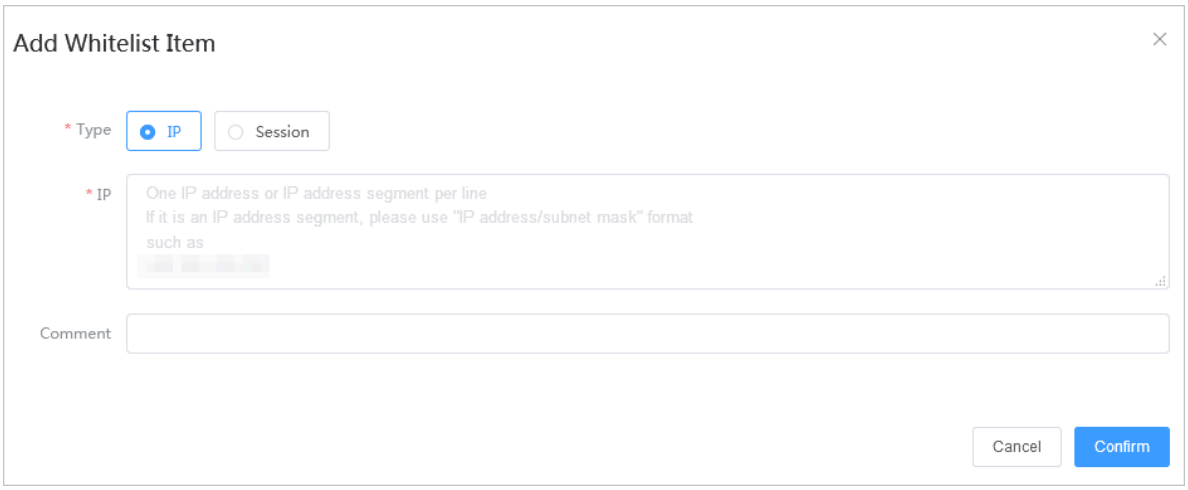
Context

If a request source is trusted, you can add this request source to an HTTP flood protection whitelist to allow the requests from this source.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Application Security > WAF. Choose Protection Configuration > HTTP Flood Detection Whitelist.

3. Click Add Whitelist Item to add a request source to the whitelist, and click Confirm.



Parameter	Description
Type	Set the type of the whitelisted request source to `IP` or `Session`.
IP/Session	Specify the IP addresses or sessions based on the selected Type. Specify one IP address or session in each line.
Comment	Enter remarks for the whitelist.

4. Manage the whitelisted users.
   - Search for a whitelisted user.

     Click Add Filter. Add filter conditions to locate specific whitelisted users.
   - Remove a whitelisted user.

     To remove a request source from the whitelist, select the specific rule, and click Delete Selected Items.

13.7.2.5 Add an Internet website for protection

This topic describes how to add an Internet website to Web Application Firewall (WAF) for protection.

Context

WAF can protect the following types of websites:

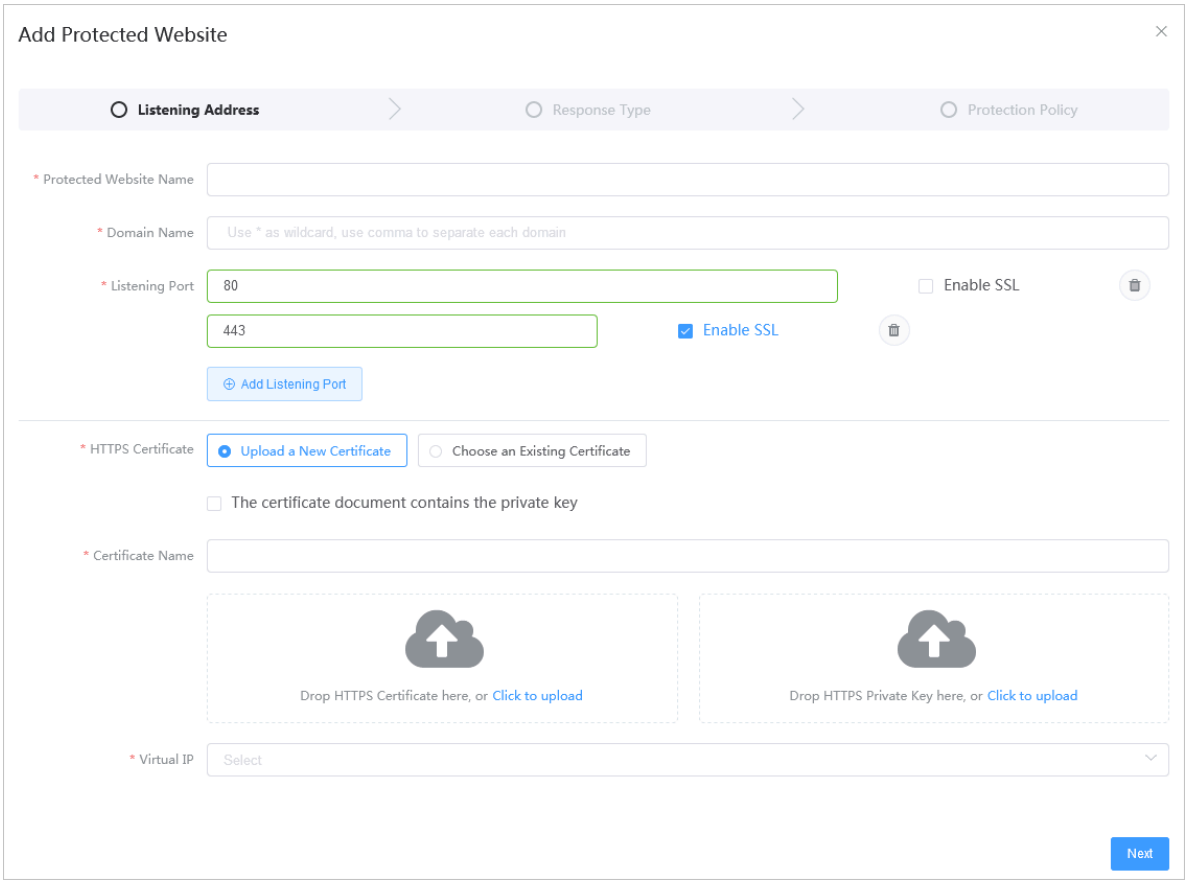
- Internet websites.

- Virtual Private Cloud (VPC) websites. For more information about how to add a VPC website for protection, see *Add a VPC website for protection*.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. In the left-side navigation pane, choose Application Security > WAF > Protection Configuration > Protected Websites. On the page that appears, click the Internet Websites tab.
3. Click Add Protected Website.
4. In the Listening Address step, set parameters and click Next.

   Set the Internet website to be protected. WAF can protect HTTP and HTTPS websites.



Parameter	Description
Protected Website Name	The name of the website to be protected.

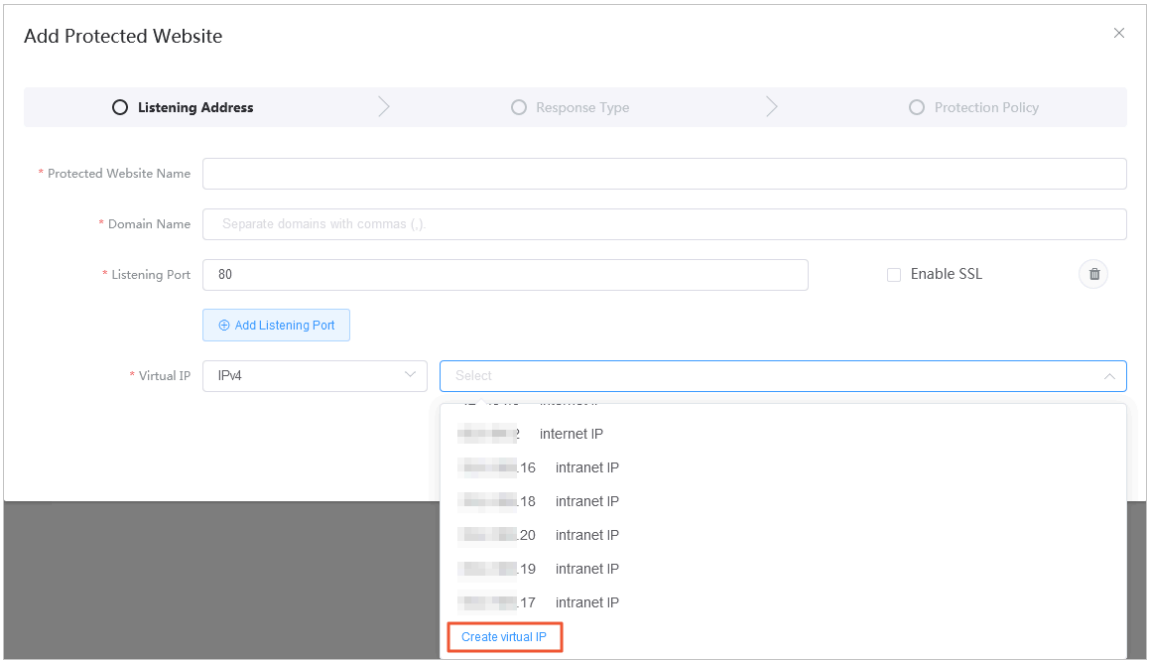
Parameter	Description
Domain Name	The domain name of the website. · Use an asterisk (*) as a wildcard domain name. · Separate multiple domain names with commas (,).
Listening Port	The port listened on WAF. · If the website can be accessed by using HTTPS, select the Enable SSL check box and upload the HTTPS certificate. · If the website can be accessed over multiple ports, click Add Listening Port to add a port.
HTTPS Certificate	The HTTPS certificate of the website. Valid values: · `Upload a New Certificate`: Select this option if the HTTPS certificate used by the website has not been uploaded to WAF before. By default, the HTTPS certificate and private key are uploaded separately. If you select the The certificate document contains the private key check box, you only need to upload a file that contains both the HTTPS certificate and private key. · `Choose an Existing Certificate`: If the HTTPS certificate used by the website has been uploaded to WAF before, select this option, and then select the HTTPS certificate from the drop-down list. Note: Set this parameter only when you select Enable SSL next to the Listening Port field.
Certificate Name	The name of the HTTPS certificate. Note: Set this parameter only when you select Enable SSL next to the Listening Port field.

Parameter	Description
Sections for uploading the HTTPS certificate and private key	Upload the HTTPS certificate and private key. By default, the HTTPS certificate and private key are uploaded separately. If you select the The certificate document contains the private key check box in the HTTPS Certificate section, you only need to upload a file that contains both the HTTPS certificate and private key. **Note:** Upload the HTTPS certificate and private key only when you select Enable SSL next to the Listening Port field.

Parameter	Description
Virtual IP	The virtual IP address of the website. WAF provides 10 virtual IP addresses by default. If there is no virtual IP address that you want to select, add one.

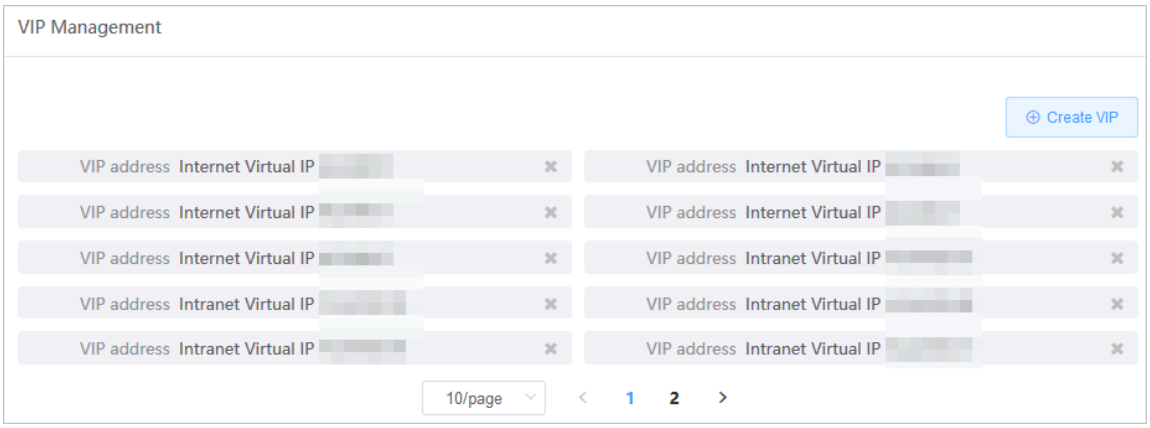
To add a virtual IP address, follow these steps:

a) In the Virtual IP drop-down list, scroll down to the bottom.

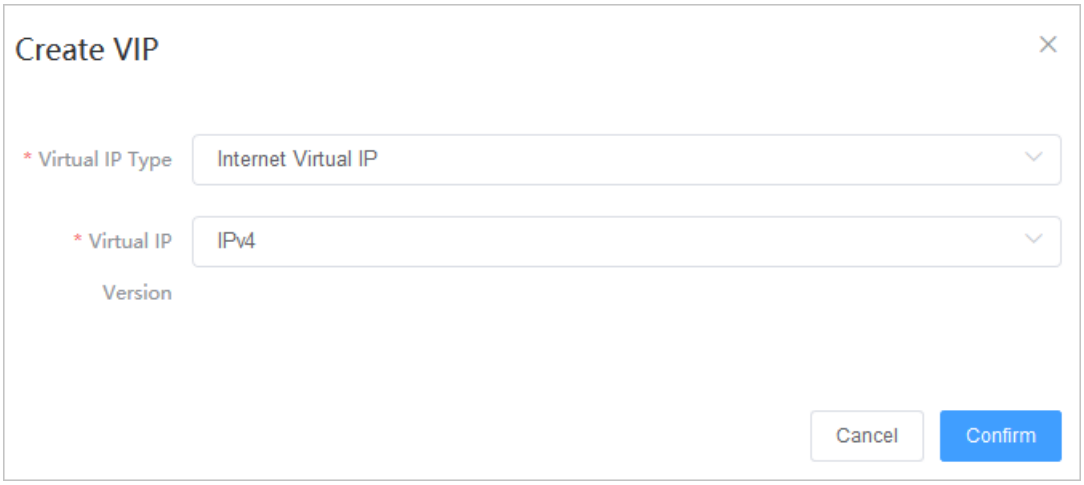


b) Click Create virtual IP.

c) On the VIP Management page, click Create VIP in the upper-right corner.

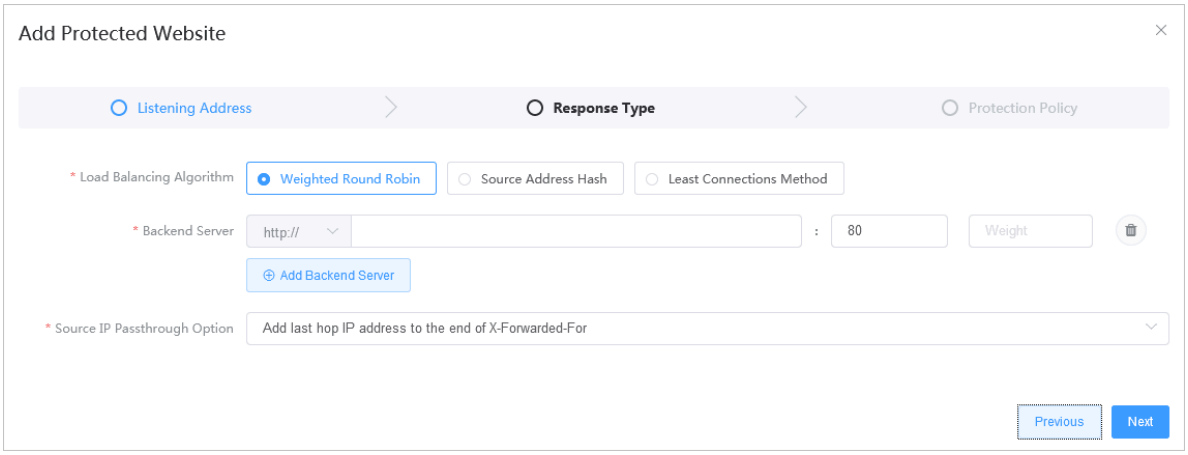


d) In the Create VIP dialog box that appears, set Virtual IP Type and Virtual IP Version.



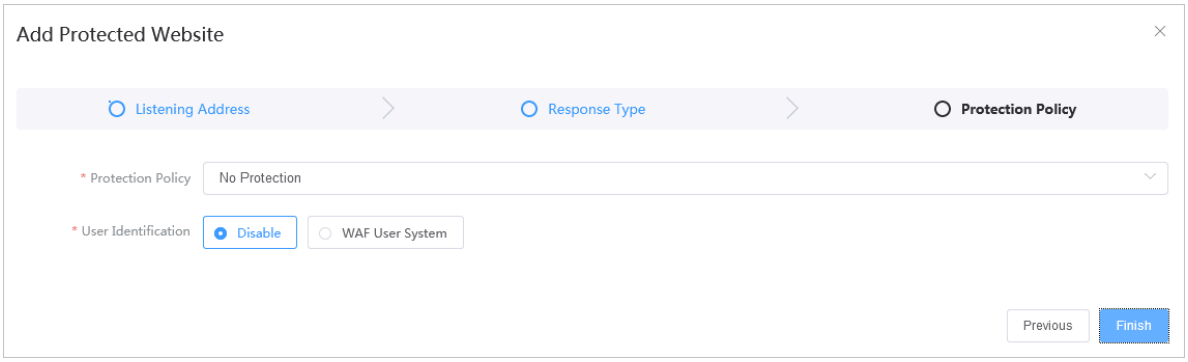
e) Click Confirm.

5. In the Response Type step, set parameters and click Next.



Parameter	Description
Load Balancing Algorithm	The algorithm for balancing load. Valid values: `Weighted Round Robin`, `Source Address Hash`, and `Least Connections Method`.
Backend Server	The address of the backend server.
Source IP Passthrough Option	The passthrough mode of the source IP address. The X-Forwarded-For (XFF) HTTP header field is a common method for identifying the original IP address of an HTTP client. It is used for request forwarding services such as HTTP proxy and load balancing.

6. In the Protection Policy step, set parameters and click Finish.



Parameter	Description
Protection Policy	The protection policy provided by WAF. For more information, see *Configure a protection policy*.
User Identification	Indicates whether to enable the user identification feature.

13.7.2.6 Add a VPC website for protection

This topic describes how to add a Virtual Private Cloud (VPC) website to Web Application Firewall (WAF) for protection.

Context

WAF can protect the following types of websites:

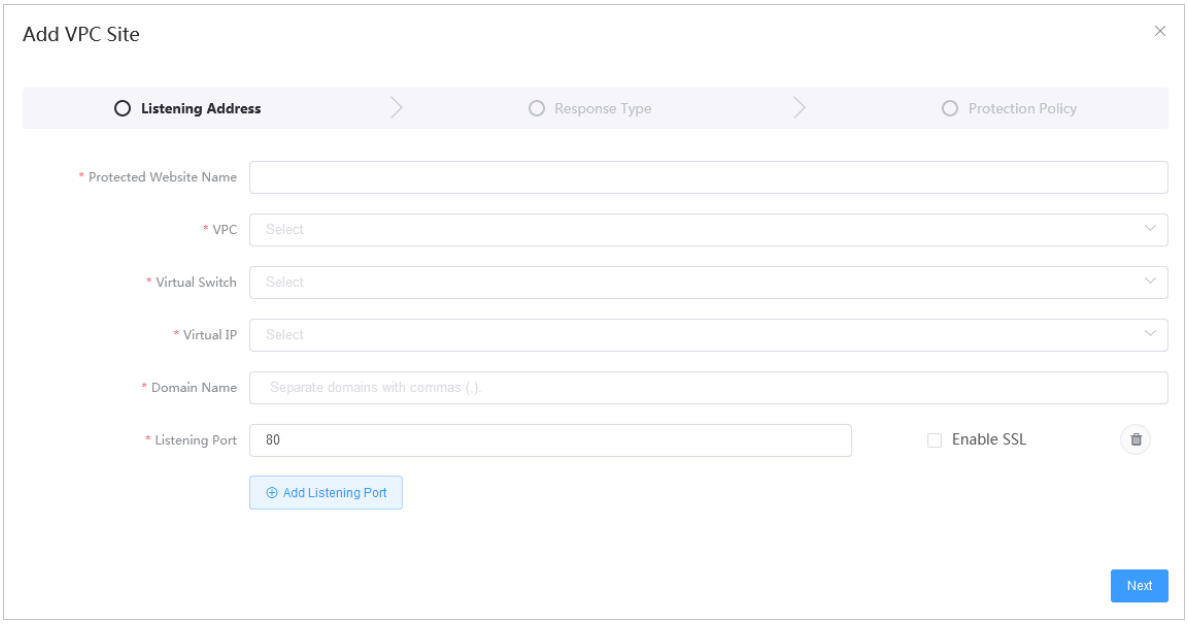
- Internet websites. For more information, see *Add an Internet website for protection*.
- VPC websites.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. Choose Application Security > WAF > Protection Configuration > Protected Websites. On the page that appears, click the VPC Websites tab.
3. Click Add VPC Website. The Add VPC Site wizard appears.

4. In the Listening Address step, set the parameters and click Next.

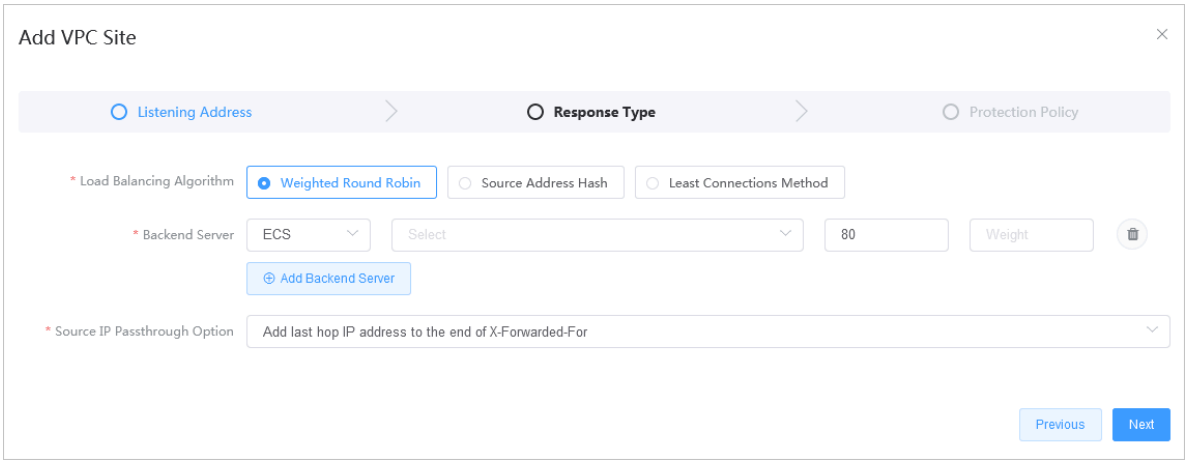
   Set the VPC website to be protected. WAF can protect HTTP and HTTPS websites.



Parameter	Description
Protected Website Name	The name of the website to be protected.
VPC	The VPC to which the website belongs.
Virtual Switch	The Vswitch to which the website belongs.
Virtual IP	The virtual IP address of the website.
Domain Name	The domain name of the website. · You can use an asterisk (*) as a wildcard. · Separate multiple domain names with commas (,).
Listening Port	The port listened by WAF. · If the website can be accessed through HTTPS, select the Enable SSL check box and upload the HTTPS certificate. · If the website can be accessed through multiple ports, click Add Listening Port to add a port.

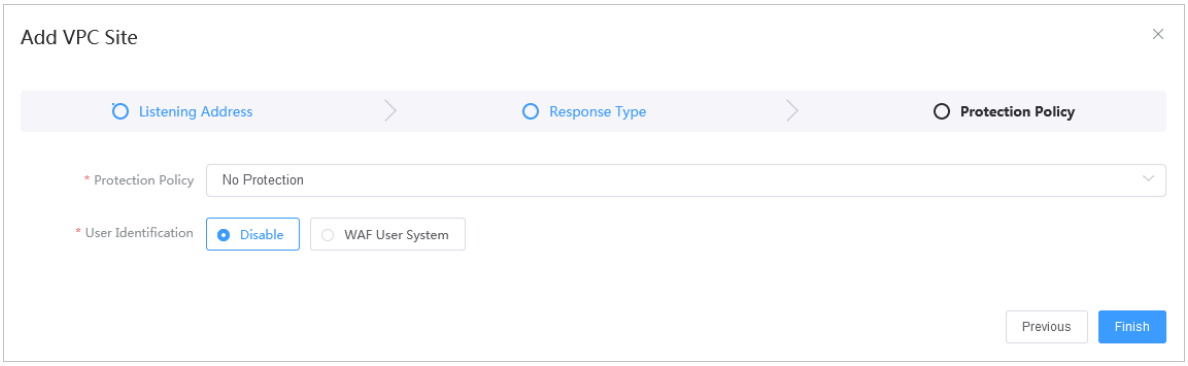
Parameter	Description
HTTPS Certificate	The HTTPS certificate of the website. Valid values: `Upload a New Certificate` and `Choose an Existing Certificate`. • `Upload a New Certificate`: Select this option if the HTTPS certificate used by the website has not been uploaded to WAF before. By default, the HTTPS certificate and private key are uploaded separately. If you select The certificate document contains the private key, you only need to upload a file that contains the HTTPS certificate and private key. • `Choose an Existing Certificate`: If the HTTPS certificate used by the website has been uploaded to WAF before, select this option, and then select the HTTPS certificate from the drop-down list. Note: Set this parameter only when you select Enable SSL in the Listening Port step.
Certificate Name	The name of the HTTPS certificate. Note: Set this parameter only when you select Enable SSL in the Listening Port step.
Areas for uploading the HTTPS certificate and private key	Upload the HTTPS certificate and private key. By default, the HTTPS certificate and private key are uploaded separately. If you select The certificate document contains the private key, you only need to upload a file that contains the HTTPS certificate and private key. Note: Upload the HTTPS certificate and private key only when you select Enable SSL in the Listening Port step.

5. In the Response Type step, set the parameters and click Next.



Parameter	Description
Load Balancing Algorithm	The algorithm for balancing load. Valid values: `Weighted Round Robin`, `Source Address Hash`, and `Least Connections Method`.
Backend Server	The address of the back-end server.
Source IP Passthrough Option	The transparent transmission mode of the source IP address. The X-Forwarded-For (XFF) HTTP header field is a common method for identifying the original IP address of an HTTP client. It is used in request forwarding services such as HTTP proxy and load balancing.

6. In the Protection Policy step, set the parameters and click Finish.



Parameter	Description
Protection Policy	The WAF protection policy. For more information, see *Configure protection policies*.
User Identification	Specifies whether to enable the user identification feature.

13.7.2.7 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 redirect 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 virtual IP 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 virtual IP and domain name to the** `hosts` **file on your local PC.**

   For example, 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 the following line at the end of the file:** `<Protected website virtual IP address><Protected website domain>`**.**

   ```
   # localhost name resolution is handled within DNS itself.
   #	127.0.0.1       localhost
   #	::1             localhost

   [redacted]4.115      example.com
   ```

   Note:
   The IP address in front of the domain name is the virtual IP address assigned by WAF.
3. **Ping the protected domain name from the local PC.**

   The resolved IP address must be the WAF virtual IP 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 connected to WAF, you can visit the website.

5. Verify the WAF protection feature.

   Simulate a Web attack request to check whether WAF blocks the request.

   For example, add `/? alert(xss)` after the URL. If you try to visit `www.example.com /? alert(xss)`, WAF must block the request.

13.7.2.8 Modify DNS resolution settings

This topic describes how to connect your businesses to WAF by modifying the DNS resolution settings.

Context

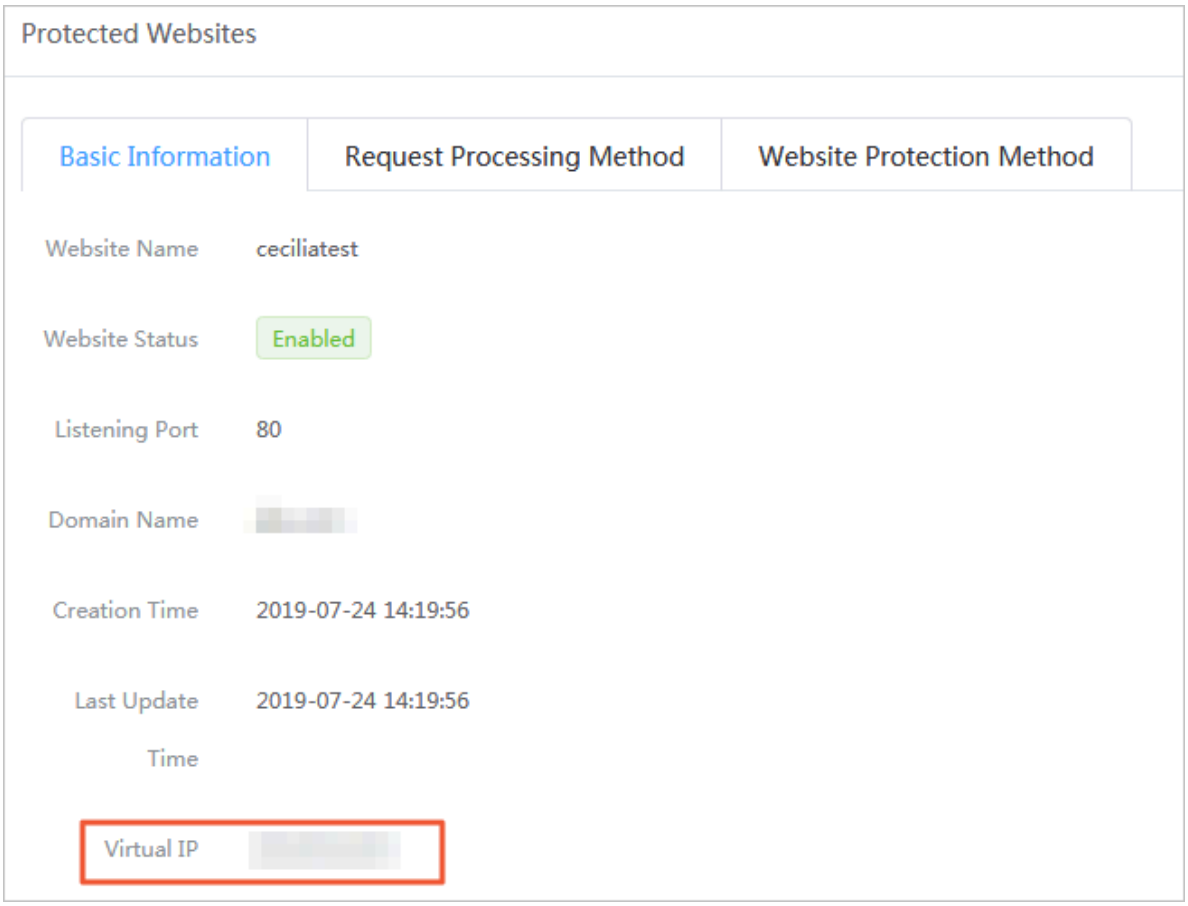
Before you modify the DNS resolution settings and redirect business traffic to WAF, make sure that you have passed local verification.

The domain name of a protected website may not be resolved by a DNS provider, for example, a website may use 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 virtual IP address of the protected website as the origin IP address of the SLB instance:

Procedure

1. *Log on to Apsara Stack Security Center.*
2. Choose Application Security > WAF. Choose Protection Configuration > Protected Website List.
3. Click the name of a website.

4. On the Basic Information tab page, obtain the virtual IP address of the protected website.



5. 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 virtual IP address of the protected site.

> **Note:**
> We recommend that you set the TTL to 600 seconds in DNS resolution settings. The greater the TTL is, the longer it takes to synchronize and update DNS records.

13.7.3 Detection overview

13.7.3.1 View protection overview

This topic describes how to view the WAF protection overview.

Context

The Protection Overview page displays the statistics of previous attacks, the geographical distribution of attackers, the numbers of total requests and blocked

requests, and other information. You can quickly learn the Web attack protection information and custom protection rules.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Application Security > WAF. Choose Detection Overview > Detection Overview.**

3. On the Detection Overview page, you can view Statistics within Last 24 Hours and Statistics within Last 30 Days.

- Attacker geographical distribution on a map

  The distribution of attackers is displayed on a map. You can select a map of China or a map of the world.

  The numbers of total requests and blocked requests are displayed.
- Distribution of top five attack types

  A pie chart is provided to display the distribution of the top five attack types and the number of attacks of each type.
- Top five attacked websites

  A bar chart is provided to display the top five attacked websites and the number of attacks on each website.

13.7.3.2 View Web service access information

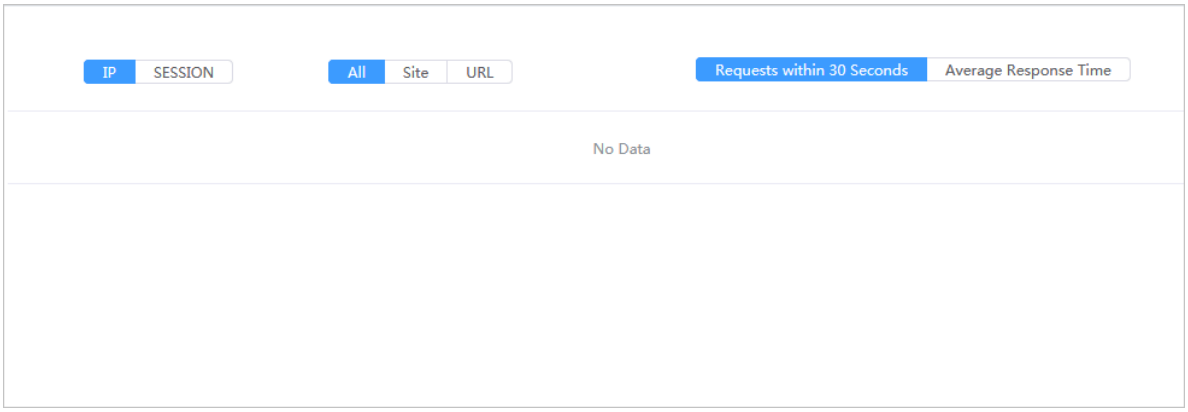
This topic describes how to view the service access information.

Context

WAF monitors the Web service access information. This allows security administrators to analyze the business access information and detect vulnerabilities.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Application Security > WAF. Choose Detection Overview > Access Status Monitor.**
3. **Filter the access records to view the details.**



13.7.4 Protection logs

13.7.4.1 View attack detection logs

This topic describes how to view attack detection logs.

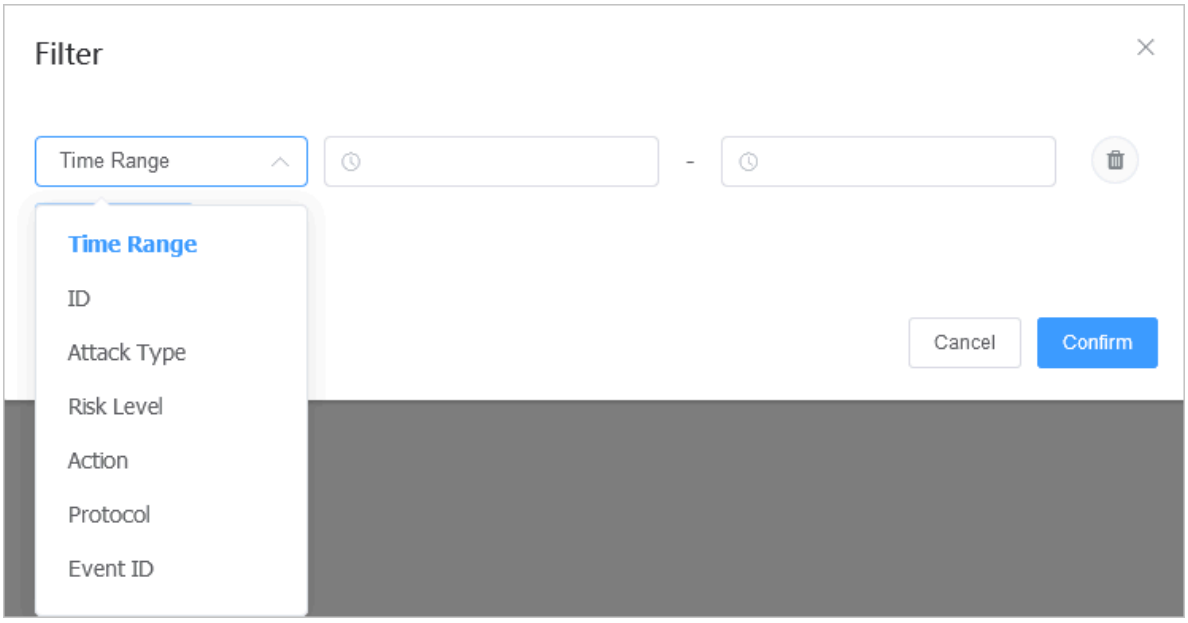
Context

These logs allow you to analyze the attacks on your Web services. Based on the analysis, you can update the attack protection policies and custom rules and fix the Web service vulnerabilities.

Procedure

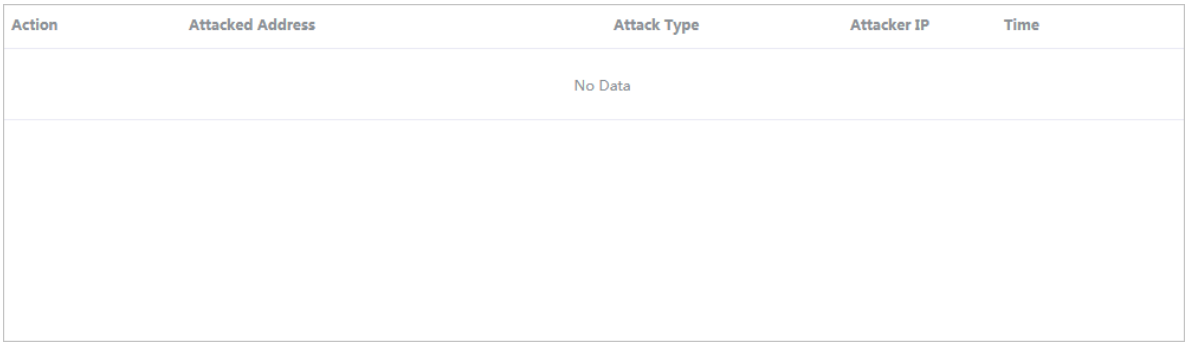
1. *Log on to Apsara Stack Security Center.*
2. **Choose Application Security > WAF. Choose Detection Logs > Attack Detection Logs.**

3. Click Add Filter, specify filter conditions, and click Confirm.



Note:
If you specify multiple conditions, all of the conditions must be met.

4. View the detected attacks.



13.7.4.2 View HTTP flood protection logs

This topic describes how to view HTTP flood protection logs.

Context

These logs allow you to analyze HTTP flood attacks on your Web services. Based on the analysis, you can update the HTTP flood protection rules and HTTP flood whitelist and fix the Web service vulnerabilities.

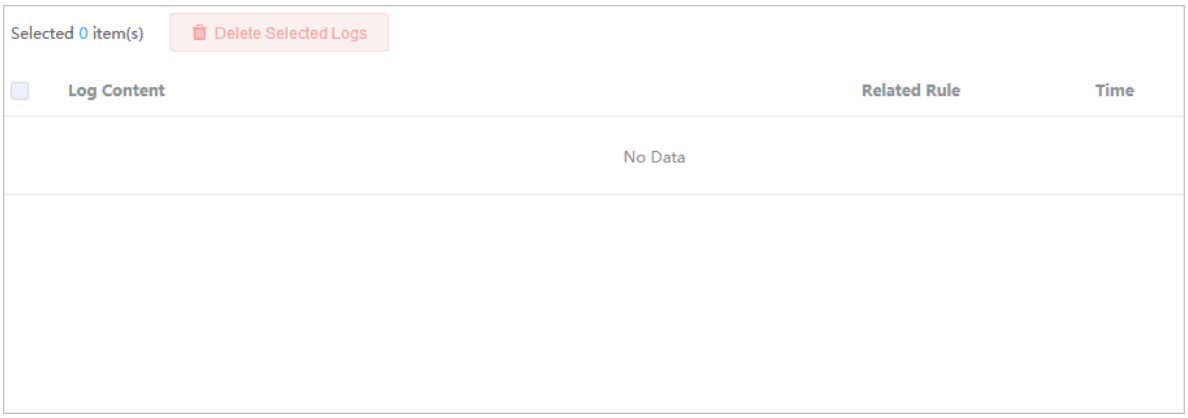
Procedure

1. *Log on to Apsara Stack Security Center.*

2. Choose Application Security > WAF. Choose Detection Logs > HTTP Flood Detection Logs.
3. Click Add Filter, specify filter conditions, and click Confirm.

Note:
If you specify multiple conditions, all of the conditions must be met.

4. View the HTTP flood detection result.



The blocked HTTP flood attacks, related rules, and attack time are displayed.

You can select a log and click Delete Selected Logs to delete a log.

13.7.5 System management

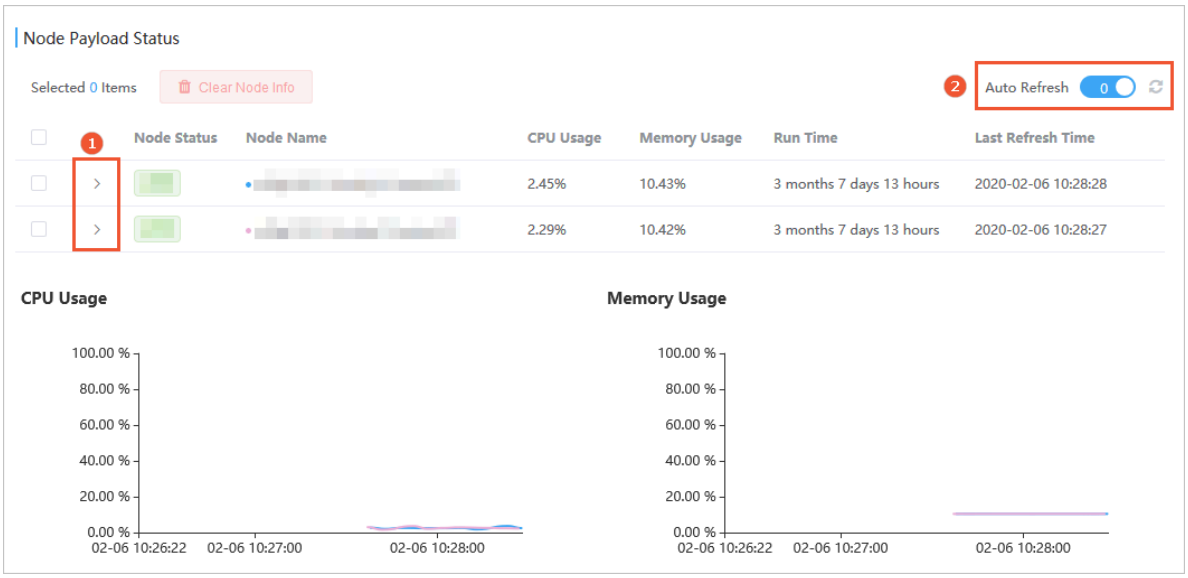
13.7.5.1 View payload status of nodes

This topic describes how to view the CPU and memory payload status of WAF nodes. You can use the status to identify faults or check whether scaling is required.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Application Security > WAF > System Management > Node Payload Status.

3. On the Node Payload Status page that appears, view the payload status of WAF nodes.



In the Node Payload Status section, you can view the CPU utilization and memory usage of each node. In the CPU Usage and Memory Usage sections, you can view changes in CPU utilization and memory usage over a period of time.

You can perform the following operations on the Node Payload Status page:

- Click the icon indicated by 1 to view the CPU utilization and memory usage of each service in a node.
- Turn off the switch indicated by 2 to enable manual update for the payload status of nodes.

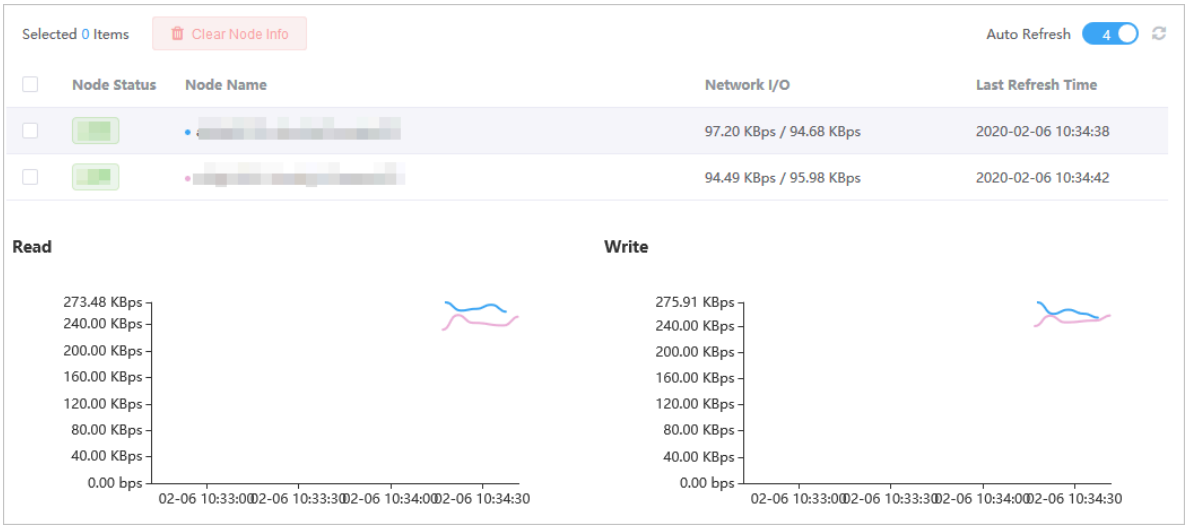
13.7.5.2 View network status of nodes

This topic describes how to view network status of WAF nodes, such as the network I/O, traffic detection status, and traffic forwarding status.

Node Network Status

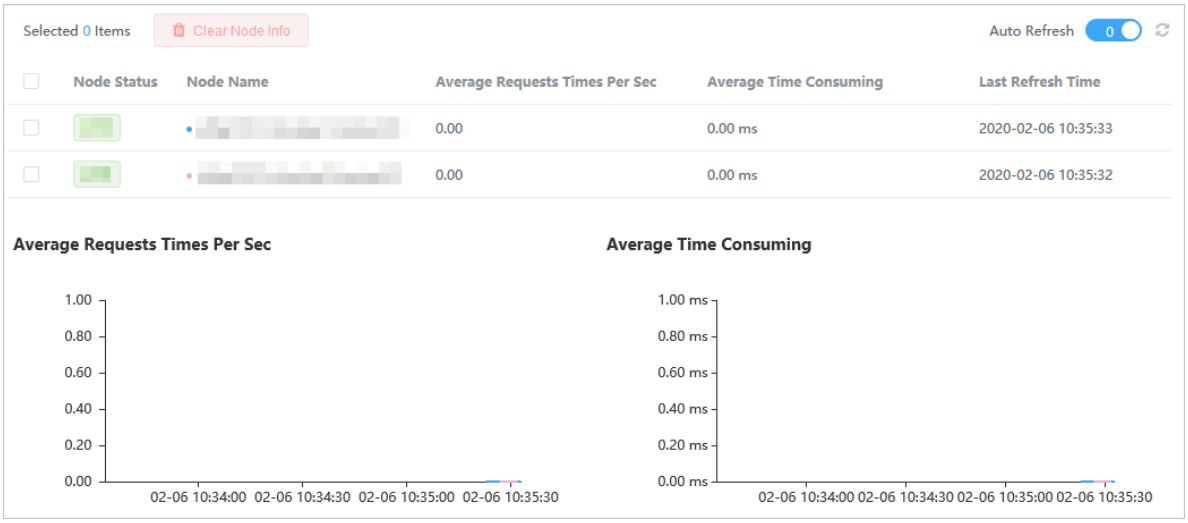
1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Application Security > WAF > System Management > Node Network Status.
3. Click the Node Network Status tab.

4. **View the network I/O of WAF nodes.**



Node Detection Status

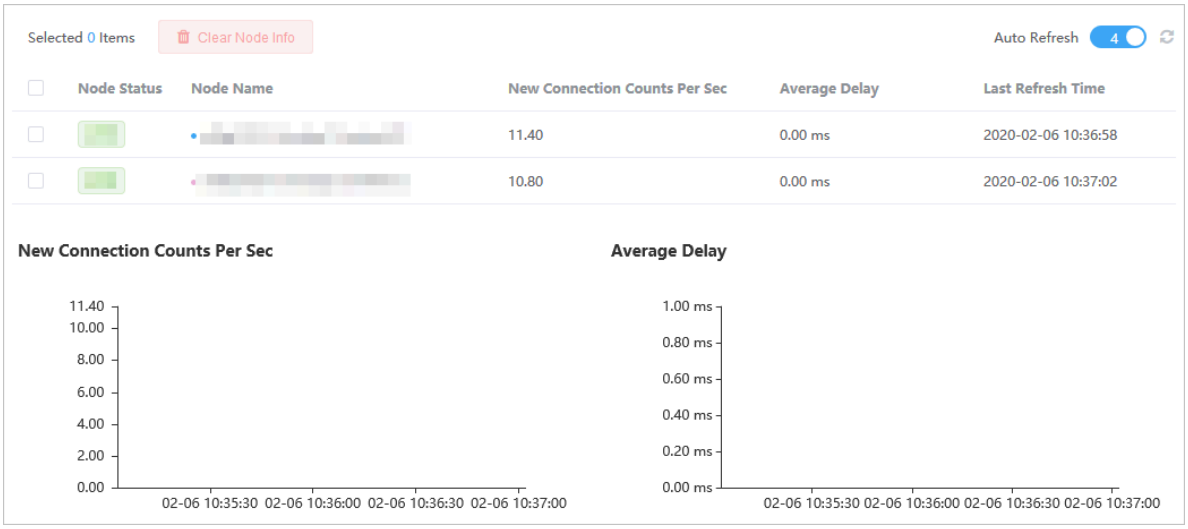
1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Application Security > WAF > System Management > Node Network Status.**
3. **Click the Node Detection Status tab.**
4. **View the network traffic detection status of WAF nodes.**



Node Forward Status

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Application Security > WAF > System Management > Node Network Status.**
3. **Click the Node Forward Status tab.**

4. View the network traffic forwarding status of WAF nodes.

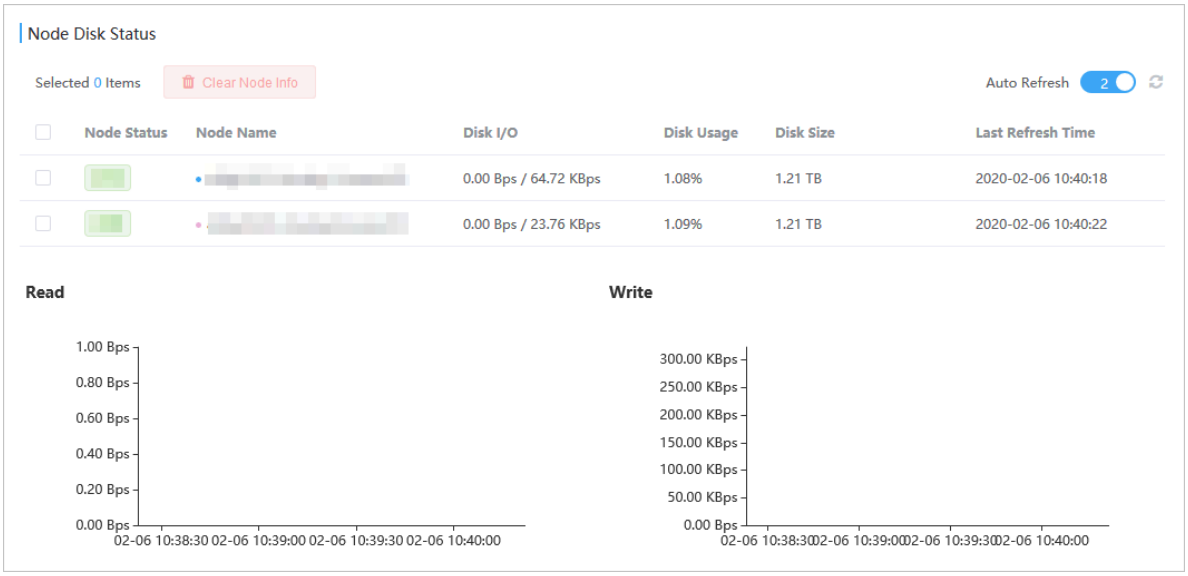


13.7.5.3 View disk status of nodes

This topic describes how to view disk status of WAF nodes. You can use the status to identify faults or check whether scaling is required.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Application Security > WAF > System Management > Node Disk Status.
3. On the Node Disk Status page that appears, view the disk status of WAF nodes.



In the Node Disk Status section, you can view the disk I/O and disk usage of nodes. In the Read and Write sections, you can view the disk read and write changes over a period of time.

13.8 Optional security products

13.8.1 Anti-DDoS settings

13.8.1.1 Overview

In Distributed Denial of Service (DDoS) attacks, attackers exploit the client-server model to combine multiple computers into a platform that can launch attacks on one or more targets. This greatly increases the threat of attacks.

Common DDoS attack types include:

- **Network-layer attacks: A typical example is UDP reflection attacks, such as NTP flood. These attacks use heavy traffic to congest the network of the victim, disabling proper responses to user requests.**
- **Transport-layer attacks: Typical examples include SYN flood and connection flood. These attacks consume a large number of connection resources of a server to cause denial of service.**
- **Session-layer attacks: A typical example is SSL flood. These attacks consume the SSL session resources of a server to cause denial of service.**
- **Application-layer attacks: Typical attack types include DNS flood, HTTP flood, and game zombie attacks. These attacks consume a large amount of application processing resources of a server to cause denial of service.**

Apsara Stack Security can redirect, scrub, and re-inject attack traffic to protect your server against DDoS attacks and ensure normal business operations.

Note:
Apsara Stack Security cannot scrub the traffic between internal networks.

13.8.1.2 View and configure anti-DDoS policies

This topic describes how to view and configure anti-DDoS policies. Anti-DDoS provides default anti-DDoS policies and DDoS traffic scrubbing policies.

Context

After an alert threshold of DDoS traffic is set for an IP address, an alert is triggered when the traffic to the IP address reaches the threshold. The alert thresholds for an IP address must be set based on the traffic volume. Excessive traffic volume

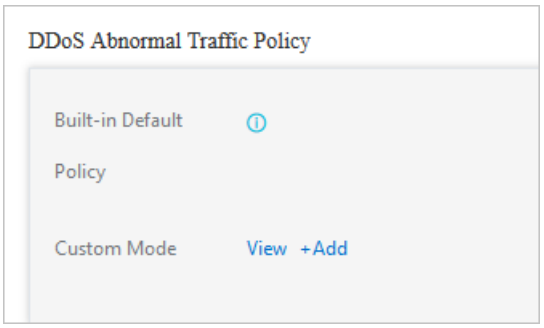
indicates a possible DDoS attack. We recommend that you set an alert threshold to a value slightly higher than the peak traffic volume.

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 alert threshold. It is set when the service is initialized.
- Alert threshold for a specific CIDR block: You can set an alert threshold for a specific CIDR block based on the traffic volume. Compared with global alert thresholds, CIDR block-specific alert thresholds allow you to control the traffic to each CIDR block.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Network Security > Policy Configuration > DDoS Defense Policy.
3. View and configure anti-DDoS policies.



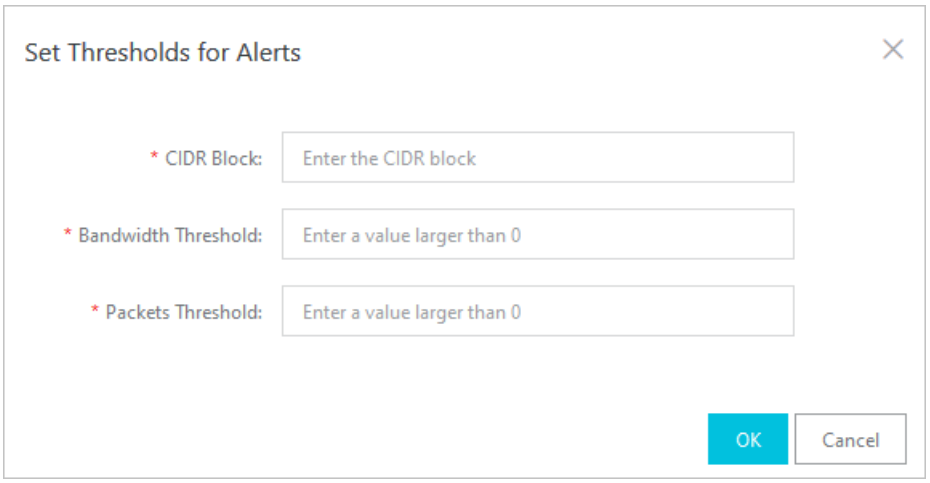
Action	Description
View default policies.	Move the pointer over the position indicated by 1 in the preceding figure to view the default anti-DDoS policy.

Action	Description
Customize a traffic scrubbing policy.	Click View to view CIDR block-specific policies, and click Add to customize a policy for a CIDR block.

To customize a policy for a CIDR block, follow these steps:

a) Click Add next to Custom Mode.

b) In the Set Thresholds for Alerts dialog box that appears, set parameters.

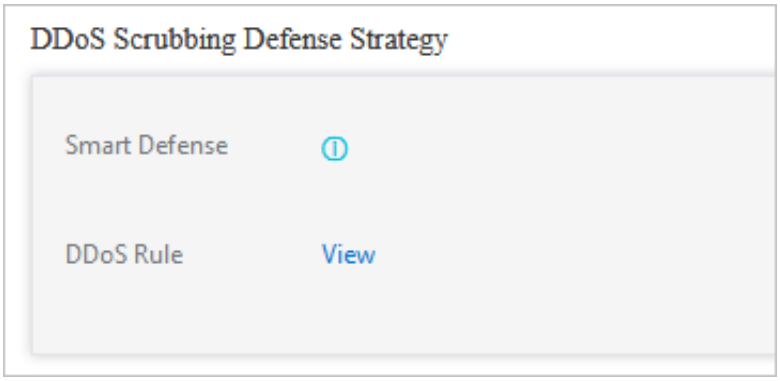


Parameter	Description
CIDR Block	The CIDR block for which the alert thresholds are used.
Bandwidth Threshold	The alert threshold for bandwidth usage in a data center. When the inbound or outbound traffic rate reaches this threshold, DDoS detection is triggered. Set this parameter to a value slightly higher than the traffic peak. We recommend that you set this parameter to 100 or higher. Unit: Mbit/s.

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. Set this parameter to a value slightly higher than the traffic peak. We recommend that you set this parameter to 20,000 or higher. Unit: packets per second (PPS).

c) Click OK.

4. In the DDoS Scrubbing Defense Strategy section, click View to view DDoS traffic scrubbing policies.



13.8.1.3 View DDoS events

This topic describes how to view distributed denial of service (DDoS) events.

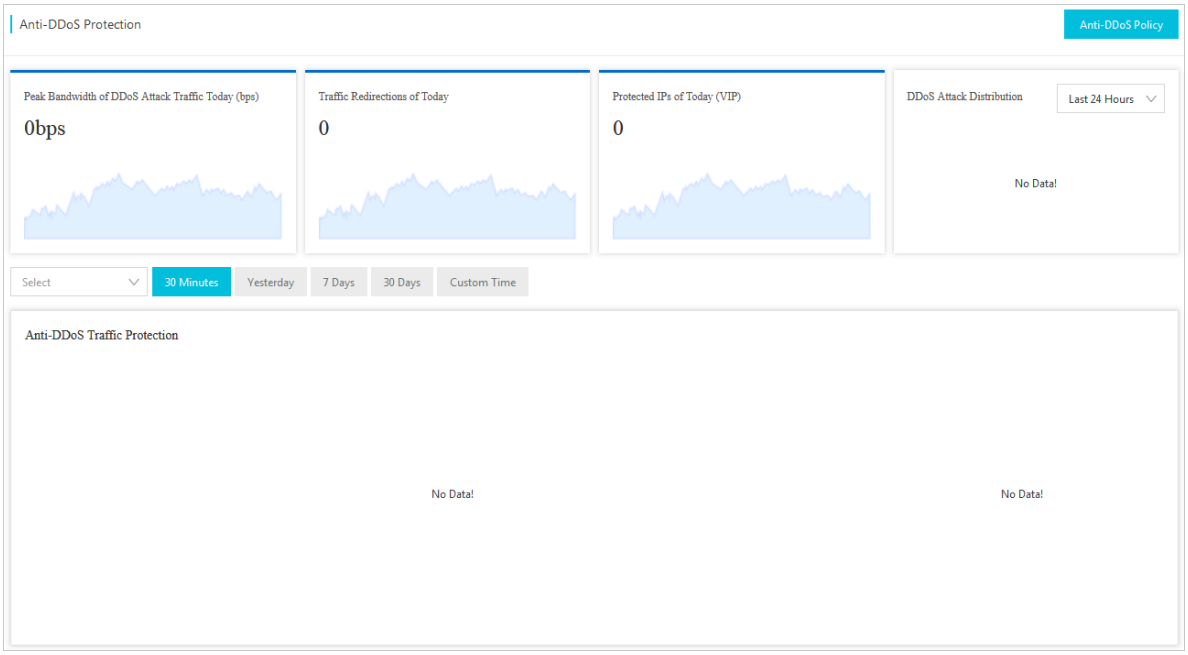
Context

During or after traffic scrubbing, Apsara Stack Security reports security events to Apsara Stack Security Center.

Procedure

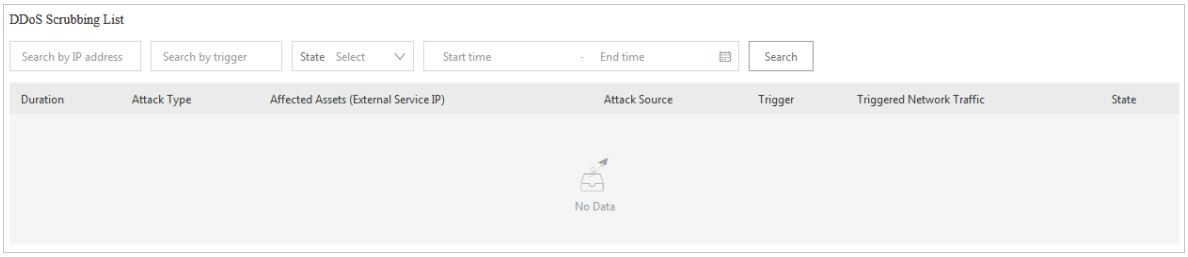
1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Network Security > Network Protection.

3. View anti-DDoS statistics.



4. Optional: In the DDoS Scrubbing List section, set search conditions and click Search.

> **Note:**
> Skip this step if you need to view all traffic scrubbing events.



Search condition	Description
Search by IP address	The IP address that was under a DDoS attack.
Search by trigger	The metric that exceeded the configured alert threshold in the DDoS attack traffic.
State	· Scrubbing: indicates that traffic scrubbing is in progress. · Scrubbing Complete: indicates that traffic scrubbing is complete.
Start time and End time	The start time and end time of DDoS traffic scrubbing.

5. In the DDoS Scrubbing List section, view details about DDoS traffic scrubbing events.

13.8.2 Sensitive Data Discovery and Protection

13.8.2.1 Grant access permissions

This topic describes how to authorize Sensitive Data Discovery and Protection (SDDP) to access data of your department before you use SDDP.

Prerequisites

The department name and AccessKey pair are obtained. For more information about how to obtain the AccessKey pair, see Obtain the AccessKey pair of an organization in *ASCM Console User Guide* . To find the topic, choose Enterprise > Organizations > Obtain the AccessKey pair of an organization.

Context

Before using SDDP, you must perform the following operations:

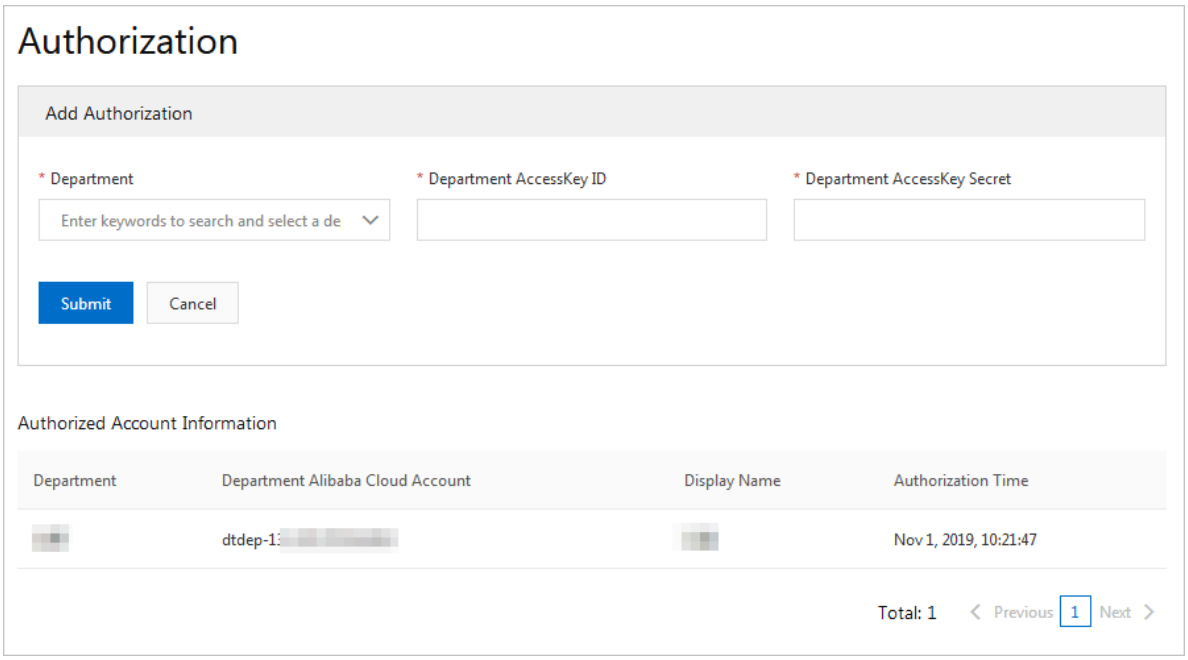
- **Authorize SDDP to access the data of your department.**
- **Authorize SDDP to access data of Apsara Stack services such as MaxCompute, Object Storage Service (OSS), and Table Store of your department.**

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Authorization.

Note:

If SDDP is not authorized to access the data of your department, set parameters on the Authorization page to authorize SDDP to do this.



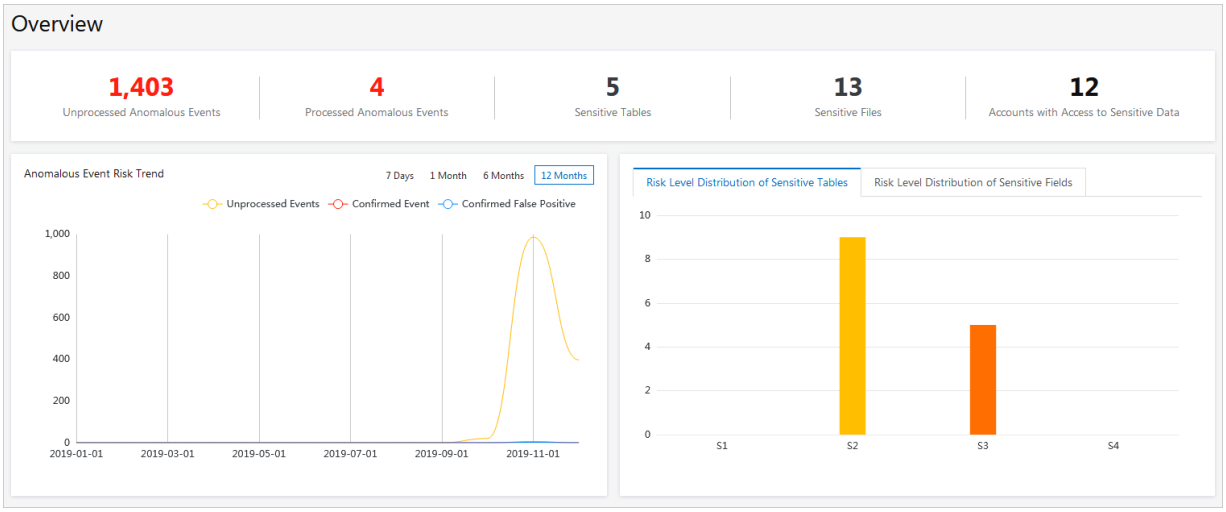
3. In the Add Authorization section, authorize SDDP to access data of your department.
   a) In the Department drop-down list, enter a keyword and select the department.
   b) Set Department AccessKey ID and Department AccessKey Secret.
   c) Click Submit.
4. In the Authorized Account Information section, view the list of authorized departments.

13.8.2.2 Overview

This topic describes the overview page of Sensitive Data Discovery and Protection (SDDP). This page displays the overall security status of data protected by SDDP, and allows a security administrator to quickly understand the current security status of sensitive data.

SDDP can detect sensitive data in your data assets based on sensitive data detection rules and track the use of sensitive data. SDDP also provides a data overview for you to obtain the security status of your data assets in real time.

Choose Data Security > Sensitive Data Discovery and Protection > Overview. On the Overview page, view the overall security status of the sensitive data.



- **Overview: displays the overall information of sensitive data, including the number of unprocessed anomalous activities, the number of anomalous activities confirmed as violations, the total number of sensitive tables, the total number of sensitive objects, and accounts that accessed sensitive data.**
- **Abnormal event risk trend: displays the trends of anomalous activities in a line chart. You can select `7 days`, `1 month`, `6 months`, or `12 months` to view the trends of unprocessed anomalous activities, anomalous activities confirmed violations, and anomalous activities excluded as false positives.**
- **Sensitive table risk level distribution: displays the distribution of sensitive tables at the S1, S2, S3, and S4 risk levels.**
- **Sensitive field risk level distribution: displays the distribution of sensitive fields at the S1, S2, S3, and S4 risk levels.**
- **Data flow situation:**
  - **Displays the dynamic statistics on core data flows in Datahub and Cloud Data Pipeline (CDP).**
  - **Provides a data flow chart that dynamically displays the data flow status and abnormal output. You can click an anomalous activity in the flow chart to go to the Abnormal data flow page.**

    Monitors the data links among entities such as data storage services MaxCompute, , Object Storage Service (OSS), and Table Store, data transmissi on services Datahub and CDP, the data flow processing service Blink, external databases, and external files.

13.8.2.3 Detect sensitive data

13.8.2.3.1 Sensitive data overview

This topic describes how to view the overall security status of your data assets.

Choose Data Security > Sensitive Data Discovery and Protection > Sensitive data identification > Sensitive data overview. On the Sensitive data overview page, you can view the overall security status of your data assets.

- **You can view the overall information about sensitive data. The information includes the total numbers of tables, objects, sensitive instances, sensitive tables , and sensitive objects.**
- **You can search for sensitive data based on conditions such as the risk level, asset type, sensitive data type, and asset name.**
- **You can view the statistics on the authorization information and sensitive data in Apsara Stack services such as MaxCompute, Object Storage Service (OSS), and Table Store in real time.**

13.8.2.3.2 View the statistics on sensitive data of MaxCompute

This topic describes how to view statistics on sensitive data of MaxCompute.

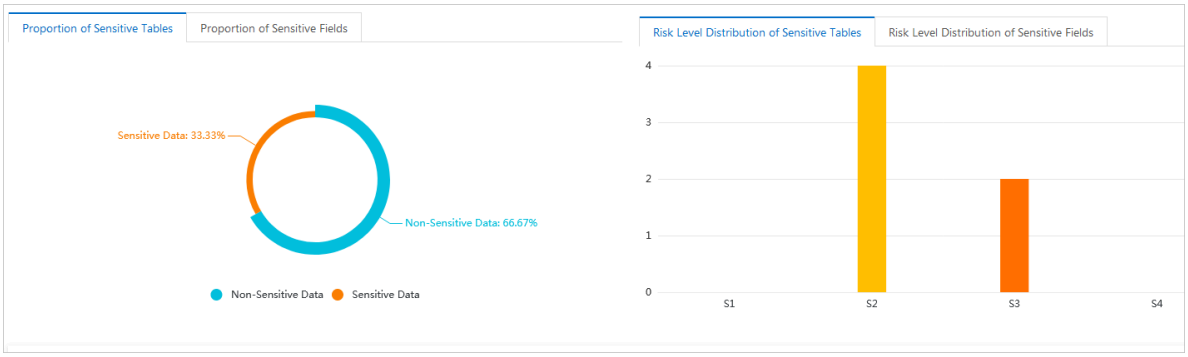
Context

MaxCompute is a rapid and fully-managed data warehouse solution that can process terabytes or petabytes of data. MaxCompute provides you with complete data import schemes and various classic distributed computing models. It supports fast computing on a large amount of data, effectively saves costs for enterprises, and guarantees data security.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Data Security > Sensitive Data Discovery and Protection > Sensitive data identification > MaxCompute.**

3. View the statistics on sensitive data of MaxCompute.



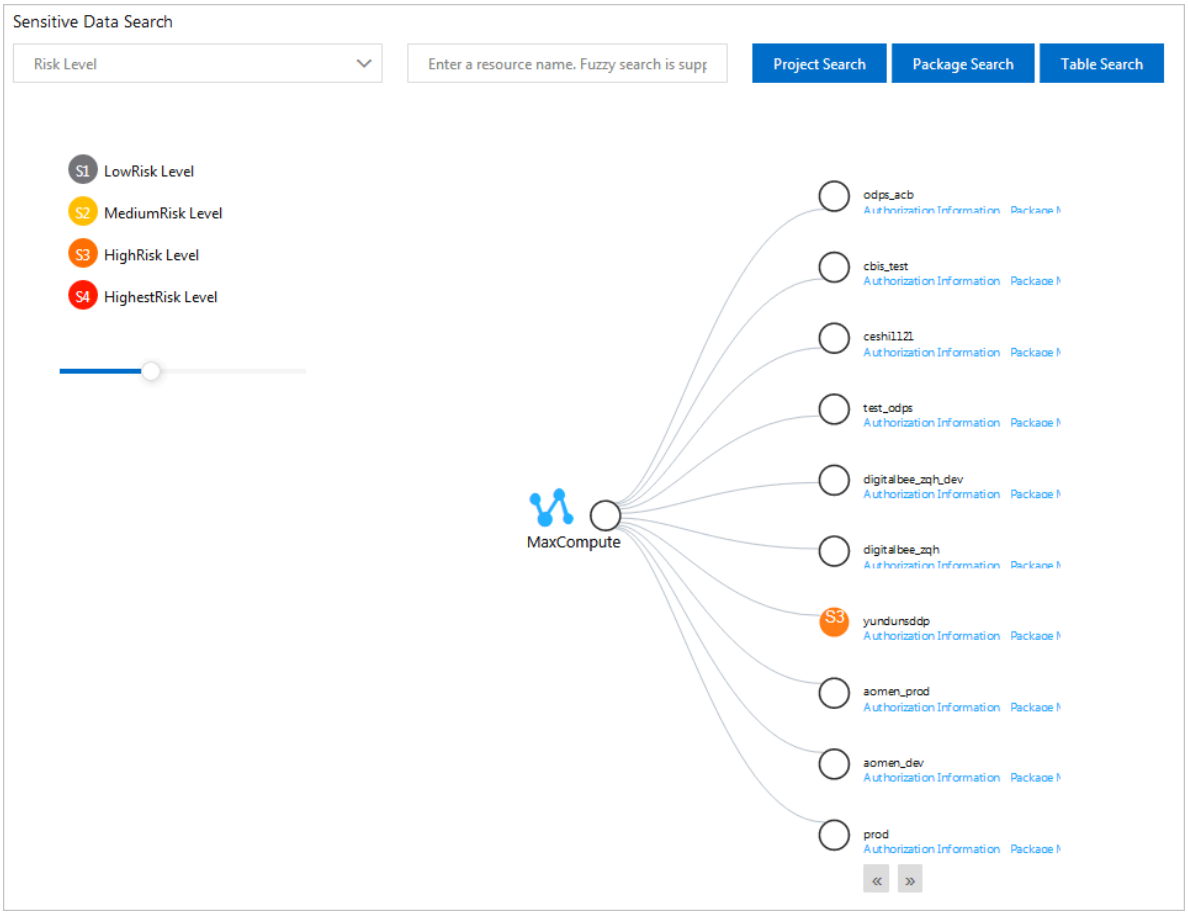
a) In the Sensitive data statistics section, enter a keyword and select the target MaxCompute project from the drop-down list.

> **Note:**
> To view the statistics on all MaxCompute projects, select `All` from the drop-down list.

b) On the Sensitive table ratio and Sensitive field ratio tabs, view the percentages of sensitive and non-sensitive tables and fields.

c) On the Sensitive table risk level distribution and Sensitive field risk level distribution tabs, view the distribution of sensitive tables and fields at the S1, S2, S3, and S4 risk levels.

4. Query the sensitive data of MaxCompute.



By default, the system displays all MaxCompute projects. The system displays different risk levels in different colors. To view the information about a specific project, package, table, or field, follow these steps:

a) Select a risk level from the Risk level drop-down list.

b) Enter a keyword of the project, package, or table in the search field.

c) Click Search project, Search package, or Search table.

You can view the relationships among the projects, packages, tables, and fields, and the related authorization information in a tree map.

- The tree map displays the distribution of sensitive data in MaxCompute.
- You can view the authorization information of a project, package, table, or field. The system displays the authorization information by category, including the authorized users and violations.
- You can click Package management under a project to view the packages in the project, including the tables and fields in the packages and related authorization information.

d) Move the pointer over the project, package, table, or field to view its details.

13.8.2.3.3 View the statistics on sensitive data of Table Store

This topic describes how to view the statistics on sensitive data of Table Store.

Context

Table Store is a multi-model NoSQL database service developed by Apsara Stack. Table Store can store a large amount of structured data and support fast query and analysis. The distributed storage and powerful index-based search engine enable Table Store to store petabytes of data while guaranteeing a 10 million TPS and a latency within milliseconds.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **Choose Data Security > Sensitive Data Discovery and Protection > Sensitive data identification > OTS.**
3. **View the statistics on sensitive data of Table Store.**

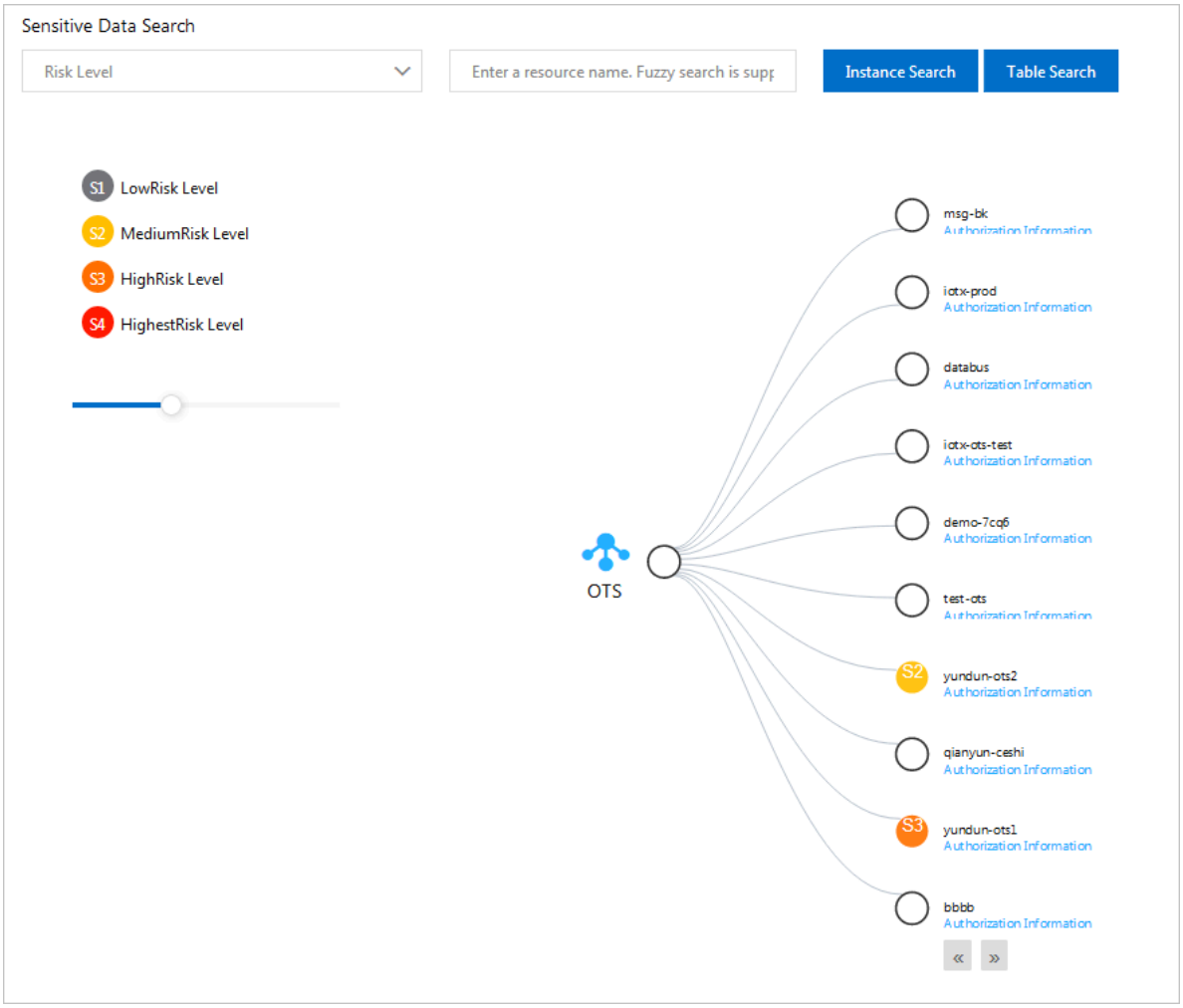
a) **In the Sensitive data statistics section, enter a keyword and select the target Table Store instance from the drop-down list.**

> **Note:**
> **To view the statistics on all Table Store instances, select `All` from the drop-down list.**

b) **On the Sensitive table ratio and Sensitive field ratio tabs, view the percentages of sensitive and non-sensitive tables and fields.**

c) **On the Sensitive table risk level distribution and Sensitive field risk level distribution tabs, view the distribution of sensitive tables and fields at the S1, S2, S3, and S4 risk levels.**

4. Query the sensitive data of Table Store.



By default, the system displays all Table Store instances. To view the information about a specific instance or table, follow these steps:

a) Select a risk level from the Risk level drop-down list.

b) Enter a keyword of the instance or table in the search field.

c) Click Search instances or Search Tables.

- The tree map displays the distribution of sensitive data in Table Store.
- You can view the authorization information of an instance or a table. The system displays the authorization information by category, including the authorized users and violations.

d) Move the pointer over the instance or table to view its details.

13.8.2.3.4 View the statistics on sensitive data of OSS

This topic describes how to view the statistics on sensitive data of Object Storage Service (OSS).

Context

OSS is a secure and reliable cloud storage service provided by Apsara Stack. It can store a large amount of data at low costs. OSS can store any type of file and is therefore suitable for various websites, enterprises, and developers.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Data Security > Sensitive Data Discovery and Protection > Sensitive data identification > OSS.**
3. **View the statistics on sensitive data of OSS.**

View the charts on the Sensitive file object ratio and Sensitive object risk level distribution tabs.

To view the charts for a specific OSS bucket, follow these steps:

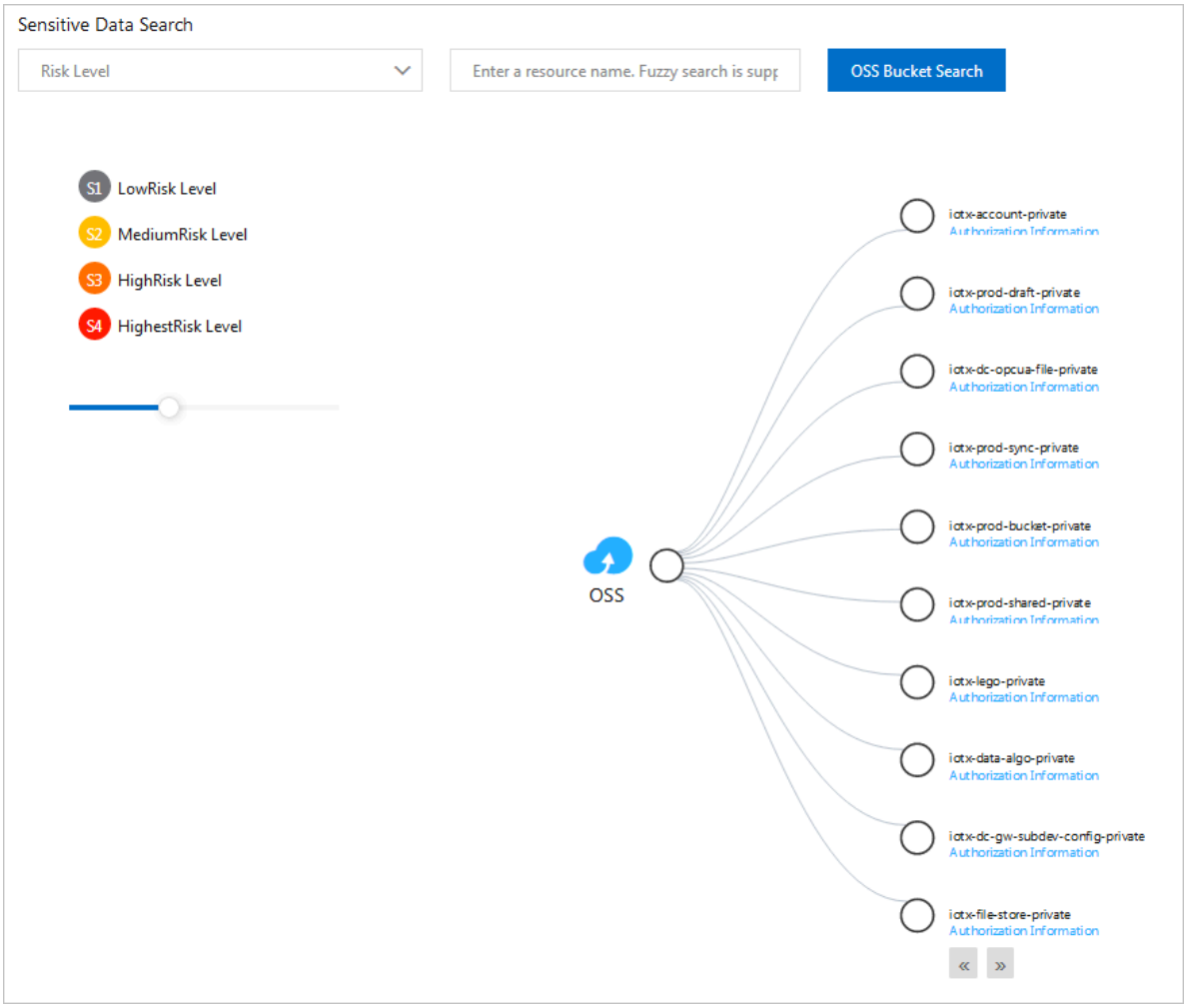
a) **In the Sensitive data statistics section, enter a keyword and select the target OSS bucket from the drop-down list.**

> **Note:**
> **To view the statistics on all OSS buckets, select `All` from the drop-down list.**

b) **On the Sensitive file object ratio tab, view the percentages of sensitive and non-sensitive objects.**

c) **On the Sensitive object risk level distribution tab, view the distribution of sensitive objects at the S1, S2, S3, and S4 risk levels.**

4. Query the sensitive data of OSS.



By default, the system displays all OSS buckets. To view the information about a specific bucket, follow these steps:

a) Select a risk level from the Risk level drop-down list.

b) Enter a keyword of the bucket in the search field and click Search bucket.

- The tree map displays the distribution of sensitive data in OSS.
- You can view the authorization information of a bucket. The system displays the authorization information by category, including the authorized users and violations.

c) Move the pointer over the bucket to view its details.

13.8.2.3.5 View statistics on sensitive data in ApsaraDB for RDS

This topic describes how to view statistics on sensitive data in ApsaraDB for RDS.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Sensitive Data Identification > RDS.
3. View statistics on sensitive data in ApsaraDB for RDS.

View Proportion of Sensitive Tables, Proportion of Sensitive Fields, Risk Level Distribution of Sensitive Tables, and Risk Level Distribution of Sensitive Fields in the Sensitive Data Statistics section.

To view the charts for a specific database, follow these steps:

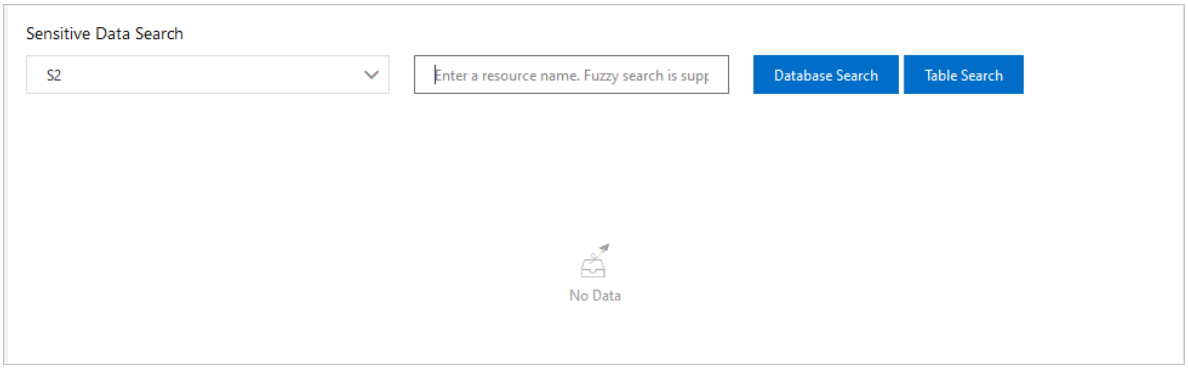
a) In the Sensitive Data Statistics section, enter a keyword and select the target database from the drop-down list.

> **Note:**
> To view the statistics on all databases, select `All` from the drop-down list.

b) On the Proportion of Sensitive Tables and Proportion of Sensitive Fields tabs, view the proportions of sensitive and non-sensitive tables and fields.

c) On the Risk Level Distribution of Sensitive Tables and Risk Level Distribution of Sensitive Fields tabs, view the distribution of sensitive tables and fields at the S1, S2, S3, and S4 risk levels.

4. Query the sensitive data in ApsaraDB for RDS.



All ApsaraDB for RDS databases are displayed by default. To view the information about a specific database or table, follow these steps:

a) Select a risk level from the Risk Level drop-down list.

b) Enter a keyword of the database or table in the search box.

c) Click Database Search or Table Search.

- The tree map displays the distribution of sensitive data in ApsaraDB for RDS.
- You can view the authorization information of a database or table. The system displays the authorization information by category, including authorized users and violations.

d) Move the pointer over the database or table to view its details.

13.8.2.4 Check data permissions

13.8.2.4.1 View permission statistics

This topic describes how to view permission statistics.

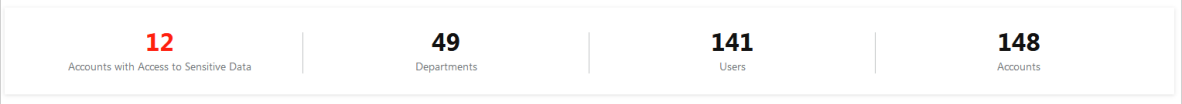
Context

On the Permissions management page, you can check the overall permission distribution of Apsara Stack. With this feature, you can quickly identify high-risk accounts and users, and troubleshoot and resolve security issues in a timely manner.

Procedure

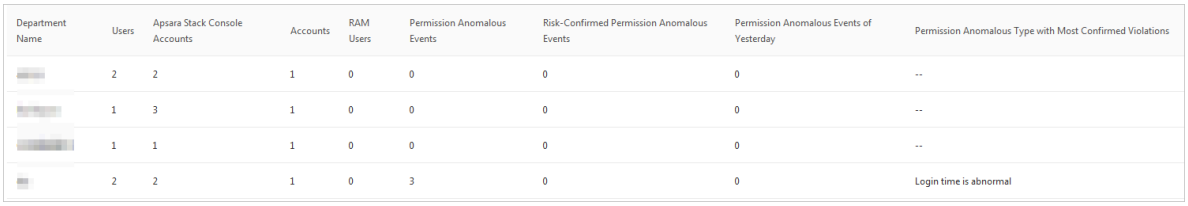
1. *Log on to Apsara Stack Security Center.*
2. Choose Data Security > Sensitive Data Discovery and Protection > data permission > Permissions management.

3. View the overall statistics on permissions.



   - Number of accounts accessible to sensitive data: the number of accounts that can access sensitive data.
   - Total number of departments: the number of departments in Apsara Stack.
   - Total number of people: the number of users in Apsara Stack.
   - Total number of accounts: the number of accounts in Apsara Stack.

4. View the department-level statistics on permissions.



Department Name	Users	Apsara Stack Console Accounts	Accounts	RAM Users	Permission Anomalous Events	Risk-Confirmed Permission Anomalous Events	Permission Anomalous Events of Yesterday	Permission Anomalous Type with Most Confirmed Violations
[redacted]	2	2	1	0	0	0	0	--
[redacted]	1	3	1	0	0	0	0	--
[redacted]	1	1	1	0	0	0	0	--
[redacted]	2	2	1	0	3	0	0	Login time is abnormal

   You can view the statistics on the users, accounts, and anomalous activities related to permission access for each department.

13.8.2.4.2 Query permissions

This topic describes how to query permissions.

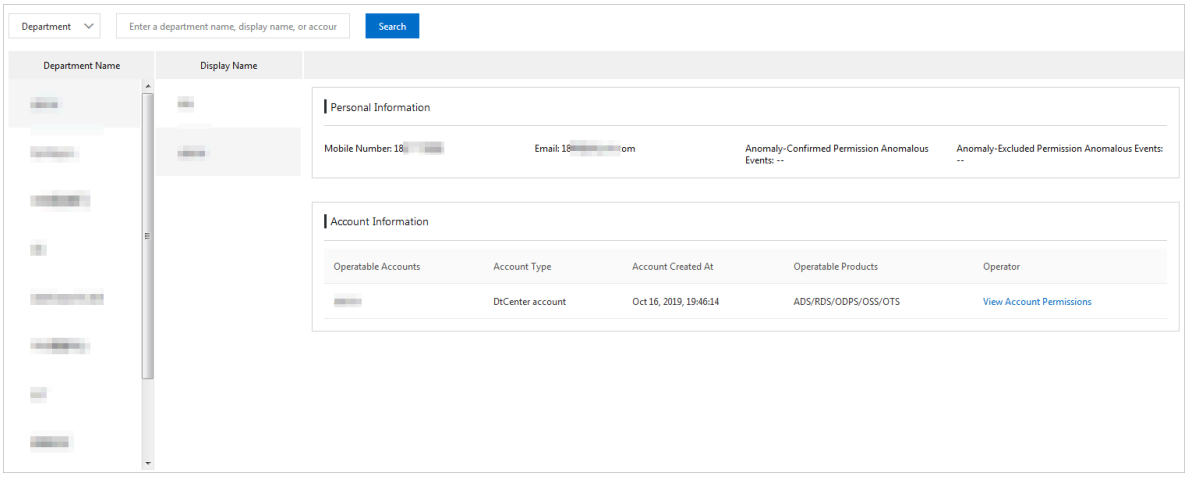
Context

You can search for an account to view the account information. With this feature, you can quickly find the owner for sensitive data.

Procedure

1. *Log on to Apsara Stack Security Center.*

2. Choose Data Security > Sensitive Data Discovery and Protection > data permission > Permissions search.



3. Search for the target account.

   To search for an account, follow these steps:

   a) Select a department or user from the drop-down list.

   b) Enter a keyword in the search field.

   c) Click Search. The accounts containing the keyword are listed in the Display name column.

   You can also click a department in the Department name column. All accounts of the department are listed in the Display name column.

4. In the Display name column, click the target account.

5. In the right pane, view the information in the Personal information and Accounts sections.

   - Personal information

     You can view the contact information of the account owner, the number of confirmed anomalous activities related to permission access, and the number of excluded anomalous activities related to permission access.

   - Accounts

     You can view the accounts that the owner can use, the type and creation time of the accounts, and Apsara Stack services that the accounts can access.

     You can click View account permissions in the Actions column for an account to view the resources, resource types, resource paths, and operation permissions of the account.

13.8.2.5 Monitor data flows

13.8.2.5.1 View data flows in Datahub

This topic describes how to view data flows in Datahub.

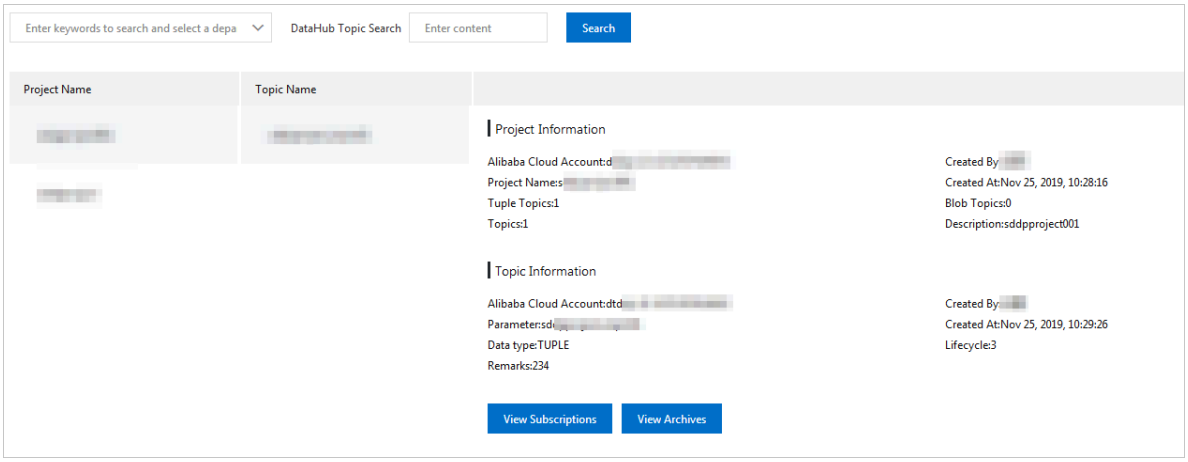
Context

Datahub is a platform designed to process streaming data. You can publish and subscribe to applications for streaming data in Datahub and distribute the data to other platforms. Datahub allows you to analyze streaming data and build applicatio ns based on the streaming data.

On the DataHub page, you can view the details of data flows in Datahub, including the relationships between Datahub projects and topics, and the relationships among topics, subscribed applications, and archive sources.

Procedure

1. *Log on to Apsara Stack Security Center.*
2. **Choose Data Security > Sensitive Data Discovery and Protection > Data flow monitoring > DataHub.**
3. **Enter a keyword and select a department from the drop-down list, enter a keyword in the DataHub topic search field, and then click Search.**



Note:

You can also click the target project in the project name column and then click the target topic in the topic name column.

You can view the information about the topic in the Project information and Topic information sections.

- Project information

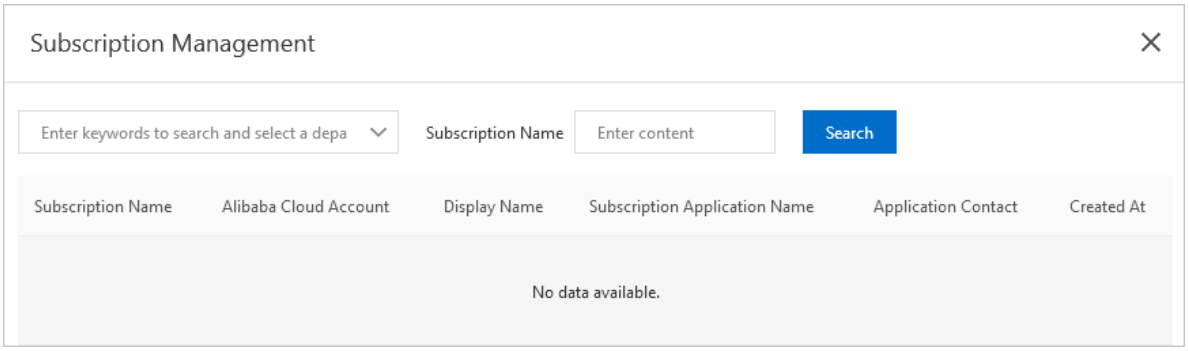
  Displays the information such as the project name, Apsara Stack account, creator, creation date, and number of topics.

- Topic information

  Displays the information such as the project name, Apsara Stack account, creator, creation date, and type.

4. Click View subscription to view the subscription list.

   The subscription list contains the information such as the subscription name, Apsara Stack account of the creator, display name, name of the subscribed application, and application contact.



   a) Enter a keyword and select a department from the drop-down list.
   b) Enter a keyword in the DataHub topic search field.
   c) Click Search to find the target Datahub topic.

5. Click View archive to view the archive list.

   The archive list contains the information such as the name of the connected instance, Apsara Stack account of the creator, display name, source service, resource path, and risk level.

   a) Enter a keyword and select a department from the drop-down list.
   b) Enter a keyword in the DataHub topic search field.
   c) Click Search to find the target Datahub topic.

13.8.2.5.2 View dataflows in CDP

This topic describes how to view dataflows in Cloud Data Pipeline (CDP).

Context

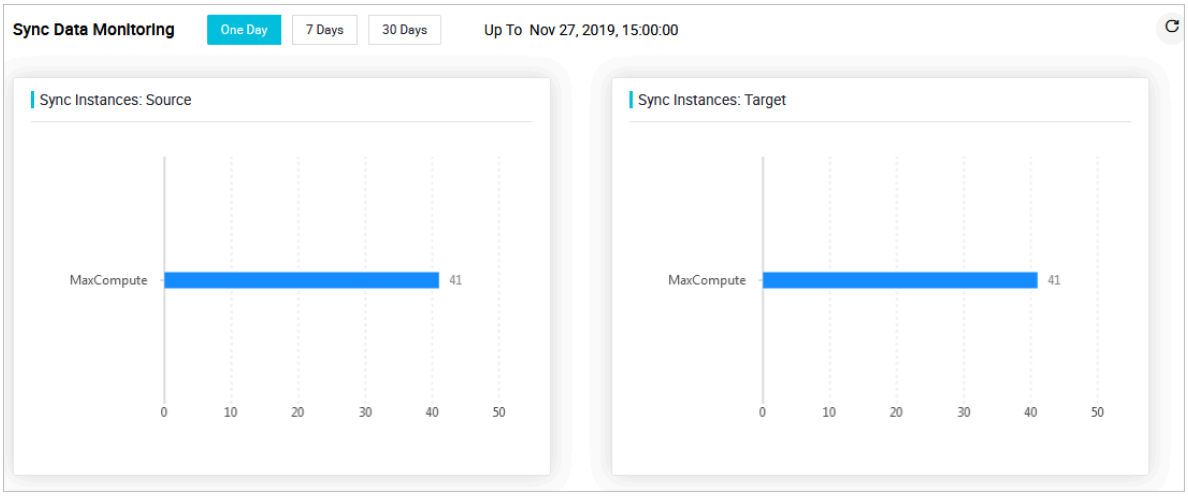
DataWorks is a comprehensive professional big data cloud R&D platform that serves as an analytics operating system and delivers intelligent, efficient, secure , and reliable big data services. It meets your requirements for data and quality management, and allows you to provide data services for external systems.

DataWorks provides the Data Integration feature, which is a stable, efficient, and scalable data synchronization platform provided by Alibaba Cloud. Data Integratio n implements fast and stable data transmission and synchronization between various data sources in complicated networks.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Data Flow Monitoring > CDP Flow Monitoring.**
3. **On the Data Integration page that appears, click Sync Data Monitoring.**
4. **View the statistical graph for the number of synchronized instances.**

> **Note:**
>
> - **The number of synchronized instances is measured from two aspects: Sync Instances: Source and Sync Instances: Target.**
> - **You can view the statistics on diverse periods (One Day, 7 Days, and 30 Days).**



5. In the Sync Instances section, view information such as the ID, time, node name, data type, and amount of synchronized data.

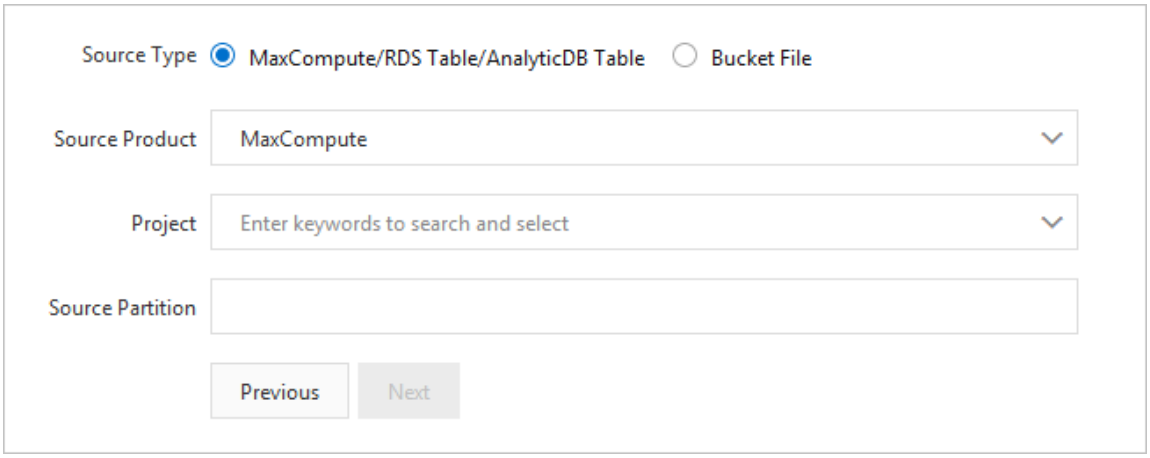
13.8.2.6 Sensitive data masking

13.8.2.6.1 Add a static desensitization task

This topic describes how to add a static desensitization task and run the task to mask sensitive data.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Sensitive Data Desensitization > Static Desensitization.
3. In the Desensitization Tasks section, click Add Desensitization Task in the upper-right corner.
4. On the Add Desensitization Task page that appears, set parameters.
   a) In the Basic Task Information step, set Task Name and Remarks, and click Next.
   b) In the Desensitization Source Configuration step, set parameters and click Next.



Service	Parameter	Description
MaxCompute	Source Type	Set the parameter to MaxCompute/RDS Table/AnalyticDB Table.
	Source Product	Set the parameter to MaxCompute.

Service	Parameter	Description
	Project	Select the project or database that contains the table with the sensitive data you want to desensitize.
	Table Name	Select the name of the table with the sensitive data you want to desensitize.
	Source Partition	Enter the name of the partition that contains the sensitive data you want to desensitize. You can configure partitions when creating a MaxCompute table. Partitions define different logical divisions of a table to help you efficiently query specific content. **Note:** If you leave this parameter unspecified, SDDP desensitizes sensitive data in all partitions of the table.
ApsaraDB for RDS	Source Type	Set the parameter to MaxCompute/RDS Table/AnalyticDB Table.
	Source Product	Set the parameter to RDS.
	Project	Select the project or database that contains the table with the sensitive data you want to desensitize.
	Table Name	Select the name of the table with the sensitive data you want to desensitize.
	Sample SQL	Optional. Enter the SQL statement that specifies the data you want to desensitize.
Object Storage Service (OSS)	Source Type	Set the parameter to Bucket File.
	File Source	Upload a bucket file or select an existing one. · Uploaded Local File: If you select this option, click Select File to select a local file for upload. · Bucket: If you select this option, select a file from the Source File drop-down list.

Service	Parameter	Description
	Source File Description	Enter the remarks of the bucket file, which help you quickly identify the task. **Note:** This parameter needs to be set if File Source is set to Uploaded Local File.
	Source File	Select an OSS bucket file. **Note:** This parameter needs to be set if File Source is set to Bucket.
	Source File Name	Optional. Enter the name of the bucket file. **Note:** This parameter needs to be set if File Source is set to Bucket.

c) In the Desensitization Algorithm Configuration step, set parameters and click Next.

You need to set the algorithm type, select an algorithm, and turn on the desensitization switch for the source field you want to desensitize.

d) In the Destination Location Configuration step, set parameters and click Next.

e) In the Confirm Process Logic step, set parameters to confirm the processing logic.

Parameter	Description
Select Trigger Method	The mode in which the desensitization task is run. Valid values: • `Manual Only`: You must manually run the desensitization task on the Static Desensitization page. • `Scheduled Only`: The desensitization task is automatically run at the specified time within an hour, day, or month. • `Manual + Scheduled`: You must manually run the desensitization task at the specified time within an hour, day, or month.

Parameter	Description
Table Name Conflict Resolution	The handling method if a table has the same name as the specified target table. Valid values: · Delete the target table and create a new table with the same name. · Insert new data to the target table.: We recommend that you select this option.
Row Conflict Resolution	The handling method for a row conflict. Valid values: · Keep conflicting rows in the target table and discard the new data.: We recommend that you select this option. · Delete conflicting rows in the target table and insert the new data.

f) Click Submit.

After the task is created, you can view it in the Desensitization Tasks section.

5. In the Desensitization Tasks section, click the switch and Start in the Operator column corresponding to the created desensitization task to run the task.
6. In the Task Execution Query section, view Execution Progress and Status of the task.

13.8.2.7 Abnormal activity detection

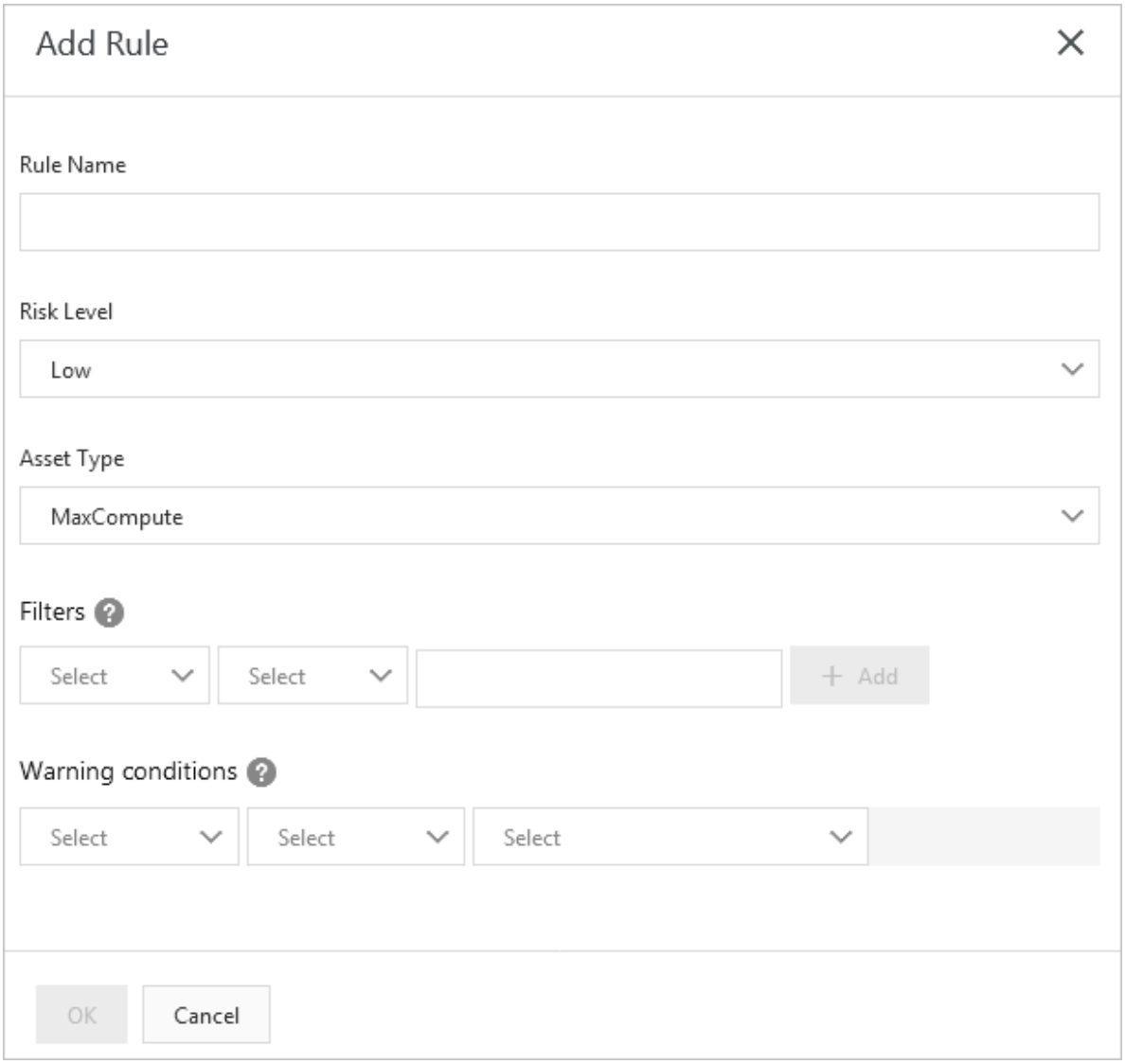
13.8.2.7.1 Add a custom rule for abnormal activities

This topic describes how to add a custom rule for abnormal activities. If Sensitive Data Discovery and Protection (SDDP) detects an activity that matches the custom rule, it identifies the activity as an abnormal activity.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Anomaly Detection > Anomalous Event Rules.
3. On the page that appears, click Add Rule.

4. In the Add Rule dialog box that appears, set parameters.



Parameter	Description
Rule Name	The name of the rule.
Risk Level	The risk level of the rule. Valid values: `Low`, `Medium`, and `High`.
Asset Type	The type of the service for which the rule is used. Valid values: `OSS`, `MaxCompute`, and `RDS`.
Filters	The filter conditions of the rule.
Warning conditions	The warning conditions of the rule.

5. Click OK.

What's next

After you create a custom rule for abnormal activities, SDDP identifies activities that match the rule as abnormal ones. You can view statistics on these abnormal activities on the Custom Anomalous Event tab of the Anomalous Event Processing page. For more information, see *Process abnormal activities*.

13.8.2.7.2 Process abnormal activities

This topic describes how to process abnormal activities in Sensitive Data Discovery and Protection (SDDP). SDDP can detect abnormal activities related to sensitive data and generate alerts. On the Anomalous Event Processing page, you can confirm abnormal activities as violations or exclude them as false positives.

Context

SDDP divides abnormal activities into the following types:

- Permission Usage Anomalous Event: Permissions are used in an inappropriate way. For example, a user logs on from an unusual IP address or by using the AccessKey pair of another user.
- Data Flow Anomalous Event: Abnormal activities are detected during dataflows. For example, a user downloads sensitive data files unnecessarily or during an unusual period.
- Data Operation Anomalous Event: Unusual operations are performed on sensitive data. For example, a user modifies sensitive fields.
- Custom Anomalous Event: Abnormal activities are detected based on custom rules.

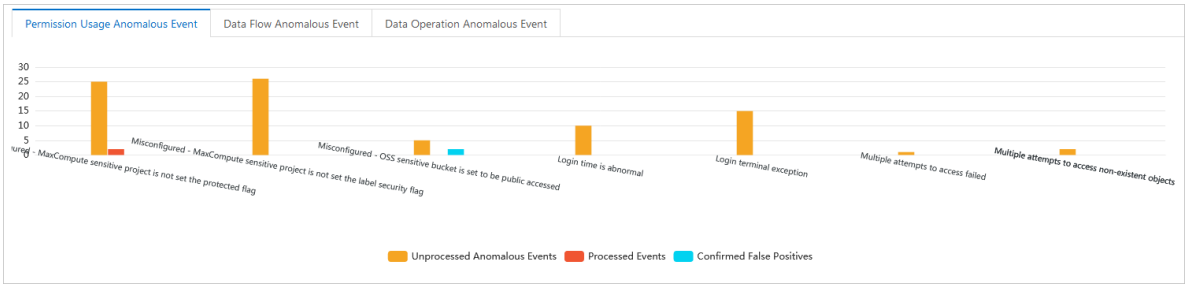
Procedure

1. *Log on to Apsara Stack Security Center.*
2. In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Anomaly Detection > Anomalous Events.
3. View the statistics on abnormal activities.

   You can view the statistics on different types of abnormal activities in the upper part of the Anomalous Event Processing page. The statistics include the types of

abnormal activities, number of processed abnormal activities, and number of unprocessed abnormal activities.

Figure 13-6: Statistics on abnormal activities



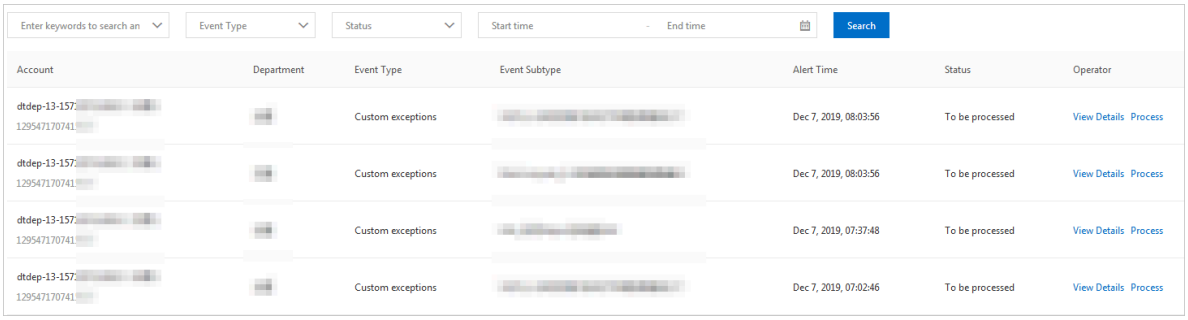
- **Click the Permission Usage Anomalous Event, Data Flow Anomalous Event, Data Operation Anomalous Event, or Custom Anomalous Event tab, and view the bar charts for the statistics.**
- **Move the pointer over an abnormal activity type to view the details.**

4. **Set search conditions and click Search to search for abnormal activities.**

   You can filter abnormal activities by keyword, department, type, status, and alerting time.

5. **Process abnormal activities.**

   You can process abnormal activities detected by SDDP in the abnormal activity list in the lower part of the Anomalous Event Processing page.



   a) **Find the target abnormal activity and click View Details in the Operator column to view the details of the abnormal activity.**

   b) **Find the target abnormal activity and click Process in the Operator column to process the abnormal activity.**

   c) **In the Anomalous Event Processing dialog box that appears, process the abnormal activity.**

Parameter	Description
Add Processing Record	Check the abnormal activity and record the verification process.
Anomalous Event Verification	· Confirmed and Processed: Confirm that the activity is an abnormal activity. If you select this option without manually processing the abnormal activity in the corresponding service, SDDP continues to generate alerts for this activity. · False Positive: Exclude the abnormal activity as a false positive. If you select this option, SDDP no longer generates alerts for this activity. In this situation, this abnormal activity will no longer appear on the Anomalous Event Processing page.

Parameter	Description
Anomalous Event Sample-based Enhancement	Select whether to use the processing result of the abnormal activity to enhance the detection of abnormal activities. **Note:** If you select this check box: · An abnormal activity that is excluded as a false positive will be returned to the current algorithm as a false positive sample. · An abnormal activity that is confirmed as a violation will be returned to the current algorithm as a positive sample. This improves the accuracy of subsequent abnormal activity detection, but may also increase the false negative rate.

d) Click Completed.

13.8.2.8 Intelligent audit

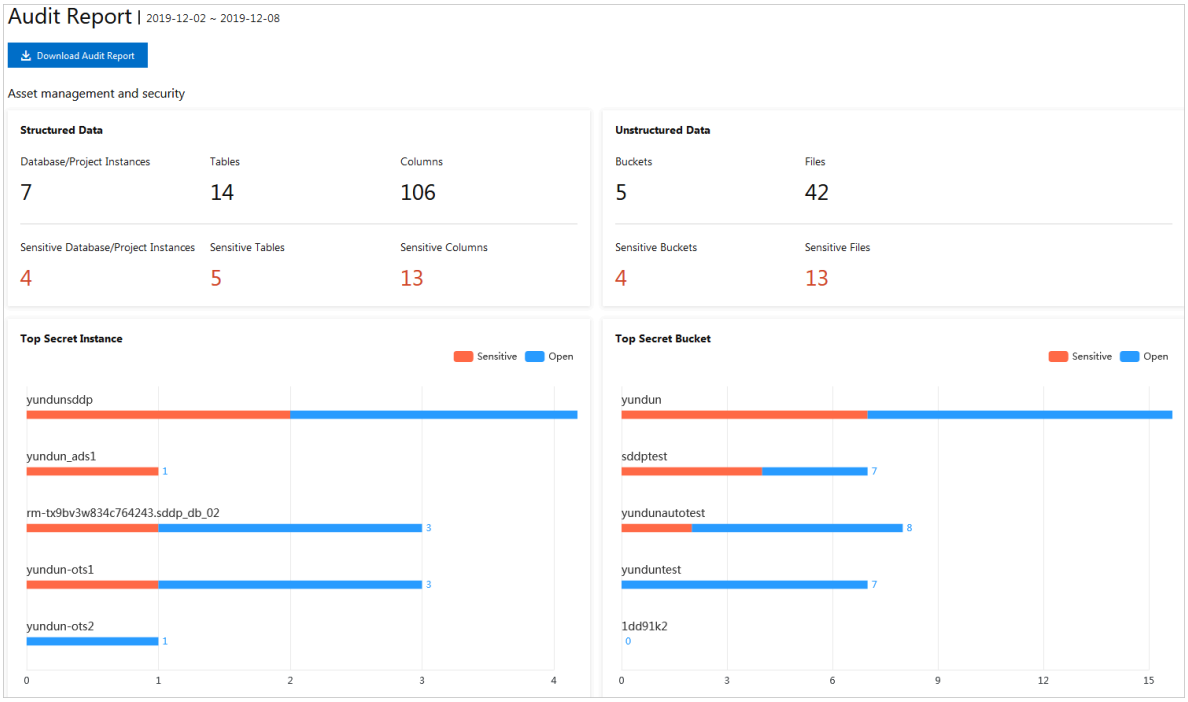
13.8.2.8.1 View and download audit reports

This topic describes how to view and download audit reports for sensitive data.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Intelligent Audit > Audit Report.

3. On the Audit Report page that appears, view an audit report.



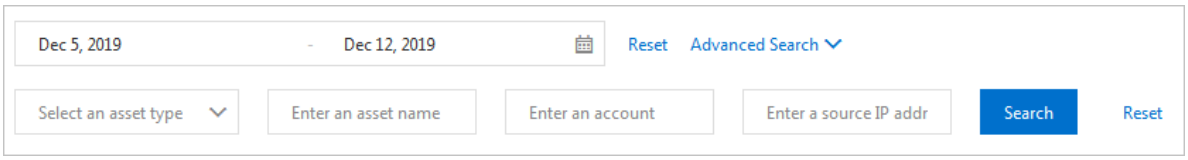
4. Click Download Audit Report to download the audit report.

13.8.2.8.2 View audit logs

This topic describes how to view audit logs.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Intelligent Audit > Audit Logs.
3. Optional: On the Audit Logs page that appears, set the time period to filter audit logs.



If you require more detailed search conditions, follow these steps:

a) Click Advanced Search.

b) Set search conditions such as the asset type, asset name, account, and source IP address.

c) Click Search.

4. View audit logs in the log list.

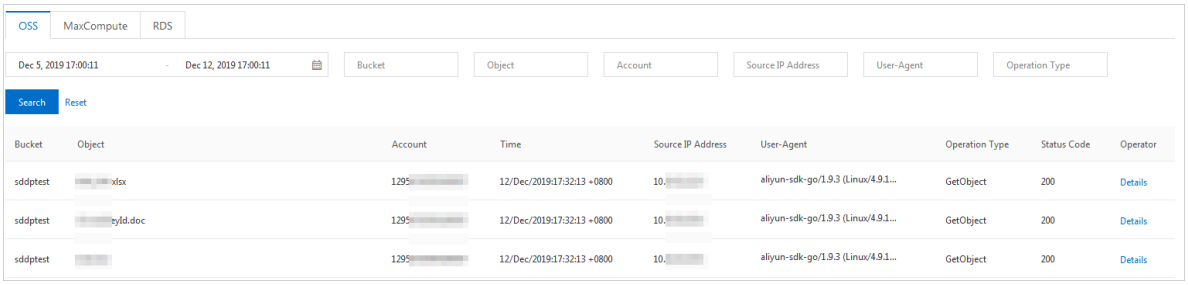
13.8.2.8.3 View raw logs

This topic describes how to view raw logs of the following services: Object Storage Service (OSS), MaxCompute, and ApsaraDB for RDS.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Intelligent Audit > Raw Logs.
3. On the Raw Logs page that appears, click a tab.

   On the Raw Logs page, you can view raw logs of OSS, MaxCompute, and RDS.
4. Optional: Set conditions to filter raw logs.
5. View information of raw logs in the log list.



6. Click Details in the Operator column corresponding to a raw log to view its information.

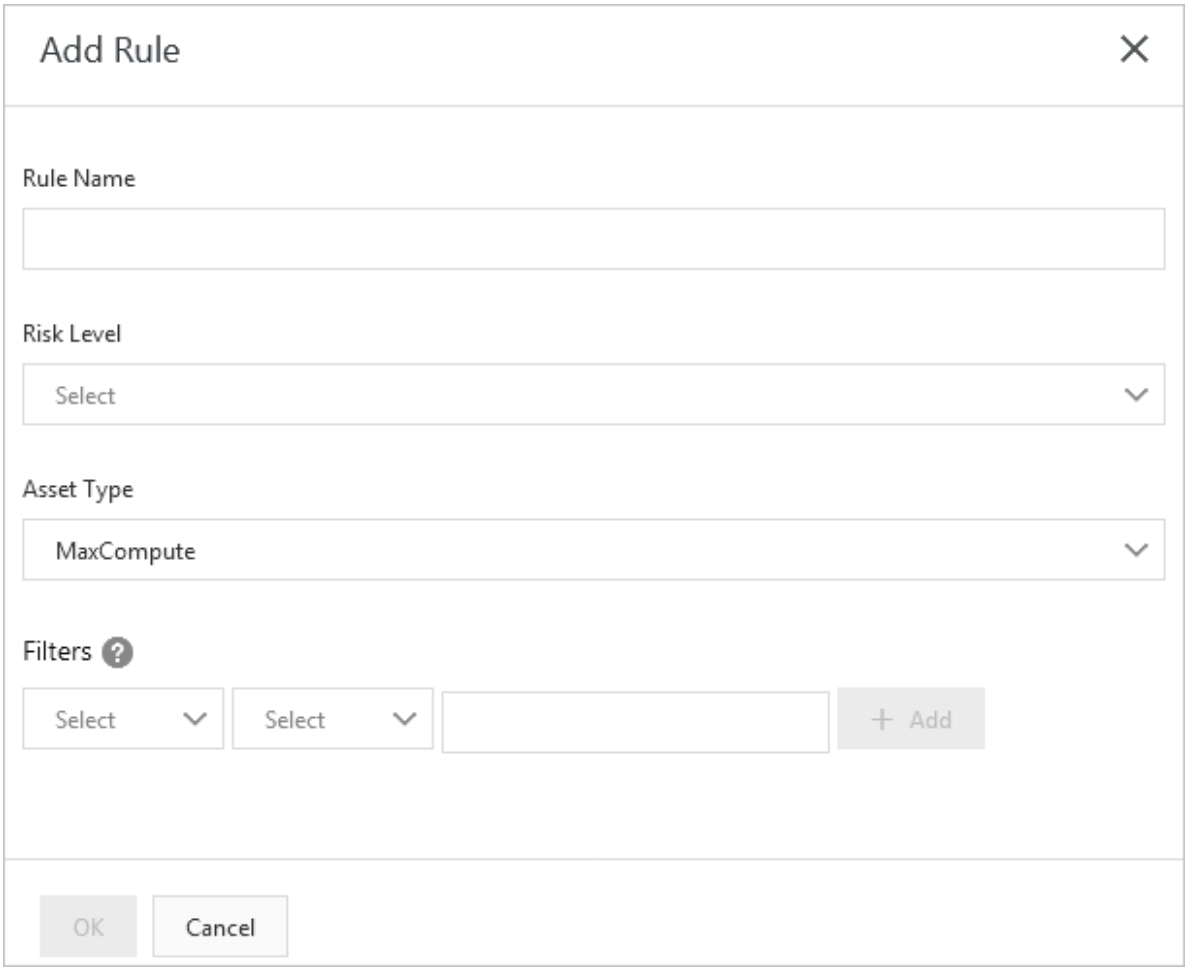
13.8.2.8.4 Add an audit rule

This topic describes how to add an audit rule. Sensitive Data Discovery and Protection (SDDP) audits raw logs that conform to audit rules.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Intelligent Audit > Audit Rules.
3. Click Add Rule on the page that appears.

4. In the Add Rule dialog box that appears, set parameters.



Parameter	Description
Rule Name	The name of the audit rule.
Risk Level	The risk level of the audit rule. Valid values: `Low`, `Medium`, and `High`.
Asset Type	The type of the asset.
Filters	The filter conditions of the rule. If a raw log meets the conditions of the rule, audit is required for the log.

5. Click OK.

Result

After you create an audit rule, it is displayed in the rule list. You can view details of a rule or edit or delete a rule by clicking the corresponding button in the Operator column.

13.8.2.9 Security configuration

13.8.2.9.1 Manage rules used to detect sensitive data

This topic describes how to create and manage rules used to detect sensitive data.

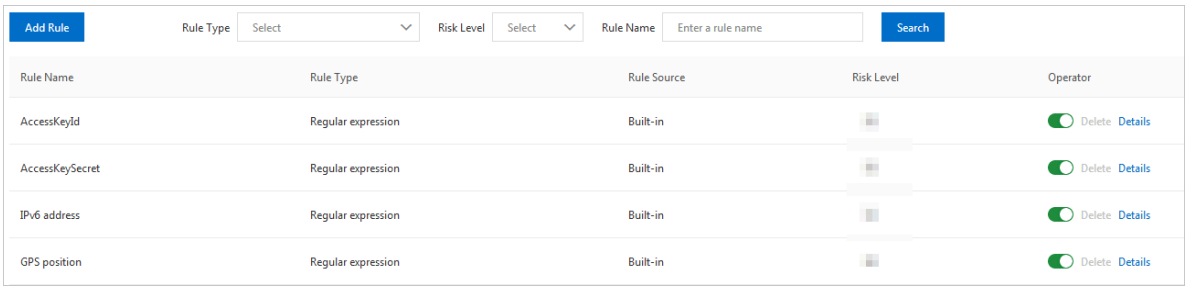
Context

Sensitive Data Discovery and Protection (SDDP) can detect and classify sensitive data in Apsara Stack services such as MaxCompute, Object Storage Service (OSS), and Table Store.

SDDP detects the sensitive data based on the rules. You can use the built-in rules of SDDP or configure custom rules based on your business needs to detect specific sensitive data.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Security Configuration > Rule Configuration.**
3. **Click the Sensitive Data Identification Rules tab to view the existing rules used to detect sensitive data.**



Set Rule Type, Risk Level, and Rule Name. Then click Search to search for rules.

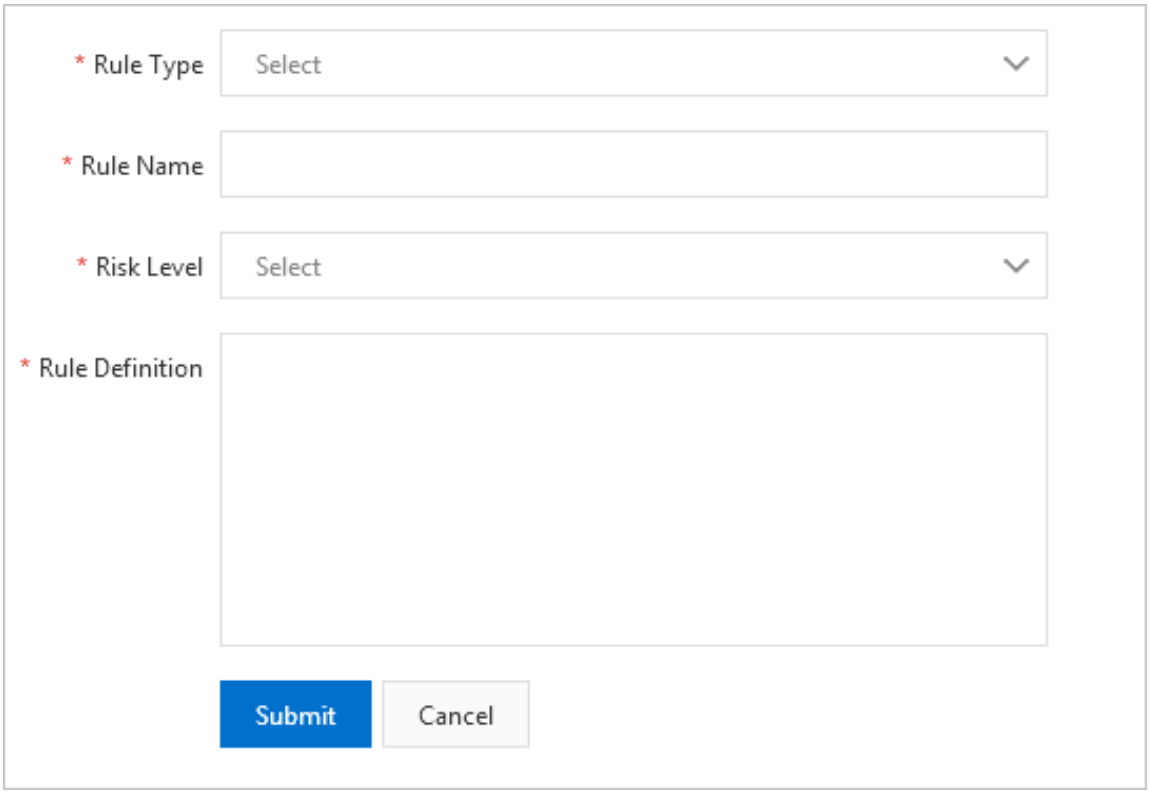
SDDP provides built-in algorithms to detect sensitive data such as ID card numbers, addresses, phone numbers, and bank card numbers. It can also use file clustering, deep neural network, and machine learning to identify sensitive images, text, and fields.

4. **Create a rule.**

   You can create a rule to detect specific sensitive data.

   a) Click Add Rule on the page that appears.

   b) In the Add Rule dialog box that appears, set parameters.



   - **Rule Type: the type of the rule. Valid values:** `Keyword` **and** `Regular expression`**.**
   - **Rule Name: the name of the rule. We recommend that you name the rule based on its purpose.**
   - **Risk Level: the risk level of the rule. Valid values: S1 (low), S2 (medium), S3 (high), and S4 (critical).**
   - **Rule Definition: the content of the rule.**

   c) Click Submit.

5. **Manage the rules.**

   In the rule list, you can click the corresponding button in the Operator column to disable, enable, or delete a rule.

   Note:

- You can delete rules but cannot modify them. After you delete a rule, SDDP no longer detects corresponding data as sensitive data. Exercise caution when deleting a rule.
- A rule is enabled by default after it is created. If you do not regard certain data as sensitive data, you can disable the corresponding rule. After you disable a rule, SDDP no longer detects corresponding data as sensitive data. We recommend that you enable all rules to reduce risks.
- Rules for which `Built-in` appears in the Rule Source column are default rules. If no custom rules are configured, SDDP can still detect sensitive data based on these default rules. The default rules cannot be modified or deleted.

13.8.2.9.2 Manage the thresholds and rules used to detect abnormal activities

This topic describes how to manage the thresholds and rules used to detect abnormal activities.

Context

On the Rule Configuration page, you can customize the thresholds and rules used to detect abnormal activities.

Procedure

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Security Configuration > Rule Configuration.**
3. **Click the Anomaly Alert Configuration tab to view the thresholds and rules used to detect abnormal activities.**
4. **Configure the thresholds used to detect abnormal activities.**

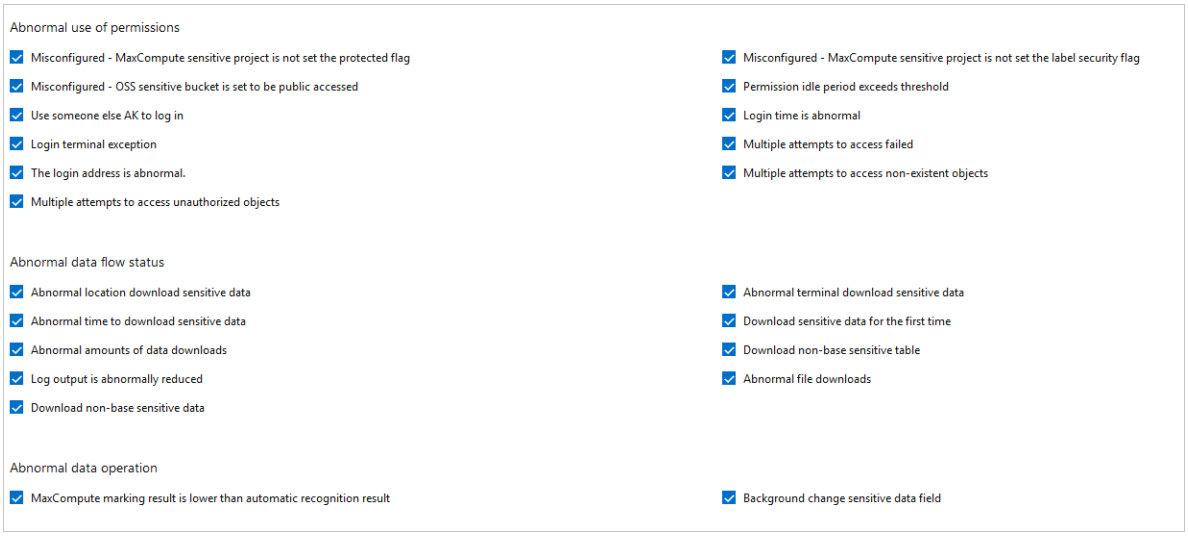
Sensitive Data Discovery and Protection (SDDP) provides default thresholds that can be customized.

a) **In the General Configuration for Anomaly Alerts section, click Modify next to a threshold.**

b) **Enter a value and click Submit.**

5. **Configure the rules used to detect abnormal activities.**

   In the Enable Anomaly Alerts section, select the types of abnormal activities that you want SDDP to detect.



   After you configure the rules, you can choose Anomaly Detection > Anomalous Events in the left-side navigation pane of the Sensitive Data Protection page, and view the statistics on suspicious events on the page that appears.

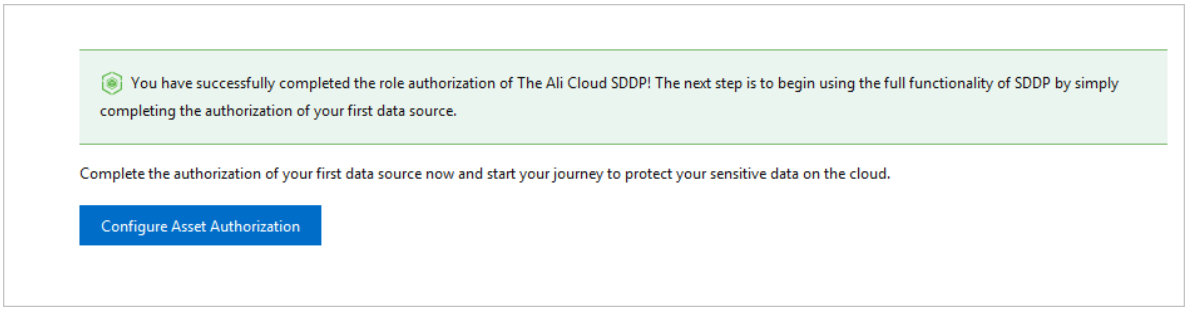
13.8.2.9.3 Configure an authorized asset

This topic describes how to configure an authorized asset for Sensitive Data Discovery and Protection (SDDP). After the configuration, SDDP scans authorized data sources. If you identify any data leak or sensitive data, choose Anomaly Detection > Anomalous Events in the left-side navigation pane of the Sensitive Data Protection page, and view suspicious events on the page that appears.

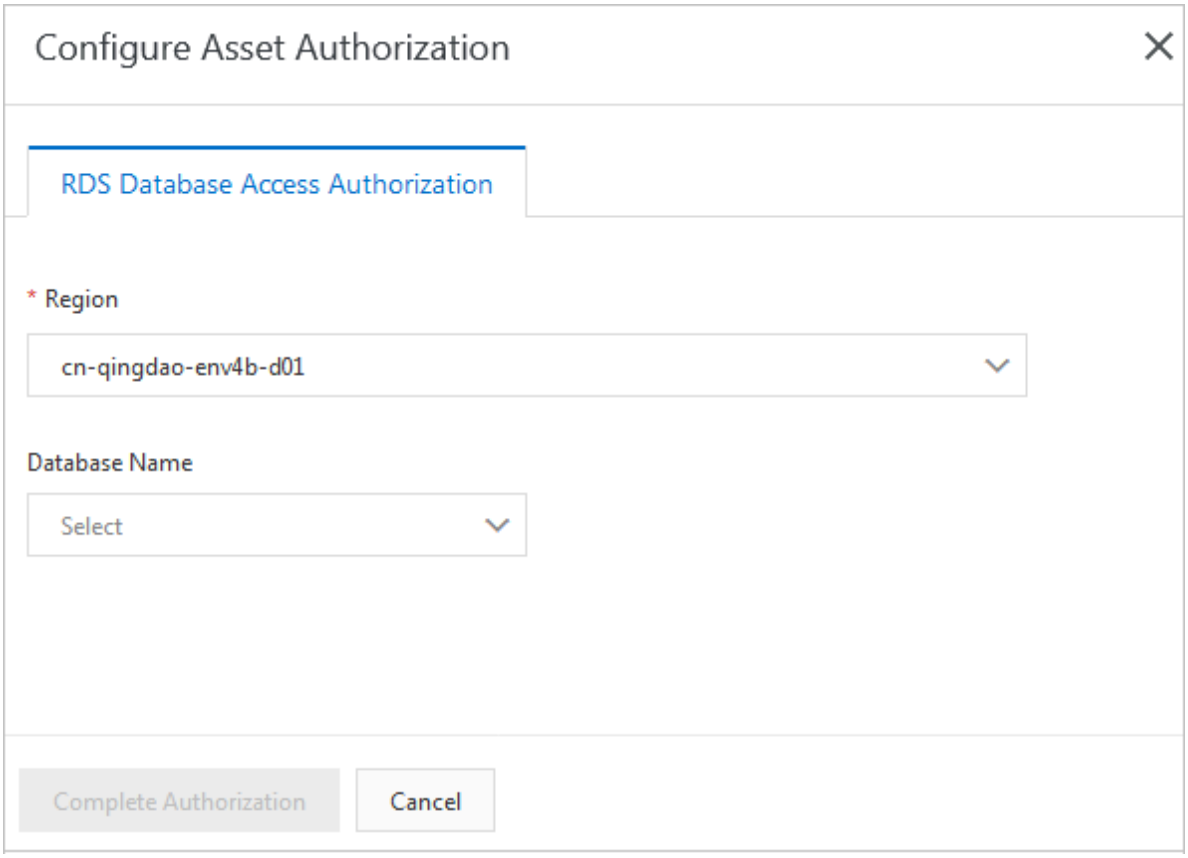
Procedure

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Security Configuration > Authorization Configuration.**

3. On the Authorization Configuration page that appears, click Configure Asset Authorization.



4. Select a region from the Please select your region drop-down list, and select a database instance and the corresponding database from the drop-down list under Select the database instance and the corresponding database that needs to be authorized.



5. Click Complete Authorization.

Result

After the configuration is complete, SDDP can scan authorized data sources.

13.8.2.9.4 Configure desensitization algorithms

This topic describes how to configure desensitization algorithms.

Context

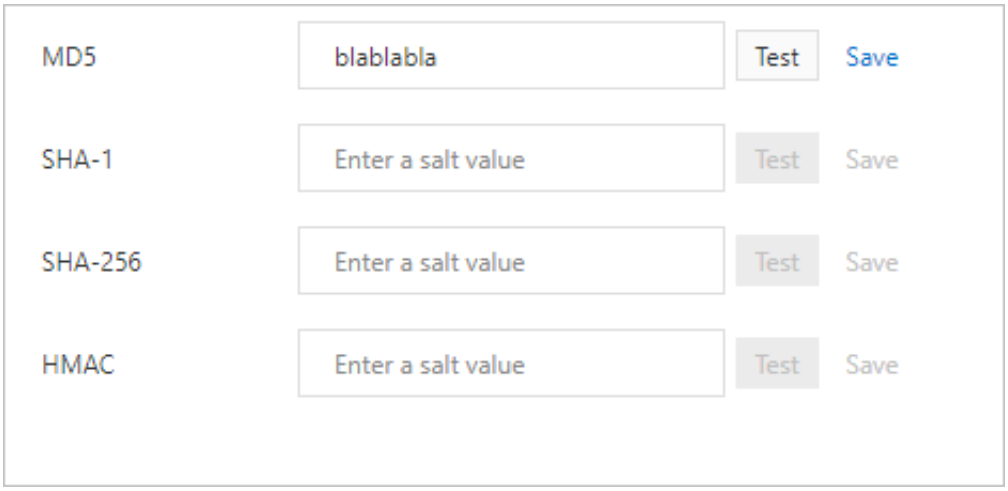
Sensitive Data Discovery and Protection (SDDP) supports the following desensitization methods:

- **Hashing: performs tokenization or masks passwords. Raw data cannot be retrieved after it is desensitized by using this method. The MD5, Secure Hash Algorithm 1 (SHA-1), SHA-256, and hash-based message authentication code (HMAC) salted algorithms are supported.**
- **Masking: replaces targeted information in sensitive data with asterisks (*) or number signs (#) in any of the following ways:**
  - **Keep the first N characters and the last M characters.**
  - **Keep characters from the Xth position to the Yth position.**
  - **Mask the first N characters and the last M characters.**
  - **Mask characters from the Xth position to the Yth position.**
  - **Mask characters before a special character.**
  - **Mask characters after a special character.**
- **Replacement: uses a mapping table or interval to randomly replace or map entire or partial field values. Raw data cannot be retrieved after it is desensitiz ed by using this method. SDDP provides multiple built-in mapping tables in the .txt or .rtf format and allows you to add custom replacement algorithms. This method is suitable for fields with a fixed format, such as ID card numbers.**
- **Transformation: rounds or offsets field values to desensitize them. You can round numbers and dates or offset characters in text based on specified parameters . Raw data can be retrieved after it is offset but cannot be retrieved after it is rounded.**
- **Encryption: uses the Data Encryption Standard (DES), Triple Data Encryption Standard (3DES), or Advanced Encryption Standard (AES) algorithm to encrypt data. Raw data can be retrieved after it is desensitized by using this method.**
- **Shuffling: shuffles values of a field within a specified range of a source table. The values can be shuffled randomly or be offset in a specified way. Raw data can be retrieved after it is offset but cannot be retrieved after it is shuffled randomly.**

Hashing

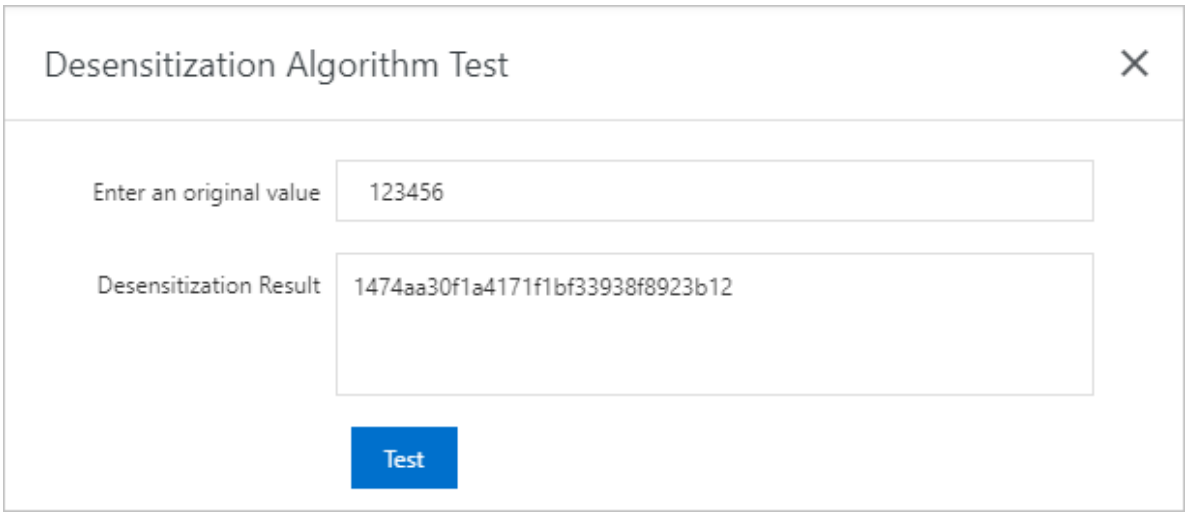
1. *Log on to Apsara Stack Security Center*.
2. In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Security Configuration > Desensitization Algorithm Configuration.
3. Click the Hashing tab.
4. Set a salt value for each algorithm.

> **Note:**
>
> In cryptography, you can insert a specific string to a fixed position of a password to generate a hash value that is different from that of the original password. This process is called salting.
>
> A salt value is the specific string that you insert.



5. **Click Test for an algorithm.**

   In the Desensitization Algorithm Test dialog box that appears, enter the original value and click Test to check whether the algorithm works.

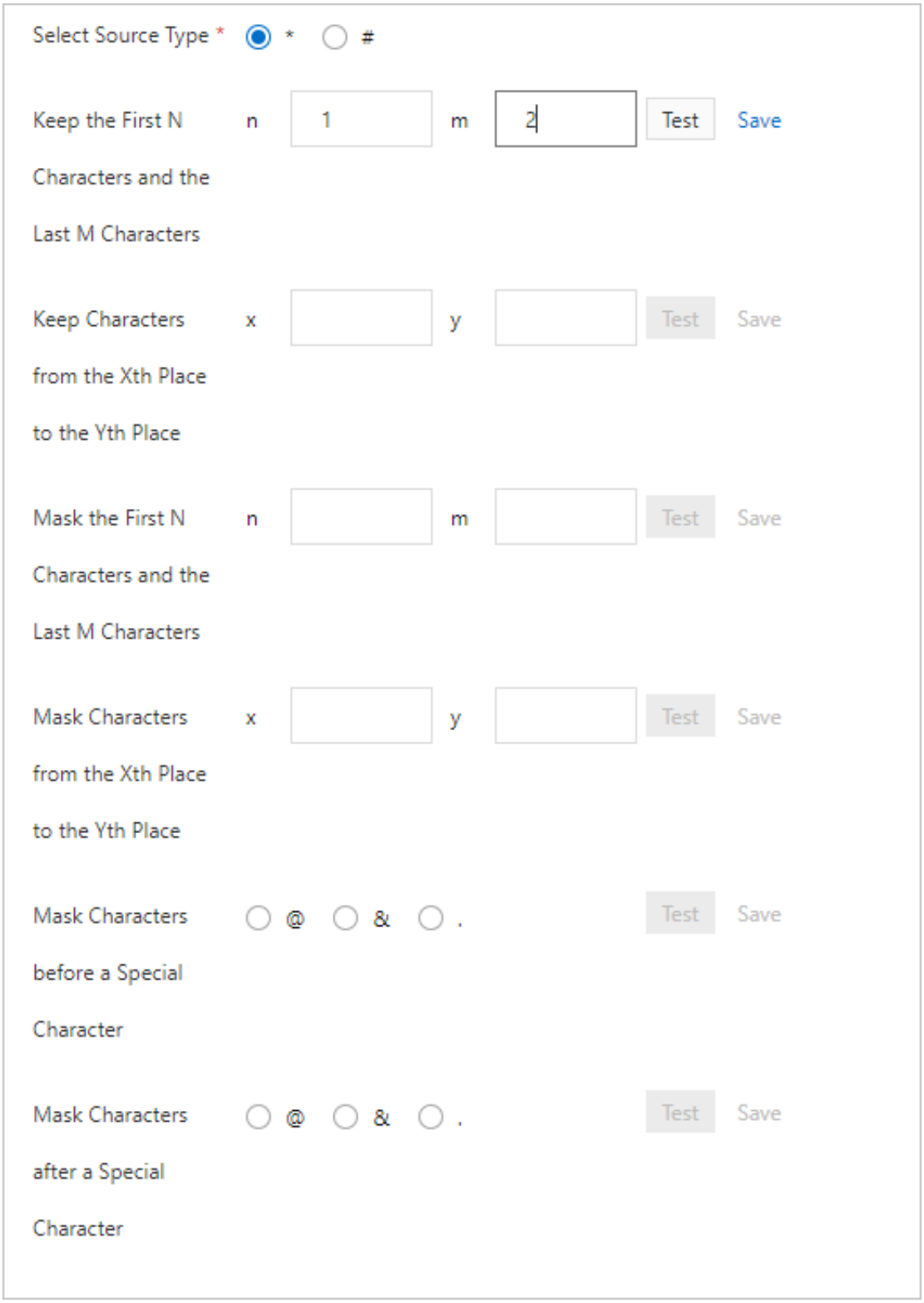


6. **After the test, click the close icon in the Desensitization Algorithm Test dialog box. Then click Save corresponding to the algorithm on the Hashing tab.**

Masking

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Security Configuration > Desensitization Algorithm Configuration.**
3. **Click the Masking tab.**

4. **Set parameters.**



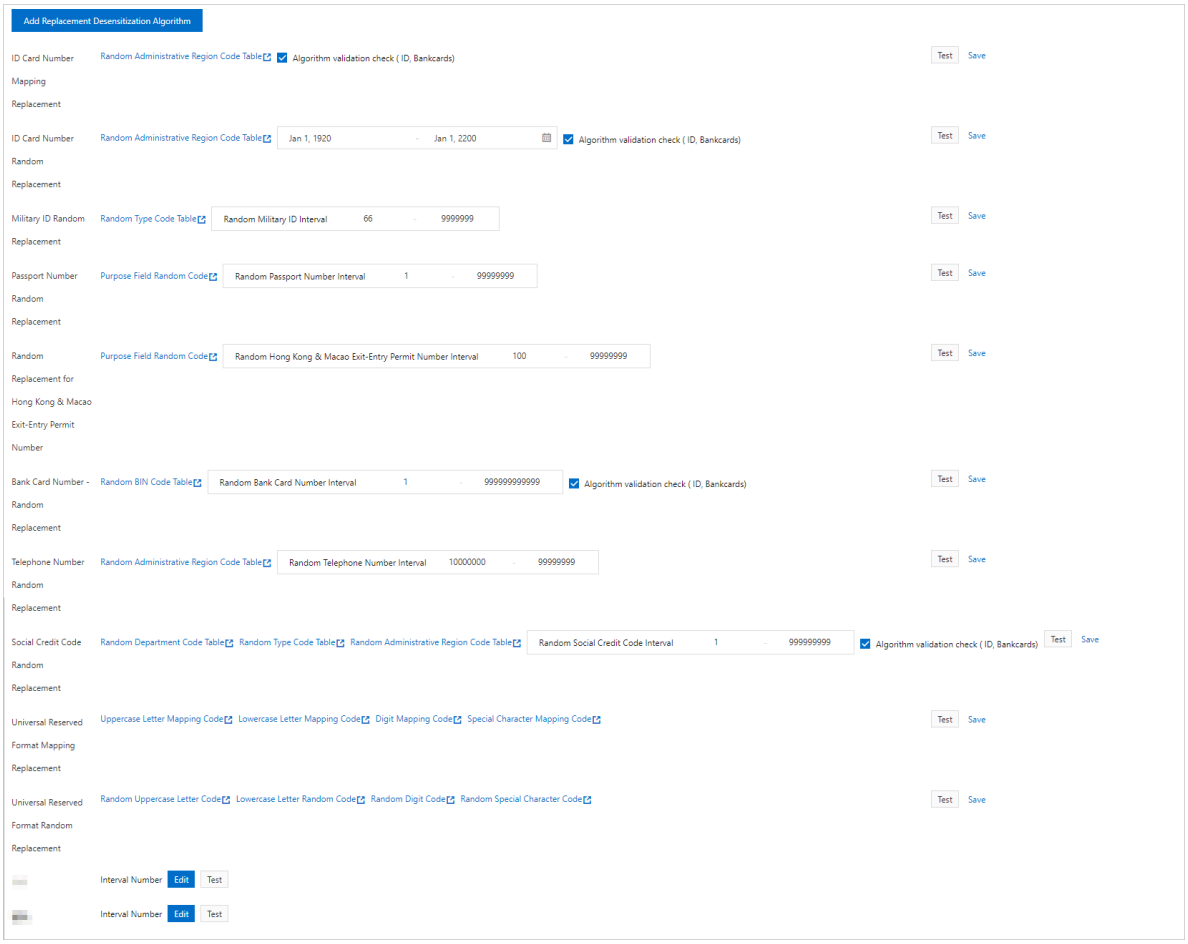
5. **Click Test for an algorithm.**

   In the Desensitization Algorithm Test dialog box that appears, enter the original value and click Test to check whether the algorithm works.

6. **After the test, click the close icon in the Desensitization Algorithm Test dialog box. Then click Save corresponding to the algorithm on the Masking tab.**

Replacement

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Security Configuration > Desensitization Algorithm Configuration.**
3. **Click the Replacement tab.**
4. **Set parameters.**



Note:

By default, SDDP provides multiple common replacement algorithms, such as ID Card Number Mapping Replacement and Telephone Number Random Replacement.

- **If you need to customize a mapping table, click the corresponding mapping table, replace the original content with your own mapping table, and click Save.**
- **If you need to customize an algorithm, click Add Replacement Desensitization Algorithm and custom the interval and mapping table.**

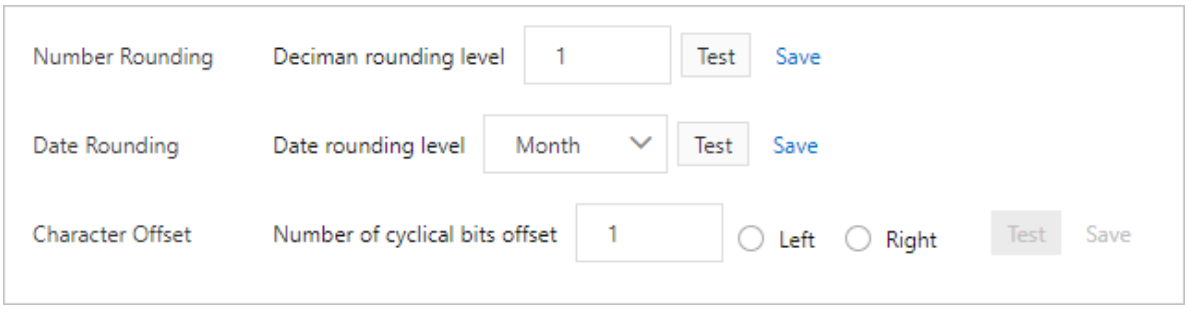
5. **Click Test for an algorithm.**

   In the Desensitization Algorithm Test dialog box that appears, enter the original value and click Test to check whether the algorithm works.

6. **After the test, click the close icon in the Desensitization Algorithm Test dialog box. Then click Save corresponding to the algorithm on the Replacement tab.**

Transformation

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Security Configuration > Desensitization Algorithm Configuration.**
3. **Click the Transformation tab.**
4. **Set parameters.**



5. **Click Test for an algorithm.**

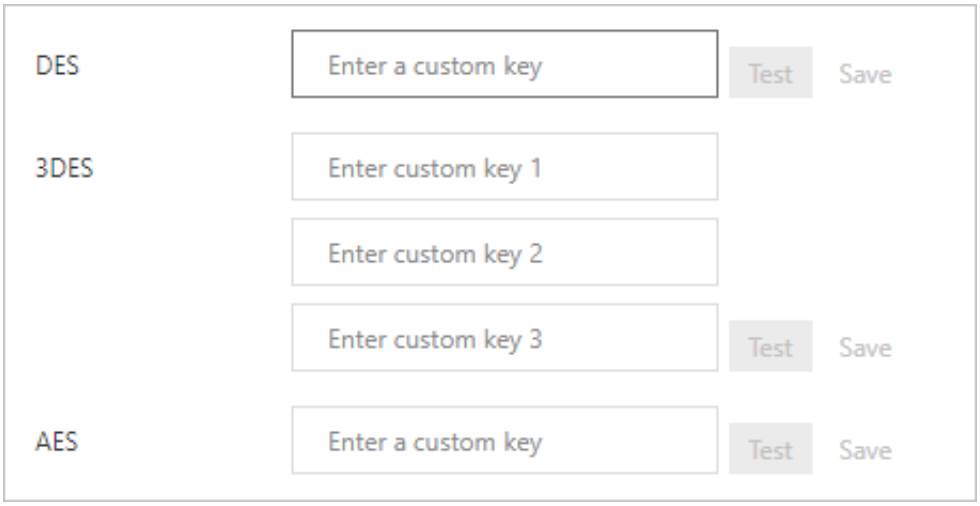
   In the Desensitization Algorithm Test dialog box that appears, enter the original value and click Test to check whether the algorithm works.

6. **After the test, click the close icon in the Desensitization Algorithm Test dialog box. Then click Save corresponding to the algorithm on the Transformation tab.**

Encryption

1. *Log on to Apsara Stack Security Center*.

2. **In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Security Configuration > Desensitization Algorithm Configuration.**
3. **Click the Encryption tab.**
4. **Set a key for an algorithm.**



5. **Click Test for an algorithm.**

   In the Desensitization Algorithm Test dialog box that appears, enter the original value and click Test to check whether the algorithm works.
6. **After the test, click the close icon in the Desensitization Algorithm Test dialog box. Then click Save corresponding to the algorithm on the Encryption tab.**

Shuffling

1. *Log on to Apsara Stack Security Center*.
2. **In the left-side navigation pane, choose Data Security > Sensitive Data Protection > Security Configuration > Desensitization Algorithm Configuration.**
3. **Click the Shuffling tab.**
4. **Select a shuffling method.**

5. **Click Save.**

14 Apsara Stack DNS

14.1 What is Apsara Stack DNS?

Apsara Stack DNS is an Apsara Stack service that resolves domain names. Apsara Stack DNS translates the requested domain names based on the rules and policies you set for domain names and IP addresses, and redirects the requests from the client to the corresponding cloud services, enterprise business systems, or services provided by Internet service providers.

Apsara Stack DNS provides basic domain name resolution and scheduling services for VPCs. You can perform the following operations on your VPC by using Apsara Stack DNS:

- **Access other ECS instances deployed in your VPC.**
- **Access cloud service instances provided by Apsara Stack.**
- **Access custom enterprise business systems.**
- **Access Internet services and businesses.**
- **Establish network connections between DNS and user-created DNS over a leased line.**
- **Manage internal domain names.**
- **Manage DNS records of internal domain names.**
- **Manage forwarding configurations.**
- **Manage recursive resolution configurations.**

14.2 User roles and permissions

Role	Permission
System administrator	Owners of this role have read, write, and execute permissions on all level -1 organization resources, global resources, and system configurations.

Role	Permission
Level-1 organization administrator	Owners of this role have read, write , and execute permissions on level-1 organization resources to which a user belongs, but do not have permission s on level-1 organization resources to which the user does not belong, global resources, or system configurations.
Lower-level organization administrator	Owners of this role do not have permissions on Apsara Stack DNS . They do not have permissions on level-1 organization resources, global resources, or system configurations.
Resource user	Owners of this role do not have permissions on Apsara Stack DNS . They do not have permissions on level-1 organization resources, global resources, or system configurations.
Other roles	Owners of this role do not have permissions on Apsara Stack DNS . They do not have permissions on level-1 organization resources, global resources, or system configurations.

14.3 Log on to the Apsara Stack DNS console

This topic uses Google Chrome as an example to demonstrate how to log on to the Apsara Stack DNS console.

Prerequisites

- Before logging on to the ASCM console, make sure that you have obtained the IP address or domain name of the ASCM console from the deployment personnel . The URL used to access the ASCM console is in the following format: http://IP address or domain name of the ASCM console/manage.
- We recommend that you use the Google Chrome browser.

Procedure

1. In the address bar, enter the URL used to access the ASCM console. Press Enter.

2. **Enter your username and password.**

   The system has a default super administrator, whose username is super. The super administrator can create system administrators. A system administrator can create system users and notify the users of the default passwords by SMS or email.

   > **Note:**
   > **When you log on to the ASCM console for the first time, you must modify the password of your username as instructed. For security concerns, your password must meet the minimum complexity requirements: The password must be 8 to 20 characters in length and must contain at least two types of the following characters: letters, digits, and special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).**

3. **Click Login to go to the ASCM console homepage.**
4. **In the top navigation bar, choose Products > Application Services > Apsara Stack DNS.**

14.4 Management of internal domain names

14.4.1 Management of tenant internal domain names (Standard Edition only)

14.4.1.1 View a domain name

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Internal Domains > Tenant Internal Domains.**
3. **In the Domain Name search bar, enter the domain name that you want to view.**
4. **Click Search.**

   The search result is displayed.

14.4.1.2 Add a domain name

Procedure

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

2. **In the left-side navigation pane, choose Internal Domains > Tenant Internal Domains.**
3. **Click Add Domain Name.**
4. **In the dialog box that appears, set** `Tenant Internal Domain Name`**.**
5. **Click OK.**

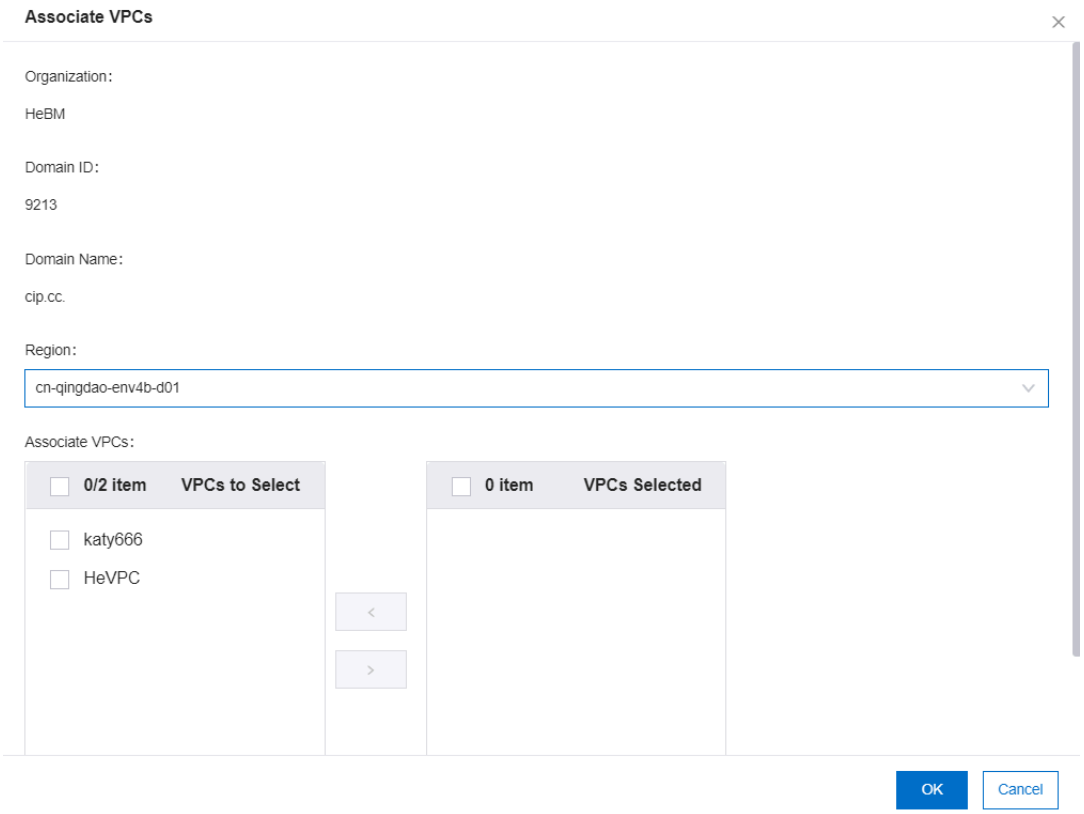
14.4.1.3 Associate a domain name with a VPC

Tenants are isolated by using VPCs. Before the DNS forwarding configurations of a domain name can take effect in a VPC, you must associate the domain name with the VPC.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Internal Domains > Tenant Internal Domains.**
3. **Select the domain name that you want to associate with a VPC, click**  **in the Actions column, and then select Associate VPCs from the shortcut menu.**

4. Select the VPC to be associated from the VPCs to Select section, click the right arrow to move the selected VPC to the VPCs Selected section, and then click OK.



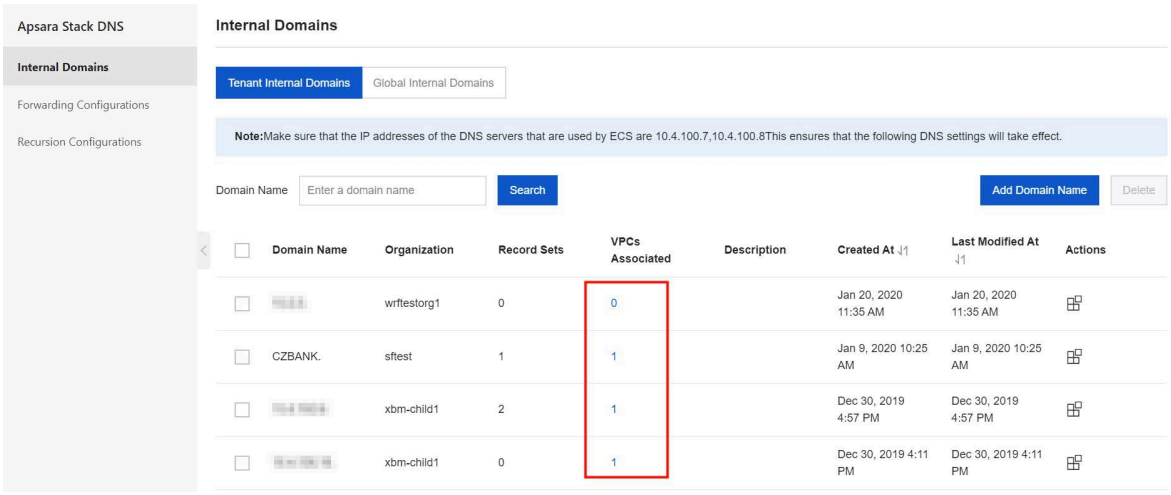
14.4.1.4 Disassociate a domain name from a VPC

This topic describes how to disassociate a domain name from a VPC.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Internal Domains > Tenant Internal Domains.

3. Select the domain name that you want to disassociate from the VPC and click the value in the VPCs Associated column.



4. On the VPCs Associated page, select the VPC from which the domain name is to be disassociated, click  in the Actions column, and then select Disassociate from the shortcut menu.

   Make sure that the disassociated VPC is no longer displayed on the VPCs Associated page.

14.4.1.5 Add a description for a domain name

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Internal Domains > Tenant Internal Domains.
3. Select the domain name for which you want to add a description, click  in the Actions column, and then select Description from the shortcut menu.
4. In the dialog box that appears, enter a description.
5. Click OK.

14.4.1.6 Delete a domain name

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Internal Domains > Tenant Internal Domains.

3. Select the domain name that you want to delete, click  in the Actions column, and then select Delete from the shortcut menu.
4. In the message that appears, click OK.

14.4.1.7 Delete domain names in batches

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Internal Domains > Tenant Internal Domains.
3. Select the domain names that you want to delete and click Delete in the upper-right corner.
4. In the message that appears, click OK.

14.4.1.8 Configure DNS records

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Internal Domains > Tenant Internal Domains.
3. Select the domain name for which you want to configure DNS records, click  in the Actions column, and then select Configure DNS Records from the shortcut menu.
4. On the Configure DNS Records page, click Add DNS Record in the upper-right corner.

5. **In the Add DNS Record dialog box, set** `Host`**,** `Type`**,** `TTL`**,** `Resolution Policy`**, and** `Record Set`**. Then, click OK.**

   The following tables describe the types of DNS records.

   - **A record**

Resolution policy	Formatting rule
No	You can enter only one IPv4 address in each row. You can enter up to 100 different IPv4 addresses in separate rows. IPv4 addresses must be in the standard IPv4 address format. Examples: - 192.168.1.1 - 192.168.1.2 - 192.168.1.3

Resolution policy	Formatting rule
Weight	You can enter only one IPv4 address in each row. You can enter up to 100 different IPv4 addresses in separate rows. Format: - [IPv4 address] [Weight]. The IPv4 address and weight must be separated with a space character. - IPv4 addresses must be in the standard IPv4 address format. - The value of the weight must be an integer ranging from 0 to 999. A greater value indicates a greater weight. Examples: - 192.168.1.1 20 - 192.168.1.1 30 - 192.168.1.1 50

- AAAA record

Resolution policy	Formatting rule
No	You can enter only one IPv6 address in each row. You can enter up to 100 different IPv6 addresses in separate rows. IPv6 addresses must be in the standard IPv6 address format. Examples: - 2400:3200::6666 - 2400:3200::6688 - 2400:3200::8888

Resolution policy	Formatting rule
Weight	You can enter only one IPv6 address in each row. You can enter up to 100 different IPv6 addresses in separate rows. Format: - [IPv6 address] [Weight]. The IPv6 address and weight must be separated with a space character. - IPv6 addresses must be in the standard IPv6 address format. - The value of the weight must be an integer ranging from 0 to 999. A greater value indicates a greater weight. Examples: - 2400:3200::6666 20 - 2400:3200::6688 20 - 2400:3200::8888 60

- CNAME record

Resolution policy	Formatting rule
No	You can enter only one domain name in each row. The domain name must be a fully qualified domain name (FQDN) that ends with a period (.). It must be 1 to 255 ASCII characters in length. Example: www.example.com.

Resolution policy	Formatting rule
Weight	You can enter only one domain name in each row. You can enter up to 100 different domain names in separate rows. Format: - [Domain name] [Weight]. The domain name and weight must be separated with a space character. - The domain name must be an FQDN that ends with a period (.). It must be 1 to 255 ASCII characters in length. - The value of the weight must be an integer ranging from 0 to 999. A greater value indicates a greater weight. Examples: - www1.example.com. 20 - www2.example.com. 20 - www3.example.com. 60

- MX record

Resolution policy	Formatting rule
No	You can enter only one MX record in each row. You can enter up to 100 different MX records in separate rows. Format: - [Priority] [Email server hostname]. The priority and email server hostname must be separated with a space character. - The value of the priority must be an integer ranging from 0 to 999. A smaller value indicates a higher priority. - The email server hostname must be an FQDN that ends with a period (.). It must be 1 to 255 ASCII characters in length. Examples: - 10 mailserver1.example.com. - 20 mailserver2.example.com.

- TXT record

Resolution policy	Formatting rule
No	You can enter only one TXT record in each row. You can enter up to 100 different TXT records in separate rows. A TXT record must be 1 to 255 ASCII characters in length. No row can be left blank. Example: "v=spf1 ip4:192.168.0.1/16 ip6:2001::1/96 ~all"

- PTR record

Resolution policy	Formatting rule
No	You can enter only one domain name in each row. You can enter up to 100 different domain names in separate rows. The domain name must be an FQDN that ends with a period (.). It must be 1 to 255 ASCII characters in length. Examples: - www1.example.com. - www2.example.com. - www3.example.com.

- SRV record

Resolution policy	Formatting rule
No	You can enter only one SRV record in each row. You can enter up to 100 different SRV records in separate rows. Format: - [Priority] [Weight] [Port number] [Application server hostname]. The priority, weight, port number, and application server hostname must be separated with space characters. - The value of the priority must be an integer ranging from 0 to 999. A smaller value indicates a higher priority. - The value of the weight must be an integer ranging from 0 to 999. A greater value indicates a greater weight. - The value of the port number must be an integer ranging from 0 to 65535. It indicates the TCP or UDP port that is used for network communication. - The application server hostname must be an FQDN that ends with a period (.). It must be 1 to 255 ASCII characters in length. Examples: - 1 10 8080 www1.example.com. - 2 20 8081 www2.example.com.

- NAPTR record

Resolution policy	Formatting rule
No	You can enter only one NAPTR record in each row. You can enter up to 100 different NAPTR records in separate rows. Format: - [Serial number] [Priority] [Flag] [Service information] [Regular expression] [Substitute domain name]. The serial number, priority, flag, service information, regular expression, and substitute domain name must be separated with space characters. - The value of the serial number must be an integer ranging from 0 to 999. A smaller serial number indicates a higher priority. - The value of the priority must be an integer ranging from 0 to 999. A smaller value indicates a higher priority. When multiple records contain the same serial number, records that have a higher priority take precedence. - The value of the flag can be null or a single character from A to Z, a to z, or 0 to 9. It is not case-sensitive and must be enclosed in double quotation marks ("). - The value of the service information can be null or a string of 1 to 32 ASCII characters. It must start with a letter and be enclosed in double quotation marks ("). - The value of the regular expression can be null or a string of 1 to 255 ASCII characters. It must be enclosed in double quotation marks ("). - The substitute domain name must be an FQDN that ends with a period (.). It must be 1 to 255 ASCII characters in length. Examples: - 100 50 "S" "Z3950+I2L+I2C" "" _z3950._tcp.example.com. - 100 50 "S" "RCDS+I2C" "" _rcds._udp.example.com. - 100 50 "S" "HTTP+I2L+I2C+I2R" "" _http._tcp.example.com.

- **CAA record**

Resolution policy	Formatting rule
No	You can enter only one CAA record in each row. You can enter up to 100 different CAA records in separate rows. Format: - [Certification authority flag] [Certificate property tag] [Authorization information]. The certification authority flag, certificate property tag, and authorization information must be separated with space characters. - The value of the certification authority flag must be an integer ranging from 0 to 255. - The value of the certificate property tag can be issue, issuewild, or iodef. - The value of the authorization information cannot be null. It must be 1 to 255 ASCII characters in length and enclosed in double quotation marks ("). Examples: - 0 issue "caa.example.com" - 0 issuewild ";" - 0 iodef "mailto:example@example.com"

- **NS record**

Resolution policy	Formatting rule
No	You can enter only one DNS server address in each row. You can enter up to 100 different DNS server addresses in separate rows. The domain name must be an FQDN that ends with a period (.). It must be 1 to 255 ASCII characters in length. Wildcard domain names are not allowed. Examples: - ns1.example.com. - ns2.example.com.

6. Perform the following operations as needed:
   - Add a description for a DNS record

     Select the DNS record for which you want to add a description, click  in the Actions column, and then select Description from the shortcut menu. In the dialog box that appears, enter a description and click OK.
   - Delete a DNS record

     Select the DNS record that you want to delete, click  in the Actions column, and then select Delete from the shortcut menu. In the message that appears, click OK.
   - Modify a DNS record

     Select the DNS record that you want to modify, click  in the Actions column, and then select Modify from the shortcut menu. In the dialog box that appears, set the required parameters and click OK.
   - Delete DNS records in batches

     Select the DNS records that you want to delete and click Delete in the upper-right corner. In the message that appears, click OK.

14.4.1.9 Query a resolution policy

This topic describes how to view the details of a resolution policy.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Internal Domains > Tenant Internal Domains.
3. Select the domain name for which you want to configure DNS records, click  in the Actions column, and then select Configure DNS Records from the shortcut menu.
4. On the page that appears, click Weight in the Resolution Policy column to view the details of the resolution policy.

14.4.2 Management of global internal domain names

14.4.2.1 Overview

All the operations of this feature require administrator privileges.

14.4.2.2 View an internal domain name

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Internal Domains > Global Internal Domains.**
3. **In the Domain Name search bar, enter the domain name that you want to view.**
4. **Click Search.**

   The search result is displayed.

14.4.2.3 Add a domain name

This topic describes how to add a domain name in the Apsara Stack Cloud Managemant (ASCM) console.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Internal Domains > Global Internal Domains.**
3. **Click Add Domain Name.**
4. **In the dialog box that appears, enter Global Internal Domains.**
5. **Click OK.**

14.4.2.4 Add a description for a domain name

This topic describes how to add a description for a domain name in the ASCM console.

Context

You can add a description for a domain name for identification. For example, you can describe a domain name by using a hostname or internal system information.

Procedure

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

2. In the left-side navigation pane, choose Internal Domains > Global Internal Domains.
3. Select the domain name for which you want to add a description, click  in the Actions column, and then select Description from the shortcut menu.
4. In the dialog box that appears, enter a description.
5. Click OK.

14.4.2.5 Delete a domain name

This topic describes how to delete a domain name in the ASCM console.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Internal Domains > Global Internal Domains.
3. Select the domain name that you want to delete, click  in the Actions column, and then select Delete from the shortcut menu.
4. In the message that appears, click OK.

14.4.2.6 Delete domain names in batches

This topic describes how to delete domain names in batches in the ASCM console.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Internal Domains > Global Internal Domains.
3. Select the domain names that you want to delete and click Delete in the upper-right corner.
4. In the message that appears, click OK.

14.4.2.7 Configure DNS records

This topic describes how to configure DNS records in the ASCM console.

Procedure

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

2. In the left-side navigation pane, choose Internal Domains > Global Internal Domains.
3. Select the domain name for which you want to configure DNS records, click  in the Actions column, and then select Configure DNS Records from the shortcut menu.
4. On the Configure DNS Records page, click Add DNS Record in the upper-right corner.
5. Perform the following operations as needed:
   - Add a description for a DNS record

     Select the DNS record for which you want to add a description, click  in the Actions column, and then select Description from the shortcut menu. In the dialog box that appears, enter a description and click OK.
   - Delete a DNS record

     Select the DNS record that you want to delete, click  in the Actions column, and then select Delete from the shortcut menu. In the message that appears, click OK.
   - Modify a DNS record

     Select the DNS record that you want to modify, click  in the Actions column, and then select Modify from the shortcut menu. In the dialog box that appears, set the required parameters and click OK.
   - Delete DNS records in batches

     Select the DNS records that you want to delete and click Delete in the upper-right corner. In the message that appears, click OK.

14.4.2.8 Query a resolution policy

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Internal Domains > Global Internal Domains.

3. **Select the domain name for which you want to configure DNS records, click**  **in the Actions column, and then select Configure DNS Records from the shortcut menu.**
4. **On the page that appears, select the domain name for which you want to configure DNS records, and click Weight in the Resolution Policy column.**
5. **On the page that appears, view the details of Resolution Policy.**

14.5 Forwarding configuration management

14.5.1 Tenant forwarding configurations (Standard Edition only)

14.5.1.1 Tenant forwarding domain names

14.5.1.1.1 View a tenant forwarding domain name

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Forwarding Domains.**
3. **In the Domain Name search bar, enter the domain name that you want to view.**
4. **Click Search.**

   The search result is displayed.

14.5.1.1.2 Add a tenant forwarding domain name

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Forwarding Domains.**
3. **Click Add Domain Name.**

4. **In the dialog box that appears, set** `Domain Name`**,** `Forwarding Mode`**, and** `Forwarder IP Addresses`**.**

Parameter	Description
Domain Name	**The domain name, which must meet the following formatting rules:** **· The domain name must be 1 to 255 characters in length. This includes the period (.) at the end of the domain name.** **· The domain name can contain multiple domain name segments separated with periods (.). A domain name segment must be 1 to 63 characters in length and cannot be empty. It cannot contain consecutive periods (.).** **· The domain name can only contain letters (a-z, A-Z), digits (0-9), hyphens (-), and underscores (_).** **· The domain name must start with a letter, digit, or underscore (_) and must end with a letter, digit, or period (.).** **· The domain name is not case-sensitive. The system saves the domain name in lowercase letters.** **· The period (.) at the end of the domain name is optional. The system adds a period (.) to the end of the domain name.**
Forwarding Mode	**For both domain name-based forwarding and default forwarding, two forwarding modes are available: forward all requests without recursion and forward all requests with recursion.** **· Forward all requests without recursion: Only a specified DNS server is used to resolve domain names. If the specified DNS server cannot resolve the domain names, a message is returned to the DNS client to indicate that the query failed.** **· Forward all requests with recursion: A specified DNS server is preferentially used to resolve domain names. If the specified DNS server cannot resolve the domain names, the local DNS is used instead. If you enter private IP addresses in the Forwarder IP Addresses field, unexpected results may occur during recursive resolution. For example, a domain name used for internal network services may be resolved to a public IP address.**

Parameter	Description
Forwarder IP Addresses	The list of destination IP addresses. **Note:** Multiple IP addresses are separated with semicolons (;).

5. Click OK.

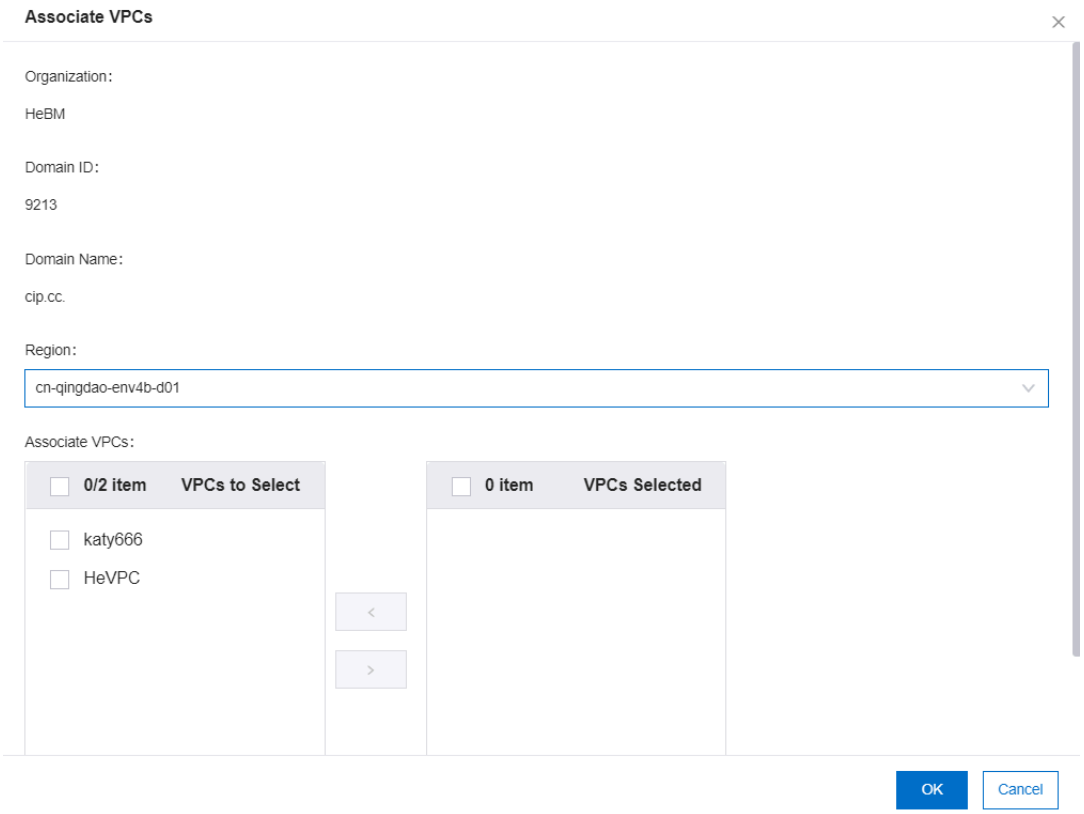
14.5.1.1.3 Associate a domain name with a VPC

Tenants are isolated by using VPCs. Before the DNS forwarding configurations of a domain name can take effect in a VPC, you must associate the domain name with the VPC.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Forwarding Domains.
3. Select the domain name that you want to associate with a VPC, click  in the Actions column, and then select Associate VPCs from the shortcut menu.

4. Select the VPC to be associated from the VPCs to Select section, click the right arrow to move the selected VPC to the VPCs Selected section, and then click OK.



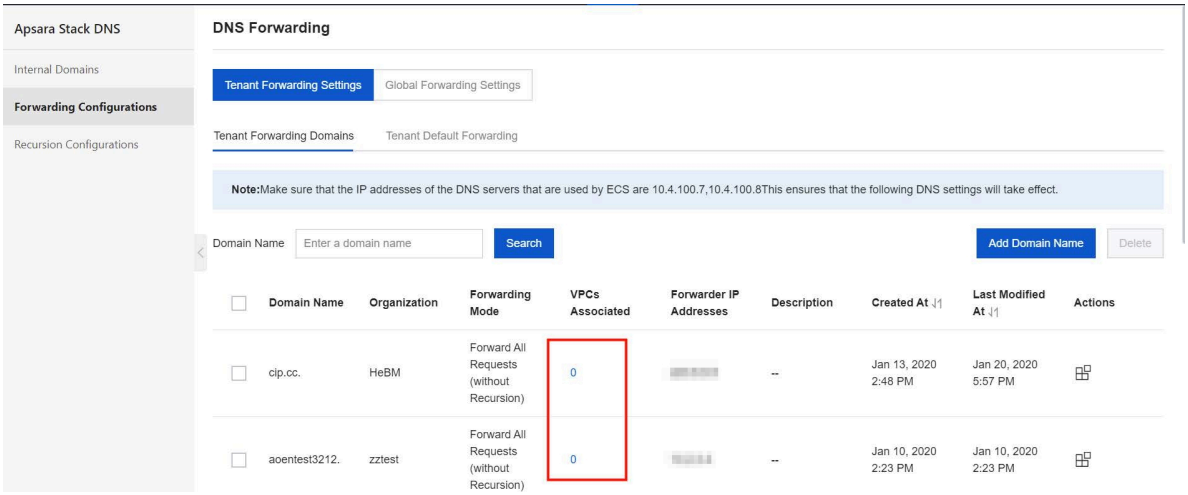
14.5.1.1.4 Disassociate a domain name from a VPC

This topic describes how to disassociate a domain name from a VPC.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Forwarding Domains.

3. Select the domain name that you want to disassociate from the VPC and click the value in the VPCs Associated column.



4. On the VPCs Associated page, select the VPC from which the domain name is to be disassociated, click  in the Actions column, and then select Disassociate from the shortcut menu.

   Make sure that the disassociated VPC is no longer displayed on the VPCs Associated page.

14.5.1.1.5 Modify the forwarding configurations of a domain name

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Forwarding Domains.
3. Select the domain name for which you want to modify the forwarding configurations, click  in the Actions column, and then select Modify from the shortcut menu.
4. In the dialog box that appears, change the value of Forwarding Mode or Forwarder IP Addresses.
5. Click OK.

14.5.1.1.6 Add a description for a tenant forwarding domain name

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Forwarding Domains.**
3. **Select the domain name that you want to add a description, click**  **in the Actions column, and then select Description from the shortcut menu.**
4. **In the dialog box that appears, enter a description.**
5. **Click OK.**

14.5.1.1.7 Delete a tenant forwarding domain name

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Forwarding Domains.**
3. **Select the domain name that you want to delete, click**  **in the Actions column, and then select Delete from the shortcut menu.**
4. **In the message that appears, click OK.**

14.5.1.1.8 Delete tenant forwarding domain names in batches

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Forwarding Domains.**
3. **Select the domain names that you want to delete and click Delete in the upper-right corner.**
4. **In the message that appears, click OK.**

14.5.1.2 Tenant default forwarding configurations

14.5.1.2.1 View default forwarding configurations

Prerequisites

You have the administrator privileges on the system and level-1 organizations.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Default Forwarding.**
3. **Move the pointer to the first drop-down list in the top navigation bar and enter the organization that you want to view in the Search by organization search box.**
4. **Press Enter.**

   The search result is displayed.

14.5.1.2.2 Add a default forwarding configuration

Prerequisites

You have the administrator privileges on the system and level-1 organizations.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Default Forwarding.**
3. **Click Add Settings.**

4. In the dialog box that appears, set `Forwarding Mode` and `Forwarder IP Addresses`.

Parameter	Description
Forwarding Mode	For both domain name-based forwarding and default forwarding, two forwarding modes are available: forward all requests without recursion and forward all requests with recursion. · Forward all requests without recursion: Only a specified DNS server is used to resolve domain names. If the specified DNS server cannot resolve the domain names, a message is returned to the DNS client to indicate that the query failed. · Forward all requests with recursion: A specified DNS server is preferentially used to resolve domain names. If the specified DNS server cannot resolve the domain names, the local DNS is used instead. If you enter private IP addresses in the Forwarder IP Addresses field, unexpected results may occur during recursive resolution. For example, a domain name used for internal network services may be resolved to a public IP address.
Forwarder IP Addresses	The list of destination IP addresses. **Note:** Multiple IP addresses are separated with semicolons (;).

5. Click OK.

14.5.1.2.3 Associate a domain name with a VPC

Tenants are isolated by using VPCs. Before the DNS forwarding configurations of a domain name can take effect in a VPC, you must associate the domain name with the VPC.

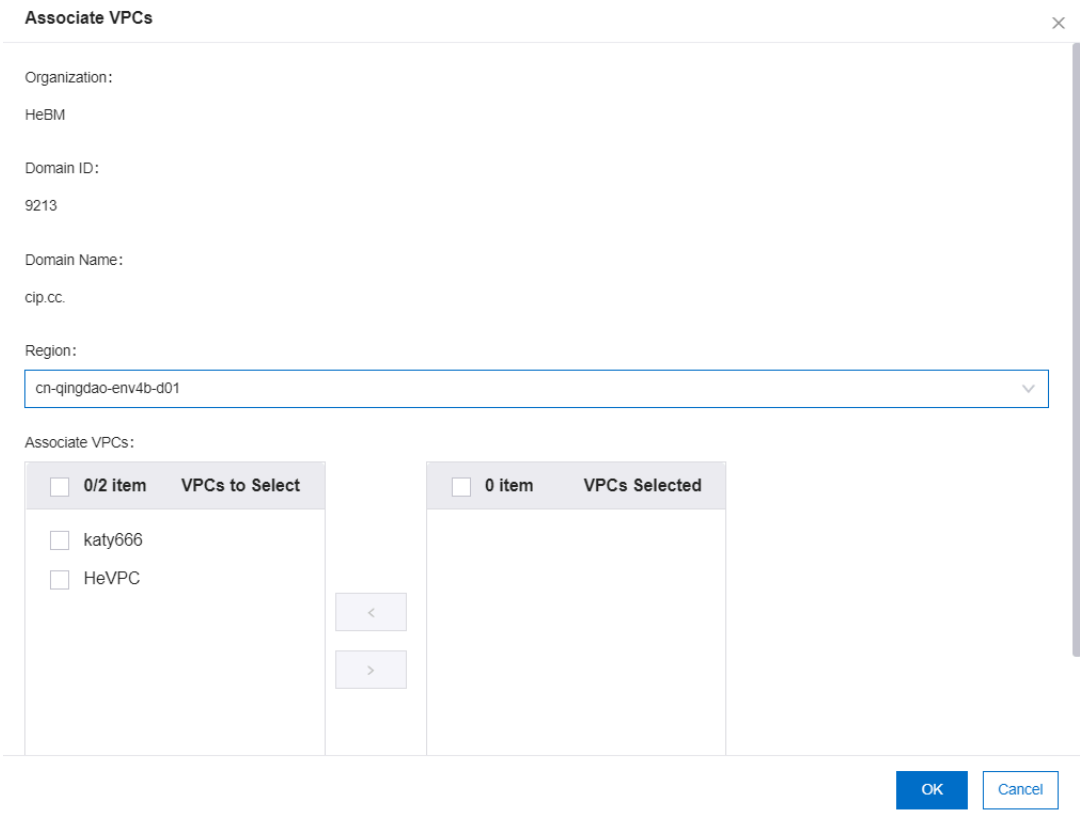
Prerequisites

You have the administrator privileges on the system and level-1 organizations.

Procedure

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

2. In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Default Forwarding.
3. Select the organization that you want to associate with a VPC, click  in the Actions column, and then select Associate VPCs from the shortcut menu.
4. Select the VPC to be associated from the VPCs to Select section, click the right arrow to move the selected VPC to the VPCs Selected section, and then click OK.



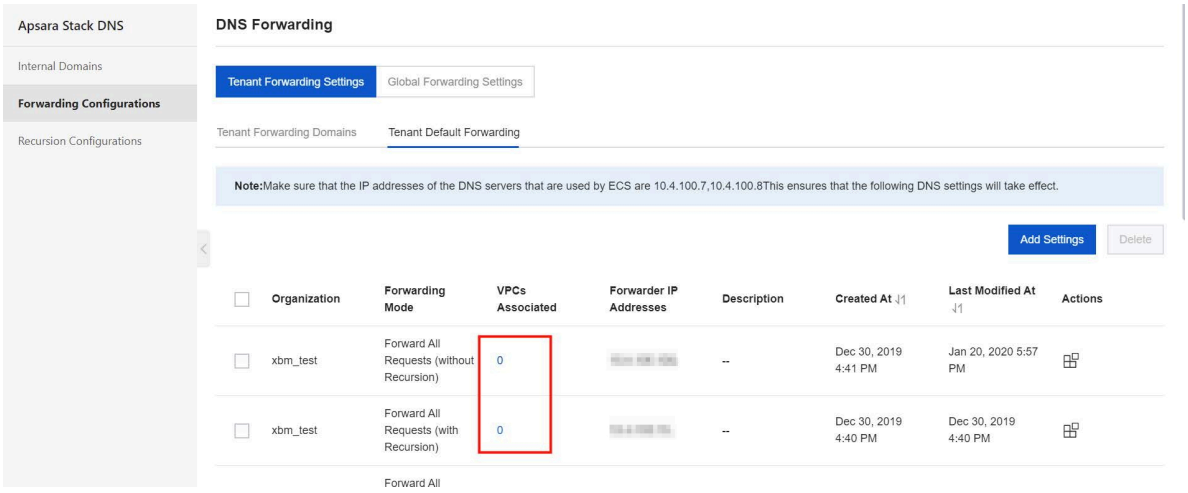
14.5.1.2.4 Disassociate a domain name from a VPC

This topic describes how to disassociate a domain name from a VPC.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Default Forwarding.

3. Select the organization that you want to disassociate from the VPC and click the value in the VPCs Associated column.



4. On the VPCs Associated page, select the VPC from which the organization is to be disassociated, click  in the Actions column, and select Disassociate from the shortcut menu.

   Make sure that the disassociated VPC is no longer displayed on the VPCs Associated page.

14.5.1.2.5 Modify default forwarding configurations

Prerequisites

You have the administrator privileges on the system and level-1 organizations.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Default Forwarding.
3. Select the organization for which you want to modify the forwarding configurations, click  in the Actions column, and then select Modify from the shortcut menu.
4. In the dialog box that appears, change the value of Forwarding Mode or Forwarder IP Addresses.
5. Click OK.

14.5.1.2.6 Add a description for a default forwarding configuration

Prerequisites

You have the administrator privileges on the system and level-1 organizations.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Default Forwarding.**
3. **Select the organization for which you want to add a forwarding configuration description, click**  **in the Actions column, and then select Description from the shortcut menu.**
4. **In the dialog box that appears, enter a description in the Description field.**
5. **Click OK.**

14.5.1.2.7 Delete default forwarding configurations

Prerequisites

You have the administrator privileges on the system and level-1 organizations.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Default Forwarding.**
3. **Select the organization for which you want to delete the forwarding configurations, click**  **in the Actions column, and then select Delete from the shortcut menu.**
4. **In the message that appears, click OK.**

14.5.1.2.8 Delete default forwarding configurations in batches

Prerequisites

You have the administrator privileges on the system and level-1 organizations.

Procedure

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

2. **In the left-side navigation pane, choose Forwarding Configurations > Tenant Forwarding Settings > Tenant Default Forwarding.**
3. **Select the organizations for which you want to delete the forwarding configurations and click Delete in the upper-right corner.**
4. **In the message that appears, click OK.**

14.5.2 Global forwarding configurations

14.5.2.1 Global forwarding domain names

14.5.2.1.1 Overview

All operations of this feature require administrator privileges.

Apsara Stack DNS forwards specific domain names to other DNS servers for resolution.

Two forwarding modes are available: forward all requests without recursion and forward all requests with recursion.

- **Forward all requests without recursion: Only a specified DNS server is used to resolve domain names. If the specified DNS server cannot resolve the domain names or the request times out, a message is returned to the DNS client to indicate that the query failed.**
- **Forward all requests with recursion: A specified DNS server is preferentially used to resolve domain names. If the specified DNS server cannot resolve the domain names, the local DNS is used instead.**

14.5.2.1.2 View global forwarding domain names

This topic describes how to view forwarding domain names in the ASCM console. This operation requires administrator privileges.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Global Forwarding Settings > Global Forwarding Domains.**
3. **In the Domain Name search bar, enter the domain name that you want to view and click Search.**

14.5.2.1.3 Add a domain name

This topic describes how to add a domain name in the ASCM console. This operation requires administrator privileges.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Global Forwarding Settings > Global Forwarding Domains.**
3. **Click Add Domain Name.**
4. **In the dialog box that appears, set** `Domain Name`, `Forwarding Mode`, **and** `Forwarder IP Addresses`. **Then, click OK.**

14.5.2.1.4 Add a description for a domain name

This topic describes how to add a description for a domain name in the ASCM console. This operation requires administrator privileges.

Context

You can add a description for a domain name for identification. For example, you can describe a domain name by using a hostname or internal system information.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Global Forwarding Settings > Global Forwarding Domains.**
3. **Select the domain name for which you want to add a description, click**  **in the Actions column, and then select Description from the shortcut menu.**
4. **In the dialog box that appears, enter a description and click OK.**

14.5.2.1.5 Modify the forwarding configurations of a domain name

This topic describes how to modify the forwarding configurations of a domain name in the ASCM console. This operation requires administrator privileges.

Procedure

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

2. In the left-side navigation pane, choose Forwarding Configurations > Global Forwarding Settings > Global Forwarding Domains.
3. Select the domain name for which you want to modify the forwarding configurations, click  in the Actions column, and then select Modify from the shortcut menu.
4. In the dialog box that appears, change the value of `Forwarding Mode` or `Forwarder IP Addresses`, and click OK.

14.5.2.1.6 Delete a domain name

This topic describes how to delete a domain name in the ASCM console. This operation requires administrator privileges.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Forwarding Configurations > Global Forwarding Settings > Global Forwarding Domains.
3. Select the domain name that you want to delete, click  in the Actions column, and then select Delete from the shortcut menu.
4. Click OK.

14.5.2.1.7 Delete domain names in batches

This topic describes how to delete domain names in batches in the ASCM console. This operation requires administrator privileges.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, choose Forwarding Configurations > Global Forwarding Settings > Global Forwarding Domains.
3. Select the domain names that you want to delete and click Delete in the upper-right corner.
4. Click OK.

14.5.2.2 Global default forwarding configurations

14.5.2.2.1 Enable default forwarding

This topic describes how to enable default forwarding in the ASCM console. This operation requires administrator privileges.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Global Forwarding Settings > Global Default Forwarding.**
3. **Click**  **in the Actions column and select Enable from the shortcut menu.**
4. **In the dialog box that appears, set** `Forwarding Mode` **and** `Forwarder IP Addresses`**, and click OK.**

   Make sure that `Enable Default Forwarding` **is set to** `ON`**.**

14.5.2.2.2 Modify default forwarding configurations

This topic describes how to modify default forwarding configurations in the ASCM console. This operation requires administrator privileges.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Global Forwarding Settings > Global Default Forwarding.**
3. **Click**  **in the Actions column and select Modify from the shortcut menu.**
4. **In the dialog box that appears, set** `Forwarding Mode` **and** `Forwarder IP Addresses`**, and click OK.**

14.5.2.2.3 Disable default forwarding

This topic describes how to disable default forwarding in the ASCM console. This operation requires administrator privileges.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. **In the left-side navigation pane, choose Forwarding Configurations > Global Forwarding Settings > Global Default Forwarding.**

3. Click  in the Actions column and select Disable from the shortcut menu.
4. In the message that appears, click OK.

14.6 Management of recursive resolution configurations

14.6.1 Enable global recursive resolution

Prerequisites

You have administrator privilegeS.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, click Recursion Configurations.
3. Click  in the Actions column and select Enable from the shortcut menu.
4. In the message that appears, click OK.

14.6.2 Disable global recursive resolution

Prerequisites

You have administrator privilegeS.

Procedure

1. *Log on to the Apsara Stack DNS console*.
2. In the left-side navigation pane, click Recursion Configurations.
3. Click  in the Actions column and select Disable from the shortcut menu.
4. In the message that appears, click OK.