

Alibaba Cloud Apsara Stack Enterprise

Product Introduction

Version: 1911, Internal: V3.10.0

Issue: 20210806









Legal disclaimer

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

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

- 5.** By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.
- 6.** Please contact Alibaba Cloud directly if you discover any errors in this document.

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 <code>cd /d C:/window</code> command to enter the Windows system folder.
Italic	Italic formatting is used for parameters and variables.	<code>bae log list --instanceid Instance_ID</code>
[] or [a b]	This format is used for an optional value, where only one item can be selected.	<code>ipconfig [-all -t]</code>

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

Legal disclaimer.....	I
Document conventions.....	I
1 Introduction to Alibaba Cloud Apsara Stack.....	1
1.1 What is Alibaba Cloud Apsara Stack?.....	1
1.2 The reasons to choose Apsara Stack.....	3
1.2.1 Hyper-scale distributed cloud operating system.....	3
1.2.2 Deployment and control system of Apsara Infrastructure Management Framework.....	4
1.2.3 High-reliability disaster recovery solutions.....	5
1.2.4 Unified operation management and automated operations capabilities.....	6
1.2.5 Open cloud service interface.....	7
1.3 Product architecture.....	7
1.3.1 Types of private cloud architectures.....	7
1.3.2 System architecture.....	8
1.3.3 Network architecture.....	10
1.3.3.1 Network architecture overview.....	10
1.3.3.2 Business service area.....	12
1.3.3.3 Integrated access area.....	14
1.3.3.4 VPC leased line access.....	16
1.3.4 Security architecture.....	17
1.3.5 Base assembly.....	18
1.4 Product panorama.....	20
1.5 Scenarios.....	21
1.6 Compliance security solution.....	23
1.6.1 Overview.....	23
1.6.2 Interpretation on key points.....	23
1.6.3 Cloud-based classified protection compliance.....	26
1.6.4 Classified protection implementation process.....	29
1.6.5 Security compliance architecture.....	30
1.6.6 Solution benefits.....	31
2 Elastic Compute Service (ECS).....	32
2.1 What is ECS?.....	32
2.2 Benefits.....	33
2.3 Architecture.....	36
2.4 Features.....	38
2.4.1 Instances.....	38
2.4.1.1 Overview.....	38
2.4.1.2 Instance families.....	38
2.4.1.3 Instance type.....	55

2.4.1.4 UserData.....	70
2.4.1.5 Instance lifecycle.....	71
2.4.1.6 EBM Instances.....	73
2.4.1.7 Super Computing Clusters.....	75
2.4.2 Block storage.....	77
2.4.2.1 Overview.....	77
2.4.2.2 Elastic block storage.....	78
2.4.2.2.1 Overview.....	78
2.4.2.2.2 Cloud disks.....	79
2.4.2.2.3 Shared block storage.....	79
2.4.2.2.4 Triplicate storage.....	80
2.4.2.3 ECS disk encryption.....	82
2.4.2.4 Local storage.....	82
2.4.3 Images.....	83
2.4.4 Snapshots.....	85
2.4.4.1 Overview.....	85
2.4.4.2 Mechanisms.....	85
2.4.4.3 Specifications of ECS Snapshot 2.0.....	86
2.4.4.4 Technical comparison.....	87
2.4.5 Deployment sets.....	88
2.4.6 Network and security.....	90
2.4.6.1 IP addresses of ECS instances of VPC type.....	90
2.4.6.2 Elastic network interfaces.....	91
2.4.6.3 Intranet.....	93
2.4.6.4 Security group rules.....	94
2.5 Scenarios.....	94
2.6 Limits.....	95
2.7 Terms.....	97
3 Auto Scaling (ESS).....	100
3.1 What is ESS?.....	100
3.2 Benefits.....	101
3.3 Architecture.....	103
3.4 Features.....	104
3.5 Usage scenarios.....	106
3.6 Limits.....	106
3.7 Terms.....	106
4 Object Storage Service (OSS).....	108
4.1 What is OSS?.....	108
4.2 Advantages.....	108
4.3 OSS architecture.....	109
4.4 Features.....	111
4.5 Scenarios.....	112
4.6 Limits.....	113
4.7 Terms.....	113

5 Table Store.....	116
5.1 What is Table Store?.....	116
5.2 Benefits.....	117
5.3 Architecture.....	118
5.4 Scenarios.....	119
5.5 Limits.....	125
5.6 Terms.....	127
5.7 Features.....	129
5.7.1 Features.....	129
5.7.2 Tunnel Service.....	130
5.7.3 Global secondary index.....	131
5.7.3.1 Feature description.....	131
5.7.3.2 Overview.....	133
5.7.3.3 Scenarios.....	135
6 ApsaraDB for RDS.....	144
6.1 What is ApsaraDB for RDS?.....	144
6.2 Benefits.....	144
6.2.1 Ease of use.....	145
6.2.2 High performance.....	145
6.2.3 High security.....	146
6.2.4 High reliability.....	147
6.3 Architecture.....	147
6.4 Features.....	148
6.4.1 Data link service.....	148
6.4.2 High-availability service.....	149
6.4.3 Backup and recovery service.....	151
6.4.4 Monitoring service.....	152
6.4.5 Scheduling service.....	152
6.4.6 Migration service.....	153
6.5 Scenarios.....	154
6.5.1 Diversified data storage.....	154
6.5.2 Read/write splitting.....	155
6.5.3 Big data analysis.....	157
6.6 Limits.....	157
6.6.1 Limits on ApsaraDB RDS for MySQL.....	157
6.7 Terms.....	159
6.8 Instance types.....	161
7 AnalyticDB for PostgreSQL.....	173
7.1 What is AnalyticDB for PostgreSQL?.....	173
7.2 Benefits.....	173
7.3 Architecture.....	175
7.4 Specification description.....	176
7.5 Features.....	178
7.5.1 Distributed architecture.....	178

7.5.2 High-performance data analysis.....	178
7.5.3 High-availability service.....	178
7.5.4 Data synchronization and tools.....	179
7.5.5 Data security.....	179
7.5.6 Supported SQL features.....	179
7.6 Scenarios.....	182
8 KVStore for Redis.....	185
8.1 What is KVStore for Redis?.....	185
8.2 Benefits.....	185
8.3 Architecture.....	187
8.4 Features.....	188
8.5 Scenarios.....	189
8.6 Limits.....	191
8.7 Terms.....	192
8.8 Instance types.....	193
9 Data Transmission Service (DTS).....	200
9.1 What is DTS?.....	200
9.2 Benefits.....	200
9.3 Environment requirements.....	201
9.4 Architecture.....	203
9.5 Features.....	207
9.5.1 Data migration.....	207
9.5.2 Data synchronization.....	211
9.5.3 Data subscription.....	215
9.6 Scenarios.....	217
9.7 Concepts.....	224
10 Data Management (DMS).....	227
10.1 What is DMS?.....	227
10.2 Advantages.....	227
10.3 Architecture.....	227
10.4 Features.....	229
10.5 Scenarios.....	230
10.5.1 Convenient data operations.....	230
10.5.2 Prohibiting data export.....	231
10.5.3 SQL statement reuse.....	231
10.6 Limits.....	232
11 Server Load Balancer (SLB).....	233
11.1 What is Server Load Balancer?.....	233
11.2 High availability.....	235
11.3 Architecture.....	238
11.4 Features.....	241
11.5 Scenarios.....	244
11.6 Limits.....	247
11.7 Terms.....	248

12 Virtual Private Cloud (VPC).....	250
12.1 What is a VPC?.....	250
12.2 Benefits.....	251
12.3 Architecture.....	252
12.4 Features.....	254
12.5 Scenarios.....	255
12.6 Terms.....	257
12.7 Limits.....	258
13 Apsara Stack Security.....	259
13.1 What is Apsara Stack Security?.....	259
13.2 Benefits.....	260
13.3 Architecture.....	263
13.4 Features.....	264
13.4.1 Apsara Stack Security Standard Edition.....	264
13.4.2 On-premises security operations services.....	267
13.4.3 Optional security services.....	268
13.5 Restrictions.....	269
13.6 Terms.....	269
14 Apsara Stack DNS.....	271
14.1 What is Apsara Stack DNS?.....	271
14.2 Benefits.....	271
14.3 Architecture.....	273
14.4 Features.....	273
14.5 Scenarios.....	274
14.6 Limits.....	275
14.7 Terms.....	276
15 MaxCompute.....	277
15.1 What is MaxCompute?.....	277
15.2 Integration with other Alibaba Cloud services.....	278
15.3 Benefits.....	281
15.4 Architecture.....	283
15.5 Features.....	285
15.5.1 Tunnel.....	285
15.5.1.1 Terms.....	285
15.5.1.2 Tunnel features.....	285
15.5.1.3 Data upload and download through Tunnel.....	285
15.5.2 SQL.....	286
15.5.2.1 Terms.....	286
15.5.2.2 SQL characteristics.....	287
15.5.2.3 Comparison with open source products.....	287
15.5.3 MapReduce.....	288
15.5.3.1 Terms.....	288
15.5.3.2 MapReduce characteristics.....	288
15.5.3.3 MaxCompute MapReduce process.....	289

15.5.3.4 Hadoop MapReduce VS MaxCompute MapReduce.....	289
15.5.4 Graph.....	290
15.5.4.1 Terms.....	290
15.5.4.2 Graph characteristics.....	290
15.5.4.3 Graph relational network models.....	291
15.5.5 Unstructured data processing (integrated computing scenarios).....	292
15.5.6 Unstructured data processing in MaxCompute.....	292
15.5.7 Enhanced features.....	292
15.5.7.1 Spark on MaxCompute.....	292
15.5.7.1.1 Terms.....	292
15.5.7.1.2 Features of Spark on MaxCompute.....	293
15.5.7.1.3 Spark features.....	293
15.5.7.1.4 Spark architecture.....	294
15.5.7.1.5 Benefits of Spark on MaxCompute.....	295
15.5.7.2 Elasticsearch on MaxCompute.....	296
15.5.7.2.1 Overview.....	296
15.5.7.2.2 Features of Elasticsearch on MaxCompute.....	297
15.5.7.2.3 Elasticsearch features.....	298
15.5.7.2.4 Elasticsearch architecture.....	299
15.5.7.2.5 Elasticsearch benefits.....	300
15.6 Scenarios.....	300
15.6.1 Scenario 1: Migrate data to the cloud cost-effectively and quickly.....	300
15.6.2 Scenario 2: Improve development efficiency and reduce storage and computing costs.....	301
15.6.3 Scenario 3: Use mass data to achieve precision marketing for millions of users.....	302
15.6.4 Scenario 4: Achieve precision marketing with big data.....	302
15.7 Limits.....	303
15.8 Terms.....	303
16 DataWorks.....	308
16.1 What is DataWorks?.....	308
16.2 Benefits.....	308
16.2.1 Powerful computing capabilities.....	308
16.2.2 End-to-end platform.....	309
16.2.3 Data integration from heterogeneous data sources.....	309
16.2.4 Web-based software.....	309
16.2.5 Multi-tenant architecture.....	309
16.2.6 Intelligent monitoring and alerts.....	309
16.2.7 Easy-to-use SQL editor.....	309
16.2.8 Comprehensive data quality monitoring.....	309
16.2.9 Convenient API development and management.....	310
16.2.10 Secure data sharing.....	310
16.3 Architecture.....	310
16.3.1 Services.....	310
16.3.2 System architecture.....	311

16.3.3 Security architecture.....	311
16.3.4 Multi-tenant architecture.....	311
16.4 Services.....	311
16.4.1 Data Integration.....	311
16.4.1.1 Overview.....	311
16.4.1.2 Various data sources.....	312
16.4.1.2.1 Metadata synchronization.....	312
16.4.1.2.2 Relational database.....	312
16.4.1.2.3 NoSQL database.....	312
16.4.1.2.4 MPP database.....	312
16.4.1.2.5 Big data product.....	312
16.4.1.2.6 Unstructured data storage.....	312
16.4.1.3 Inbound data control.....	312
16.4.1.4 High transmission rate.....	313
16.4.1.5 Accurate flow control.....	313
16.4.1.6 Synchronization agents.....	313
16.4.1.7 Transmission in complicated networks.....	313
16.4.2 DataStudio.....	313
16.4.2.1 Overview.....	313
16.4.2.2 Business flows.....	314
16.4.2.2.1 Description.....	314
16.4.2.2.2 Nodes.....	314
16.4.2.2.3 Configure nodes.....	314
16.4.2.2.4 Versions.....	314
16.4.2.2.5 Publish business flows.....	314
16.4.2.3 Solutions.....	314
16.4.2.4 Code editor.....	315
16.4.2.4.1 ODPS SQL nodes.....	315
16.4.2.4.2 ODPS MR nodes.....	315
16.4.2.4.3 Resources.....	315
16.4.2.4.4 Register user-defined functions.....	315
16.4.2.4.5 Shell nodes.....	315
16.4.2.5 Code repository and team collaboration.....	316
16.4.3 Administration.....	316
16.4.3.1 Overview.....	316
16.4.3.2 Overview.....	316
16.4.3.3 Task instances.....	316
16.4.3.4 Monitor.....	317
16.4.4 Workspace Management.....	317
16.4.4.1 Description.....	317
16.4.4.2 Organization.....	317
16.4.4.3 Workspaces.....	317
16.4.4.4 Members.....	318
16.4.4.5 Authorizations.....	318
16.4.5 Realtime Analysis.....	319

16.4.6 Data Service.....	320
16.4.7 Data Asset Management.....	320
16.4.8 Data Protection.....	321
16.4.8.1 Overview.....	321
16.4.8.2 Terms.....	321
16.4.8.2.1 Organization.....	321
16.4.8.2.2 Workspace.....	321
16.4.8.2.3 Regular expression.....	322
16.4.8.2.4 Data classification.....	322
16.4.8.2.5 Data recognition.....	322
16.4.8.2.6 Data masking.....	322
16.4.8.2.7 MaxCompute.....	322
16.4.8.3 Data recognition rules.....	322
16.4.8.4 Data recognition.....	322
16.4.8.5 Data activities.....	323
16.4.8.6 Data masking.....	323
16.4.8.7 Levels.....	323
16.4.8.8 Manual check.....	324
16.4.8.9 Data risks.....	324
16.4.8.10 Risk rules.....	324
16.4.8.11 Data auditing.....	324
16.5 Scenarios.....	324
16.5.1 Cloud-based data warehouse.....	324
16.5.2 Business intelligence.....	325
16.5.3 Data-driven management.....	325
16.6 Limits.....	325
16.7 Terms.....	325
17 Realtime Compute.....	327
17.1 What is Realtime Compute?.....	327
17.2 End-to-end real-time computing.....	328
17.3 Differences between real-time computing and batch computing.....	330
17.3.1 Overview.....	330
17.3.2 Batch computing.....	330
17.3.3 Real-time computing.....	331
17.3.4 Comparison between real-time computing and batch computing.....	333
17.4 Benefits.....	333
17.5 Product architecture.....	335
17.5.1 Business process.....	335
17.5.2 Business architecture.....	339
17.5.3 Technical architecture.....	340
17.6 Features.....	341
17.7 Product positioning.....	344
17.8 Scenarios.....	345
17.8.1 Overview.....	345
17.8.2 Management of e-commerce activities.....	346

17.8.3 Multidimensional analysis of data from IoT sensors.....	347
17.8.4 Big screen service for the Tmall Double 11 Shopping Festival.....	352
17.8.5 Mobile data analysis.....	353
17.9 Restrictions.....	354
17.10 Terms.....	354
18 Machine Learning Platform for AI.....	357
18.1 What is machine learning?.....	357
18.2 Benefits.....	358
18.3 Architecture.....	360
18.4 Features.....	362
18.4.1 Visualized modeling.....	362
18.4.2 All-in-one experience.....	364
18.4.3 Multiple templates on the homepage.....	365
18.4.4 Data visualization.....	366
18.4.5 Model management.....	366
18.4.6 Algorithm components.....	367
18.5 Scenarios.....	378
18.6 Limits.....	379
18.7 Terms.....	380
19 Quick BI.....	382
19.1 What is Quick BI?.....	382
19.2 Benefits.....	382
19.3 Architecture.....	383
19.4 Features.....	384
19.5 Scenarios.....	385
19.5.1 Instant data analysis and effective decision-making.....	385
19.5.2 Integration with existing systems.....	386
19.5.3 Permission control of transaction data.....	387
19.6 Limits.....	388
19.7 Terms.....	389
20 Apsara Big Data Manager (ABM).....	391
20.1 What is Apsara Big Data Manager?.....	391
20.2 Benefits.....	391
20.3 Architecture.....	392
20.3.1 System architecture.....	392
20.4 Features.....	394
20.4.1 Dashboard.....	394
20.4.2 Repository.....	397
20.4.3 Business.....	399
20.4.4 O&M.....	400
20.4.5 Management.....	407
20.5 Scenarios.....	408
20.6 Limits.....	408
20.7 Concepts.....	408

1 Introduction to Alibaba Cloud Apsara Stack

1.1 What is Alibaba Cloud Apsara Stack?

Based on Alibaba Cloud's distributed architecture and tailored to the enterprise-level market features, Alibaba Cloud Apsara Stack is an open, unified, and trusted enterprise-level full-stack cloud platform. Apsara Stack and Alibaba Cloud public cloud share the same roots. You can locally deploy public cloud products and services in any environment and expand to the public cloud with one click, which allows you to enjoy hybrid cloud services anytime and anywhere.

Private cloud

Private cloud is a locally deployed cloud computing system for governments or enterprises built by cloud computing service providers. It places cloud infrastructures, software, and hardware resources within firewalls to allow departments within an organization or an enterprise to share resources in their data centers. It can be managed by an organization or a third party and located within the organization or outside the organization. Compared with public cloud, private cloud provides better privacy and exclusivity.

Private cloud is divided into two types by the enterprise scale or business requirements:

- Multi-tenant comprehensive private cloud for industries and large groups: A full stack cloud system created in a top-down manner, which takes the hyper-scale digital applications as a business driver and satisfies IT requirements, such as the continuous integration and development of DevOps applications and operation support of production environments.
- Single-tenant basic private cloud for small- and medium-sized enterprises and scenarios: A cloud system that hosts technical systems including large-scale Software as a Service (SaaS) applications, industrial clouds, and large group clouds. It also performs local computing tasks.

Alibaba Cloud Apsara Stack

During the evolution from IT architecture to clouds, more and more enterprises want to have the service experience that is brought by large-scale cloud computing in their own data centers, which is based on the construction requirements, such as security compliance, reuse of existing data centers, and experience localization.

Alibaba Cloud Apsara Stack is an extension of Alibaba Cloud public cloud, which brings the technologies of public cloud to Apsara Stack. By helping enterprises deliver complete and customizable Alibaba Cloud software solutions in their own data centers, Apsara Stack allows you to experience the same characteristics as the hyper-scale cloud computing and big data products provided by Alibaba Cloud public cloud in the local environment. Apsara Stack also provides enterprises with the consistent hybrid cloud experience where you can obtain IT resources as required and guarantee the business continuity.

Apsara Stack provides an on-premises deployment and can continue to operate, be managed, and provide services disconnected from Alibaba Cloud for an indefinite period of time.

Service values

Supported by various products and services, based on successful digital practice cases of Alibaba Group, and integrated with the mature solutions and rich experience in various industries, Apsara Stack helps governments and enterprises digitally transform their businesses and services. Apsara Stack provides service values in the following four aspects:

- Elastic

Combines all resources into a supercomputer and flexibly scales out resources to minimize costs and maximize performance and stability.

- Agile

Integrates business by using Internet and microservices to speed up the innovation of traditional enterprises.

- Data

Uses digitalization to allow data to flow between vertical businesses and forms a data shared service to deal with large amounts of data.

- Smart

Allows smart transformation of businesses globally and helps reinvent business models.

Platform features

As an enterprise-level cloud platform, Apsara Stack has the following three features:

- Software-defined platform: masks underlying hardware differences, enables resources to scale up or out as required, and does not affect the performance of upper-layer applications.

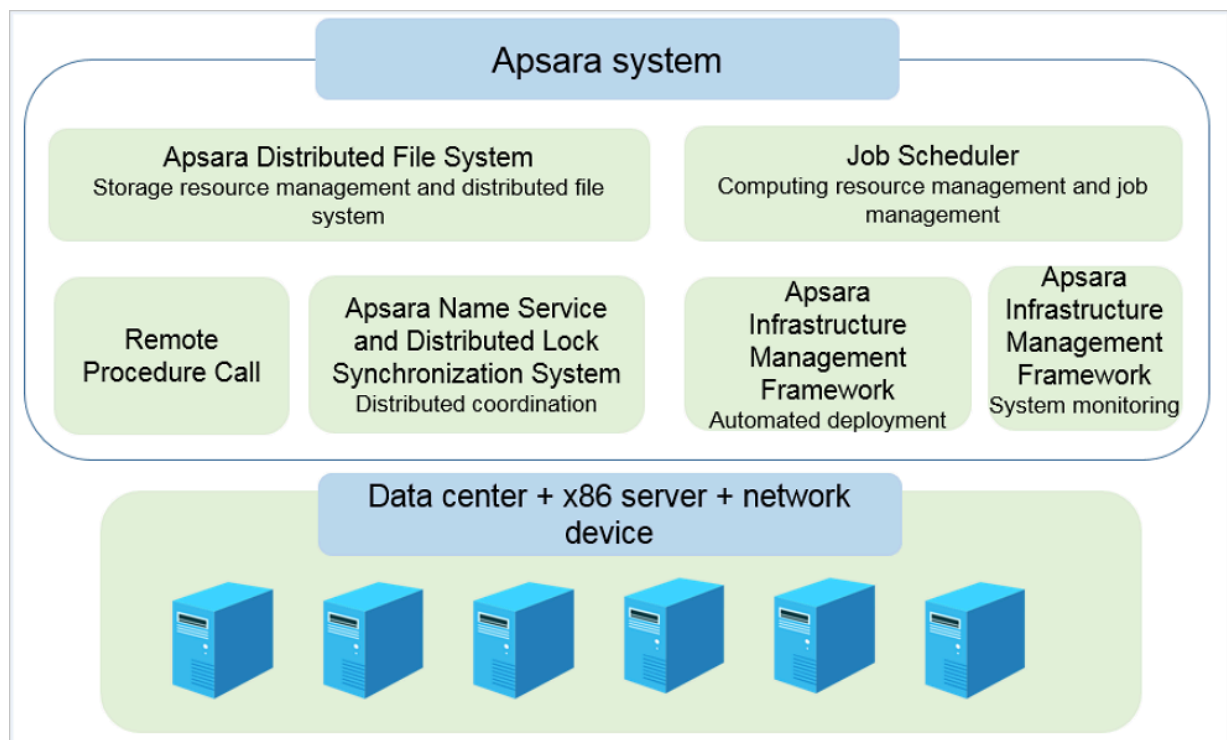
- Production-level reliability and security compliance: guarantees the continuity and security of enterprise data.
- Unified access management: isolates permissions of different roles for easy subsequent operations management.

1.2 The reasons to choose Apsara Stack

1.2.1 Hyper-scale distributed cloud operating system

Apsara Stack is based on the same underlying architecture (large-scale distributed computing system kernel of Apsara) as Alibaba Cloud public cloud. It provides underlying support for upper-layer services in terms of storage, computing, and scheduling. It is a hyper-scale and universal computing operating system that is independently developed by Alibaba Cloud for the global market. Apsara can connect millions of servers all over the world into a supercomputer and provide the community with computing capabilities in the form of online public services. The computing capabilities provided by Apsara are powerful, universal, and beneficial to everyone.

Figure 1-1: Apsara system kernel architecture



The modules of the Apsara platform kernel have the following primary functions:

- Underlying services for distributed systems

The modules provide the underlying services required in a distributed environment, such as coordination, remote procedure call, security management, and resource management. These services provide support for the upper-layer modules, such as the distributed file system and job scheduling.

- Distributed file system

The modules aggregate storage capabilities from different nodes in a cluster to construct a massive, reliable, and scalable data storage service. The modules also protect against software and hardware faults automatically to guarantee uninterrupted data access. With the support for incremental expansion and automatic data balancing, the modules provide APIs that are similar to Portable Operating System Interface of UNIX (POSIX) for accessing the files in the user space. The modules also perform random read/write and append write operations.

- Job scheduling

The modules schedule jobs in cluster systems and support online services that rely heavily on the response speed and offline jobs that require high data processing throughput. The modules detect faults and hot spots in systems automatically and guarantee a stable and reliable job completion in various methods, such as error retries and issuing concurrent backup jobs for long-tail jobs.

- Cluster monitoring and deployment

The modules monitor the running status and performance metrics of upper-layer application services and the cluster status to send alert notifications of exception events and keep a record. The modules enable the operations personnel to manage the deployment and configuration of Apsara platform and upper-layer applications. The modules also support online cluster scaling and online update of application services.

1.2.2 Deployment and control system of Apsara Infrastructure Management Framework

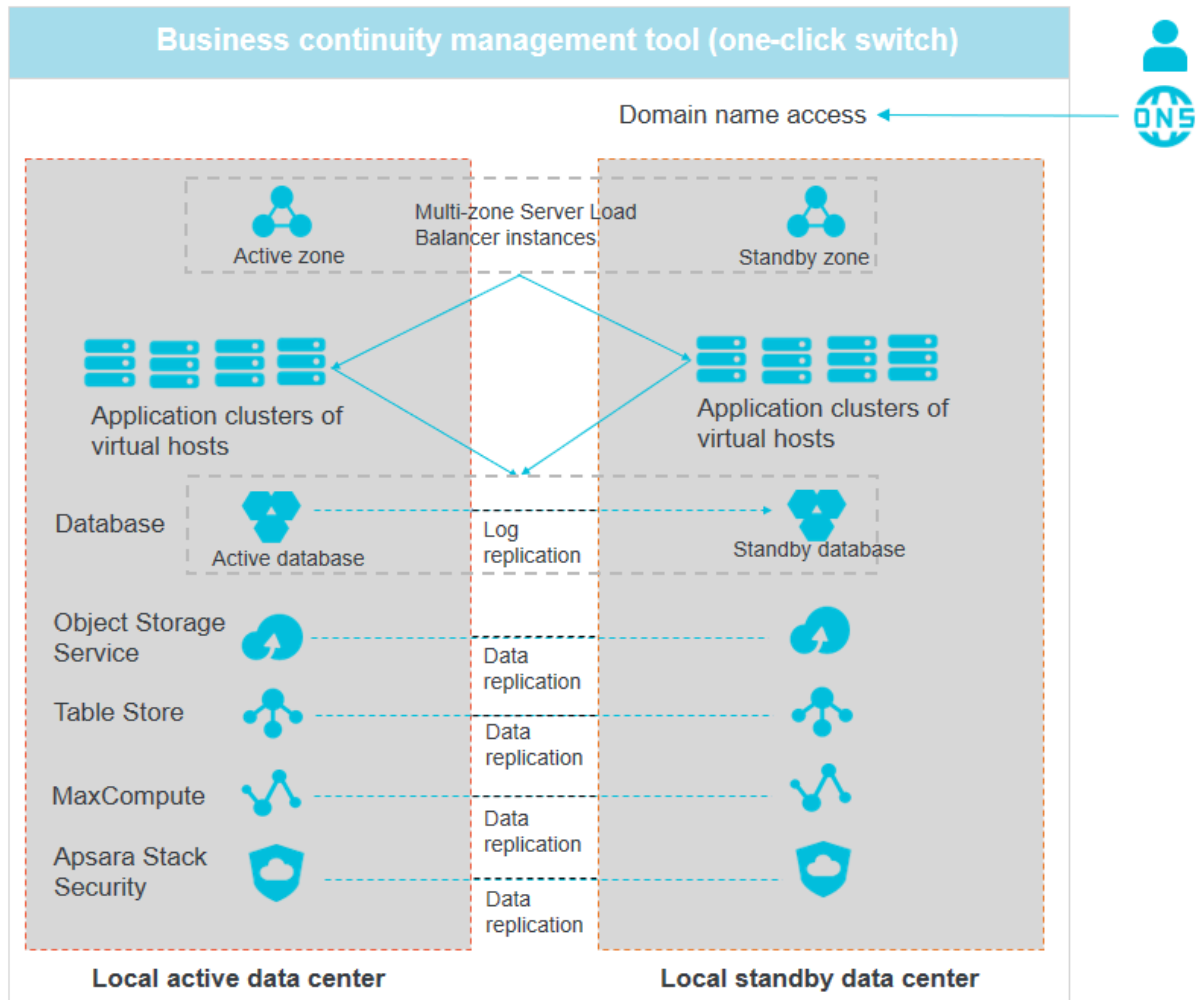
Apsara Infrastructure Management Framework provides the cloud services with basic support by supplying unified deployment, authentication, authorization, and control capabilities for cloud service products.

Apsara Infrastructure Management Framework contains various modules, including deployment framework, resource library, meta database, authentication and authorization component, interface gateway, Log Service, and control service module.

- The deployment framework provides all cloud services with unified functions that can deploy access platform and manage the dependencies among services.
- The resource library stores the execution files of all cloud services and their dependent components.
- The authentication and authorization component provides access control for cloud services and supports isolation of multiple tenants.
- The interface gateway provides a unified API management console for all cloud services.
- Log Service stores, retrieves, and obtains logs of cloud services.
- The control service module monitors the basic health status of cloud services and supports the operations system of the cloud platform.

1.2.3 High-reliability disaster recovery solutions

Apsara Stack disaster recovery solutions are designed and developed based on the cloud computing capabilities of Alibaba Cloud. The solutions comply with common international disaster recovery standards. When the network conditions meet the design requirements, the Apsara Stack platform implements the active-active mode of the network access layer and user application layer, and the active-standby mode of the data persistence layer.

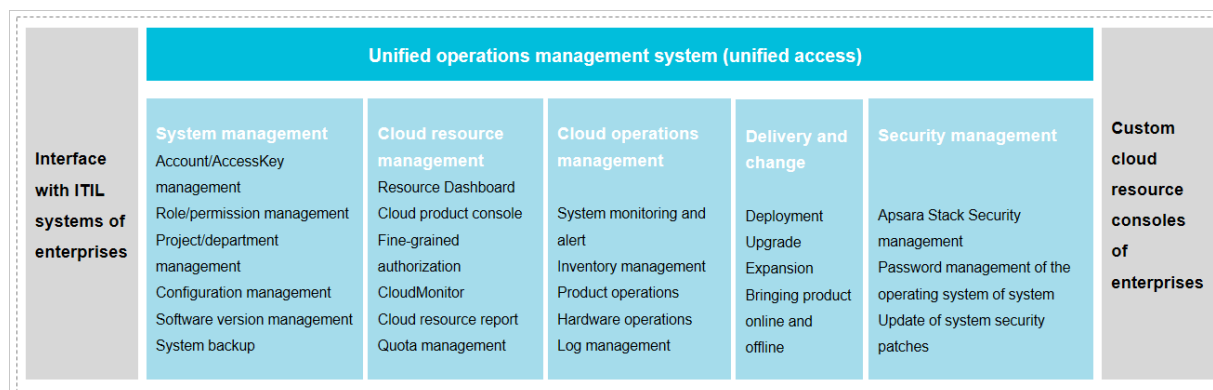


1.2.4 Unified operation management and automated operations capabilities

Apsara Stack provides a unified entry for management system to configure different management permissions for different user roles. You can gain operations management capabilities by using open interfaces and customize your own cloud resource consoles . Apsara Stack also provides the capability to interface with the Information Technology

Infrastructure Library (ITIL) systems of enterprises, which aims to interface and integrate synchronously with the existing IT systems of various enterprises.

Figure 1-2: Unified operations management



1.2.5 Open cloud service interface

Cloud services provide a wide variety of SDKs and RESTful APIs on an OpenAPI platform. You can use these APIs to flexibly access various cloud services provided by Apsara Stack. You can also obtain basic control information about the cloud platform by using these APIs and connect the Apsara Stack platform to your unified control system.

1.3 Product architecture

1.3.1 Types of private cloud architectures

Private cloud architectures have two types: native cloud architecture and integrated cloud architecture.

- Native cloud architecture

The native cloud architecture evolves from the open architecture of Internet and is based on the distributed system framework. It is initially designed to handle big data and host Web applications, and subsequently expands to run basic services.

- Integrated cloud architecture

The integrated cloud architecture focuses on virtualization of computing services. As a breakthrough from the traditional architecture, it is open-sourced by the OpenStack and becomes the mainstream private cloud architecture.

Apsara Stack adopts the native cloud architecture and is based on self-developed distributed technologies and products of Alibaba Cloud. The single system supports all

cloud products and services, and enables complete openness of the cloud platform. It comes with comprehensive service features for enterprises, a complete backup capability, and full autonomous control capability.

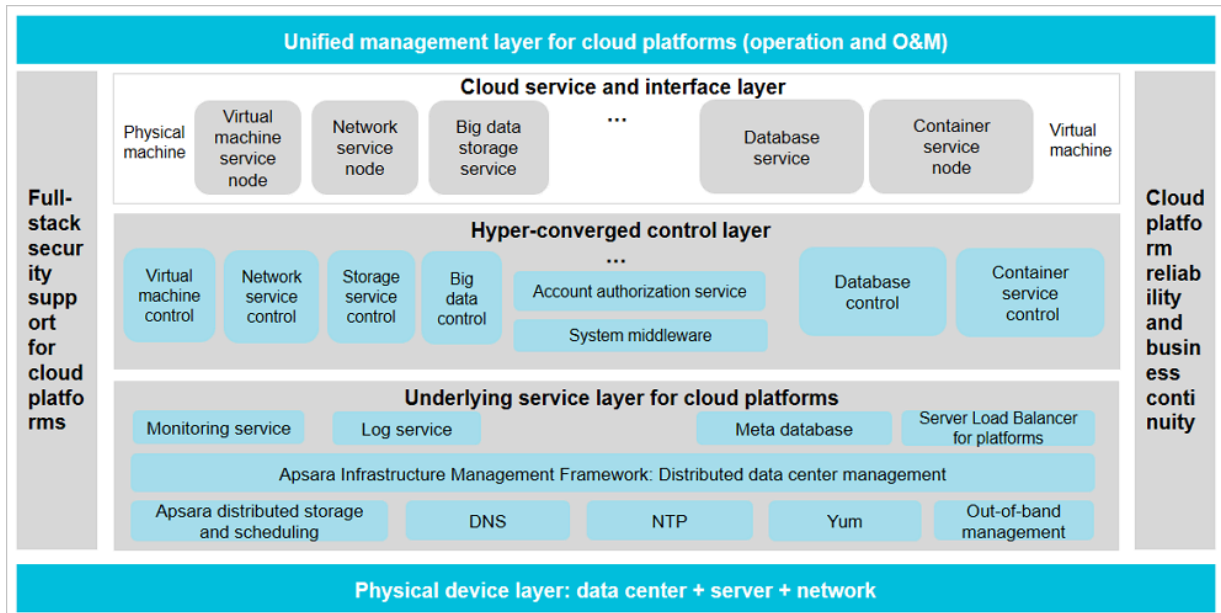
1.3.2 System architecture

The system architecture of Apsara Stack consists of the following parts, as shown in [Figure 1-3: The system architecture of Apsara Stack](#):

- Physical device layer: includes hardware devices for cloud computing, such as physical data centers, servers, and network.
- Underlying service layer for cloud platforms: bases on the underlying physical environment to provide underlying services for upper-layer applications.
- Hyper-converged control layer: provides unified scheduling for upper-layer applications or services by using the hyper-converged control architecture.
- Cloud service and interface layer: provides unified management and Operation and Maintenance (O&M) for virtual machines and physical machines by using converged service nodes management, and uses the API platform to unify the interfaces and support customized development.
- Unified management layer for cloud platforms: provides unified operation and O&M management.

Apsara Stack also provides full-stack security support and guarantees the reliability of cloud platforms and business continuity.

Figure 1-3: The system architecture of Apsara Stack



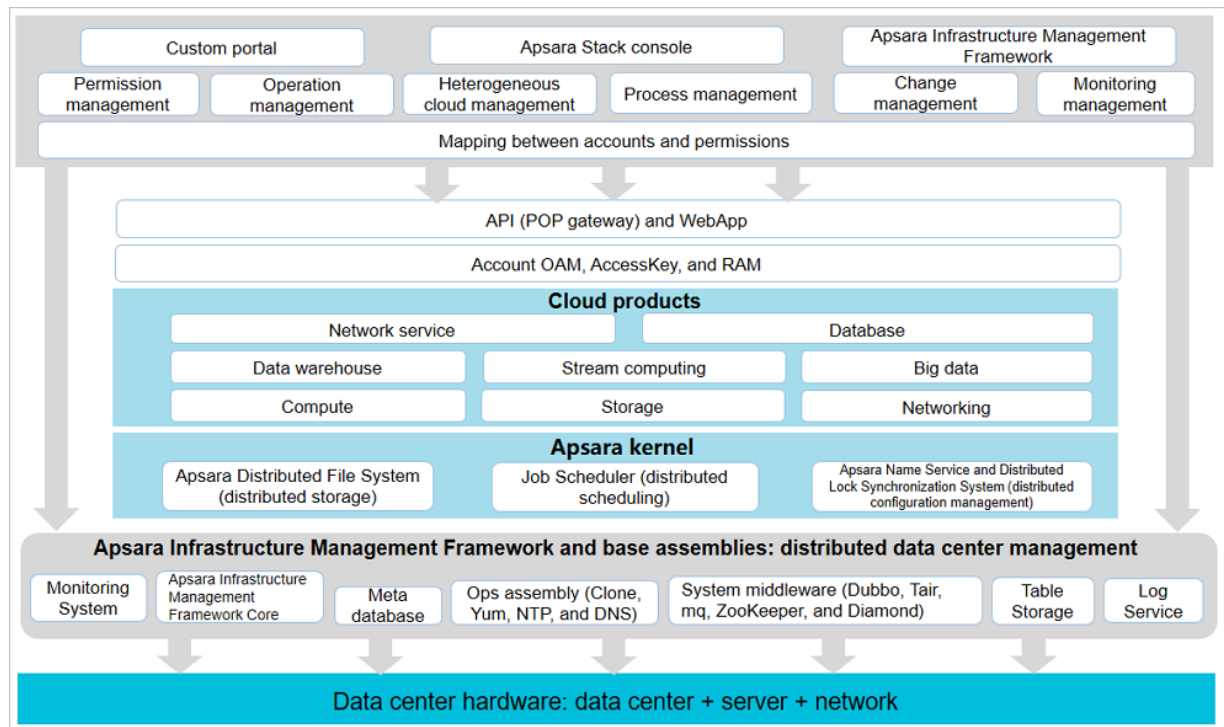
Logical architecture

Apsara Stack virtualizes the computing and storage capabilities of physical servers and network devices to achieve virtual computing, distributed storage, and software-defined networks. On this basis, Apsara Stack provides ApsaraDB, big data processing, and distributed middleware services. Apsara Stack also provides the supporting capabilities of underlying IT services for your applications, and can be interconnected with your existing account systems and monitoring operations systems. The logical architecture of Apsara Stack has the following characteristics:

- With data center + x86 server + network device as the hardware basis
- Based on the Apsara kernel (distributed engine) to provide various cloud products
- All cloud products are required to follow a unified API framework, O&M (accounts, authorization, monitoring, and logs) and management system, and security system.

- Make sure that all cloud products provide a consistent user experience.

Figure 1-4: The logical architecture of Apsara Stack



1.3.3 Network architecture

1.3.3.1 Network architecture overview

The network architecture of Apsara Stack defines two logical areas, namely the business service area and the integrated access area, as shown in [Figure 1-5: Logical areas](#).

- **Business service area**

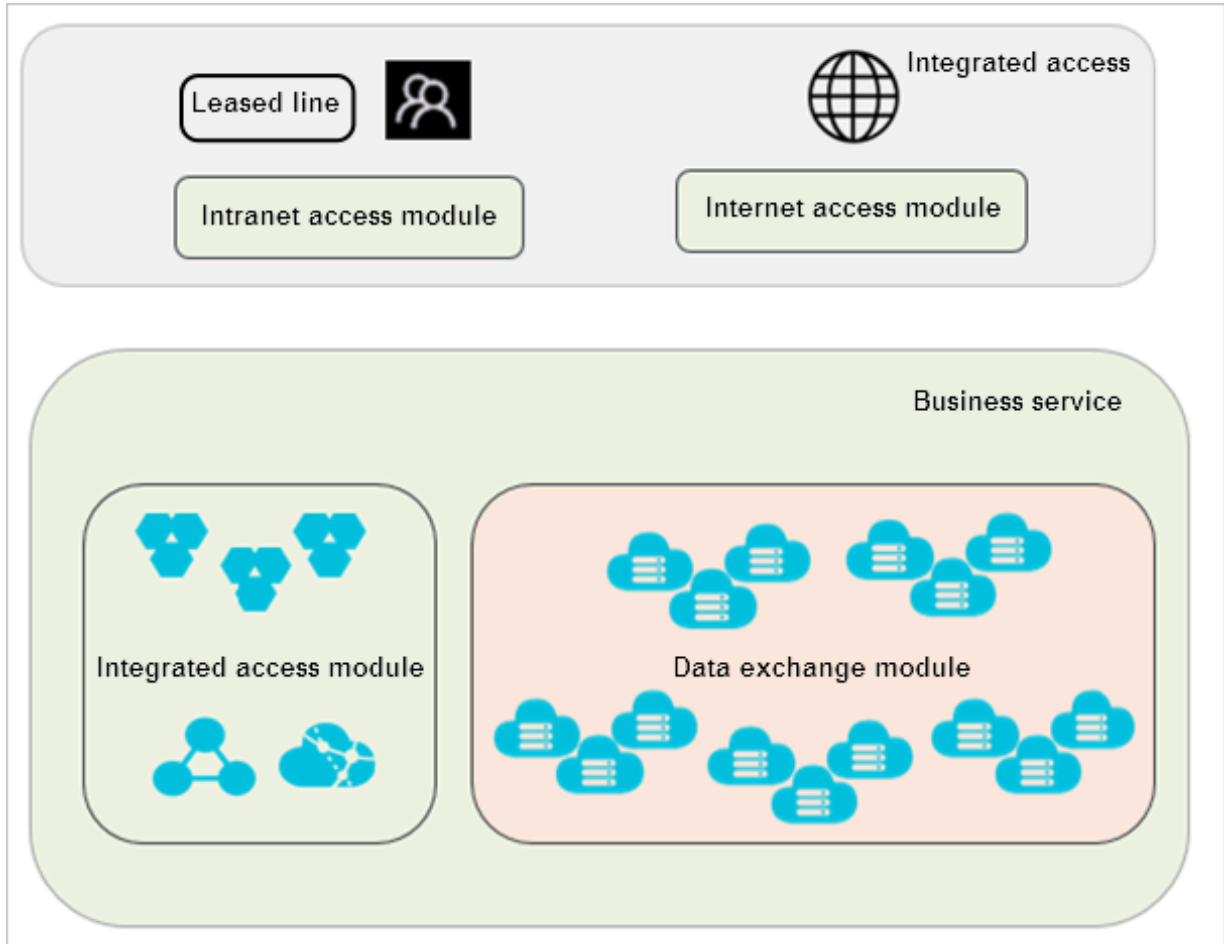
This area provides the networks of all cloud services and all cloud service systems exchange internal traffic in this area. This is the core area of Apsara Stack networks.

- **Integrated access area**

As an extension of the business service area, the integrated access area provides a channel for user management, and the access to Apsara Stack networks by using

Internet and user private networks. This area can be tailored based on the actual deployment requirements.

Figure 1-5: Logical areas



The roles and purposes of the switches in each area are as follows:

Role	Module	Function
ISW (interconnected switch)	Internet access module	ISW is an egress switch and provides access to Internet service providers (ISPs) or users' backbone networks.

Role	Module	Function
CSW (intranet access switch)	Intranet access module	CSW facilitates the access to users' internal backbone networks, including the access to Virtual Private Cloud (VPC) by using leased lines. It performs route distribution and interaction between the inside and outside of cloud networks.
DSW (distributed layer switch)	Data exchange module	DSW functions as a core switch to connect each access switch (ASW).
ASW (access switch)	Data exchange module	ASW provides access to Elastic Compute Service (ECS) and is uplinked with the core switch DSW.
LSW (integrated access switch)	Integrated access module	LSW provides access to cloud product services, such as VPC and Server Load Balancer (SLB).

1.3.3.2 Business service area

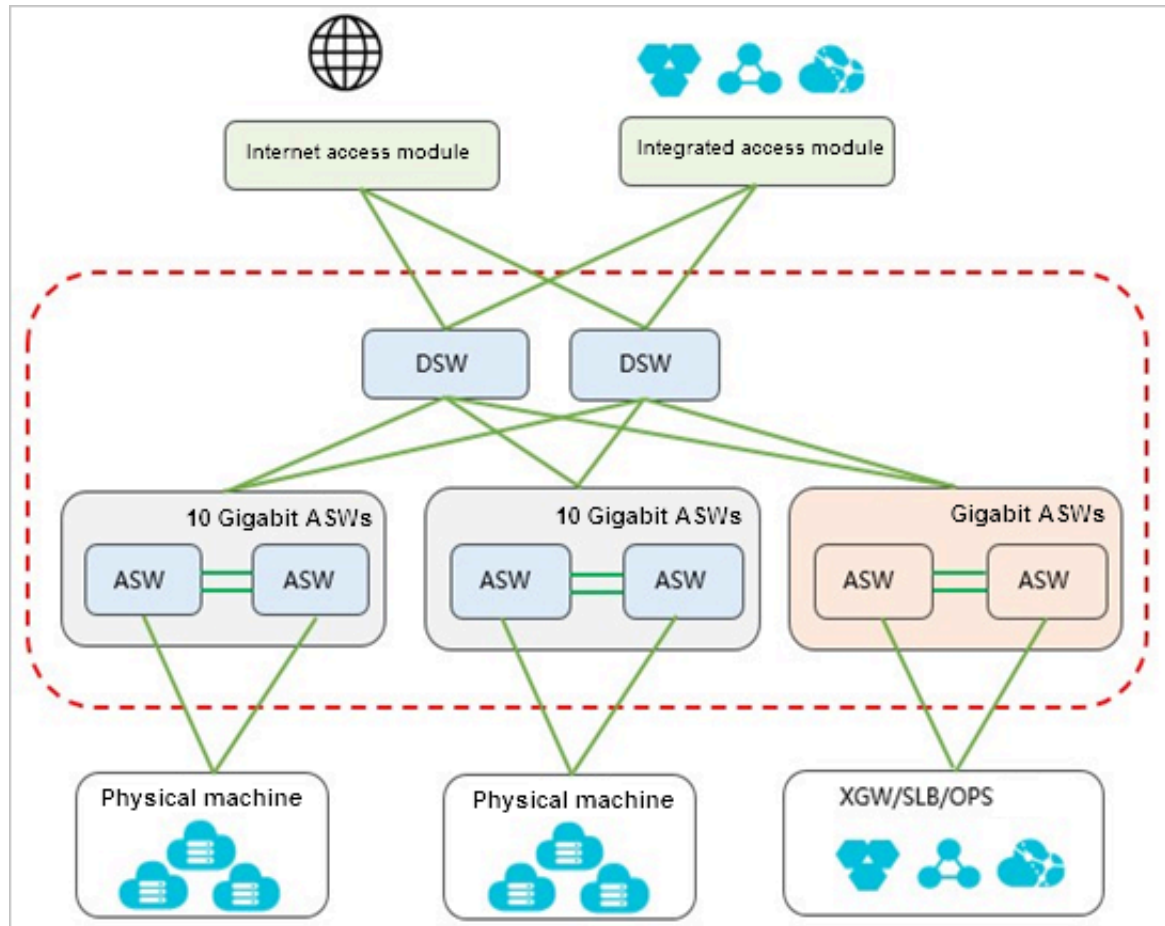
The business service area consists of the data exchange module and the integrated service module.

- **Data exchange module**

The data exchange module has a typical layer-2 CLOS architecture that consists of DSWs and ASWs. Each ASW pair forms a stack as a leaf node. According to the network sizes, this node can select data exchange models that have different applicable scopes. All cloud service servers are uplinked with the devices on the ASW stacks. ASWs are connected to DSWs by using External Border Gateway Protocol (EBGP). The DSWs are isolated from each other. The data exchange module is connected to other modules

by using EBGp, receives the Internet routes from ISWs, and releases the Classless Inter-Domain Routing (CIDR) block of cloud products to the ISWs.

Figure 1-6: Data exchange module

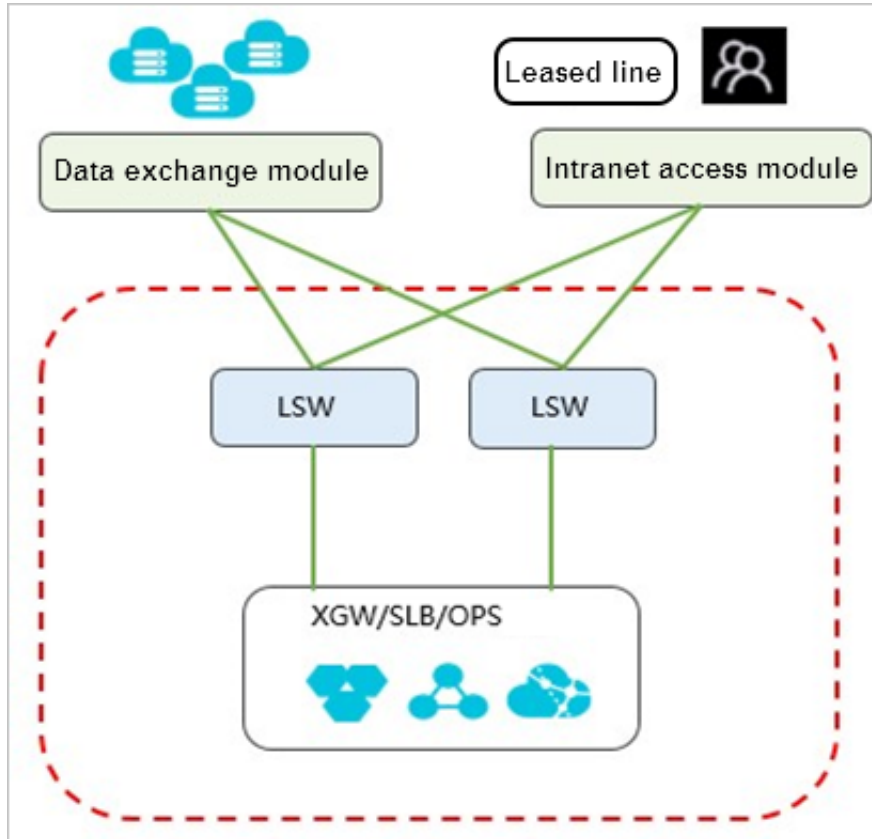


- **Integrated service module**

Each cloud service server (XGW/SLB/OPS) is connected to two LSWs. These servers exchange routing information by using Open Shortest Path First (OSPF). The two LSWs exchange routing information between each other by using Internal Border Gateway

Protocol (IBGP), and LSW exchange routing information with DSWs and CSWs by using EBGP.

Figure 1-7: Integrated service module



1.3.3.3 Integrated access area

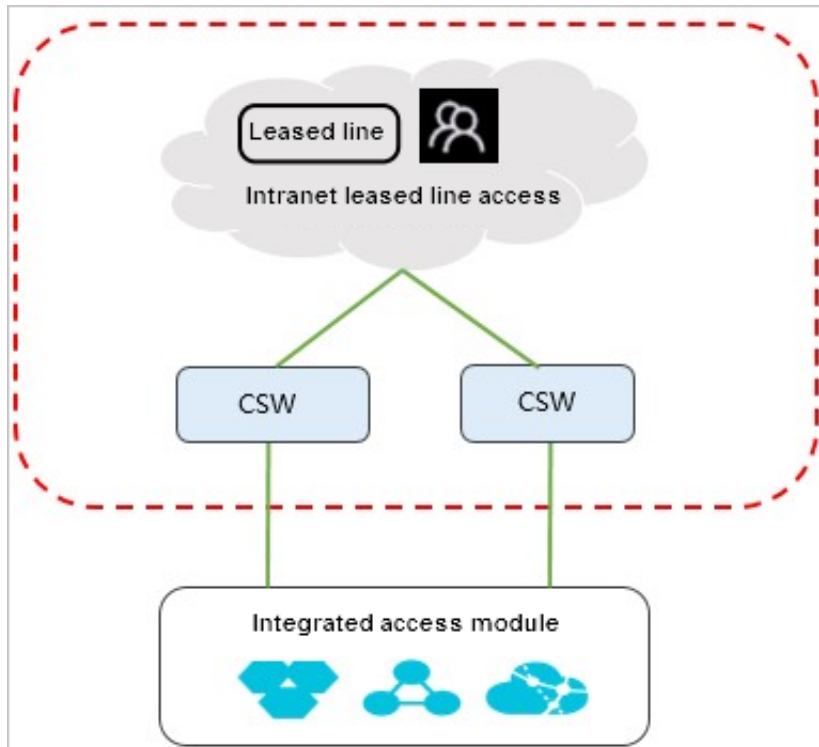
The integrated access area consists of the intranet access module and Internet access module.

- **Intranet access module**

In the intranet access module, two CSWs provide internal users with access to VPC (Virtual Private Cloud) and general cloud services. For access to VPC, CSWs set up a map from internal users to VPCs and import these users into different VPCs. Different user groups are isolated from each other on CSWs. For access to general cloud services,

CSWs are connected to the integrated service module by using External Border Gateway Protocol (EBGP) and allow direct access to all resources in the business service area.

Figure 1-8: Intranet access module



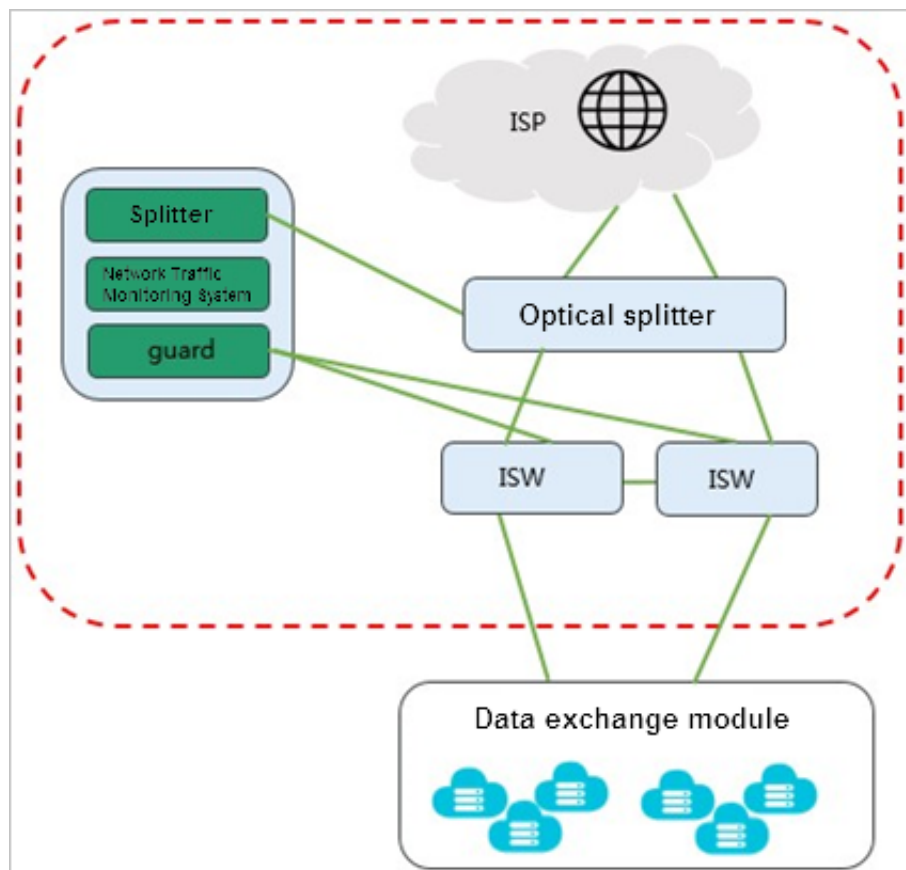
- **Internet access module**

The Internet access module consists of two ISWs. It facilitates the access to ISPs or users' public backbone networks and performs route distribution and interaction between the inside and outside of cloud networks. The two ISWs run Internal Border Gateway Protocol (IBGP) to back up routes between each other. Based on actual conditions, ISWs can use static routing or EBGP to uplink with Internet service providers (ISPs) or users' public backbone networks. The link bandwidth is defined based on the size of users' Alibaba Cloud networks and the bandwidth of their public backbone networks. We recommend that ISWs can use BGP to connect with multiple carriers to improve the reliability. Each carrier has 2×10 GE lines. The Internet access module also uses EBGP to exchange routes with the data exchange module, releases relevant Internet routes to the data exchange module, and receives the internal cloud service routes that are sent by the data exchange module to implement the interaction between the inside and outside of cloud networks.

The Internet access module is parallel to an Alibaba Cloud security protection system . The traffic generated by the Internet to cloud networks is diverted to Network Traffic

Monitoring System by using an optical splitter. When Network Traffic Monitoring System detects malicious traffic, it releases the corresponding routes by using Apsara Stack Security to divert the malicious traffic to Apsara Stack Security for scrubbing. The scrubbed traffic is injected back into the Internet access module.

Figure 1-9: Internet access module



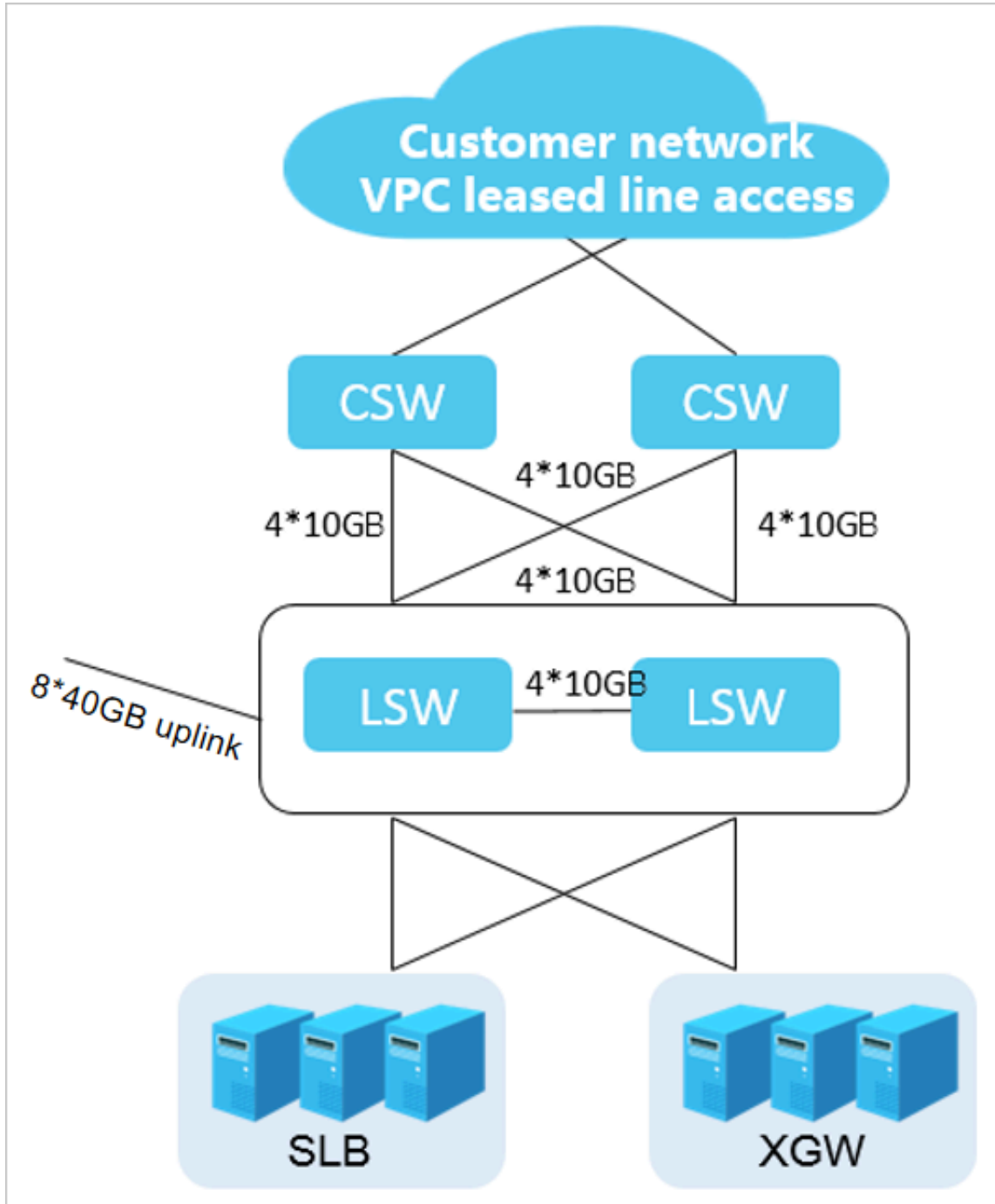
1.3.3.4 VPC leased line access

The leased line access solution of Virtual Private Cloud (VPC) allows you to control over your own virtual networks, such as selecting your own IP address ranges and configuring route tables and gateways. You can also connect your VPC to a traditional data center by using leased lines or VPN connections to create a customized network environment. This enables smooth migration of applications to the cloud.

Each cloud service server (XGW/SLB) is connected to two LSWs. These servers exchange routing information by using Open Shortest Path First (OSPF). Two LSWs exchange routing information between each other by using Internal Border Gateway Protocol (IBGP), and

LSWs exchange routing information with CSWs by using External Border Gateway Protocol (EBGP).

Figure 1-10: VPC leased line access



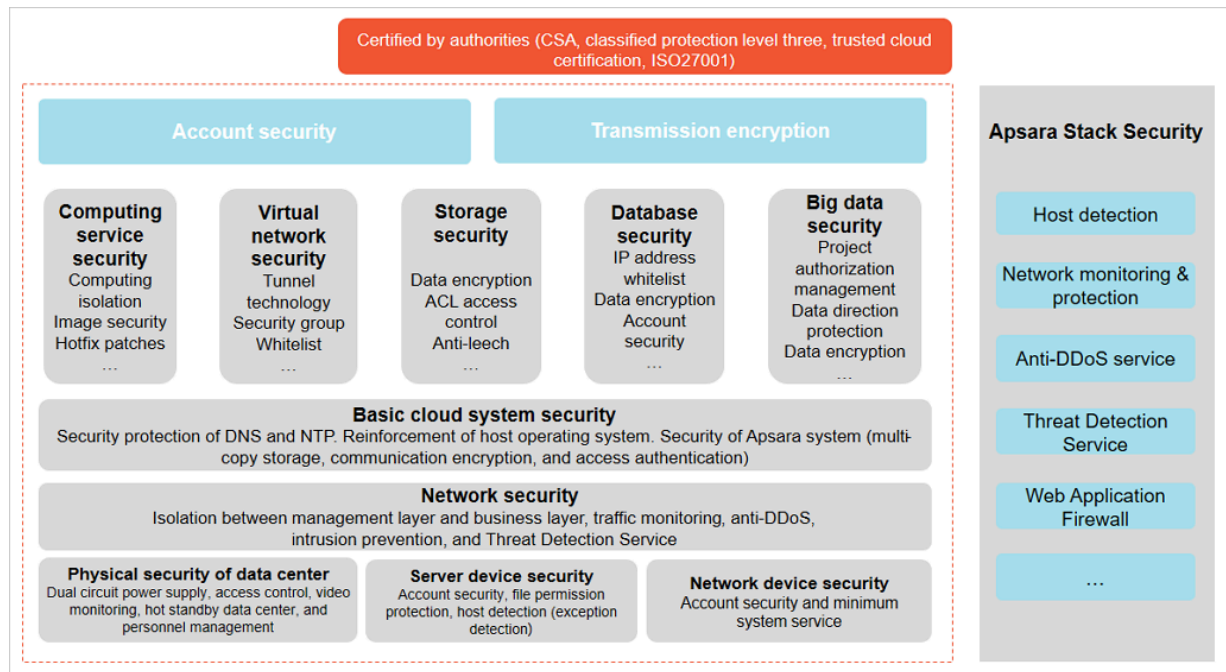
1.3.4 Security architecture

Apsara Stack provides all-around security capabilities from underlying communication protocols to upper-layer applications to guarantee the security of your access and data . Access to every console in Apsara Stack is allowed only with HTTPS certificates. Apsara

Stack provides a comprehensive role authorization mechanism to guarantee a secure and controllable access to resources in multi-tenant mode. It supports different security roles, such as a security administrator, system administrator, and security auditor.

Apsara Stack has incorporated Apsara Stack Security since the V3 version and provides you with a multi-level and integrated cloud security protection solution.

Figure 1-11: Hierarchical security architecture of Apsara Stack



1.3.5 Base assembly

Apsara Stack base consists of three types of assemblies, which provides support for the deployment and operations of the cloud platform.

Table 1-1: Base assembly

Assembly		Function
Ops assemblies	Yum	<p>Install package</p> <p>The software source is deployed during the initial installation phase. This package is mainly used to install the operating system and deploy application software packages and their dependent components of Apsara Stack, such as the Apsara platform and Elastic Compute Service (ECS), on physical machines.</p>
	Clone	Machine cloning service
	NTP	<p>Clock source service</p> <p>The physical machines deployed on Apsara Stack synchronize time from a standard NTP time source and provide the time to other hosts.</p>
	DNS	<p>Domain name resolution service</p> <p>DNS provides forward and reverse resolution of domain names for the internal Apsara Stack environment. It runs a bind instance on each of the two OPS machines and uses keepalived to provide high-availability services. When one machine fails, the other machine automatically takes over its work.</p>
Base middleware	Dubbo	Distributed RPC service
	Tair	Cache service
	mq	Message Queue service
	ZooKeeper	Distributed collaboration

Assembly		Function
	Diamond	Configuration management service
	SchedulerX	Timing task service
Basic base assemblies	Apsara Infrastructure Management Framework	Data center management
	Monitoring System	Data center monitoring
	OTS-inner	Table Store service
	SLS-inner	Log Service of cloud platform
	Meta database	Meta database
	POP	APIs on the cloud platform
	OAM	Account system
	RAM	Authentication and authorization system
	WebApps	Support for the Apsara Stack Operations console

1.4 Product panorama

Apsara Stack provides a variety of products to meet requirements of different users.

Infrastructure

Apsara Stack provides a wide variety of basic virtual resources, such as virtual computing, virtual network, and virtual scheduling. The main products include Elastic Compute Service (ECS), Virtual Private Cloud (VPC), Server Load Balancer (SLB), Container Service, Auto Scaling, and Key Management Service (KMS).

Storage products

Apsara Stack provides various storage products for different storage objects. The main products include Object Storage Service (OSS), Network Attached Storage (NAS), Table Store , and Apsara File Storage for HDFS (HDFS).

Internet middleware and applications

Apsara Stack provides middleware services and can host various customer applications. This facilitates the conversion of applications to services and encourages applications to evolve into a microservice architecture. The main products include API Gateway, Log Service , Apsara Stack DNS, Enterprise Distributed Application Service (EDAS).

Database

Apsara Stack provides diversified data engines. These data engines can interoperate with each other. The main products include ApsaraDB for RDS, KVStore for Redis, KVStore for Memcache, ApsaraDB for MongoDB, ApsaraDB RDS for PostgreSQL, Data Transmission Service (DTS), and Data Management Service (DMS).

Big data processing

Apsara Stack provides various functions of big data analysis, application, and visualization , which maximizes the value of data. The main products include MaxCompute, DataWorks, Realtime Compute, Quick BI, E-MapReduce (EMR), DataQ - Smart Tag Service, Dataphin, and Apsara Bigdata Manager (ABM).

Artificial intelligence

Apsara Stack provides a machine learning algorithm platform based on the distributed computing engine developed by Alibaba Cloud, such as Machine Learning Platform for AI (PAI).

Security

Apsara Stack provides an all-around protection from underlying communication protocols to upper-layer applications, which guarantees the security of your access and data. For example, the product Apsara Stack Security.

1.5 Scenarios

Apsara Stack provides flexible and scalable industrial solutions for users who are of different scales and from different sectors. Based on the business traits of different sectors , such as industry, agriculture, transportation, government, finance, and education, Apsara Stack creates customized solutions to provide users with one-stop products and services. This topic focuses on introducing the following two scenarios.

City Brain

Urban management is a field that involves one of the largest volumes of data in China. This marks the transition of governmental information from a closed-flow model to an open-flow online model. With more time and space to flow in, urban data has a higher value. Cloud computing becomes an urban infrastructure, data becomes a new means of production and a strategic resource, and AI technology becomes the nerve center of a smart city. All of these forms the City Data Brain.

The values and features are as follows:

- A breakthrough of urban governance mode. With the urban data as a resource, City Brain improves the government management capabilities, resolves prominent issues of urban governance, and achieves an intelligent, intensive, and humane form of governance.
- A breakthrough of urban service mode. City Brain provides more accurate and convenient services for enterprises and individuals, makes the urban public services more efficient, and saves more public resources.
- A breakthrough of urban industrial development. City Brain lays down an industrial AI layout, takes open urban data as an important fundamental resource, drives the development of industries, and promotes the transformation and upgrade of traditional industries.

Finance Cloud

Finance Cloud is an industrial cloud that serves financial organizations, such as banks, security agencies, insurance companies, and funds. It relies on a cluster of independent data centers to provide cloud products that meet the regulatory requirements of the People's Bank of China, China Banking Regulatory Commission (CBRC), China Securities Regulatory Commission (CSRC), and China Insurance Regulatory Commission (CIRC). It also provides more professional and comprehensive services for financial customers. Enterprises can build Finance Cloud independently or with Alibaba Cloud. Finance Cloud meets the requirements of large- and medium-sized financial organizations for independent cloud data centers that are completely physically isolated. It can also output the cloud computing and big data platform to customers' data centers.

The values and features are as follows:

- Independent resource clusters
- Stricter data center management

- Better disaster recovery capability
- Stricter requirements for network security isolation
- Stricter access control
- Compliance with the security supervision requirements and compliance requirements of banks
- Dedicated security operation team, security compliance team, and security solution team of the Finance Cloud sector
- Dedicated account managers and cloud architects of Finance Cloud
- Stricter user access mechanism

1.6 Compliance security solution

1.6.1 Overview

On June 1, 2017, the Cybersecurity Law of the People's Republic of China was officially implemented, which has made clear provisions for classified protection compliance. Drawing on its technical advantages on Apsara Stack Security products, Alibaba Cloud builds a classified protection compliance ecosystem to help you quickly align with the provisions for classified protection compliance. Alibaba Cloud works with its cooperative assessment agencies and security consulting providers based worldwide to offer one-stop classified protection assessment services. It offers complete attack protection, data auditing, encryption, and security management that make it easier for you to quickly pass the classified protection compliance assessment.

1.6.2 Interpretation on key points

Network and communication security

Interpretation on clauses

- Divide the network into different security domains according to different server roles and server importance.
- Set access control policies at the security domain boundary between the intranet and Internet, which must be configured on specific ports.
- Deploy intrusion prevention measures at the network boundary to prevent against and record intrusion behaviors.
- Record and audit the user behavior logs and security events in the network.

Coping strategies

- We recommend that you use Virtual Private Cloud (VPC) and security group of Alibaba Cloud to divide a network into different security domains and perform reasonable access control.
- You can use Web Application Firewall (WAF) to prevent network intrusion.
- You can use the log feature to record, analyze, and audit user behavior logs and security events.
- If the system is frequently threatened by DDoS attacks, you can use Anti-DDoS Pro to filter and scrub abnormal traffic.

Device and computing security

Interpretation on clauses

- Avoid account sharing, record, and audit operations actions, which is an elementary security requirement.
- Secure system layer with necessary security measures and prevent servers from intrusions.

Coping strategies

- You can audit the server and data actions, and create an independent account for each operations personnel to avoid account sharing.
- You can use Server Guard to conduct complete vulnerability management, baseline check and intrusion prevention on servers.

Application and data security

Interpretation on clauses

- An application directly implements specific business and is not like the network and system with relative standard characteristics. The functions of most applications such as identity authentication, access control and operation audit are difficult to be replaced by third-party products.
- Encryption is the most effective method to secure data integrity and confidentiality except security prevention methods at other levels.
- Remote data backup is one of the most important requirements that distinguishes the third level of classified protection from the second level. It is also the most basic technical safeguard measure for business sustainability.

Coping strategies

- At the beginning of the application development, application functions such as identity authentication, access control, and security audit must be considered.
- For online systems, you can add functions such as account authentication, user permission identification, and log auditing to satisfy classified protection requirements.
- For data security, HTTPS can be used to guarantee that data remains encrypted in the transmission process.
- For data backup, we recommend that you can use remote disaster recovery instance of ApsaraDB for RDS to automatically back up data and manually synchronize backup files of database to Alibaba cloud servers in other regions.

Security management policies

Interpretation on clauses

- Security policy, regulation, and management personnel are significant bases for sustainable security. Policy guides the security direction. Regulation specifies the security process. Management personnel fulfills the security responsibilities.
- Classified protection requirements provide a methodology and best practice. You can perform continuous security construction and management according to the classified protection methodology.

Coping strategies

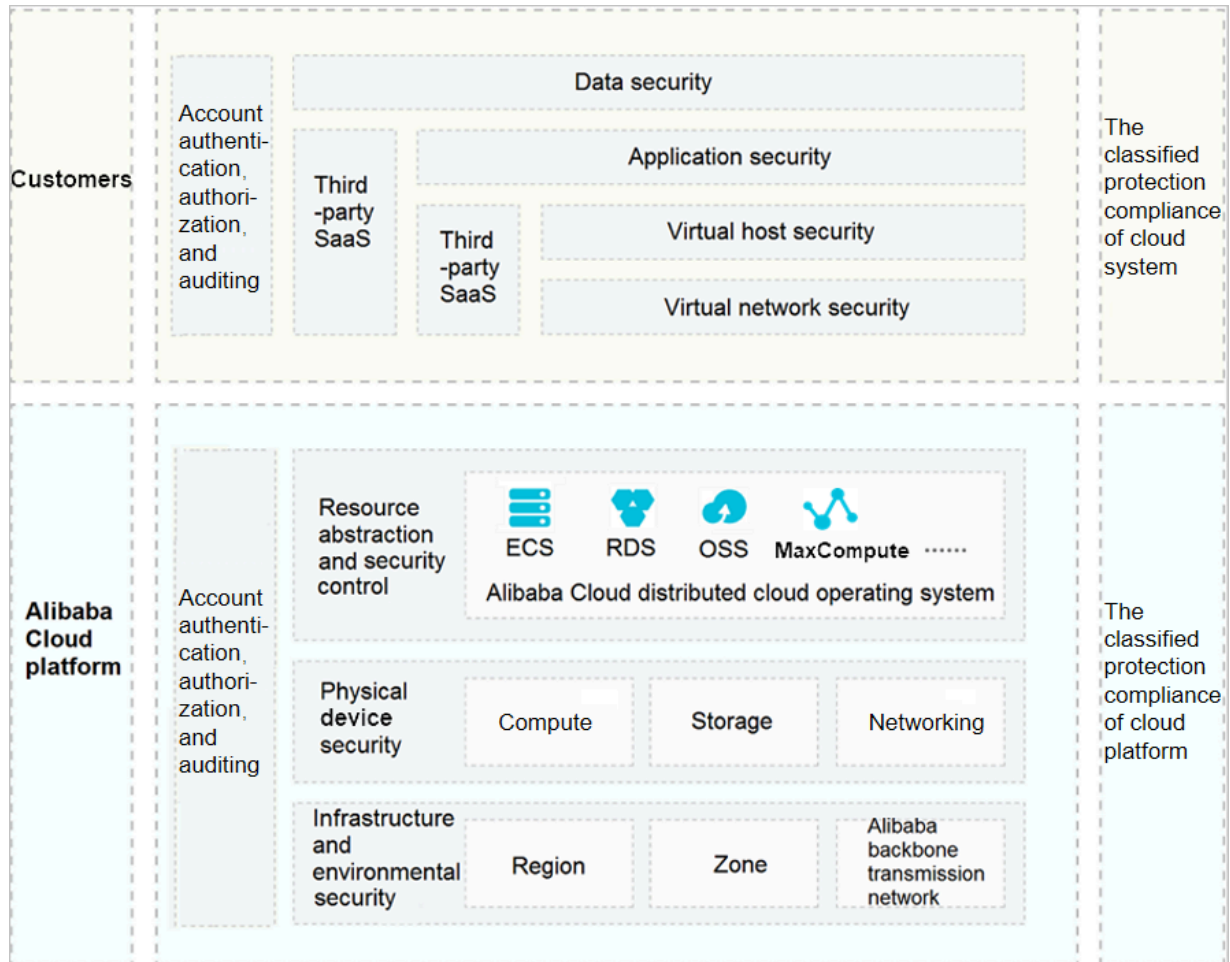
- The customer management staff can arrange, prepare, and fulfill the security policy, regulation, and management personnel according to the actual condition of enterprise and form specialized documents.
- For the technical means required in the process of vulnerability management, we recommended that you can use Alibaba Cloud Server Guard to quickly detect the vulnerabilities of cloud system and resolve them in time.

1.6.3 Cloud-based classified protection compliance

Shared compliance responsibilities

The Alibaba Cloud platform and the cloud tenant systems are classified and assessed respectively. You can use the assessment conclusions of the Alibaba Cloud platform when assessing the tenant systems.

Figure 1-12: Shared compliance responsibilities



Alibaba Cloud provides the following contents:

- Classified protection filing certification of the Alibaba Cloud platform
- Key pages of the Alibaba Cloud assessment report
- Sales license of Apsara Stack Security
- Description of partial assessment items of Alibaba Cloud

More details about shared responsibilities are as follows:

- Alibaba Cloud is the unique cloud service provider in China that participates in and passes the pilot demonstration of cloud computing classified protection standards.

Public Cloud and E-Government Cloud pass the filing and assessment of the third level of classified protection. Finance Cloud passes the filing and assessment of the fourth level of classified protection.

- According to the regulatory authority, you can use the assessment conclusions of physical security, partial network security, and security management for the classified protection assessment of the tenant systems on Alibaba Cloud, and Alibaba Cloud can provide supporting details.
- With the complete security technology, management architecture, and protection system of Apsara Stack Security, Alibaba Cloud platform makes it easy for tenants to pass the classified protection assessment.

Classified protection compliance ecology

Current conditions of cloud-based classified protection are as follows:

- Most tenants do not know classified protection.
- Most tenants do not know how to start with classified protection.
- Most tenants are not good at communicating with supervision authorities.
- Security systems lag behind business development.

Alibaba Cloud establishes Classified Protection Compliance Ecology to provide one-stop classified protection compliance solutions for cloud-based systems to quickly pass classified protection assessment.

Figure 1-13: Classified protection compliance ecology



Work division of classified protection:

- Alibaba Cloud: integrates capabilities of service agencies and provides security products
- Consulting firm: provides technical support and consulting services in the whole process
- Assessment agency: provides assessment services
- Public security organ: reviews filing and supervises services

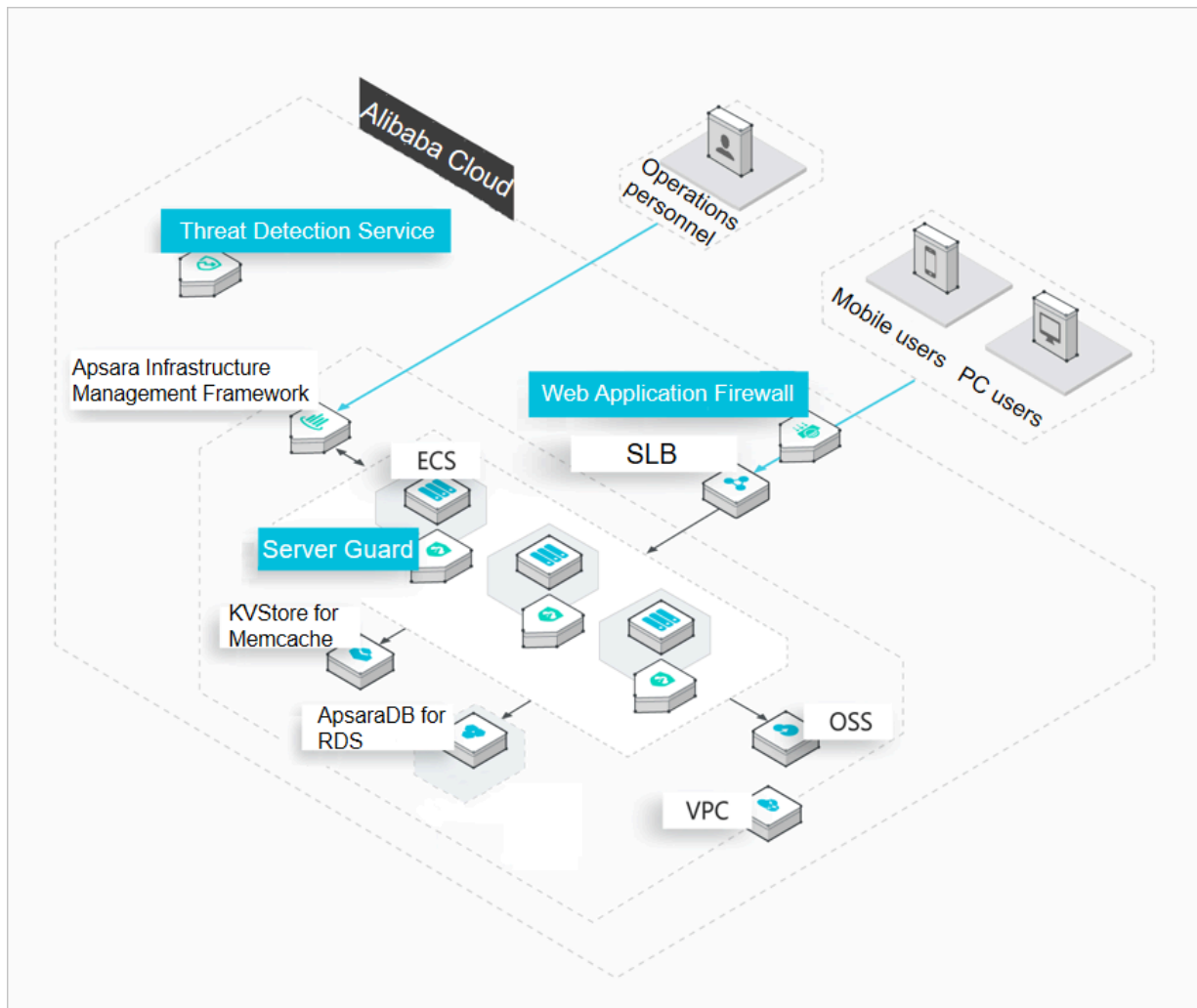
1.6.4 Classified protection implementation process

	Operating unit	Alibaba Cloud	Consulting or assessment agency	Public security organ
System rating	Determine the class of security protection and write rating report	Coordinate the third party agency to provide counseling services for operating unit	Counseling the operating unit to prepare the rating materials and organize expert review (level three)	None
System filing	Prepare and present the filing materials to the local public security organ	Coordinate the third party agency to provide counseling services for operating unit	Counseling the operating unit to prepare the filing materials and to issue filing	None
Construction rectification	Construct the security technology and management system in line with class requirements	Provide the obligatory security products and services that meet the class requirements	Counseling the operating unit to carry out system security reinforcement and develop security management regulation	The local public security organ reviews and accepts the filing materials
Rating assessment	Prepare for and accept the assessment from the assessment agency	Provide the cloud service provider's security qualification and the proof that the cloud platform has passed the classified protection	The assessment agency assesses the system class conformity	None
Supervision & inspection	Accept the regular inspection of public security organ	None	None	Supervise and inspect the operating unit to carry out the class protection work

1.6.5 Security compliance architecture

With the security compliance architecture, Alibaba Cloud can fast connect to Apsara Stack Security, quickly improve the security, and comply with basic technical requirements for classified protection at minimal security costs.

Figure 1-14: Security compliance architecture



Basic requirements of classified protection are as follows:

- Physical and environmental security: includes data center power supply, temperature and humidity control, and prevention of wind, rain, and lightning. You can use the assessment conclusions of Alibaba Cloud.
- Network and communication security: includes network architecture, boundary protection, access control, intrusion prevention, and communication encryption.
- Device and computing security: includes intrusion prevention, malicious code prevention, identity authentication, access control, centralized control, and security auditing.

- Application and data security: includes security auditing, data integrity, and data confidentiality.

1.6.6 Solution benefits

One-stop assessment service of classified protection

Select high-performance consulting and assessment partners to provide one-stop compliance support throughout, allowing the operators to achieve significant cost savings.

- Eliminates multi-level communication and work redundancy to help the operators reduce investment.
- Improves efficiency by shortening the assessment cycle to as short as two weeks.
- Alibaba Cloud provides best practices of security and compliance on the cloud.

A complete security protection system

With a complete Apsara Stack Security architecture, operators can locate corresponding products on Alibaba Cloud, rectify non-conformances, and meet all requirements of classified protection.

2 Elastic Compute Service (ECS)

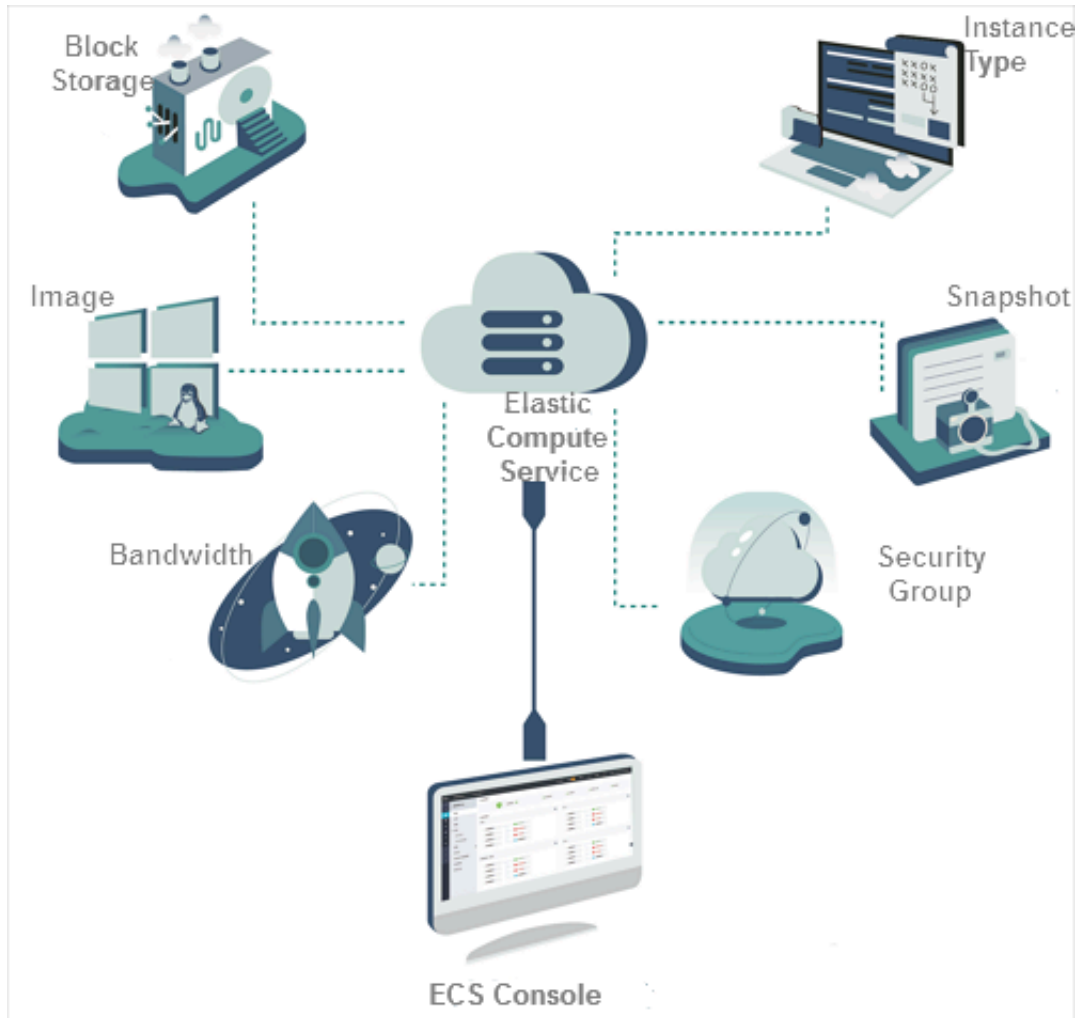
2.1 What is ECS?

Elastic Compute Service (ECS) is a computing service that features elastic processing capabilities. Compared with physical servers, ECS instances are more user-friendly and can be managed more efficiently. 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 through the ECS console. Other resources, such as block storage, images,

and snapshots, can only be used after they are integrated with ECS instances. For more information, see [Figure 2-1: ECS components](#).

Figure 2-1: ECS components



2.2 Benefits

ECS has the following benefits compared with the services provided by other service vendors and common IDCs:

- [High availability](#)
- [Security](#)
- [Scalability](#)

High availability

Compared with the services provided by common IDCs and server vendors, ECS adopts more stringent IDC standards, server access standards, and O&M standards to ensure data reliability, and high availability at the infrastructure and instance levels.

Apsara Stack provides you with the following support services:

- Products and services to improve availability. These include ECS, Server Load Balancer (SLB), multi-backup database, and Data Transformation.
- Industry partners and ecosystem partners that help you build a more advanced and stable architecture and ensure service continuity.
- Diverse training services to help you achieve high availability from the business end to the underlying basic service end.

Security

Security and stability are two of the primary concerns for any cloud service user. Alibaba Cloud has recently passed a host of international information security certifications which demand strict confidentiality of user data and user information and user privacy protection , including ISO27001 and MTCS.

- **With a simple configuration to connect your business environment to global IDCs** , Apsara Stack Virtual Private Cloud (VPC) can increase the flexibility, scalability, and stability of your business.
- **You can connect your own IDC to** Apsara Stack VPC through a leased line to implement a hybrid cloud. You can use a variety of hybrid cloud architectures to provide network services and robust networking. A superior business ecosystem is made possible with the ecosystem of Apsara Stack.
- **VPCs are more stable and secure.**
 - **Stability:** After constructing your VPC, you can update your network architecture and obtain new functions daily to constantly evolve your network infrastructure and ensure your business is always running smoothly. VPCs allow you to divide, configure, and manage your network as needed.
 - **Security:** VPCs feature traffic isolation and attack isolation to protect your services against endless attacks on the Internet. The first line of defense against malicious attacks and traffic is established when you build your VPC.

VPCs provide a stable, secure, controllable, and fast-deliverable network environment. The capability and architecture of VPC hybrid cloud bring the technical advantages of cloud computing to enterprises in traditional industries not engaged in cloud computing.

Elasticity

Elasticity is a key benefit of cloud computing.

- **Elastic computing**

- **Vertical elasticity involves modifying the configurations of a server.** In a traditional IDC, it is difficult to change the configuration of a single server. However, you can change the capacity of your ECS instance or storage service based on the actual needs of your business.
- **Transverse elasticity** allows for re-division of resources between applications. For example, a traditional IDC may not be able to immediately provide sufficient resources for online gaming or live video streaming applications during peak hours. The elasticity of cloud computing makes it possible to provide the resources required in peak hours. When the load returns to normal levels, you can release unnecessary resources to reduce operation costs. The combination of ECS vertical and transverse elasticity enables resources to scale up and down by specified quantity as scheduled or against business load.

- **Elastic storage**

Apsara Stack provides elastic storage. For a traditional IDC, you must add servers to increase the storage space. However, the number of servers that you can add is limited. Cloud computing can provide you with large-capacity storage. You can purchase the storage you need at any time.

- **Elastic network**

Apsara Stack features elastic network as well. Apsara Stack VPCs can be configured to match the specifications of your own IDCs. In addition, VPCs provide the following benefits: The IDCs can communicate with each other while being isolated through security domains. VPC configurations and planning are flexible.

To sum up, Apsara Stack provides elastic computing, storage, networking, and business architecture planning and allows you to combine your businesses as needed.

2.3 Architecture

ECS is built on the Apsara system that was developed by Alibaba Cloud. The individual ECS instances are virtualized by using KVM while storage is implemented on Apsara Distributed File System.

ECS architecture

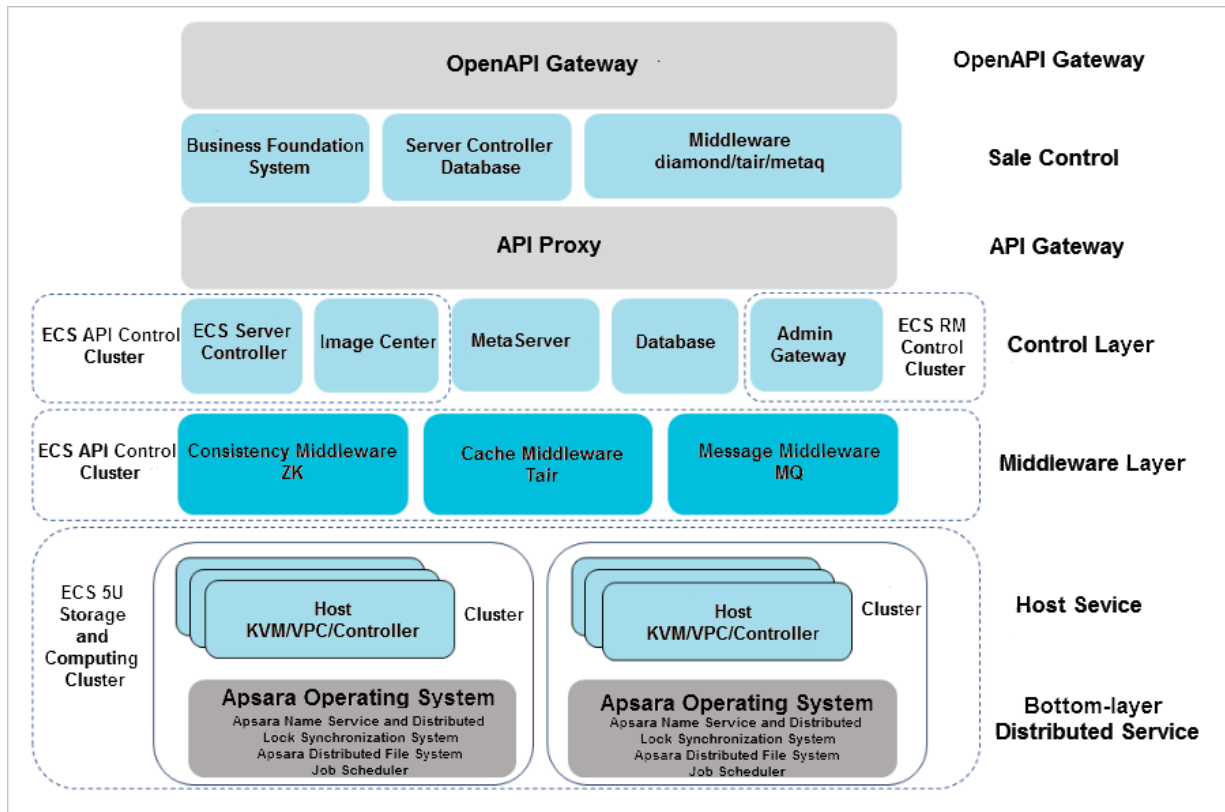


Table 2-1: Architecture description

Component	Description
Apsara Name Service and Distributed Lock Synchronization System	A basic module that provides services related to distribution consistency in Apsara Stack. As a key distributed coordination system of Apsara Stack, it provides three types of basic services: distributed lock services, subscription and notification services , and lightweight metadata storage services.

Component	Description
Apsara Distributed File System	A distributed storage system developed by Alibaba Cloud. As of 2017, hundreds of clusters and hundreds of thousands of storage nodes using Apsara Distributed File System have been deployed in the production environments. Apsara Distributed File System manages tens of exabytes of disk space.
Job Scheduler	A distributed resource scheduler that manages and allocates resources in the distributed systems.
Server Controller	The ECS scheduling system that schedules storage, network, and computing resources in a unified manner and produces virtual machines (ECS instances) deliverable to users.
Scheduling process	API > Business layer > Control system > Host service
OpenAPI Gateway	Provides services such as authentication and request forwarding.
Business Foundation System	Creates and releases instances and snapshot policies, processes sales requests, and provides APIs to users.
API Proxy	Forwards requests to services of the region corresponding to region_id.
Server Controller databases	Stores control data and status data.
Server Controller	A center for scheduling storage, network, and computing resources.
Tair	Provides cache services for Server Controller.
Zookeeper	Provides the distributed lock service for Server Controller.
MQ	Virtual machine status message queue service.
Image Center	Provides image management services such as import and copy.
MetaServer	Provides metadata management services for ECS instances.
Host service	Provides services such as KVM for computing virtualization, VPC for network virtualization, and control through interaction with Libvirt.

Component	Description
Admin Gateway (AG)	Functions as the bastion host used to log on to an NC during O&M management.
ECS Decider	Determines on which NC to deploy ECS.

2.4 Features

2.4.1 Instances

2.4.1.1 Overview

An ECS instance is a virtual machine that contains basic computing components such as the CPU, memory, operating system, and network. You can fully customize and modify all configurations of an ECS instance. After logging on to Apsara Stack console, you can manage resources and configure the environment of your ECS instances.

2.4.1.2 Instance families

An ECS instance is the smallest unit offered by Alibaba Cloud that provides compute capabilities and services for your business. The compute capabilities vary with the instance type. ECS instances are categorized into different instance families based on different business scenarios.



Note:

The instance families and instance types described in this topic are for reference only. The types of instances that are available for purchase are determined by the physical servers where the instances are hosted.

Instance family	Feature	Scenario
n4	<p>n4 is a general purpose shared instance family with the following features:</p> <ul style="list-style-type: none"> • CPU-to-memory ratio of 1:2 • 2.5 GHz Intel Xeon E5-2680 v3 (Broadwell) processors • Latest DDR4 memory • I/O optimized by default 	Small and medium-sized web servers, batch processing , advertising services, and distributed analysis

Instance family	Feature	Scenario
mn4	<p>mn4 is a balanced and shared instance family with a greater CPU-to-memory ratio. mn4 has the following features:</p> <ul style="list-style-type: none"> • CPU-to-memory ratio of 1:4 • 2.5 GHz Intel Xeon E5-2680 v3 (Broadwell), Intel Xeon E5-2680 v4 (Haswell), Intel Xeon E5-2682 v4 (Broadwell), or Intel Xeon E5-2650 v2 (Haswell) processors • Latest DDR4 memory • I/O optimized by default 	Medium-sized web servers, batch processing, advertising services, distributed analysis, and Hadoop clusters
xn4	<p>xn4 is a compact and shared instance family with the following features:</p> <ul style="list-style-type: none"> • CPU-to-memory ratio of 1:1 • 2.5 GHz Intel Xeon E5-2680 v4 (Haswell) or Intel Xeon E5-2682 v4 (Broadwell) processors • Latest DDR4 memory • I/O optimized by default 	Minisite web applications, small databases, development or testing environments, and code storage servers
e4	<p>e4 is a memory optimized instance family with the following features:</p> <ul style="list-style-type: none"> • CPU-to-memory ratio of 1:8 • 2.5 GHz Intel Xeon E5-2680 v4 (Broadwell), Intel Xeon E5-2680 v3 (Broadwell), Intel Xeon E5-2650 v2 (Haswell), or Intel Xeon E5-2682 v4 (Broadwell) processors • I/O optimized by default 	Applications that involve a large number of memory operations, queries, and computations, such as Cache, Redis, search applications, and in-memory databases

Instance family	Feature	Scenario
sn1ne, compute optimized instance family with enhanced network performance	<ul style="list-style-type: none"> • CPU-to-memory ratio of 1:2 • Ultra high packet forwarding rate • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors with consistent computing performance • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Scenarios where large volumes of packets are received and transmitted , such as live commenting on videos and telecom data forwarding • Web frontend servers • Frontends of MMO games • Data analysis, batch computing, and video encoding • High-performance science and engineering applications
g5, general purpose instance family	<ul style="list-style-type: none"> • I/O-optimized • Support for enhanced SSDs , standard SSDs, and ultra disks • CPU-to-memory ratio of 1:4 • Ultra high packet forwarding rate • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) or 8269CY (Cascade Lake) processors with consistent computing performance • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Scenarios where large volumes of packets are received and transmitted , such as live commenting on videos and telecom data forwarding • Enterprise-grade applications of various types and sizes • Small and medium-sized database systems, caches, and search clusters • Data analysis and computing • Computing clusters and memory-intensive data processing

Instance family	Feature	Scenario
sn2ne, general purpose instance family with enhanced network performance	<ul style="list-style-type: none"> • CPU-to-memory ratio of 1:4 • Ultra high packet forwarding rate • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors with consistent computing performance • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Scenarios where large volumes of packets are received and transmitted , such as live commenting on videos and telecom data forwarding • Enterprise-grade applications of various types and sizes • Small and medium-sized database systems, caches, and search clusters • Data analysis and computing • Computing clusters and memory-intensive data processing
se1ne, memory optimized instance family with enhanced network performance	<ul style="list-style-type: none"> • CPU-to-memory ratio of 1:8 • Ultra high packet forwarding rate • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors with consistent computing performance • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Scenarios where large volumes of packets are received and transmitted , such as live commenting on videos and telecom data forwarding • High-performance and in-memory databases • Data analysis and mining , and distributed memory caches • Hadoop clusters, Spark clusters, and other memory-intensive enterprise applications

Instance family	Feature	Scenario
se1, memory optimized instance family	<ul style="list-style-type: none"> • Consistent computing performance • CPU-to-memory ratio of 1:8 • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors • Latest DDR4 memory • Higher network performance along with more computing capacity • I/O optimized by default 	se1 is a memory-dedicated instance family with a greater CPU-to-memory ratio. se1 is applicable to computing scenarios with fixed performance, such as Cache, Redis, search applications, in-memory databases, databases with high I/O (Oracle and MongoDB), Hadoop clusters, and large-volume data processing.
ebmg5s, general purpose ECS Bare Metal Instance family with enhanced network performance	<ul style="list-style-type: none"> • Support for enhanced SSDs, standard SSDs, and ultra disks • CPU-to-memory ratio of 1:4 • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96 vCPUs, Turbo Boost up to 2.7 GHz • High network performance: 4,500 Kpps packet forwarding rate • Supports VPCs only • Provides dedicated hardware resources and physical isolation 	<ul style="list-style-type: none"> • Workloads or licensing schemes that require direct access to physical resources • Third-party virtualization (including Xen and KVM), and AnyStack (including OpenStack and ZStack) • Containers (including Docker, Clear Containers, and Pouch) • Enterprise-grade applications, such as large and medium-sized databases • Video encoding

Instance family	Feature	Scenario
ebmg5, general purpose ECS Bare Metal Instance family	<ul style="list-style-type: none"> • CPU-to-memory ratio of 1:4 • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96 vCPUs, Turbo Boost up to 2.9 GHz • High network performance: 4 ,500 Kpps packet forwarding rate • Support for standard SSDs and ultra disks 	<ul style="list-style-type: none"> • Deployment of Apsara Stack services such as OpenStack and ZStack • Deployment of services such as Docker containers • Scenarios where large volumes of packets are received and transmitted , such as live commenting on videos and telecom data forwarding • Enterprise-grade applications of various types and sizes • Large and medium-sized databases, caches, and search clusters • Data analysis and computing • Computing clusters and memory-intensive data processing
i2, instance family with local SSDs	<ul style="list-style-type: none"> • High-performance local NVMe SSDs with high IOPS, high I/O throughput, and low latency • CPU-to-memory ratio of 1:8, designed for high-performance databases • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Online transaction processing (OLTP) and high -performance relational databases • NoSQL databases such as Cassandra and MongoDB • Search scenarios such as Elasticsearch

Instance family	Feature	Scenario
d1, big data instance family	<ul style="list-style-type: none"> • High-capacity local SATA HDDs with high throughput and up to 17 Gbit/s of bandwidth among instances • CPU-to-memory ratio of 1:4, designed for big data scenarios • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Hadoop MapReduce, HDFS , Hive, and HBase • Spark in-memory computing and MLlib • Enterprises in Internet , finance, and other industries that need to compute, store, and analyze large volumes of data • Elasticsearch and logging
d2, big data instance family	<ul style="list-style-type: none"> • High-capacity local SATA HDDs with high throughput and bandwidth of up to 10 Gbit/s between instances • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Hadoop MapReduce, HDFS , Hive, and HBase • Spark in-memory computing and MLlib • Enterprises in Internet , finance, and other industries that need to compute, store, and analyze large volumes of data • Elasticsearch and logging
sccg5ib, general purpose SCC instance family	<ul style="list-style-type: none"> • CPU-to-memory ratio of 1:8 • Ultra high packet forwarding rate • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors with consistent computing performance • 100G InfiniBand (IB) network with ultra high bandwidth and ultra low latency 	<ul style="list-style-type: none"> • Data analysis and computing • Artificial intelligence computation • Manufacturing simulation • High-performance computing clusters • Genetic analysis • Pharmaceutical analysis

Instance family	Feature	Scenario
scch5ib, SCC instance family with high clock speed	<ul style="list-style-type: none"> • CPU-to-memory ratio of 1:6 • Ultra high packet forwarding rate • 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors with consistent computing performance • 100G InfiniBand (IB) network with ultra high bandwidth and ultra low latency 	<ul style="list-style-type: none"> • Data analysis and computing • Artificial intelligence computation • Manufacturing simulation • High-performance computing clusters • Genetic analysis • Pharmaceutical analysis
re5, memory optimized instance family with enhanced performance	<ul style="list-style-type: none"> • Optimized for high-performance databases, in-memory databases, and other memory-intensive enterprise applications • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors with consistent computing performance • CPU-to-memory ratio of 1:16.5, memory of up to 2,970 GiB 	<ul style="list-style-type: none"> • High-performance and in-memory databases • Memory-intensive applications • Big data processing engines such as Apache Spark or Presto
sn1, compute optimized instance family	<ul style="list-style-type: none"> • CPU-to-memory ratio of 1:2 • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v3 (Haswell) processors with consistent computing performance • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Web frontend servers • Frontends of MMO games • Data analysis, batch computing, and video encoding • High-performance science and engineering applications

Instance family	Feature	Scenario
sn2, general purpose instance family	<ul style="list-style-type: none"> CPU-to-memory ratio of 1:4 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v3 (Haswell) processors with consistent computing performance Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> Enterprise-grade applications of various types and sizes Small and medium-sized database systems, caches, and search clusters Data analysis and computing Computing clusters and memory-intensive data processing

Instance family	Feature	Scenario
f1, compute optimized instance family with FPGAs	<ul style="list-style-type: none"> Intel ARRIA 10 GX 1150 FPGAs CPU-to-memory ratio of 1:7.5 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> Deep learning inference Genomics research Financial analysis Image transcoding Computational workloads such as real-time video processing and security management
f3, compute optimized instance family with FPGAs	<ul style="list-style-type: none"> Xilinx Virtex UltraScale+ VU9P CPU-to-memory ratio of 1:4 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> Deep learning inference Genetic computation Video encoding and decoding Chip prototype verification Database acceleration

Instance family	Feature	Scenario
gn5, compute optimized instance family with GPU capabilities	gn5 instances use GPU processing and have the following features: <ul style="list-style-type: none"> • NVIDIA P100 GPU processors • Multiple CPU-to-memory ratios • High-performance NVMe SSDs • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Deep learning • Scientific computing applications such as fluid dynamics, finance, genomics, and environmental analysis • Server-side GPU computational workloads such as high-performance computation, rendering, and multi-media coding and decoding
gn4, compute optimized instance family with GPU capabilities	gn4 instances use GPU processing and have the following features: <ul style="list-style-type: none"> • NVIDIA M40 GPU processors • Multiple CPU-to-memory ratios • 2.5 GHz Intel Xeon E5-2680 v4 (Haswell) processors • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Deep learning • Scientific computing applications such as fluid dynamics, finance, genomics, and environmental analysis • Server-end GPU computational workloads such as high-performance computing, rendering, and multi-media encoding and decoding
ga1, compute-optimized instance family with GPU capabilities	ga1 instances use GPU processing and have the following features: <ul style="list-style-type: none"> • AMD S7150 GPUs • CPU-to-memory ratio of 1:2.5 • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors • High-performance NVMe SSDs • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Rendering and multimedia encoding and decoding • Machine learning, high-performance computing, and high-performance databases • Server-side workloads that require powerful concurrent floating-point compute capabilities

Instance family	Feature	Scenario
gn5i, compute optimized instance family with GPU capabilities	gn5i instances use GPU processing and have the following features: <ul style="list-style-type: none"> • NVIDIA P4 GPU processors • CPU-to-memory ratio of 1:4 • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Deep learning inference • Server-side GPU computational workloads such as multimedia encoding and decoding
gn5e, compute optimized instance family with GPU capabilities	<ul style="list-style-type: none"> • I/O-optimized • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors with consistent computing performance • NVIDIA P4 GPU processors • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Deep learning and reasoning • Video and image processing, such as noise reduction, encoding, and decoding

Instance family	Feature	Scenario
gn6i, compute optimized instance family with GPU capabilities	<p>gn6i instances use GPU processing and have the following features:</p> <ul style="list-style-type: none"> • I/O-optimized • CPU-to-memory ratio of 1:4 • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors • Support for standard SSDs, ultra disks, and enhanced SSDs that deliver millions of IOPS • Better performance with the new-generation X-Dragon compute architecture • NVIDIA T4 GPU computing accelerators <ul style="list-style-type: none"> - New NVIDIA Turing architecture - Up to 320 Turing Tensor Cores - Up to 2,560 CUDA Cores - Mixed-precision Tensor Cores supporting 65 FP16 TFLOPS, 130 INT8 TOPS, and 260 INT4 TOPS - 16 GB memory capacity (320 GB/s bandwidth) - Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • AI (deep learning and machine learning) inference for computer vision, voice recognition, speech synthesis, natural language processing (NLP), machine translation, and reference systems • Real-time rendering for cloud gaming • Real-time rendering for AR and VR applications • Graphics workstations or overloaded graphics computing • GPU-accelerated databases • High-performance computing

Instance family	Feature	Scenario
gn6v, compute optimized instance family with GPU capabilities	<ul style="list-style-type: none"> • I/O-optimized • Support for enhanced SSDs , standard SSDs, and ultra disks • NVIDIA V100 GPU processors • CPU-to-memory ratio of 1:4 • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors • Uses NVIDIA V100 GPU computing accelerators (SXM2-based) <ul style="list-style-type: none"> - New NVIDIA Turing architecture - 16 GB of HBM2 memory (900 GB/s bandwidth) - 5,120 CUDA Cores - 640 Tensor Cores - Up to six NVLink connections for a total bandwidth of 300 GB/s with 25 GB/s each • Higher network performance along with more computing capacity 	<ul style="list-style-type: none"> • Deep learning applications such as training and inference applications of AI algorithms used in image classification, autonomous vehicles, and speech recognition • Scientific computing applications such as fluid dynamics, finance, genomics, and environmental analysis

Instance family	Feature	Scenario
sccgn6p, compute optimized SCC instance family with GPU capabilities	<ul style="list-style-type: none"> • I/O-optimized • CPU-to-memory ratio of 1:8 • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors with consistent computing performance • With all features of ECS Bare Metal Instances • Storage: <ul style="list-style-type: none"> - Support for standard SSDs , ultra disks, and enhanced SSDs that deliver millions of IOPS - Supports high-performance Cloud Parallel File System (CPFS) - Supports NAS • Networking: <ul style="list-style-type: none"> - Supports VPC networks equipped with two 25 Gbps ports - Supports 100 Gbit/s InfiniBand (IB) networks for low-latency RDMA communication • Uses NVIDIA V100 GPU computing accelerators (SXM2-based) <ul style="list-style-type: none"> - New NVIDIA Turing architecture - 32 GB HBM2 memory - 5,120 CUDA Cores - 640 Tensor Cores - Memory bandwidth of 900 GB/s - Up to six NVLink connections and total bandwidth of 300 GB/s (25 GB/s per connection) 	<ul style="list-style-type: none"> • Ultra-large-scale training for machine learning on a distributed GPU cluster • Large-scale high performance scientific computing and simulations • Large-scale data analysis , batch computing, and video encoding

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

Instance family	Feature	Scenario
n1, entry-level computing instance family	<ul style="list-style-type: none"> • CPU-to-memory ratio of 1:2 • 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors • Higher network performance along with more computing capacity • I/O-optimized • Support for standard SSDs and ultra disks 	<ul style="list-style-type: none"> • Small and medium-sized web servers • Batch processing • Distributed analysis • Advertising services
n2, entry-level general purpose instance family	<ul style="list-style-type: none"> • CPU-to-memory ratio of 1:4 • 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors • Higher network performance along with more computing capacity • I/O-optimized • Support for standard SSDs and ultra disks 	<ul style="list-style-type: none"> • Medium-sized web servers • Batch processing • Distributed analysis • Advertising services • Hadoop clusters
e3, entry-level memory optimized instance family	<ul style="list-style-type: none"> • CPU-to-memory ratio of 1:8 • 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors • Higher network performance along with more computing capacity • I/O-optimized • Support for standard SSDs and ultra disks 	<ul style="list-style-type: none"> • Cache or Redis • Search applications • In-memory databases • Databases with high I/O requirements, such as Oracle and MongoDB • Hadoop clusters • Large-volume data processing

Instance family	Feature	Scenario
c1, generation I instance family	<ul style="list-style-type: none"> Intel Xeon E5-2420 processors with a minimum operating frequency of 1.9 GHz. Latest DDR3 memory. I/O optimized and non-I/O optimized instances are available. I/O optimized instances support standard SSDs and ultra disks. Non-I/O optimized instances only support basic disks. 	Instance types of Generation I are legacy shared instance types. They are still categorized by the number of cores (1, 2, 4, 8, and 16 cores) and are not sensitive to instance families.
c2, generation I instance family	<ul style="list-style-type: none"> Intel Xeon E5-2420 processors with a minimum operating frequency of 1.9 GHz. Latest DDR3 memory. I/O optimized and non-I/O optimized instances are available. I/O optimized instances support standard SSDs and ultra disks. Non-I/O optimized instances only support basic disks. 	Instance types of Generation I are legacy shared instance types. They are still categorized by the number of cores (1, 2, 4, 8, and 16 cores) and are not sensitive to instance families.
m1, generation I instance family	<ul style="list-style-type: none"> Intel Xeon E5-2420 processors with a minimum operating frequency of 1.9 GHz. Latest DDR3 memory. I/O optimized and non-I/O optimized instances are available. I/O optimized instances support standard SSDs and ultra disks. Non-I/O optimized instances only support basic disks. 	Instance types of Generation I are legacy shared instance types. They are still categorized by the number of cores (1, 2, 4, 8, and 16 cores) and are not sensitive to instance families.

Instance family	Feature	Scenario
m2, generation I instance family	<ul style="list-style-type: none"> Intel Xeon E5-2420 processors with a minimum operating frequency of 1.9 GHz. Latest DDR3 memory. I/O optimized and non-I/O optimized instances are available. I/O optimized instances support standard SSDs and ultra disks. Non-I/O optimized instances only support basic disks. 	Instance types of Generation I are legacy shared instance types. They are still categorized by the number of cores (1, 2, 4, 8, and 16 cores) and are not sensitive to instance families.
s1, generation I instance family	<ul style="list-style-type: none"> Intel Xeon E5-2420 processors with a minimum operating frequency of 1.9 GHz Latest DDR3 memory Non-I/O optimized Support for basic disks only 	Instance types of Generation I are legacy shared instance types. They are still categorized by the number of cores (1, 2, 4, 8, and 16 cores) and are not sensitive to instance families.
s2, generation I instance family	<ul style="list-style-type: none"> Intel Xeon E5-2420 processors with a minimum operating frequency of 1.9 GHz. Latest DDR3 memory. I/O optimized and non-I/O optimized instances are available. I/O optimized instances support standard SSDs and ultra disks. Non-I/O optimized instances only support basic disks. 	Instance types of Generation I are legacy shared instance types. They are still categorized by the number of cores (1, 2, 4, 8, and 16 cores) and are not sensitive to instance families.

Instance family	Feature	Scenario
s3, generation I instance family	<ul style="list-style-type: none"> Intel Xeon E5-2420 processors with a minimum operating frequency of 1.9 GHz. Latest DDR3 memory. I/O optimized and non-I/O optimized instances are available. I/O optimized instances support standard SSDs and ultra disks. Non-I/O optimized instances only support basic disks. 	Instance types of Generation I are legacy shared instance types. They are still categorized by the number of cores (1, 2, 4, 8, and 16 cores) and are not sensitive to instance families.
t1, generation I instance family	<ul style="list-style-type: none"> Intel Xeon E5-2420 processors with a minimum operating frequency of 1.9 GHz Latest DDR3 memory Non-I/O optimized Support for basic disks only 	Instance types of Generation I are legacy shared instance types. They are still categorized by the number of cores (1, 2, 4, 8, and 16 cores) and are not sensitive to instance families.

2.4.1.3 Instance type

An ECS instance is the smallest unit that can provide compute capabilities and services for your business. The compute capabilities vary with instance type.

The ECS instance type defines the basic properties of an ECS instance: CPU (including CPU model and clock speed) and memory. When you create an instance, you must also configure the disk type, image, and network type in addition to the instance type. The following table describes all instance families and their types.

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
n4	ecs.n4.small	1	2.0	None	0.5	50	1	1

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	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	500	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
	ecs.e4.xlarge	4	32.0	None	0.8	150	1	2

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	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
	ecs.g5.2xlarge	8	32.0	None	2.5	800	2	4

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	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
	ecs.sn2ne.8xlarge	32	128.0	None	6.0	2,500	8	8

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	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
	ecs.se1.xlarge	4	32.0	None	0.8	200	1	3
	ecs.se1.2xlarge	8	64.0	None	1.5	400	1	4

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	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
d1	ecs.d1.2xlarge	8	32.0	4 × 5,500	3.0	300	1	4
	ecs.d1.3xlarge	12	48.0	6 × 5,500	4.0	400	1	6
	ecs.d1.4xlarge	16	64.0	8 × 5,500	6.0	600	2	8
	ecs.d1.6xlarge	24	96.0	12 × 5,500	8.0	800	2	8

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	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 × 7500	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
	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

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.d2-gab.8xlarge	32	128.0	12 x 1150	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
	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

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including 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
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

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including 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
	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

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPUs
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
	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

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPUs
	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
	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

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPUs
gn5e	ecs.gn5e-c11g1.3xlarge	100	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	150	8 × NVIDIA P4
gn6i	ecs.gn6i-c10g1.2xlarge	100	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

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPUs
	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
	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

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPUs
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
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

Instance family	Instance type	vCPUs	Memory (GiB)
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

2.4.1.4 UserData

UserData allows you to customize the startup behavior of instances and import data to ECS instances. It is the basis for ECS instance customization.

UserData is implemented through different types of scripts. Before UserData is implemented on an instance, all ECS instances will have the same initial environment and configurations when started for the first time. After enterprises or individuals enter valid UserData information based on their scenarios and needs, required ECS instances are provided after the first startup.

Methods

- **UserData-Scripts:** are applicable to users who need to initialize instances by executing the shell scripts. The UserData-Scripts begin with `#!/bin/sh`. A review of user data shows that most users input UserData by running UserData-Scripts. UserData-Scripts are also suitable for complicated deployment scenarios.

- **Cloud-Config:** is a special script supported by cloud-init. It packs frequently-used personalized configurations into YAML files, which enable you to complete the frequently-used configurations more conveniently. The script starts with `# Cloud-config` in the first line and is followed by an array containing `ssh_authorized_keys`, `hostname`, `write_files`, and `manage_etc_hosts`.

Scenarios

- SSH authentication
- Software source updates and configuration
- DNS configuration
- Application installation and configuration

2.4.1.5 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 2-2: Lifecycle description](#).

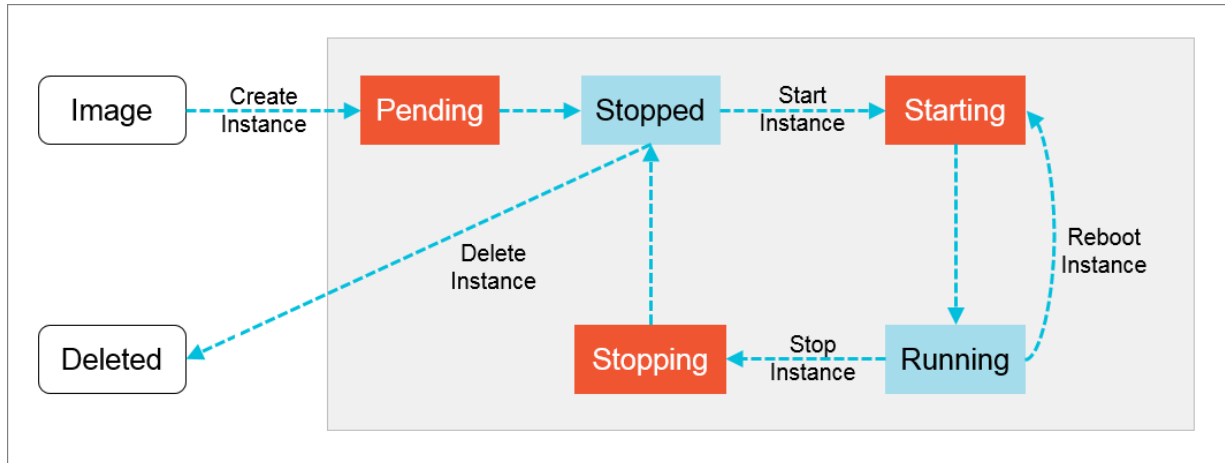
Table 2-2: Lifecycle description

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

Status	Status attribute	Description	Corresponding API status
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
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 2-2: 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



2.4.1.6 EBM Instances

ECS Bare Metal (EBM) Instance is a new computing service that combines the elasticity of virtual machines with the performance and features of physical machines. EBM Instances are designed based on the state-of-the-art virtualization technology developed by Alibaba Cloud.

The virtualization used by EBM Instances is optimized to support common ECS instances and nested virtualization, maintaining elastic performance with the user experience of physical servers.

Benefits

EBM Instances provides the following benefits through technological innovation:

- **Exclusive computing resources**

As a cloud-based elastic computing service, EBM Instances surpass the performance and isolation of physical servers, enabling you to exclusively occupy computing resources without virtualization performance overhead or feature loss. EBM Instances support ultrahigh-frequency instances and can contain 8, 16, 32, or 96 CPU cores. An EBM Instance with eight CPU cores supports ultrahigh frequency processing from 3.7 to 4.1 GHz, providing better performance and response for gaming and finance businesses than peer services.

- **Encrypted computing**

For security, EBM Instances use a chip-level trusted execution environment (Intel® SGX) in addition to physical server isolation to ensure that encrypted data can only be computed within a secure and trusted environment. This chip-level hardware security protection provides a safe box for the data of cloud users and allows users to control all data encryption and key protection processes.

- **Any Stack on Alibaba Cloud**

An EBM Instance combines the performance strengths and complete features of physical machines and the ease-of-use and cost-effectiveness of cloud servers. It can effectively meet the demands of high-performance computing and help you build new hybrid clouds. Thanks to the flexibility, elasticity, and all the other strengths inherited from both virtual and physical machines, EBM Instances are endowed with re-virtualization ability. Offline private clouds can be seamlessly migrated to Alibaba Cloud without the performance overhead that may arise from nested virtualization, giving you a new approach for moving businesses onto the cloud.

- **Heterogeneous instruction set processor support**

The virtualization 2.0 technology used by EBM Instances is developed independently by Alibaba Cloud and supports ARM and other instruction set processors at no additional cost.

Configuration features

The following table lists the configuration features of EBM Instances.

Table 2-3: Features

Item	Description
CPU configuration	Only supports the ebmg5 general-purpose EBM Instance type family.
Memory configuration	Supports expansion from 32 GiB to 384 GiB as needed. The ratio of CPU to memory is 1:2 or 1:4 to provide better computing performance.
Storage configuration	Supports startup from virtual machine images or cloud disks, achieving delivery in several seconds.
Network configuration	Supports Virtual Private Clouds (VPCs), and interoperability with ECS, GPU, and other cloud services. Delivers performance and stability comparable to physical machine networks.

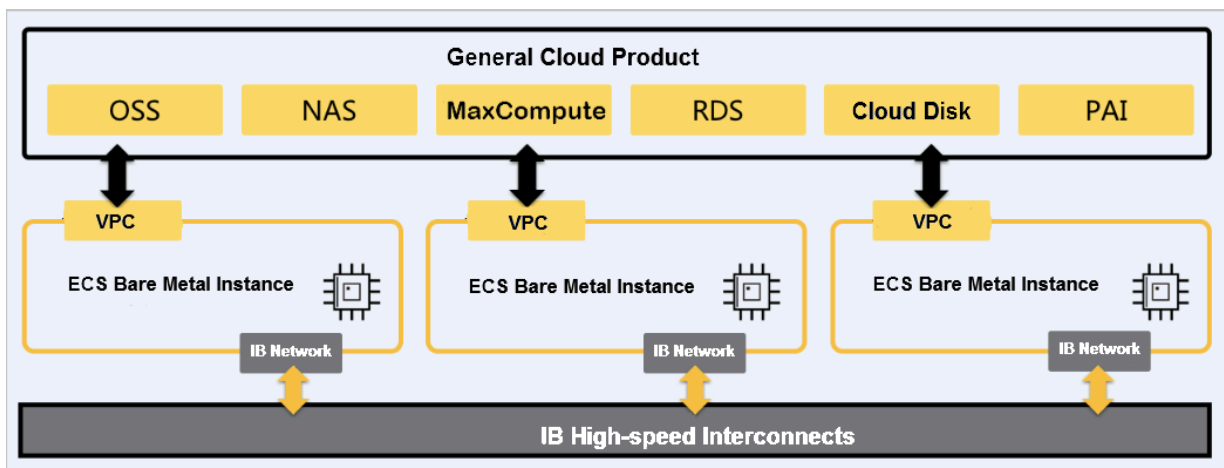
Item	Description
Image configuration	Supports ECS images.
Security configuration	Maintains the same security policies and flexibility as existing ECS instances.

2.4.1.7 Super Computing Clusters

Super Computing Clusters (SCCs) provide computing cluster services with ultimate computing performance and parallel efficiency by integrating CPUs and heterogeneous accelerators such as GPUs that are interconnected through the high-speed InfiniBand (IB) network. SCCs are suited for scenarios such as high-performance computing, artificial intelligence, machine learning, scientific and engineering computing, data analysis, and audio and video processing.

SCC architecture

The following figure shows the SCC architecture.



SCCs are based on ECS Bare Metal Instances. By integrating the high-speed interconnects of InfiniBand technology and heterogeneous accelerators such as GPUs, SCCs have the following features:

- SCCs have all the benefits of ECS Bare Metal Instances. The underlying architecture allows you to use exclusive cloud servers or physical servers to create a secure and controllable underlying environment where you can configure security groups and VPCs for your SCC instances to implement traffic control.
- SCCs adopt InfiniBand, a conversion cable technology that supports multiple concurrent connections. InfiniBand is the next-generation I/O standard for compute server platforms and features high scalability, high bandwidth, and low latency. InfiniBand is ideal for establishing communication between servers such as replication servers

and distributed servers, between servers and storage devices such as SAN and direct-attached storage, and between servers and networks such as LANs, WANs, and the Internet. The InfiniBand architecture is commonly used in high-performance computing and provides higher bandwidth, lower latency, and more reliable connections than the Ethernet architecture.

You can build your High Performance Computing (HPC) system based on SCCs.

Scenarios

Apsara Stack SCCs offer mature and flexible industry solutions and are suited for the following scenarios:

Scenario	Description
Large-scale AI computing	SCC-based HPC provides the computing capabilities that are required by large-scale AI computing for quick handling of problems associated with data and models. HPC and large-scale AI computing scenarios have similar requirements during planning, designing, and deployment phases.
Hybrid cloud for HPC	Both Apsara Stack and Alibaba Cloud public cloud have the same architecture for HPC, bringing a consistent hybrid cloud experience. You can use the E-HPC service provided by the Alibaba Cloud public cloud to migrate computing workloads from Apsara Stack to the public cloud. Before migrating your workload, you must activate resources on the public cloud and schedule tasks based on the E-HPC scheduler to build a hybrid cloud for HPC.
Supercomputing center in the cloud	<p>Supercomputing centers are the earliest and most mature IT service developed in China. Traditional supercomputing centers integrate technologies from multiple vendors, which requires lengthy and complex technical solution consultation, equipment selection, and verification. The construction cycle is long, and operations and maintenance are complex.</p> <p>Apsara Stack supercomputing solutions can be used to build cloud supercomputing centers that provide large-scale supercomputing services in an agile and flexible manner. The all-in-one cloud for supercomputing and automated O&M reduce the complexity of operations and maintenance.</p>

Scenario	Description
Industry verticals	<p>The application of HPC in traditional industries is mature and continues to evolve. Apsara Stack SCCs are suited for all the industries where traditional supercomputing is applied. The industries include the following:</p> <ul style="list-style-type: none"> • Petrochemical: seismic data processing and reservoir simulation • Finance: financial derivative analysis, actuarial analysis, asset and liability management, and investment risk analysis • Industrial manufacturing: collision analysis, failure analysis, and thermodynamic analysis • Life science: drug discovery, protein folding, DNA sequencing, and medical imaging • Media and entertainment: video post-production and animation rendering • Government and higher education: weather forecast, high-energy physics, and geophysics

2.4.2 Block storage

2.4.2.1 Overview

This topic describes the diverse types of block storage. This includes elastic block storage services based on a distributed storage architecture and local storage services based on the local hard disks of physical machines.

Description of elastic block storage and local storage:

- [Elastic block storage](#) provides ECS with persistent and highly reliable random block-level data storage with low latency. Data is stored in triplicate over a distributed system to ensure data reliability. Elastic block storage can be created, released, and expanded at any time.
- [Local storage](#), also known as local disks, refers to temporary disks mounted on the physical machine where an ECS instance resides. Local storage is designed for business scenarios that require high storage I/O performance. Local storage provides block-level data access for instances with low latency, high random IOPS, and high throughput.

Block storage, OSS, and NAS

Currently, Apsara Stack provides three types of data storage services: block storage, Network Attached Storage (NAS), and Object Storage Service (OSS).

Table 2-4: Storage service comparison

Type	Description	Scenario
Block storage	A high-performance and low-latency block-level storage device provided by Apsara Stack for ECS instances. It supports random reads and writes. You can format the block storage and create a file system on it as you would with a hard disk.	Block storage can be used for data storage in most common business scenarios .
OSS	A storage space suited for high volume unstructured data such images, video, audio, and other data generated online. Data stored in OSS can be accessed anytime and anywhere through APIs.	OSS is suited for business scenarios such as website construction, separation between dynamic and static resources, and CDN acceleration.
NAS	NAS is a storage system similar to OSS, and as such is also suited for storing high volume unstructured data. Data must be accessed through standard file access protocols, such as the Network File System (NFS) protocol for Linux systems and the Common Internet File System (CIFS) protocol for Windows systems. You can set permissions to allow different clients to access the same file at the same time.	NAS is suited for business scenarios such as file sharing across departments , non-linear editing of radio and television data, high-performance computing, and Docker.

2.4.2.2 Elastic block storage

2.4.2.2.1 Overview

Elastic block storage can be divide into the following types based on whether it can be attached to multiple ECS instances.

- **Cloud disks:** A cloud disk can be attached to a single ECS instance that resides in the same zone and region.
- **Shared block storage:** A shared block storage can be attached to up to four ECS instances that belong to the same zone and region.

2.4.2.2.2 Cloud disks

Cloud disks can be classified in either of the following ways:

- **Performance-based classification**

Cloud disks are divided by performance into basic cloud disks, ultra cloud disks, and SSD cloud disks.

- Basic cloud disks are ideal for the minimally I/O-intensive scenarios, and only provide several hundreds of IOPS for ECS instances.
- Ultra cloud disks are ideal for medium I/O load scenarios and provide a storage performance of up to 3,000 random IOPS for ECS instances.
- SSD cloud disks are ideal for I/O-intensive scenarios and provide stable and high random IOPS performance.

- **Function-based classification**

Cloud disks can be divided by their functions into system disks and data disks.

- System disks: has a lifecycle that is tied to the ECS instance to which it is attached. A system disk is created and released along with the instance. Shared access is not allowed.
- Data disks: can be created separately or together with ECS instances. Shared access is not allowed. A data disk created with an ECS instance has a lifecycle that is tied to that of the instance, and is created and released along with the instance. Data disks that are created independently can be released independently or in conjunction with the ECS instance to which it is attached. The capacity of a data disk is determined by its type.

2.4.2.2.3 Shared block storage

Shared block storage is a block-level data storage service that supports concurrent read/write operations to multiple ECS instances, giving it a high level of performance and reliability.

A single shared block storage device can be attached to a maximum of four ECS instances. Shared block storage can only be used as data disks and must be created individually.

Shared access is allowed. You can configure a shared block storage device to be released when its associated ECS instance is released.

Shared block storage can be divided into the following types based on performance:

- **SSD shared block storage:** adopts an SSD storage medium to provide stable and high-performance storage with enhanced random I/O and data reliability.
- **Ultra shared block storage:** adopts a hybrid SSD and HDD storage medium.

When used as data disks, shared block storage allows up to 16 data disks to be attached to each ECS instance.

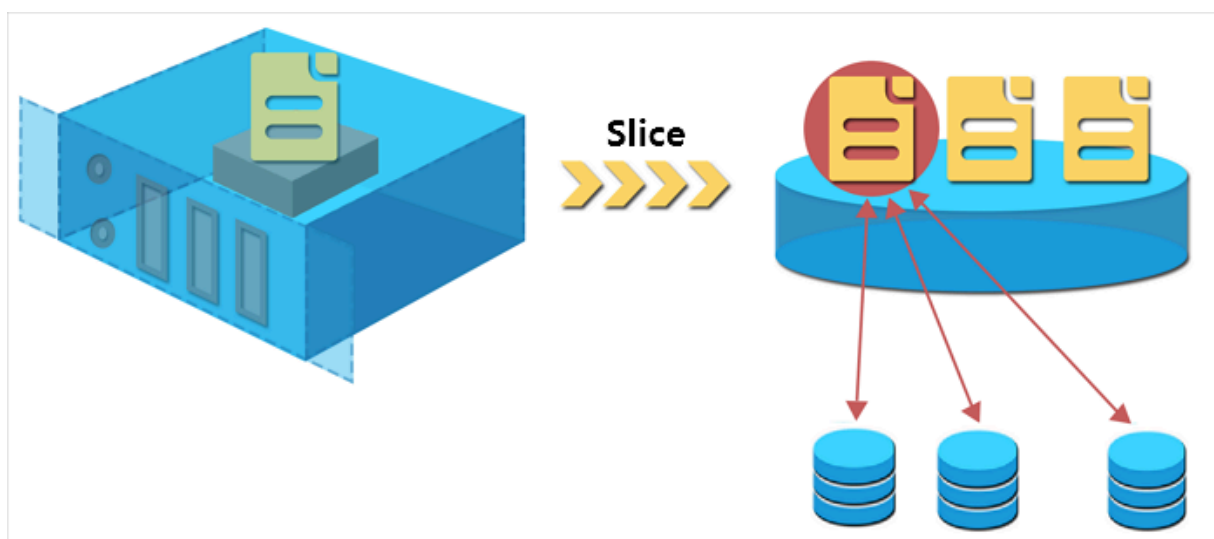
2.4.2.2.4 Triplicate storage

Apsara Distributed File System provides stable, efficient, and reliable data access to ECS instances.

Chunks

When ECS users perform read and write operations on virtual disks, the operations are translated into the corresponding processes on the files stored in Apsara Stack data storage system. Apsara Stack uses a flat design in which a linear address space is divided into slices called chunks. Each chunk is replicated into three copies. Each copy is stored on a different node in the cluster, which ensures data reliability.

Figure 2-3: Triplicate backup



How triplicate technology works

Triplicate storage is made up of three components: master, chunk server, and client. Each write operation performed by an ECS user is converted into an operation executed by the client. The execution process is as follows:

1. The client determines the location of a chunk corresponding to the write operation.
2. The client sends a request to the master to query the chunk servers where the three chunk replicas are each stored.
3. The client sends write requests to the chunk servers based on the results returned from the master.
4. If the three replicas of the chunk are all successfully written as requested, the client returns a message to indicate the success of the operation. If the write operation fails, a failure message is returned.

The master component distributes chunks based on the disk usage, rack distribution, power supply, and machine workloads of chunk servers. This ensure that chunk replicas are each distributed to chunk servers on different racks and that data does not become unavailable due to the failure of a single server or rack.

Data protection mechanism

When a data node is damaged or disk faults occur on a data node, the total number of valid replicas of some chunks in a cluster becomes less than three. In these cases, the master replicates data between chunk servers to ensure that there are always three valid replicas of chunks in the cluster.

Figure 2-4: Automatic replication



All user-level operations for data on cloud disks are synchronized across the three chunk replicas at the underlying layer. Operations that are synchronized include adding,

modifying, and deleting data. This mode ensures the reliability and consistency of user data.

To prevent data losses caused by viruses, accidental deletion, or malicious attacks, we recommend that you use other protection methods such as backing up data and taking snapshots in addition to triplicate storage. Implement all appropriate measures to ensure the security and availability of your data.

2.4.2.3 ECS disk encryption

ECS disk encryption is a simple and secure encryption method that can be used to encrypt new cloud disks.

With ECS disk encryption, there is no need to create or maintain your own key management infrastructure, change existing applications and maintenance procedures, or add additional encryption operations. Disk encryption does not have any negative impact on your business processes. After an encrypted ECS disk is created and attached to an ECS instance, the following types of data can be encrypted:

- Data on the cloud disk.
- Data transmitted between the cloud disk and instance. Data in the instance operating system is not encrypted again.
- All snapshots created from the encrypted cloud disk. These snapshots are called encrypted snapshots.

The data transmitted from the ECS instance to the cloud disk is encrypted on the host where the ECS instance resides.

Disk encryption is supported on all available cloud disks (basic cloud disks, ultra cloud disks, and SSD cloud disks) and shared block storage (ultra and SSD) for all Apsara Stack instances.

2.4.2.4 Local storage

Local storage, also known as local disks, are disks that reside on the same physical machines as their ECS instances. Local disks provide temporary block storage for instances and are designed for scenarios that require extremely high I/O performance.

Local storage provides block-level data access for instances with high random IOPS, high throughput, and low latency. The reliability of data stored in local disks depends on the reliability of the physical server to which the disks are attached. This is a single

point of failure risk which may cause data loss. We recommend that you implement data redundancy at the application layer to ensure the availability of the data.

**Note:**

Storing data on local disks poses a risk for data persistence, such as when the host server is down. We recommend that you do not use local disks to store data for long periods of time. If no data reliability architecture is available for your applications, we recommend that you use disks or shared block storage for your ECS instances.

Local disk types

Apsara Stack provides two types of local disks:

- NVMe SSDs: are used together with gn5 and ga1 instance families.
- SATA HDDs: are used together with d1ne and d1 instance families. This type of local disks is suitable for customers from Internet, finance, and other industries that require large storage capacity with storage analysis and offline computing. SATA HDDs satisfy the performance, capacity, and bandwidth requirements of distributed computing models such as Hadoop.

2.4.3 Images

An image is a template for running environments in one or more ECS instances, and generally includes an operating system and preinstalled software.

An image works as a copy of all of the data stored on one or more disks. These disks can be a single system disk, or a combination of system disk and data disks. You can use an image to create an ECS instance or change the system disk of an ECS instance.

Image types

ECS provides a variety of image types to easily access image resources.

Table 2-5: Image description

Type	Description
Public image	Public images officially provided by Apsara Stack support Windows and most popular versions of Linux operating systems, including: <ul style="list-style-type: none">• Windows• CentOS• CoreOS• Debian• Gentoo• FreeBSD• OpenSUSE• SUSE Linux• Ubuntu
Custom image	Custom images created based on your existing physical servers , virtual machines, or cloud hosts. This image type is flexible enough to meet all of your specific business needs.

Obtain an image

ECS allows you to obtain images through the following methods:

- Create a custom image based on an existing ECS instance.
- Choose an image shared by another Apsara Stack tenant account.
- Import an offline image file to an ECS cluster to generate a custom image.
- Copy a custom image to another region to achieve consistent environment or application deployment across regions.

Image formats

Currently, ECS supports VHD and RAW images. Images in other formats must be converted to VHD or RAW images before they can be run in ECS. For more information about format conversion, see **Convert image format** in ECS User Guide.

2.4.4 Snapshots

2.4.4.1 Overview

A snapshot is a copy of data on a disk created at a specific point in time. When using a disk, you may encounter the following scenarios:

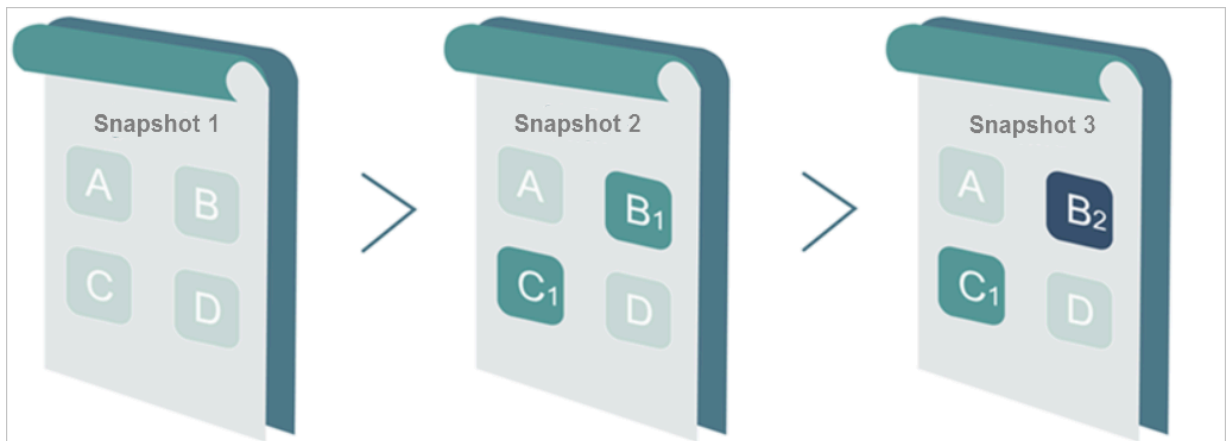
- When writing or storing data to a disk, you may want to use snapshot data from another disk as the basis for the target disk.
- While cloud disks represent a secure way to store data, their data may be subject to errors caused by application errors or malicious read and write operations, and require additional safeguard mechanisms. For this reason, you may want to use snapshots to restore data to a previous point in time in case of data errors.

2.4.4.2 Mechanisms

This topic describes snapshots. Snapshots retain a copy of data stored on a disk at a certain point in time. You can schedule disk snapshots to be created periodically to ensure continuous operation of your business.

Snapshots are created incrementally such that only data changes between two snapshots are copied instead of all of the data, as shown in [Figure 2-5: Snapshots](#).

Figure 2-5: Snapshots



Snapshot 1, Snapshot 2, and Snapshot 3 are the first, second, and third snapshots of a disk. When a snapshot is created, the file system checks each block of data stored on the disk, and only copies the blocks of data that differ from those on the previous snapshots. The changes between snapshots in the preceding figure are described as follows:

- All data on the disk is copied to Snapshot 1 because it is the first disk snapshot.

- The changed blocks B1 and C1 are copied to Snapshot 2. Blocks A and D are referenced from Snapshot 1.
- The changed block B2 is copied to Snapshot 3. Blocks A and D are referenced from Snapshot 1, and block C1 is referenced from Snapshot 2.
- When the disk needs to be restored to the status of Snapshot 3, snapshot rollback will copy blocks A, B2, C1, and D to the disk, which will be restored to the status at the time of Snapshot 3.
- If Snapshot 2 is deleted, block B1 in the snapshot is deleted, but block C1 is retained because it is referenced by other snapshots. When you roll back a disk to Snapshot 3, block C1 is recovered.

**Note:**

Snapshots are stored on the Object Storage Service (OSS), but are hidden from users. Snapshots do not consume bucket space in OSS. Snapshot operations can only be performed from the ECS console or through APIs.

2.4.4.3 Specifications of ECS Snapshot 2.0

Built on the features of the original snapshot function, ECS Snapshot 2.0 data backup service provides a higher snapshot quota and a more flexible automatic task policy, further reducing its impact on business I/O.

Table 2-6: Comparison of snapshot specifications

Item	Traditional snapshot specifications	Snapshot 2.0 specifications	User benefits	Example
Snapshot quota	The snapshot quota is limited to: the number of disks \times 6 + 6.	Each disk can have up to 64 snapshots.	Longer protection cycle, and smaller protection granularity.	<ul style="list-style-type: none"> • A snapshot is created for the data disks of non-core business at 00:00 every day. Snapshots for the last two months are retained. • A snapshot is created for the data disks of core business every four hours. Snapshots for the last ten days are retained.

Item	Traditional snapshot specifications	Snapshot 2.0 specifications	User benefits	Example
Automatic task policy	By default, the task is scheduled to be triggered once a day and cannot be modified manually.	You can customize the time of day and days of the week that snapshots are scheduled to be created and the retention period of snapshots. The disk quantity and related details associated with an automatic snapshot policy can be queried.	More flexible protection policy	<ul style="list-style-type: none"> You can schedule snapshots to be created on the hour several times in a single day. You can choose which days of the week to recur snapshot creation. You can specify the snapshot retention period or choose to retain it permanently. When the number of automatic snapshots reaches the upper limit, the oldest automatic snapshot will be automatically deleted.
Implementation	Copy-on-write (COW)	Redirect-on-write (ROW)	Mitigates the impact of snapshot tasks on I/O write performance.	Snapshots can be taken at any time without interruptions to your business.

2.4.4.4 Technical comparison

Alibaba Cloud ECS Snapshot 2.0 has many advantages over the snapshot feature of traditional storage products.

Table 2-7: Comparison of technical advantages

Item	ECS Snapshot 2.0	Traditional snapshot
Capacity	Unlimited capacity, meeting the data protection needs of extra-large businesses.	Capacity limited by the initial storage device capacity, merely meeting the data protection needs for a few core services.

2.4.5 Deployment sets

A deployment set is a tool that allows you to view the physical topology of hosts, racks, and switches and select a deployment policy that best suits the reliability and performance requirements of your business.

There may be increased reliability or performance requirements when you use multiple ECS instances in the same zone.

- **Improve business reliability**

To avoid the impacts caused by the failure of physical hosts, racks, or Switches, multiple copies of application instances must be distributed across different physical hosts, racks, or Switches.

- **Improve network performance**

For scenarios that involve frequent network interactions between instances, lower latency and higher bandwidth can be achieved by aggregating corresponding instances onto a single Switch.

Deployment granularities and policies

- **Deployment granularities**

- Host: indicates physical-server-level scheduling.
- Rack: indicates rack-level scheduling.
- Switch: indicates Switch-level scheduling.

- **Deployment policies**

- LooseAggregation
- StrictAggregation
- LooseDispersion
- StrictDispersion

LooseAggregation and StrictAggregation are intended for higher performance, while LooseDispersion and StrictDispersion are intended for higher reliability.

[Table 2-8: Granularities and policies](#) lists the deployment policies and business scenarios corresponding to each deployment granularity.

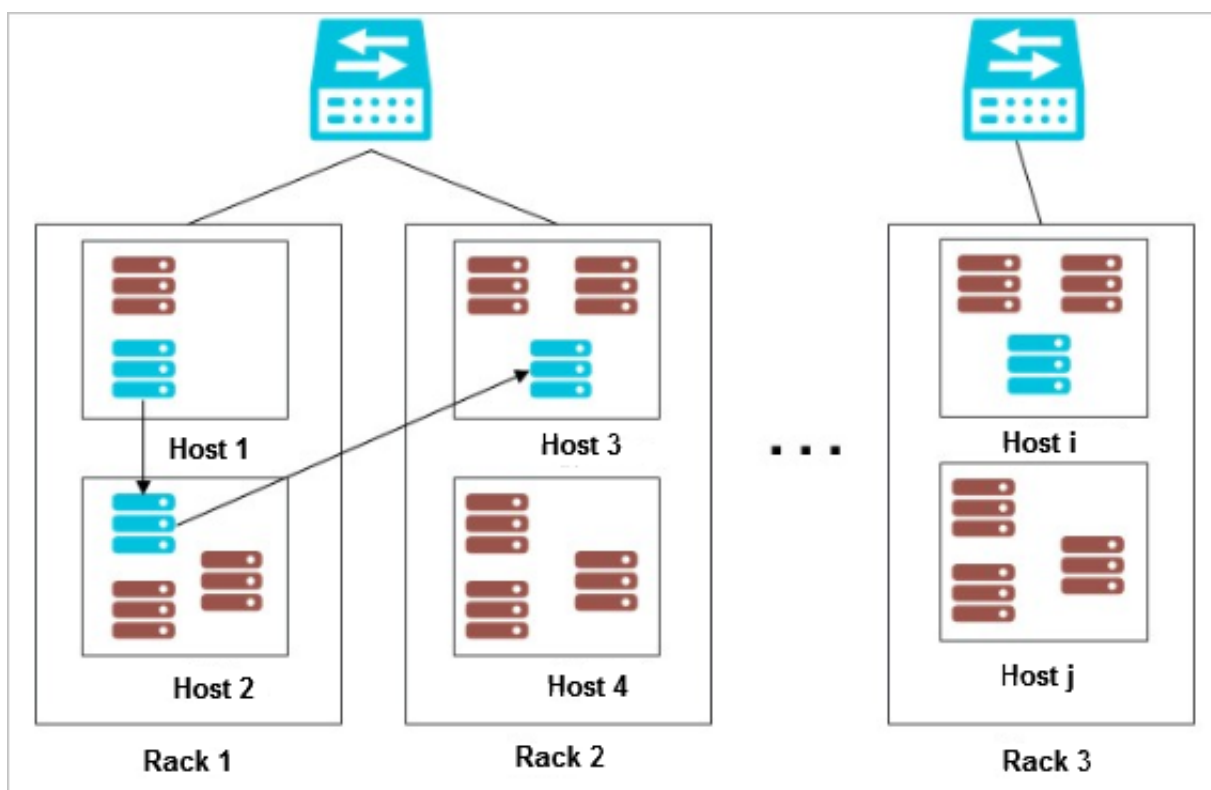
Table 2-8: Granularities and policies

Deployment granularity	Deployment policy	Business scenario
Host	StrictDispersion	General purposes
	LooseDispersion	
Rack	StrictDispersion	Big data and databases
	LooseDispersion	Game customers
Switch	StrictDispersion	VPN
	LooseDispersion	Game customers
	StrictAggregation	Big data and databases
	LooseAggregation	Game customers

Typical examples

The following figure shows a typical case where business reliability is improved by using deployment sets. Three ECS instances of a tenant are distributed on three different physical hosts, which are distributed on at least two different racks.

Figure 2-6: Typical example



Note:

For more information about the deployment set APIs, see **Deployment sets** in ECS Developer Guide .

2.4.6 Network and security

2.4.6.1 IP addresses of ECS instances of VPC type

This topic describes the IP address types supported by ECS instances and the corresponding scenarios.

IP address types

ECS instances have the following IP address types:

- **Private IP addresses**

When you create an ECS instance, a private IP address is assigned based on the VPC and the CIDR block of the VSwitch to which the instance belongs.

- **Elastic IP (EIP)**

An EIP is a public IP address. You can apply for an EIP as necessary.

Scenarios

- **Private IP:** A private IP address is used to access the intranet. When creating an instance, you can directly configure the private IP address.

**Note:**

If the private IP address is not configured, the system automatically allocates a private IP address for the instance.

- **EIP:** An EIP is used to access the Internet. You can separately bind an EIP to an instance after it has been created. For more information, see **EIP** in VPC User Guide . EIPs can be applied for and retained long-term. You can bind and unbind an EIP to and from an instance, delete the EIP, or modify its bandwidth.

2.4.6.2 Elastic network interfaces

This topic describes Elastic Network Interfaces (ENIs) and their application scenarios.

An ENI is a virtual network card that can be attached to an ECS instance on a VPC network. ENIs help you implement high-availability clusters, low-cost failover, and refined network management. ENIs are supported in all regions.

ENIs can be used in the following scenarios:

- **High-availability clusters**

An ENI can meet the demands for multiple NICs on a single instance in a high-availability architecture.

- **Cost-effective failover**

You can detach an ENI from a failed ECS instance and reattach it to another instance to quickly redirect traffic intended for a failed instance to a backup instance and immediately recover service.

- **Lean network management**

You can configure multiple ENIs for an instance. For example, you can use some ENIs for internal management and others for Internet business access, so as to isolate management data from business data. You can also configure precisely-targeted security group rules for each ENI based on the source IP address, protocols, and ports, so as to achieve traffic control.

ENI types

ENIs are classified into two types:

- **Primary ENI**

The ENI created by default upon the creation of an instance in a VPC is called the **primary ENI**. The lifecycle of the primary ENI is tied to that of the instance, and the primary ENI cannot be removed from the instance.

- **Secondary ENI**

You can create a secondary ENI and attach it to or detach it from the instance. The maximum number of ENIs that can be attached to a single instance varies with the instance type. For more information, see [Instance families](#).

ENI attributes

The following table displays the attributes of an ENI.

Table 2-9: Attribute description

Attribute	Quantity
Primary private IP addresses	1
MAC addresses	1
Security groups	1 to 5
Description	1
ENI name	1

Limits

ENIs have the following limits:

- A single account can own up to 100 ENIs in a single region.

- The ECS instance must belong to the same zone and same region as the ENI, but does not have to use the same VSwitch.
- For instance types that support ENI attaching and the number of ENIs supported by each instance type, see [Instance type](#).
- Attaching multiple ENIs does not increase the instance bandwidth.

**Note:**

The instance bandwidth varies according to the instance type.

2.4.6.3 Intranet

ECS instances communicate through the intranet. Non-I/O-optimized instances share 1 GiB of bandwidth and I/O-optimized instances share 10 GiB of bandwidth. The intranet is a shared network, so the bandwidth may fluctuate.

**Note:**

Currently, most mainstream instances are I/O-optimized instances, and the actual bandwidth is related to the physical hardware.

If you need to transmit data between two ECS instances in the same region, we recommend that you transmit data through an intranet connection. Intranet connections can also be implemented between ApsaraDB for RDS, SLB, and OSS services. Intranets can share up to 1 GiB of bandwidth.

ECS can communicate with RDS, SLB, and OSS in the same region through the intranet.

The following rules apply to **VPC**-type ECS instances in the intranet:

- Intranet communication is permitted by default for instances in the same security group of the same account in the same region. If instances with the same account in the same region are in different security groups, intranet communication can be implemented by authorizing mutual access between the two security groups.
- For instances that belong to the same account and same region but do not belong to the same VPC network, you can use Express Connect to implement their intranet communication.
- The intranet IP address of an instance can be modified or changed as needed.
- Virtual IP (VIP) addresses cannot be configured as the intranet or Internet addresses of instances.

- Instances of different network types cannot communicate with each other over an intranet.

2.4.6.4 Security group rules

Security group rules permit or deny Internet or intranet traffic to or from the ECS instances associated with the security group.

You can add or delete security group rules at any time. Changes in security group rules are automatically applied to ECS instances associated with the security group.

Be sure to configure concise security group rules. If you associate an instance with multiple security groups, hundreds of rules may apply to the instance. This may cause connection errors when you access the instance.

2.5 Scenarios

ECS can be used either independently as a simple Web server or with other Apsara Stack services such as OSS, to provide advanced flexible solutions. ECS is typically applied in the following scenarios:

Official websites for enterprises and simple Web applications

Initially, official websites for enterprises do not have high volumes of traffic and only require low-configuration ECS instances to run applications, databases, and store files. As your website develops, you can increase the ECS specifications and the number of ECS instances at any time without the need to worry about low resources during traffic spikes.

Multimedia and high-traffic applications or websites

When ECS is used with OSS, static images, videos, and downloaded packages can be stored on OSS to reduce storage costs. In addition, ECS can be used with Server Load Balancer (SLB) to greatly shorten user response time, reduce bandwidth fees, and improve availability.

Applications or websites with large traffic fluctuations

Some applications and websites may encounter large fluctuations in traffic within a short period. ECS also features elastic processing capabilities. The number of ECS instances automatically increases or decreases in response to spikes and dips in traffic for the purpose of meeting resource requirements and preserving cost efficiency. ECS can be used with SLB to achieve a high availability architecture.

Databases

Databases with high I/O requirements are supported. A high-configuration I/O-optimized ECS instance can be used with an SSD cloud disk to support high I/O concurrency with higher data reliability. Alternatively, multiple lower-configuration I/O-optimized ECS instances can be used with SLB to achieve a high availability architecture.

2.6 Limits

The limits of ECS are as follows:

- 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 can use a maximum of 4 cores in its CPU.
- Windows operating systems support a maximum of 64 vCPUs in their instance specifications.
- Virtualization software installation and subsequent virtualization such as VMware are not supported.
- Currently, 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.

The following table lists additional limits to ECS.

Table 2-10: Other limits

Type	Description
Instance type	For more information, see Instance families and Instance type .

Type	Description
Block storage	<p>Specification limits</p> <ul style="list-style-type: none"> • Number of system disks per instance: 1. • Number of data disks per instance: 16. • Default quota of instances to which one shared block storage can be attached: 4. • System disk capacity: 40-500 GiB. • Capacity of one basic cloud disk: 5-2,000 GiB. • Capacity of one SSD disk: 20-32,768 GiB. • Capacity of one ultra cloud disk: 20-32,768 GiB. • Total capacity of one ultra block storage: 32,768 GiB. <p>Limits</p> <ul style="list-style-type: none"> • Only data disks can be encrypted. System disks cannot be encrypted. • Unencrypted disks cannot be directly converted into encrypted disks. • Encrypted disks cannot be directly converted into unencrypted disks. • Unencrypted snapshots cannot be directly converted into encrypted snapshots. • Encrypted snapshots cannot be directly converted into unencrypted snapshots. • Images with encrypted snapshots cannot be shared. • Images with encrypted snapshots cannot be exported.
Snapshot quota	The number of disks × 64.
Image	<ul style="list-style-type: none"> • Maximum number of users that a single image can be shared with: 50. • Instances with 4 GiB of memory or higher memory do not support 32-bit images.

Type	Description
Security group	<ul style="list-style-type: none"> • A single security group cannot contain more than 1,000 instances. If more than 1,000 instances need to access each other over the intranet, you can distribute them to different security groups and authorize mutual access among the security groups. • Each instance can belong to a maximum of five security groups. • Each user can have a maximum of 100 security groups. • Each security group can have a maximum of 100 security group rules. • Adjusting security groups will not affect the continuity of services. • Security groups are stateful. If a security group permits outbound traffic over a link, it also permits inbound traffic over this link.
Elastic network interface	The number of elastic network cards that can be bound to different instance type families. For details, see Instance type .
Instance user data	Currently, ECS UserData supports VPC + I/O optimized instances. You can use the UserData function when creating such instances . Because UserData depends on the cloud-init service, cloud-init must be installed in the image.

2.7 Terms

ECS

A simple and efficient cloud computing service that provides elastic processing capabilities and supports operating systems such as Linux and Windows.

instance

An independent resource entity that contains basic resource elements.

security group

A virtual firewall that provides status detection and packet filtering functions and is used to control the network access of one or more ECS instances. Instances in the same security group are able to communicate with each other, while instances in different security groups are isolated from each other. You can configure the rules of two security groups to authorize mutual access between them.

image

A running environment template for ECS instances. An image includes an operating system and preinstalled software. Images can be divided into public images and custom images. You can use an image to create an ECS instance or change the system disk of an ECS instance.

snapshot

Data backup of a disk at a certain point in time. Includes automatic snapshots and manual snapshots.

cloud disk

An independent disk that can be attached to any ECS instance in the same zone of the same region. Cloud disks are divided by performance into ultra disks, SSD disks, and basic disks.

block storage

A low-latency and high-reliability persistent random block-level data storage service provided by Apsara Stack for ECS.

throughput

The amount of data successfully transmitted through a network, device, port, virtual circuit, or another facility within a given period of time.

performance test

A world-leading SaaS performance test platform, with powerful distributed stress test capability. It can simulate real business scenarios with large amounts of users to locate all application performance problems.

Virtual Private Cloud (VPC)

A virtual private cloud built and customized based on Apsara Stack. Full logical isolation is achieved between VPCs. Users can create and manage cloud services, such as ECS instances, Intranet Server Load Balancer (SLB) instances, and RDS instances in their own VPCs.

intranet IP address

A service connection address for a client that uses a private IP address as the source.

GPU instance

A GPU-based computing service used in scenarios such as video decoding, graphics rendering, deep learning, and scientific computation. GPU instances feature real-time and high-speed computation and provide powerful concurrent and floating point computing capabilities.

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:

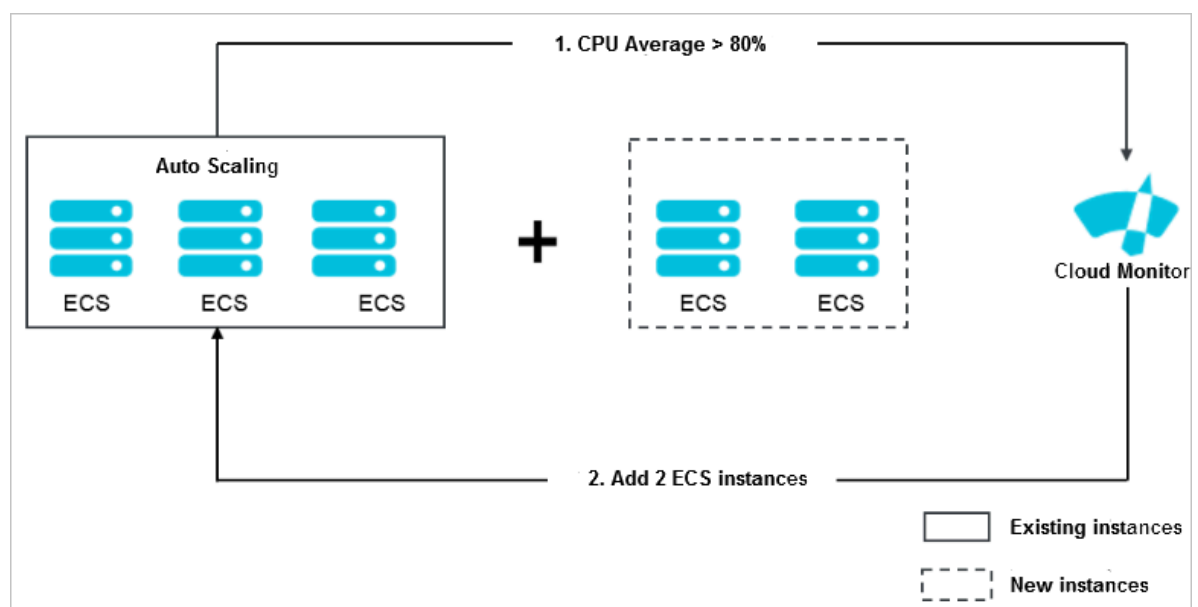
- **Elastic scale-out**

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.

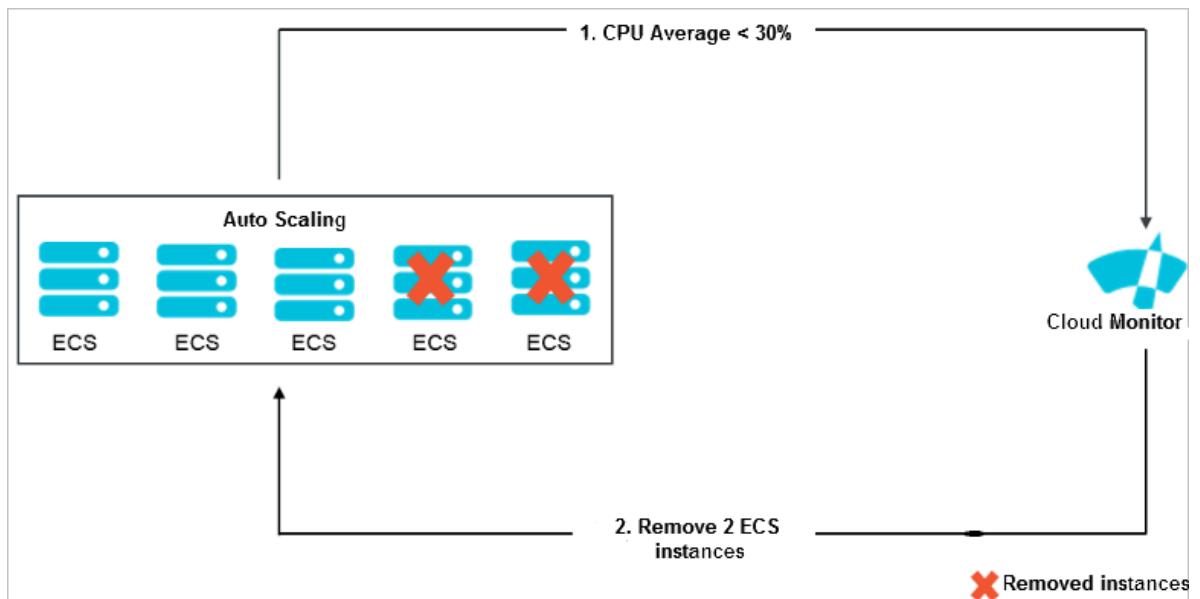
Figure 3-1: Elastic scale-out



- **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 Benefits

ESS has the following benefits:

- **Automatic scaling of instances on-demand**

ESS can automatically add ECS instances during peak traffic hours, and remove ECS instances during off-peak hours to scale with actual business needs. This helps to lower infrastructure costs because you only pay for what you actually use.

- **Real-time instance monitoring and automatic replacement of unhealthy instances**

ESS performs real-time monitoring on instances and automatically replaces unhealthy instances that are discovered, reducing operations and maintenance (O&M) overheads.

- **Intelligent whitelist management and control, no user intervention required**

ESS is integrated with Server Load Balancer (SLB) and ApsaraDB for Relational Database Service (RDS). It automatically manages SLB backend servers and RDS whitelists, eliminating the need to perform manual O&M.

- **Various scaling modes for you to mix and match**

ESS allows you to schedule, customize, fix the minimum number of instances, and configure automatic replacement of unhealthy instances. It also provides APIs to allow you to monitor instances through external monitoring systems.

3.3 Architecture

Figure 3-3: Auto Scaling architecture shows the Auto Scaling architecture details.

Figure 3-3: Auto Scaling architecture

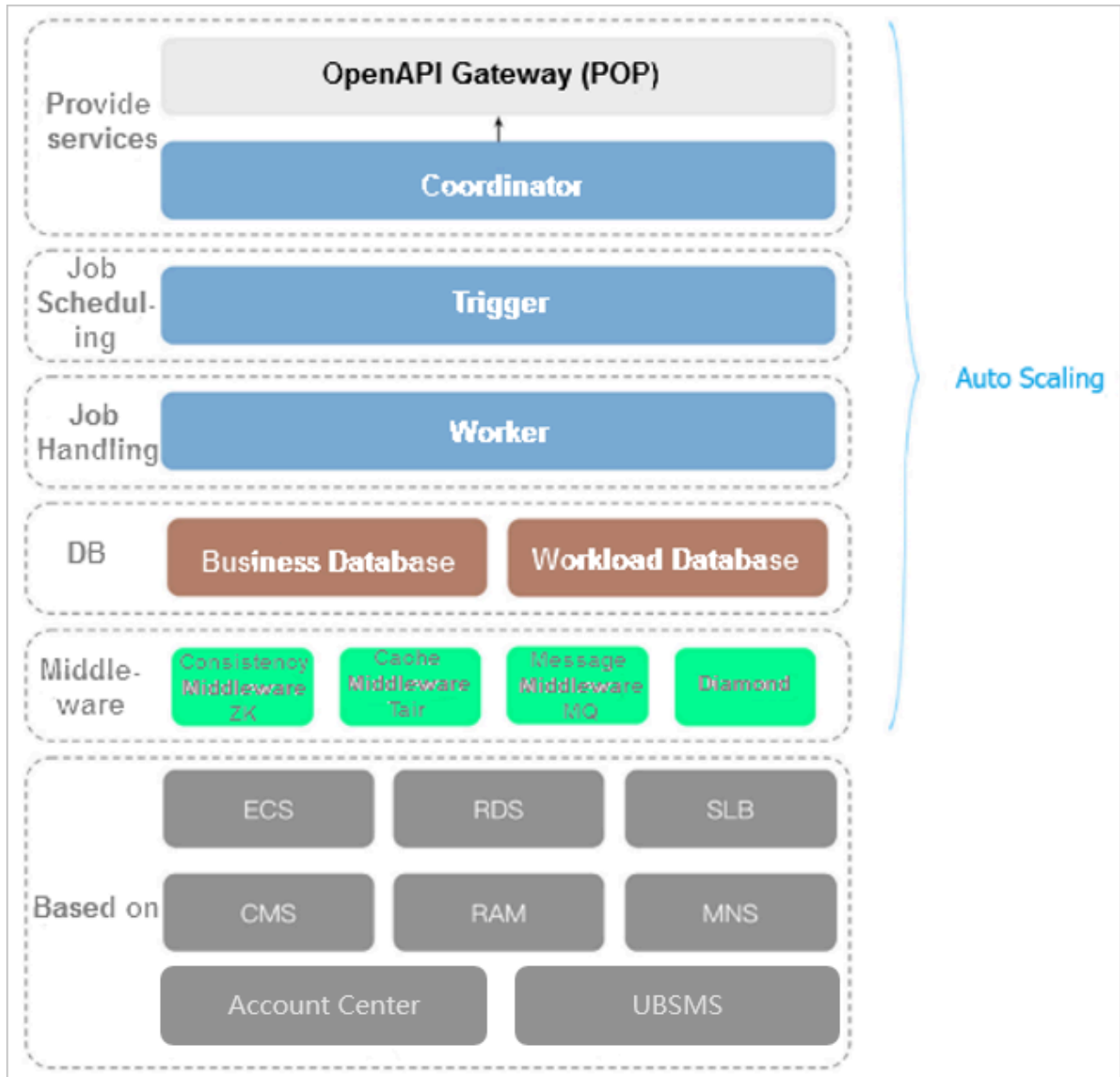


Table 3-1: Architecture description

Component	Description
OpenAPI Gateway	Provides basic services such as authentication and parameter passthrough.
Coordinator	Serves as the ingress of the ESS architecture. It provides external management and control for services, processes API calls, and triggers tasks.

Component	Description
Trigger	Obtains information from the health checks of instances and scaling groups, scheduled tasks, and CloudMonitor to perform tasks scheduling.
Worker	Functions as the core part of ESS. After receiving a task, it handles the entire life cycle of the task, including splitting, executing, and returning the execution results.
Database	Includes the business database and workload database.
Middleware layer	ZooKeeper: ensures consistency by implementing distributed locks for Server Controller.
	Tair: provides caching services for Server Controller.
	Message Queue (MQ): provides message queuing services of VM statuses.
	Diamond: manages persistent configurations.

3.4 Features

ESS has the following features:

- **Automatically adding or removing ECS instances based on your business demands**

You can use the following scaling modes to adjust the number of ECS instances:

- Scheduled mode: Configure periodic tasks to add or remove ECS instances at a specified point in time, such as 13:00 every day.
- Custom mode: Call APIs to manually adjust the number of ECS instances based on monitoring system statistics.
 - You can manually implement scaling rules.
 - You can manually add or remove existing ECS instances.
 - After you have manually adjust MinSize (the minimum number of instances) and MaxSize (the maximum number of instances), ESS automatically creates or

releases ECS instances to ensure that the number of instances remains within the MinSize and MaxSize range.

- Fixed-number mode: Maintain a fixed number of healthy ECS instances by specifying the MinSize attribute. This mode can be used to ensure day-to-day business availability.
- Health mode: Automatically remove or release ECS instances when they are detected as unhealthy (such as they are not in the running state).
- Multimode: Combine any of the preceding modes to meet your own business requirements. For example, if you predict that business peak hours are between 13:00 to 14:00, you can configure a scaling mode that creates 20 ECS instances at the scheduled time. If you are not sure whether the actual demand during peak hours will exceed the number of scheduled resources (for example, the actual load requires 40 ECS instances), another scaling mode can be configured to handle unexpected business loads.
- **Automatically adding or removing ECS instances to or from the SLB backend server group**

The health status of an ECS instance 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 removes the instance and creates a new one. ESS then adds the new instance to the SLB instance and RDS whitelist.



Note:

ECS instances used for automatic scaling can be removed. Therefore, these instances cannot be used to store application status information (such as sessions) and related data (such as databases and logs). If applications deployed on these ECS instances require data to be saved, you can save the status information to independent ECS instances, databases to RDS, and logs to Log Service.

- **Automatically adding or removing IP addresses of ECS instances to or from the RDS whitelist**

When an ECS instance is automatically added to or removed from an SLB backend server group, the IP address of the ECS instance is also automatically added or removed from the RDS whitelist. This mechanism automatically maintains the RDS whitelist and effectively controls access to the RDS instance.

3.5 Usage scenarios

ESS can be used in the following scenarios:

- Video streaming: Traffic loads surge during holidays and festivals. Cloud computing resources must be automatically scaled out to meet the increased demands.
- Live streaming and broadcast: Traffic loads are ever-changing and difficult to predict. Cloud computing resources must be scaled based on CPU utilization, application load, and bandwidth usage.
- Gaming: Traffic loads increase during lunch and after work hours. Cloud computing resources must be scaled out in advance.

3.6 Limits

Auto Scaling has the following limits:

- Instances added by Auto Scaling cannot be automatically added to the whitelist of KVStore for Memcache. For more information about KVStore for Memcache, see KVStore for Memcache Product Introduction .
- Auto Scaling does not support vertical scaling. It can only scale the number of ECS instances. The CPU, memory, and bandwidth configurations of the ECS instances cannot be automatically adjusted.
- The following quotas apply to a scaling group:

Item	Quota
Scaling configuration	You can create a maximum of 10 scaling configurations for a scaling group.
Scaling rule	You can create a maximum of 50 scaling rules for a scaling group.
ECS instance	A scaling group can contain a maximum of 1,000 ECS instances.

3.7 Terms

Auto Scaling

Auto Scaling (ESS) is a management service that automatically adjusts the number of elastic computing resources based on your business demands and strategies. It automatica

lly creates ECS instances during high business loads, and automatically releases ECS instances during low business loads.

Scaling group

A scaling group is a group of ECS instances that is dynamically scaled based on the configured scenario. You can specify the maximum and minimum number of ECS instances in a scaling group, as well as SLB and RDS instances associated with the group.

Scaling configuration

A scaling configuration specifies the configurations of ECS instances in ESS.

Scaling rule

A scaling rule defines the specific scaling activity, for example, the number of ECS instances to be added or removed.

Scaling activity

After a scaling rule is triggered, a scaling activity is performed. A scaling activity shows the changes to the ECS instances in a scaling group.

Scaling trigger task

A scaling trigger task is a task that triggers a scaling rule, such as scheduled tasks.

Cooldown period

The cooldown period indicates a period of time after the completion of a scaling activity in a scaling group. During this period, no other scaling activities can be executed.

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 outstanding advantages in reliability, security, cost, and data processing capabilities. Using OSS, you can store and retrieve a variety of unstructured data files, such as text files, images, audios, and videos, over the network at any time.

OSS uploads data files as objects to buckets. OSS is an object storage service that uses a key-value pair format. You can retrieve object content based on unique object names (keys).

On OSS, you can:

- Create a bucket and upload objects to the bucket.
- Obtain an object URL from OSS to share or download an object.
- Complete the ACL settings of a bucket or object by modifying its properties or metadata.
- Perform basic and advanced OSS tasks through the OSS console.
- Perform basic and advanced OSS tasks using the Alibaba Cloud SDKs or directly calling the RESTful APIs in your application.

4.2 Advantages

Advantages of OSS over user-created server storage

Item	OSS	User-created server storage
Reliability	<ul style="list-style-type: none">• Automatically expands capacities without affecting your services.• Supports automatic redundant data backup.	<ul style="list-style-type: none">• Prone to errors due to low hardware reliability. If a disk has a bad sector, data may be irretrievably lost.• Manual data recovery is complex and requires a lot of time and technical resources.

Item	OSS	User-created server storage
Security	<ul style="list-style-type: none"> Provides hierarchical security protection for enterprises. Provides user resource isolation mechanisms and supports zone-disaster recovery. Provides various authentication and authorization mechanisms. It also provides features such as whitelisting, hotlink protection, RAM, and Security Token Service (STS) for temporary access. 	<ul style="list-style-type: none"> Additional scrubbing and black hole equipment is required. A separate security mechanism is required.

More advantages of OSS

- Ease of use

The standard RESTful APIs (some compatible with Amazon S3 APIs) are supported. A wide range of SDKs, client tools, and console are provided. You can upload, download, retrieve, and manage large amounts of data for websites and mobile apps in the same way you use regular files systems.

- There is no limit on the number and size of objects. Therefore, you can expand your buckets in OSS as required.
 - Streaming writes and reads are supported, which is suitable for business scenarios where you need to simultaneously read and write videos and other large objects.
 - Lifecycle management is supported. You can delete multiple expired data.
- Powerful and flexible security mechanisms

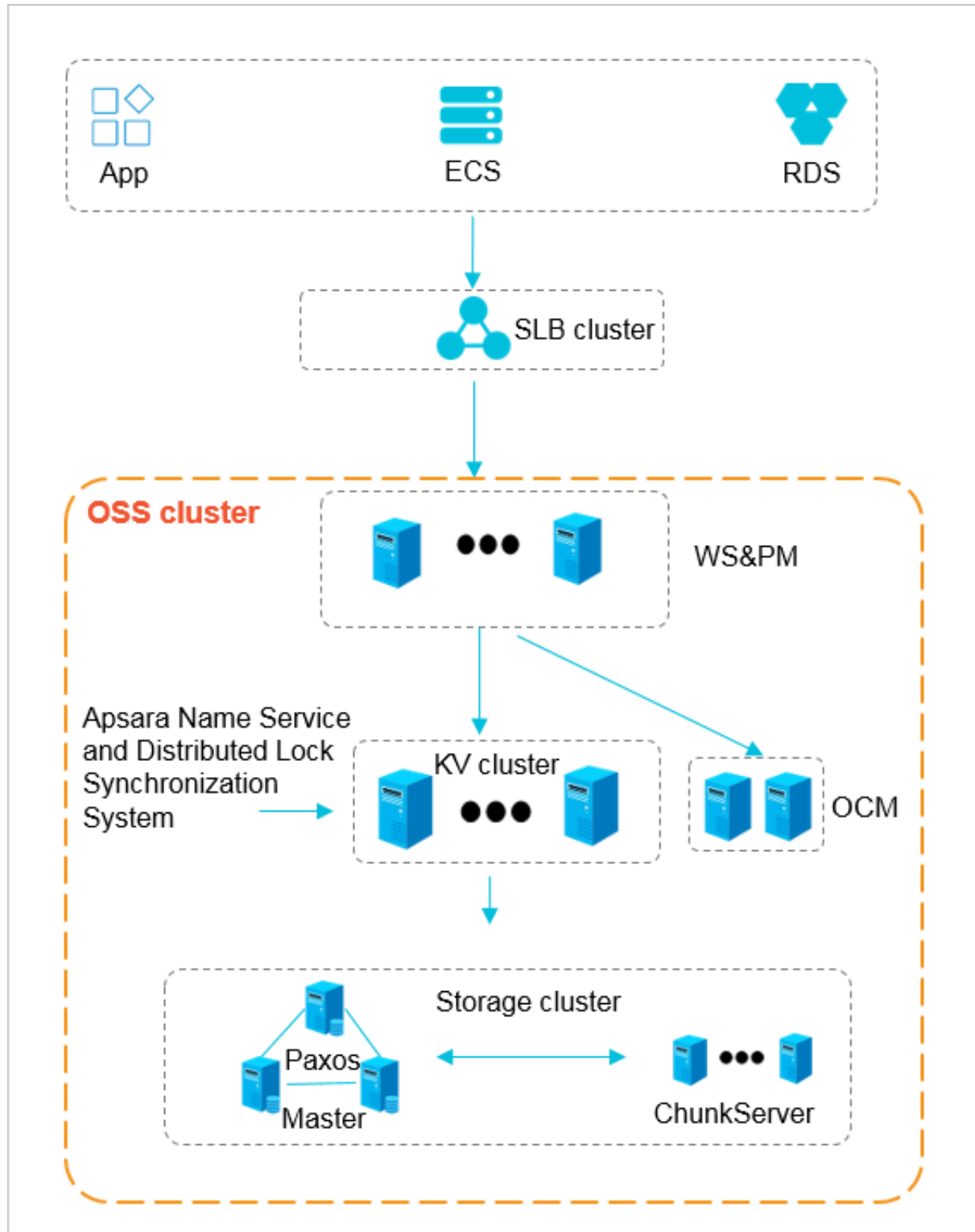
Flexible authentication and authorization mechanisms are available. OSS provides STS and URL-based authentication and authorization mechanisms, as well as whitelisting, hotlink protection, and RAM.

4.3 OSS architecture

Object Storage Service (OSS) is a storage solution built on the Alibaba Cloud Apsara platform. It is based on infrastructure such as Apsara Distributed File System and SchedulerX. Such infrastructure provides OSS and other Alibaba Cloud services with

distributed scheduling, high-speed networks, and distributed storage features. The following figure shows the OSS architecture.

Figure 4-1: OSS architecture



- WS & PM (the protocol layer): is used for receiving users' requests sent through the REST protocol and performing authentication. If the authentication succeeds, users' requests

are forwarded to the key-value engine for further processing. If the authentication fails, an error message is returned.

- **KV cluster:** is used for processing structured data, including reading and writing data based on keys. The KV cluster also supports large-scale concurrent requests. When a service has to operate on a different physical server due to a change in the service coordination cluster, the KV cluster can quickly coordinate and find the access point.
- **Storage cluster:** Metadata is stored on the master node. A distributed message consistency protocol (Paxos) is adopted between master nodes to ensure the consistency of metadata. This ensures efficient distributed storage and access of objects.

4.4 Features

The following table describes features of OSS.

Table 4-1: OSS features

Category	Feature	Description
Bucket	Create a bucket	Before uploading an object to OSS, you must have a bucket to contain the object.
	Delete a bucket	If you no longer need a bucket, delete it to avoid incurring further fees.
	Modify ACL configurations for a bucket	OSS provides ACL for access control. You can configure ACL when creating a bucket and modify the ACL after creating the bucket.
	Configure static website hosting	You can configure static website hosting for your bucket and access this static website through the bucket domain name.
	Configure hotlink protection	To prevent fees incurred by hotlinked OSS data, OSS provides hotlink protection based on the Referer field in the HTTP header.
	Manage CORS	OSS provides CORS in HTML5 to implement cross-origin access.
	Configure lifecycle	You can define and manage the lifecycle of all or a subset objects in a bucket. Lifecycle is configured to manage multiple objects and automatically delete parts.
Object	Upload an object	You can upload all types of objects to a bucket.

Category	Feature	Description
	Create a folder	You can manage OSS folders in the same way you manage folders in Windows.
	Search for objects	You can search for objects whose names contain the same prefix in a bucket or folder.
	Obtain an object URL	You can obtain an object URL from OSS to share or download an object.
	Delete objects	You can delete a single object or multiple objects.
	Delete a folder	You can delete a single folder or multiple folders.
	Modify ACL configurations for an object	You can configure ACL when uploading an object and modify the ACL after uploading the object.
	Manage parts	You can delete all or some parts from a bucket.
API operation	API operation	RESTful API operations are supported and relevant examples are provided.
SDK	SDK	SDK-based development operations and relevant examples for various programming languages are provided.

4.5 Scenarios

Massive storage for image, audio, and video applications

OSS can be used to store large amounts of data, such as images, audios, videos, and logs. OSS supports various devices. Websites and mobile applications can directly read or write OSS data. OSS supports file writing and streaming writing.

Dynamic and static content separation for websites and mobile applications

OSS leverages the BGP bandwidth to achieve ultra-low latency of direct data download.

Offline data storage

OSS is cheap and highly available, enabling enterprises to store data that needs to be archived offline for a long time to OSS.

4.6 Limits

Item	Description
Bucket	<ul style="list-style-type: none"> You can create a maximum of 100 buckets. After a bucket is created, its name and region cannot be modified.
Object upload	<ul style="list-style-type: none"> Objects larger than 5 GB cannot be uploaded by using the following methods: console upload, simple upload, form upload, and append upload. To upload an object that is larger than 5 GB, you must use multipart upload. The size of an object uploaded by using multipart upload cannot exceed 48.8 TB. If you upload an object that has the same name of an existing object in OSS, the new object will overwrite the existing object.
Object deletion	<ul style="list-style-type: none"> Deleted objects cannot be recovered. You can delete up to 50 objects at a time in the Apsara Stack Cloud Management (ASCM) 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.

4.7 Terms

This topic describes several basic terms used in OSS.

object

Files that are stored in OSS. They are the basic unit of data storage in OSS. An object is composed of Object Meta, object content, and a key. An object is uniquely identified by a key in the bucket. Object Meta defines the properties of an object, such as the last modification time and the object size. You can also specify User Meta for the object.

The lifecycle of an object starts when it is uploaded, and ends when it is deleted.

Throughout the lifecycle of an object, Object Meta cannot be changed. Unlike the file system, OSS does not allow you to modify objects directly. If you want to modify an object, you must upload a new object with the same name as the existing one to replace it.

**Note:**

Unless otherwise stated, objects and files mentioned in OSS documents are collectively called objects.

bucket

A container that stores objects. Objects must be stored in the bucket they are uploaded to. You can set and modify the properties of a bucket for object access control and lifecycle management. These properties apply to all objects in the bucket. Therefore, you can create different buckets to implement different management functions.

- OSS does not have the hierarchical structure of directories and subfolders as in a file system. All objects belong to their corresponding buckets.
- You can have multiple buckets.
- A bucket name must be globally unique within OSS and cannot be changed after a bucket is created.
- A bucket can contain an unlimited number of objects.

strong consistency

A feature of operations in OSS. Object operations in OSS are atomic, which indicates that operations are either successful or failed. There are no intermediate states. OSS never writes corrupted or partial data.

Object operations in OSS are strongly consistent. For example, after you receive a successful upload (PUT) response, the object can be read immediately, and the data is already written in triplicate. Therefore, OSS avoids the situation where no data is obtained when you perform the read-after-write operation. An object also has no intermediate states when you delete the object. After you delete an object, that object no longer exists.

Similar to traditional storage devices, modifications are immediately visible in OSS while consistency is guaranteed.

Comparison between OSS and the file system

OSS is a distributed object storage service that uses a key-value pair format. You can retrieve object content based on unique object names (keys). Although you can use names like test1/test.jpg, this does not necessarily indicate that the object is saved in a directory named test1. In OSS, test1/test.jpg is only a string, which is no different from a.jpg. Therefore, similar resources are consumed when you access objects that have different names.

A file system uses a typical tree index structure. Before accessing a file named test1/test.jpg, you must access directory test1 and then locate test.jpg. This makes it easy for a file system to support folder operations, such as renaming, deleting, and moving directories, because these operations are only directory node operations. System performance depends on the capacity of a single device. The more files and directories that are created in the file system, the more resources are consumed, and the lengthier your process becomes.

You can simulate similar functions in OSS, but this operation is costly. For example, if you want to rename test1 directory test2, the actual OSS operation would be to replace all objects whose names start with test1/ with copies whose names start with test2/. Such an operation would consume a large amount of resources. Therefore, try to avoid such operations when using OSS.

You cannot modify objects stored in OSS. A specific API must be called to append an object, and the generated object is of a different type from that of normally uploaded objects. Even if you only want to modify a single Byte, you must re-upload the entire object. A file system allows you to modify files. You can modify the content at a specified offset location or truncate the end of a file. These features make file systems suitable for more general scenarios. However, OSS supports sporadic bursts of access, whereas the performance of a file system is subject to the performance of a single device.

Therefore, mapping OSS objects to file systems is inefficient, which is not recommended. If attaching OSS as a file system is required, we recommended that you perform only the operations of writing data to new files, deleting files, and reading files. You can make full use of OSS capabilities. For example, you can use OSS to store and process large amounts of unstructured data such as images, videos, and documents.

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 Benefits

Scalability

There is no upper limit to the amount of data that can be stored in Table Store tables. As data increases, Table Store adjusts partitions to provide more storage space for tables and improve the capability of handling access request bursts.

High performance

If you use a high-performance instance, its average access latency of single rows is measured in single-digit milliseconds. The read/write performance is not affected by the size of data in a table.

Reliability

Table Store provides high data reliability. It stores multiple data copies and restores data when some copies become invalid.

High availability

Through automatic failure detection and data migration, Table Store shields applications from host- and network-relevant hardware faults to achieve high availability.

Ease of management

Table Store automatically performs complex O&M tasks, such as the management of data partitions, software and hardware upgrades, configuration updates, and cluster scale-out.

Access security

Table Store provides multiple permission management mechanisms. It verifies and authenticates the identity of the request to prevent unauthorized data access, improving the data security.

Strong consistency

Table Store ensures strong data consistency for data writes. A successful write operation indicates that the data is written to three copies and stored in disks. Applications can read the latest data immediately.

Flexible data models

Table Store tables do not require a fixed format. Each row can contain a different number of columns. Table Store supports multiple data types, such as Integer, Boolean, Double, String, and Binary.

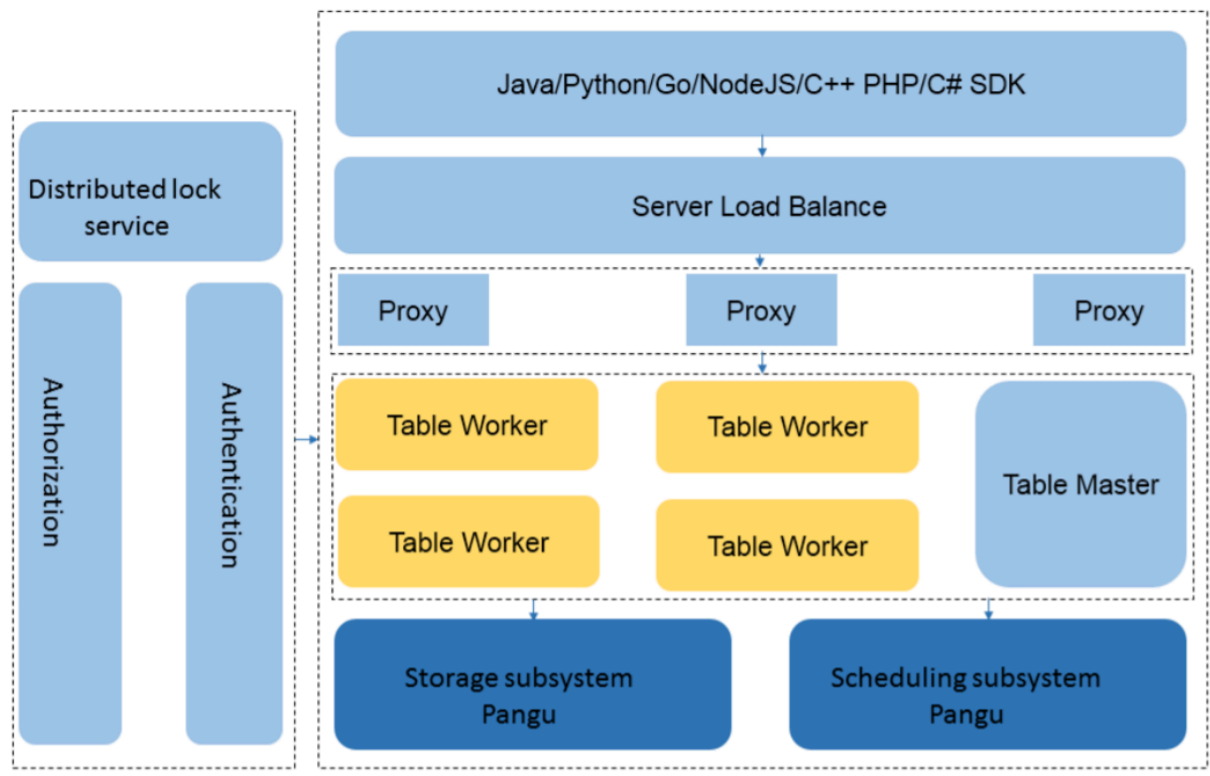
Monitoring integration

You can log on to the Table Store console to obtain monitoring information in real time, including the requests per second and average response latency.

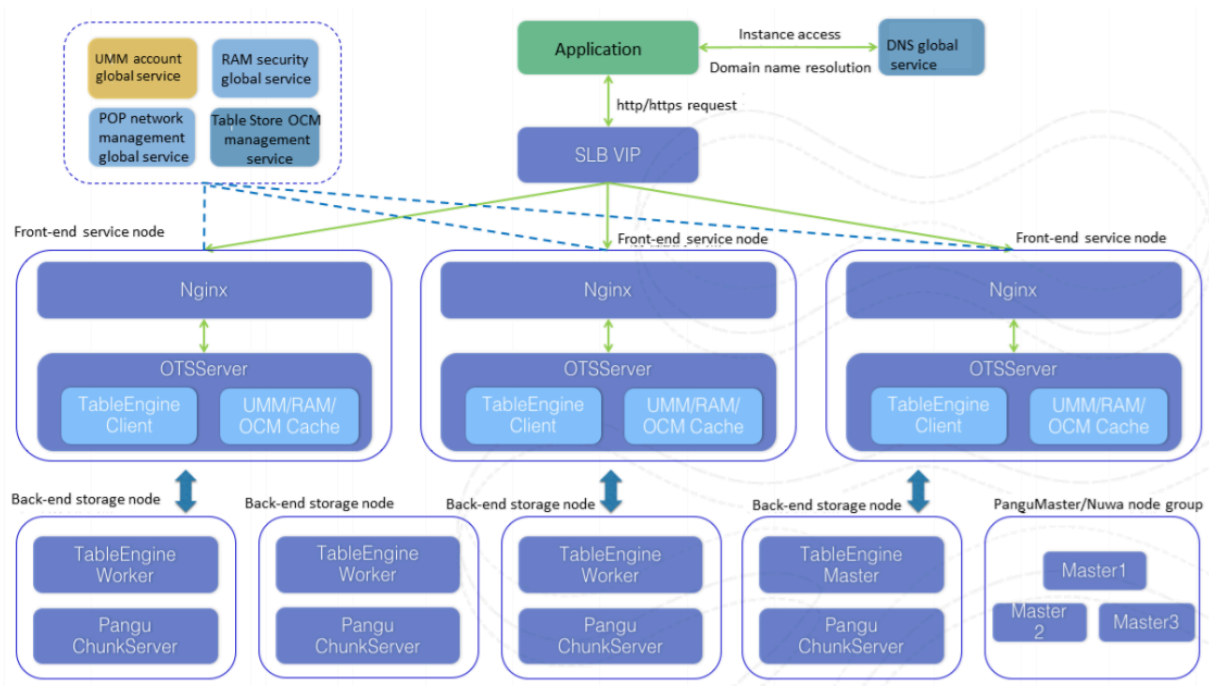
5.3 Architecture

The architecture of Table Store is referenced from Bigtable (one of the three core technologies of Google) and uses the log-structured merge-tree (LSM) storage engine to provide high performance writes. The performance of primary key-based single-row queries and range queries is stable and predictable. The performance is not affected by the volume of data and access concurrency.

The following figure shows the basic architecture of Table Store.



The following figure shows the detailed architecture of Table Store.



- The top layer is the protocol access layer. SLB distributes user requests to various proxy nodes. The proxy nodes receive requests that are sent through the RESTful protocol and implement security authentication. If the authentication succeeds, the user requests are forwarded to the corresponding data engine based on the value of the first primary key column for further operations. If the authentication fails, an error message is directly returned to the user.
- Table Worker is the data engine layer that processes structured data. It uses a primary key to search for or store data. Table Worker supports large-scale access request bursts.
- The bottom layer is the storage layer. Apsara Distributed File System is deployed at this layer. Metadata is stored in Master server roles. The distributed message consistency protocol Paxos is adopted between Master service roles to ensure metadata consistency. In this scenario, efficient distributed file storage and access are achieved. This method guarantees three copies of data stored in the system and system recovery from any hardware or software faults.

5.4 Scenarios

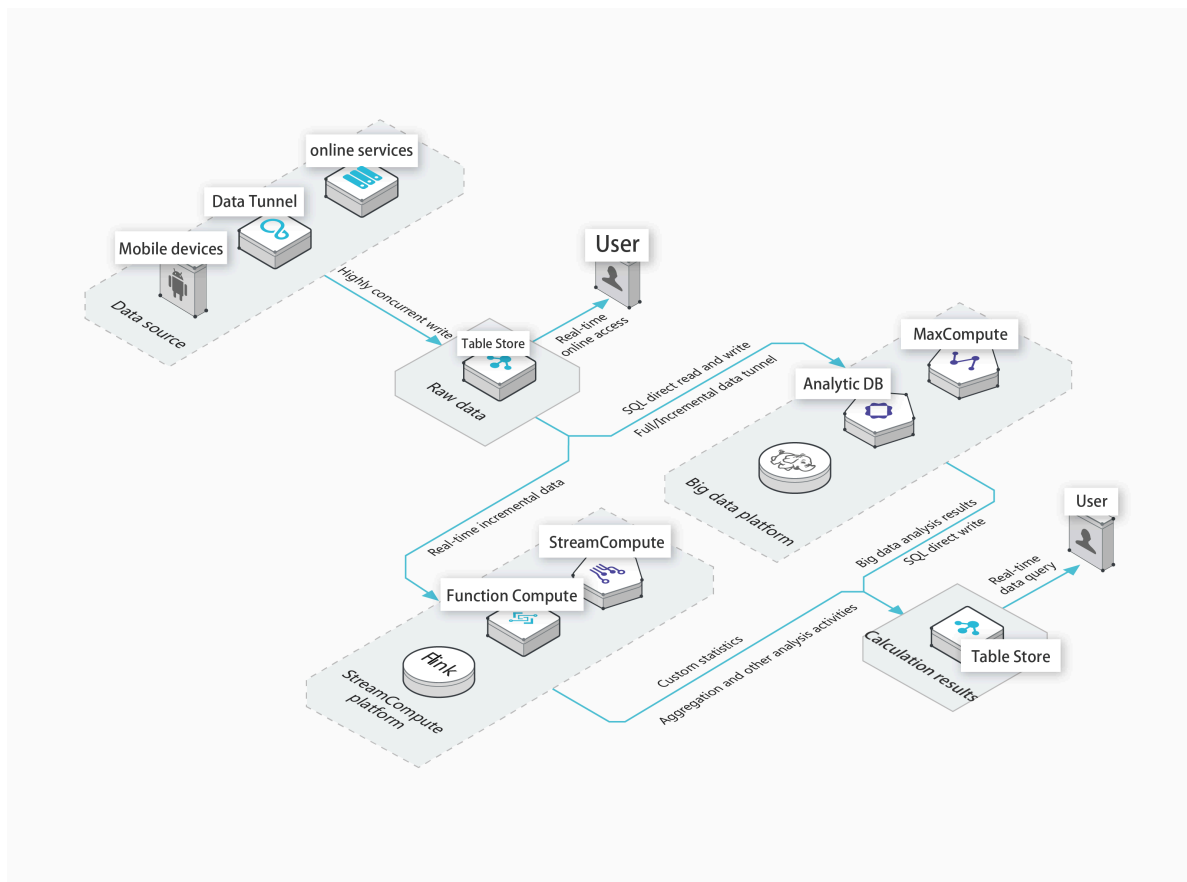
- Scenario 1: Big data storage and analytics

Table Store provides cost-friendly, highly concurrent, and low-latency storage, whose data can be accessed online. It provides tunnels for full and incremental data, and supports direct SQL read and write operations on various big data analytics platforms

such as MaxCompute. It provides API operations to directly read incremental data streams, which allows you to implement real-time stream computing.

Table Store provides the following features:

- Table Store supports various big data computing platforms, stream computing services, and real-time computing services that are provided by Alibaba Cloud.
- Table Store provides high-performance and capacity instances to meet the requirements of different business.



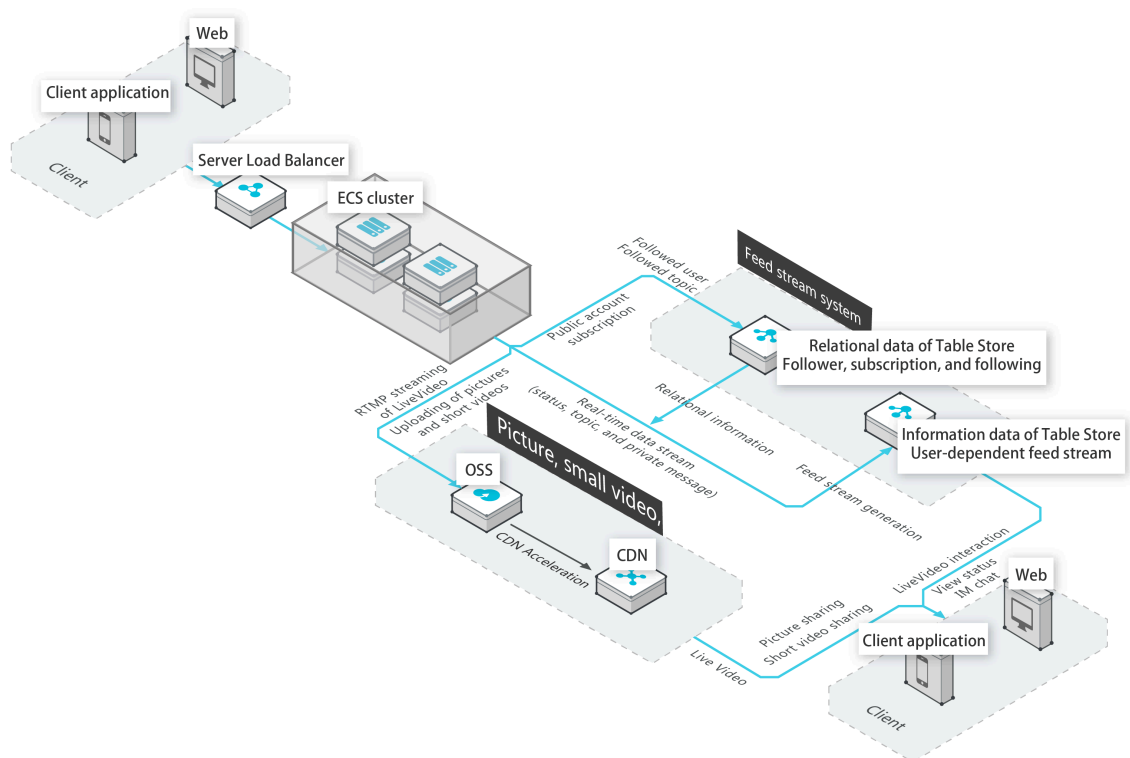
- Scenario 2: Social media feeds on the Internet

You can use Table Store to store large amounts of instant messaging (IM) messages and social media feed information such as comments, posts, and likes. The elastic resources

available for Table Store can meet application requirements, such as handling significant traffic fluctuations, high concurrency, and low latency, at relatively low costs.

Table Store provides the following features:

- Built-in auto-increment primary key columns reduce the number of external system dependencies.
- Average read and write performance of high-performance instances are not affected by volumes.
- Highly available storage for large amounts of messages, and multi-terminal message synchronization are supported.



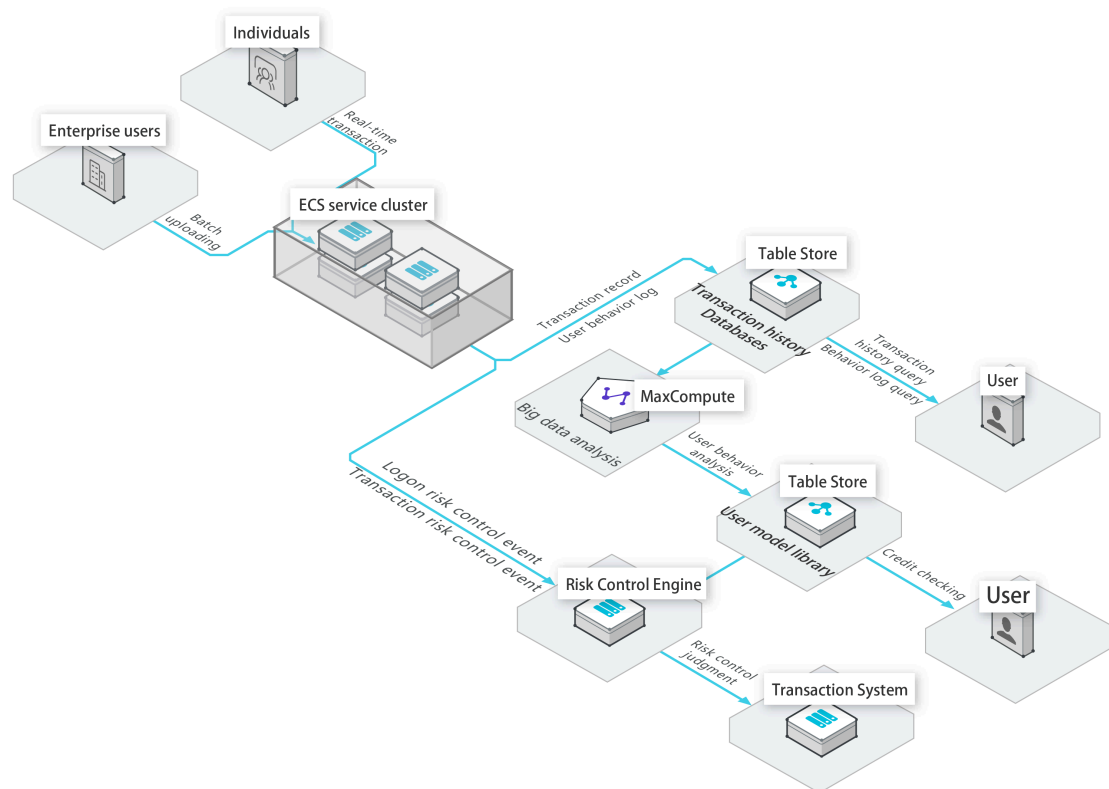
- Scenario 3: Storage and real-time queries of large amounts of transaction records and user models

Table Store instances are elastic, low latency, and highly concurrent, which provides optimal running conditions for risk control systems. This helps you control transaction

risks. Furthermore, the flexible data structure allows your business model to rapidly evolve to meet market demands.

Table Store provides the following features:

- A table can store full historical transaction records.
- Data is stored in three copies to ensure high consistency and data security.
- Rapid service development is made available by a schema-free model and attribute columns that can be added as required.



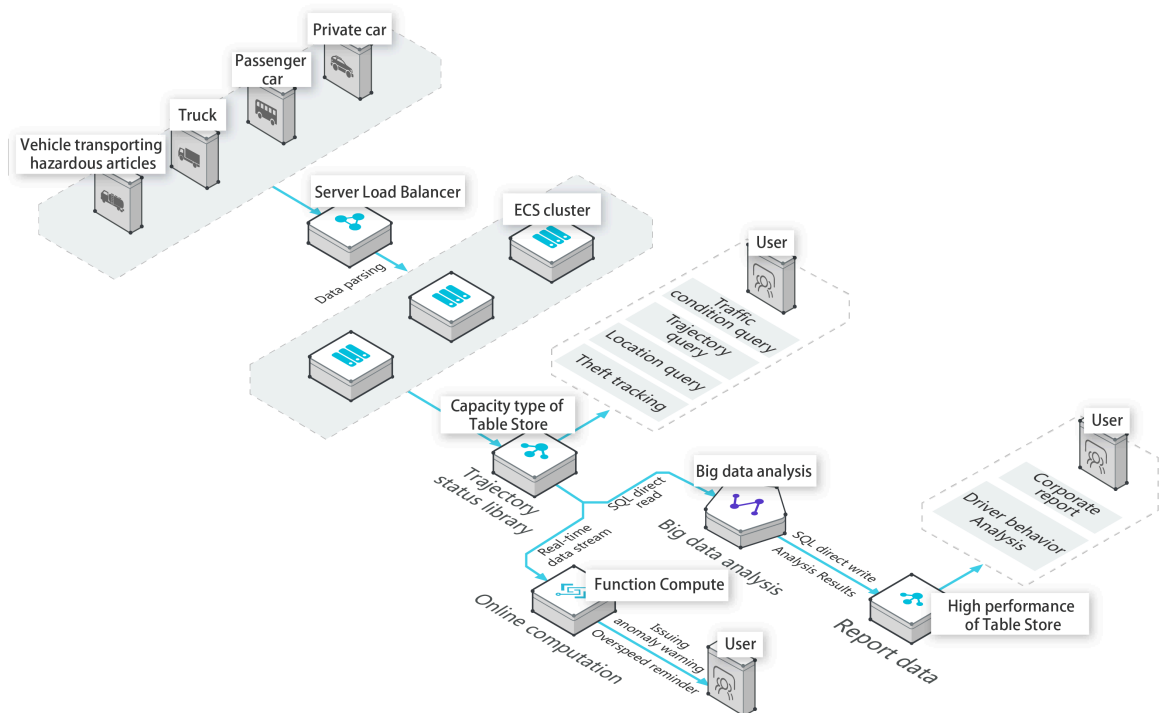
- Scenario 4: Efficient and flexible storage of large amounts of IoV data

The schema-free data model simplifies access to the data collected from different vehicle-mounted devices. Table Store can be seamlessly integrated with multiple big

data analytics platforms and real-time computing services to implement real-time online queries and business report analysis.

Table Store provides the following features:

- Data is stored in a table without being partitioned or sharded, simplifying business logic.
- The query performance for vehicle conditions and recommended routes is stable and predictable.
- The schema-free model allows you to easily store data collected from different vehicle-mounted devices.



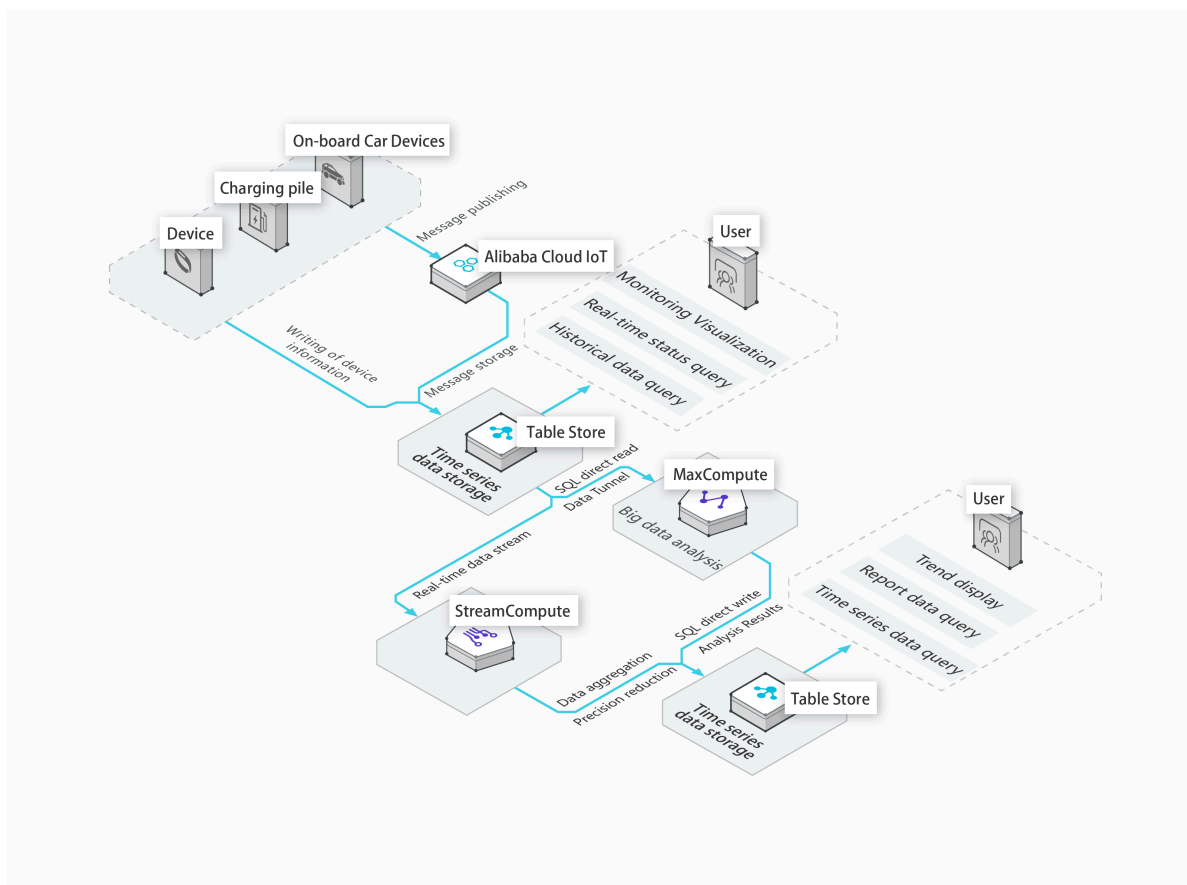
- Scenario 5: Storage of large amounts of IoT data for efficient queries and analysis

Table Store can be used to store time series data from IoT devices and monitoring systems. It provides API operations to directly read SQL data and incremental data

streams, which allow you to implement offline data analysis and real-time stream computing.

Table Store provides the following features:

- Table Store can meet the data write and storage requirements of ultra-large-scale IoT devices and monitoring systems.
- Table Store can integrate with a variety of offline or stream data analysis platforms . This allows you to use a single piece of data for multiple analysis and computing operations.
- Table Store supports data lifecycle management.



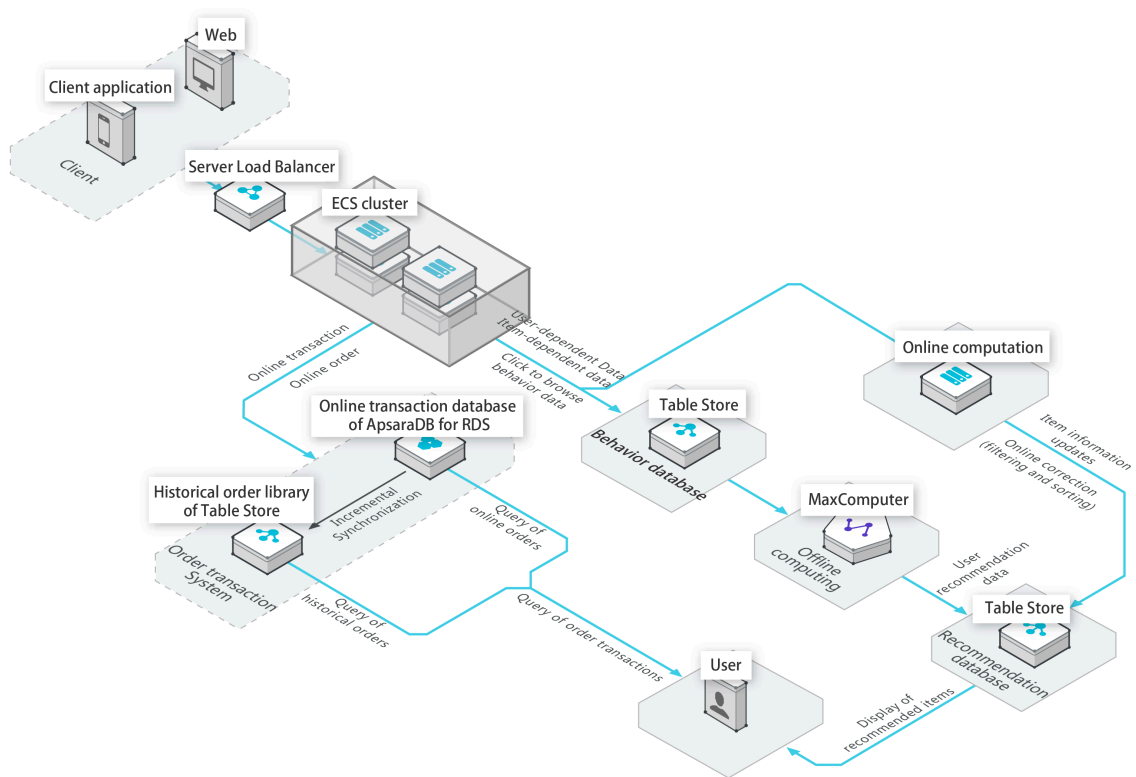
- Scenario 6: Databases for large-scale e-commerce transaction orders and user-specific recommendations

Table Store can easily manage large amounts of historical transaction data and improve access performance. Table Store can be used together with MaxCompute to implement

precision marketing and elastic resource storage. This allows you to handle service requests during peak hours when all users go online.

Table Store provides the following features:

- Resources can be scaled based on data volumes and access concurrency, allowing the service to handle scenarios with high access fluctuations during various periods.
- Various big data analytics platforms are supported for direct analysis of user behavior .
- Single-digit millisecond latency for queries on large amounts of data.



5.5 Limits

[Table 5-1: Limits](#) lists the limits for Table Store. Some limit ranges indicate the maximum values that can be used rather than the suggested values. Table structures and row data sizes can be tailored to improve performance.

Table 5-1: Limits

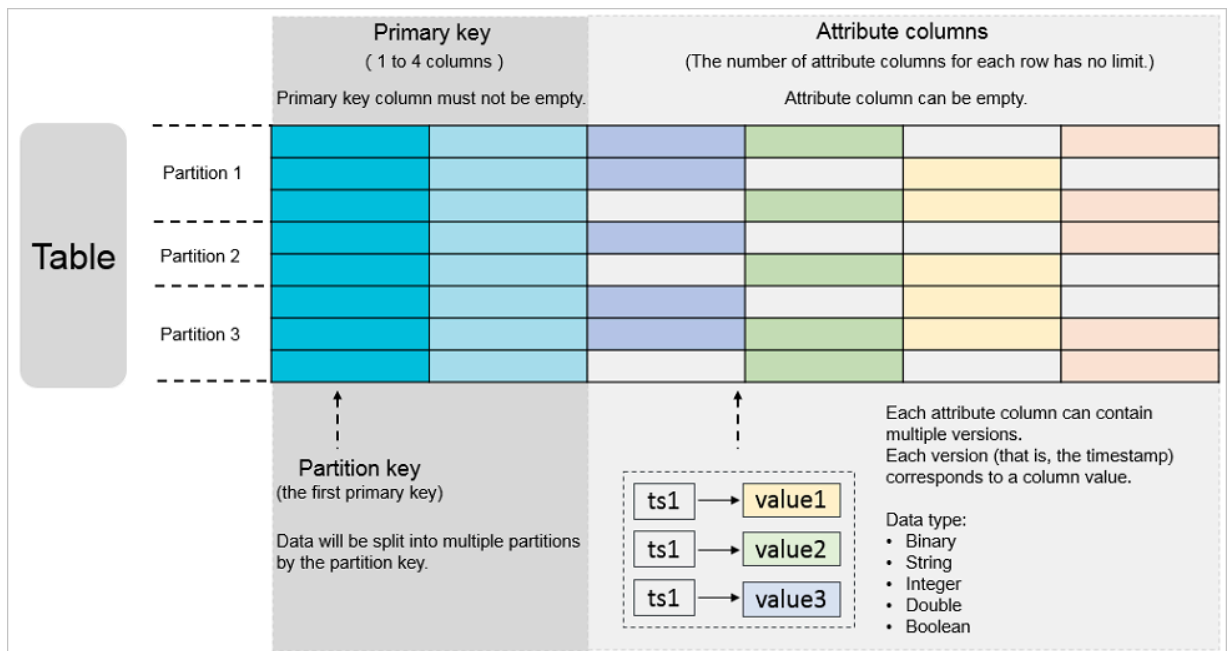
Limit	Limit range	Description
The number of instances created under an Apsara Stack tenant account	1024	If you need to increase the maximum number of instances, contact an administrator.
The number of tables in an instance	1024	If you need to increase the maximum number of tables, contact an administrator.
The number of columns of a primary key	1-4	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.
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 amount of attribute columns.
The number of attribute columns written by one request	1024	The number of attribute columns written by a PutRow, UpdateRow, or BatchWriteRow request to a row cannot exceed 1,024.
The size of a single row	Unlimited	The total size of all column names and column value data for a row is unlimited.
The number of columns involved in columns_to_get in a read request	0-128	The maximum number of columns obtained from a row of data in a read request cannot exceed 128.
The number of UpdateTable operations for a table	Upper limit: unlimited Lower limit: unlimited	The frequency of UpdateTable operations for a table is limited.
UpdateTable frequency for a table	Once every 2 minutes	The reserved read/write throughput for a table can be adjusted once every two minutes at most.

Limit	Limit range	Description
The number of rows read by one BatchGetRow request	100	N/A
The number of rows written by one BatchWriteRow request	200	N/A
The size of data written by one BatchWriteRow request	4 MB	N/A
Data returned by one GetRange request	5,000 rows or 4 MB	The data returned by a request cannot exceed 5,000 rows or 4 MB. When either of the limits is exceeded, data that exceeds the limits is truncated at the row-level. The data primary key information in the next row is returned.
The data size of an HTTP request body	5 MB	N/A

5.6 Terms

data model

A model that consists of tables, rows, primary keys, and attributes, as shown in the following figure.



max versions

A data table attribute that indicates the maximum number of data versions that can be stored in each attribute column of a data table. If the number of versions in an attribute column exceeds the max versions value, the oldest version is deleted. This operation is performed asynchronously.

time To live (TTL)

A data table attribute measured in seconds. It indicates the validity period of data. To save space and reduce costs for data storage, the Table Store automatically clears any data that exceeds its TTL.

max version offset

A data table attribute that describes the maximum allowable difference between the version to be written and the current time, in seconds.

To prevent the writing of unexpected data, a server checks the versions of attribute columns when processing writing requests. Writing data to a row fails if the row has an attribute column in which: its version is earlier than the current writing time minus the max version offset value, or its version is later than or equal to the current writing time plus the max version offset value.

The valid version range for an attribute column is calculated based on the formula: Valid version range = [Data write time - Max version offset value, Data write time + Max version offset value). Data write time is the number of seconds that have elapsed since 1970-01-01 00:00:00 UTC. The versions of the attribute columns are stored in milliseconds, and must fall within the valid version range after it is converted to seconds (divide by 1,000).

primary key and attribute

A primary key is the unique identifier of each row in a table. It can consist of one to four primary key columns. When you create a table, you must define a primary key. Specifically, you must specify the name, data type, and sequence of each primary key column. Data types of values in primary key columns include String, Integer, and Binary. For a primary key column of the String or Binary type, the size of the column value must be smaller than 1 KB.

An attribute is a Table Store attribute that stores data in a row. You can create an unlimited number of attribute columns for each row.

read/write throughput

A Table Store attribute that is measured by read/write capacity units (CUs).

region

An Apsara Stack physical data center. Table Store is deployed across multiple Apsara Stack regions. Select a region that suits your business requirements.

instance

A logical entity in Table Store. It is used to manage tables, which are equivalent to databases in traditional relational databases. An instance is the basic unit of the Table Store resource management system. Table Store allows you to control access and meter resources at the instance level.

endpoint

The connection URL (also known as an endpoint) for each instance. You need to specify the endpoint before you perform any operations on Table Store tables and data.

stream

A data table attribute used for real-time analysis of incremental data streams and incremental data synchronization.

Serial ATA (SATA)

A disk that is based on serial connections and provides stronger error-correcting capabilities. It aims to improve the reliability of data during transmission.

5.7 Features

5.7.1 Features

- Data partition and load balancing

The first column of a primary key in each row of a table is the partition key. The system splits a table into multiple partitions based on the range of the partition key. These partitions are evenly scheduled across different storage nodes. When the data in a partition exceeds a certain size, the partition is automatically split into two smaller partitions. The data and access loads are distributed across these two partitions. The partitions are scheduled to different nodes. As a result, access loads are distributed to different nodes. This allows single-table data and access loads to scale linearly. A

partition is a logical organization of data based on the shared storage mechanism. No migration of physical data is involved when a partition is split. However, this may cause the partition to be unable to provide services for 100 milliseconds.

- Automatic recovery from single point of failures (SPOFs)

Each node in the storage engine of Table Store provides services for multiple data partitions of different tables. The master node manages partition distribution and scheduling, and also monitors the health of each service node. If a service node fails, the master node migrates data partitions from the faulty node to other healthy nodes. The migration is performed on the logical level, and does not involve physical entities, so services can rapidly recover from the single point of failure (SPOF).

5.7.2 Tunnel Service

Tunnel Service is a centralized service that uses the Table Store API to allow you to consume full and incremental data. You can use Tunnel Service to create tunnels from which you can consume distributed data in real time. Tunnels allow you to consume incremental data, full data, and full and incremental data. After you create a tunnel, you can use it to consume historical and incremental data that is exported from a specified table.

Background

Table Store is applicable to scenarios such as metadata management, time series data monitoring, and message systems. These scenarios often make use of incremental or full and incremental data streams to trigger operations such as:

- Data synchronization: synchronizes data to a cache, search engine, or data warehouse.
- Event triggering: triggers Function Compute, sends a notification when data is consumed, or calls an API operation.
- Stream data processing: connects to a stream-computing engine or a stream- and batch-computing engine.
- Data migration: backs up data to OSS or migrates data to a Table Store capacity instance.

You can use Tunnel Service to build efficient and elastic solutions to consume full data, incremental data, and full and incremental data in the preceding scenarios.

Features

The following table lists the features provided by Tunnel Service.

Feature	Description
Tunnels for full and incremental data consumption	Tunnel Service allows you to consume incremental data, full data, and full and incremental data simultaneously.
Orderly incremental data consumption	Tunnel Service sequentially distributes incremental data to one or more logical partitions based on the write time. Data in different partitions can be consumed simultaneously.
Consumption latency monitoring	Tunnel Service allows you to call the DescribeTunnel operation to view the recovery point objective (RPO) information of the consumed data on each client. Tunnel Service also allows you to monitor data that is consumed through tunnels in the Table Store console.
Horizontal scaling of data consumption capabilities	Tunnel Service supports automatic load balancing among logical partitions to accelerate data consumption.

5.7.3 Global secondary index

5.7.3.1 Feature description

This topic describes features of the global secondary index of Table Store.

- Supports asynchronous data synchronization between a base table and indexes. Under normal network conditions, the data synchronization can reach single-digit millisecond latency.
- Supports single-column indexes, compound indexes, and covered indexes. Predefined columns are specified in advance in a base table. You can create an index on any predefined column or a primary key of a base table. In addition, you can specify multiple predefined columns of a base table as attribute columns of an index or choose not to specify attribute columns. If you specify predefined columns of a base table as the attribute columns of an index, you can directly query this index instead of querying the

base table. For example, a base table includes the primary keys PK0, PK1, and PK2. Additionally, the table has the predefined columns Defined0, Defined1, and Defined2.

- You can create an index on PK2 without specifying an attribute column.
- You can create an index on PK2 and specify Defined0 as an attribute column.
- You can create an index on PK1 and PK2 without specifying an attribute column.
- You can create an index on PK1 and PK2 and specify Defined0 as an attribute column.
- You can create an index on PK2, PK1, and PK0 and specify Defined0, Defined1, and Defined2 as attribute columns.
- You can create an index on Defined0 without specifying an attribute column.
- You can create an index on Defined0 and PK1 and specify Defined1 as an attribute column.
- You can create an index on Defined1 and Defined0 without specifying an attribute column.
- You can create an index on Defined1 and Defined0 and specify Defined2 as an attribute column.
- Supports sparse indexes. You can specify a base table predefined column as an index attribute column. A row will be indexed even when all primary keys exist despite the predefined column being excluded from the row of the base table. However, a row will not be indexed when the row excludes one or more indexed columns. For example, a base table includes the primary keys PK0, PK1, and PK2. Additionally, the table has the predefined columns Defined0, Defined1, and Defined2. You can create an index on Defined0 and Defined1 and specify Defined2 as an attribute column.
 - When a row in the base table excludes Defined2 and includes Defined0 and Defined1, the row will be included in the index.
 - When a row in the base table excludes Defined1 but includes Defined0 and Defined2, the row is excluded from the index.
- Supports creating and deleting indexes on an existing base table. Supports adding the existing data in a base table to a new index.
- When you query an index, the query is not automatically performed on the base table of the created index. You need to query the base table as needed. This feature will be supported in later versions.

5.7.3.2 Overview

This topic describes the basic concepts, limits, and precautions about global secondary indexes.

Terms

Term	Description
index	You can create an index for some columns in a base table. The index is read-only.
predefined column	Table Store uses a schema-free model. You can write an unlimited number of columns to a row. You do not need to specify a fixed number of attributes in a schema. You can also predefine some columns and specify their data types when you create a table.
single-column index	You can create an index only for one column.
compound index	You can create an index for multiple columns in a table. A compound index can have indexed columns 1 and 2.
indexed attribute column	You can map predefined columns in a base table to non-primary key columns in an index.
autocomplete	Table Store automatically adds all primary key columns of the base table to the index.

Limits

- The index table names must be unique in an instance.
- You can create a maximum of five indexes in a base table. You cannot create more indexes if you have created five indexes.
- You can create a maximum of 15 predefined columns in a base table. You cannot create a base table if you have created 15 predefined columns.
- An index can contain a maximum of four indexed columns. These indexed columns include a combination of primary keys and predefined columns of the base table. If you specify more than four indexed columns, you cannot create the index.
- An index can contain a maximum of eight attribute columns. If you specify more than eight attribute columns, you cannot create the index.
- You can specify the type of an indexed column as Integer, String, or Binary. The limits for indexed columns are the same as those for primary key columns of the base table.
- If an index combines multiple columns, the size limit for the columns is the same as that for primary key columns of the base table.

- When you specify a column of the String or Binary type as an attribute column of an index, the limits for attribute columns are the same as those for attribute columns of the base table.
- You cannot create an index on a table that has the time to live (TTL) parameter configured. If you want to create indexes on a table that has TTL configured, use DingTalk to request technical support.
- You cannot create an index in a table that has Max Versions configured. If a table has Max Versions configured, you cannot create an index for the table. You cannot configure Max Versions for an indexed base table.
- You cannot customize versions when writing data to an indexed base table. Otherwise, you cannot write data to the base table.
- You cannot use the Stream feature in an index.
- An indexed base table cannot contain repeated rows that have the same primary key during the same BatchWrite operation. Otherwise, you cannot write data to the base table.

Precautions

- Table Store automatically adds all primary key columns of the base table to the index. When you scan an index, you must specify the range of primary key columns. The range can be from negative infinity to positive infinity. For example, a base table includes the primary keys PK0 and PK1 and a predefined column Defined0.

When you create an index for the Defined0 column, Table Store generates an index that has the primary keys Defined0, PK0, and PK1. When you create an index for the Defined0 and PK1 columns, Table Store generates an index that has the primary keys Defined0, PK1, and PK0. When you create an index for the PK column, Table Store generates an index that has the primary keys PK1 and PK0. When you create an index, you can only specify the column that you want to index. Table Store automatically adds all primary key columns of the base table to the index. For example, a base table contains the primary keys PK0 and PK1 and a predefined column Defined0.

- When you create an index for the Defined0 column, Table Store generates the index that has the primary keys Defined0, PK0, and PK1.
- When you create an index for the PK1 column, Table Store generates the index that has the primary keys PK1 and PK0.
- You can specify predefined columns as attribute columns in the base table. When you specify a predefined column of the base table as an attribute column of the index, you

can search this index for the column value instead of searching the base table. However, this increases storage costs. Otherwise, you have to query the base table based on the index. You can choose between these methods based on your requirements.

- We recommend that you do not specify a column whose values are date or time as the first primary key column of an index. This may slow down the index updates. We recommend that you hash the column related to the time or date and create an index for the hashed column. If you have such requirements, use DingTalk to request technical support.
- We recommend that you do not define a column of low cardinality or a column that contains enumerated values, as the first primary key column of an index. For example, the gender column restricts the horizontal scalability of the index and leads to poor write performance.

5.7.3.3 Scenarios

Global secondary index allows you to create an index based on a specified column. Data in the generated index table is sorted by the specified indexed column. All data written to the base table is automatically synchronized to the index table asynchronously. If you only write data to a base table and query indexes created on the table, the query performance can be improved in most scenarios. This topic describes how to use a global secondary index to query phone records.

For example, the following table contains a number of phone records.

CellNumber	StartTime (Unix timestamps)	CalledNumber	Duration	BaseStationNumber
123456	1532574644	654321	60	1
234567	1532574714	765432	10	1
234567	1532574734	123456	20	3
345678	1532574795	123456	5	2
345678	1532574861	123456	100	2
456789	1532584054	345678	200	3

- The CellNumber and StartTime columns act as the primary key. They represent the caller and call start time.
- The CalledNumber, Duration, and BaseStationNumber columns represent the call recipient, call duration, and base station number.

When you end a phone call, information about the call is written to this table. You can create global secondary indexes for different query scenarios. For example, you can create global secondary indexes whose primary key is `CalledNumber` or `BaseStationNumber`.

Assume that you have the following query requirements:

- You want to obtain the rows where the value of `CellNumber` is `234567`.

Table Store uses a global ordering model, which sorts all rows by their primary keys and provides the `getRange` operation to perform sequential scans. When you use `getRange` to scan the base table for this example, you only need to set the minimum and maximum values of PK0 to `234567`, and set the minimum value of PK1 (call start time) to `0` and the maximum value of PK1 to `INT_MAX`

```
private static void getRangeFromMainTable(SyncClient client, long cellNumber)
{
    RangeRowQueryCriteria rangeRowQueryCriteria = new RangeRowQueryCriteria(
        TABLE_NAME);

    // Specify the primary key to start from.
    PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrim
    aryKeyBuilder();
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKey
    Value.fromLong(cellNumber));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKey
    Value.fromLong(0));
    rangeRowQueryCriteria.setInclusiveStartPrimaryKey(startPrimaryKeyBuilder.build());

    // Specify the primary key to end with.
    PrimaryKeyBuilder endPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuil
    der();
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKey
    Value.fromLong(cellNumber));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKey
    Value.INF_MAX);
    rangeRowQueryCriteria.setExclusiveEndPrimaryKey(endPrimaryKeyBuilder.build());

    rangeRowQueryCriteria.setMaxVersions(1);

    String strNum = String.format("%d", cellNumber);
    System.out.println("The cell number" + strNum + "makes the following calls:");
    while (true) {
        GetRangeResponse getRangeResponse = client.getRange(new GetRangeRequest(
            rangeRowQueryCriteria));
        for (Row row : getRangeResponse.getRows()) {
            System.out.println(row);
        }

        // If the nextStartPrimaryKey value is not null, continue to read data from the base
        table.
        if (getRangeResponse.getNextStartPrimaryKey() != null) {
            rangeRowQueryCriteria.setInclusiveStartPrimaryKey(getRangeResponse.
            getNextStartPrimaryKey());
        } else {
            break;
        }
    }
}
```

```
}
```

- You want to obtain the rows where the value of `CalledNumber` is 123456.

Table Store sorts all rows based on primary keys. Because `CalledNumber` is a predefined column, queries that involve this column are slow and inefficient. Therefore, you create an index based on `CalledNumber` to improve query speed and efficiency.

`IndexOnBeCalledNumber`:

PK0	PK1	PK2
CalledNumber	CellNumber	StartTime
123456	234567	1532574734
123456	345678	1532574795
123456	345678	1532574861
654321	123456	1532574644
765432	234567	1532574714
345678	456789	1532584054



Note:

Table Store automatically adds all primary key columns of the base table to the index. In other words, the global secondary index contains the columns you specify as the primary keys, and the primary key columns of the base table. Therefore, the global secondary index contains three primary key columns.

The primary key of the `IndexOnBeCalledNumber` is `CalledNumber`. You can perform a query on this index to obtain the rows where the value of `CalledNumber` is 123456.

```
private static void getRangeFromIndexTable(SyncClient client, long cellNumber) {
    RangeRowQueryCriteria rangeRowQueryCriteria = new RangeRowQueryCriteria(
        INDEX0_NAME);

    // Specify the primary key to start from.
    PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimary
        KeyBuilder();
    startPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_1, PrimaryKey
        Value.fromLong(cellNumber));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKey
        Value.INF_MIN);
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKey
        Value.INF_MIN);
    rangeRowQueryCriteria.setInclusiveStartPrimaryKey(startPrimaryKeyBuilder.build());

    // Specify the primary key to end with.
    PrimaryKeyBuilder endPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimary
        KeyBuilder();
}
```

```

endPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_1, PrimaryKey
Value.fromLong(cellNumber));
endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKey
Value.INF_MAX);
endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKey
Value.INF_MAX);
rangeRowQueryCriteria.setExclusiveEndPrimaryKey(endPrimaryKeyBuilder.build());

rangeRowQueryCriteria.setMaxVersions(1);

String strNum = String.format("%d", cellNumber);
System.out.println("The cell number" + strNum + "was called by the following
numbers:");
while (true) {
    GetRangeResponse getRangeResponse = client.getRange(new GetRangeRequest(
rangeRowQueryCriteria));
    for (Row row : getRangeResponse.getRows()) {
        System.out.println(row);
    }

    // If the nextStartPrimaryKey value is not null, continue to read data from the base
table.
    if (getRangeResponse.getNextStartPrimaryKey() != null) {
        rangeRowQueryCriteria.setInclusiveStartPrimaryKey(getRangeResponse.
getNextStartPrimaryKey());
    } else {
        break;
    }
}
}

```

- You want to obtain the rows where the value of BaseStationNumber is 002 and the value of StartTime is 1532574740.

This query specifies BaseStationNumber and StartTime as conditions. Therefore, you can create a compound index based on the BaseStationNumber and StartTime columns.

IndexOnBaseStation1:

PK0	PK1	PK2
BaseStationNumber	StartTime	CellNumber
001	1532574644	123456
001	1532574714	234567
002	1532574795	345678
002	1532574861	345678
003	1532574734	234567

PK0	PK1	PK2
003	1532584054	456789

The following code provides an example on how to query the `IndexOnBaseStation1` index table:

```
private static void getRangeFromIndexTable(SyncClient client,
                                           long baseStationNumber,
                                           long startTime) {
    RangeRowQueryCriteria rangeRowQueryCriteria = new RangeRowQueryCriteria(
        INDEX1_NAME);

    // Specify the primary key to start from.
    PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrim
    aryKeyBuilder();
    startPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKey
    Value.fromLong(baseStationNumber));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKey
    Value.fromLong(startTime));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKey
    Value.INF_MIN);
    rangeRowQueryCriteria.setInclusiveStartPrimaryKey(startPrimaryKeyBuilder.build());

    // Specify the primary key to end with.
    PrimaryKeyBuilder endPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBui
    lder();
    endPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKey
    Value.fromLong(baseStationNumber));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKey
    Value.INF_MAX);
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKey
    Value.INF_MAX);
    rangeRowQueryCriteria.setExclusiveEndPrimaryKey(endPrimaryKeyBuilder.build());

    rangeRowQueryCriteria.setMaxVersions(1);

    String strBaseStationNum = String.format("%d", baseStationNumber);
    String strStartTime = String.format("%d", startTime);
    System.out.println("All called numbers forwarded by the base station" + strBaseSta
    tionNum + "that start from" + strStartTime + "are listed as follows:");
    while (true) {
        GetRangeResponse getRangeResponse = client.getRange(new GetRangeRequest(
            rangeRowQueryCriteria));
        for (Row row : getRangeResponse.getRows()) {
            System.out.println(row);
        }

        // If the nextStartPrimaryKey value is not null, continue to read data from the base
        table.
        if (getRangeResponse.getNextStartPrimaryKey() != null) {
            rangeRowQueryCriteria.setInclusiveStartPrimaryKey(getRangeResponse.
            getNextStartPrimaryKey());
        } else {
            break;
        }
    }
}
```

```
}
```

- You want to obtain the rows where the value of `BaseStationNumber` is `003` and the value of `StartTime` ranges from `1532574861` to `1532584054` and return only the `Duration` column.

In this query, you specify both `BaseStationNumber` and `StartTime` as conditions, but only the `Duration` column is returned. You can initiate a query on the previous index, and then obtain `Duration` by querying the base table.

```
private static void getRowFromIndexAndMainTable(SyncClient client,
                                                long baseStationNumber,
                                                long startTime,
                                                long endTime,
                                                String colName) {
    RangeRowQueryCriteria rangeRowQueryCriteria = new RangeRowQueryCriteria(
INDEX1_NAME);

    // Specify the primary key to start from.
    PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrim
aryKeyBuilder();
    startPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKey
Value.fromLong(baseStationNumber));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKey
Value.fromLong(startTime));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKey
Value.INF_MIN);
    rangeRowQueryCriteria.setInclusiveStartPrimaryKey(startPrimaryKeyBuilder.build());

    // Specify the primary key to end with.
    PrimaryKeyBuilder endPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuil
der();
    endPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKey
Value.fromLong(baseStationNumber));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKey
Value.fromLong(endTime));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKey
Value.INF_MAX);
    rangeRowQueryCriteria.setExclusiveEndPrimaryKey(endPrimaryKeyBuilder.build());

    rangeRowQueryCriteria.setMaxVersions(1);

    String strBaseStationNum = String.format("%d", baseStationNumber);
    String strStartTime = String.format("%d", startTime);
    String strEndTime = String.format("%d", endTime);

    System.out.println("The duration of calls forwarded by the base station" +
strBaseStationNum + "from" + strStartTime + "to" + strEndTime + "is listed as follows:");
    while (true) {
        GetRangeResponse getRangeResponse = client.getRange(new GetRangeRequest(
rangeRowQueryCriteria));
        for (Row row : getRangeResponse.getRows()) {
            PrimaryKey curIndexPrimaryKey = row.getPrimaryKey();
            PrimaryKeyColumn mainCalledNumber = curIndexPrimaryKey.getPrimary
KeyColumn(PRIMARY_KEY_NAME_1);
            PrimaryKeyColumn callStartTime = curIndexPrimaryKey.getPrimaryKeyColumn(
PRIMARY_KEY_NAME_2);
            PrimaryKeyBuilder mainTablePKBuilder = PrimaryKeyBuilder.createPrim
aryKeyBuilder();
```



```

        mainTablePKBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, mainCalled
Number.getValue());
        mainTablePKBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, callStartT
ime.getValue());
        PrimaryKey mainTablePK = mainTablePKBuilder.build(); // Specify primary keys
for the base table.

        // Query the base table.
        SingleRowQueryCriteria criteria = new SingleRowQueryCriteria(TABLE_NAME,
mainTablePK);
        criteria.addColumnsToGet(colName); // Read the Duration column value of the
base table.
        // Set MaxVersions to 1 to read the latest data version.
        criteria.setMaxVersions(1);
        GetRowResponse getRowResponse = client.getRow(new GetRowRequest(
criteria));
        Row mainTableRow = getRowResponse.getRow();

        System.out.println(mainTableRow);
    }

    // If the nextStartPrimaryKey value is not null, continue to read data from the base
table.
    if (getRangeResponse.getNextStartPrimaryKey() != null) {
        rangeRowQueryCriteria.setInclusiveStartPrimaryKey(getRangeResponse.
getNextStartPrimaryKey());
    } else {
        break;
    }
}
}

```

To improve query performance, you can create a compound index based on `BaseStationNumber` and `StartTime` and specify `Duration` as an attribute column of the index table.

The following index table is created.

IndexOnBaseStation2:

PK0	PK1	PK2	Defined0
BaseStationNumber	StartTime	CellNumber	Duration
001	1532574644	123456	60
001	1532574714	234567	10
002	1532574795	345678	5
002	1532574861	345678	100
003	1532574734	234567	20

PK0	PK1	PK2	Defined0
003	1532584054	456789	200

The following code provides an example on how to query the IndexOnBaseStation2 index table:

```
private static void getRangeFromIndexTable(SyncClient client,
                                           long baseStationNumber,
                                           long startTime,
                                           long endTime,
                                           String colName) {
    RangeRowQueryCriteria rangeRowQueryCriteria = new RangeRowQueryCriteria(
        INDEX2_NAME);

    // Specify the primary key to start from.
    PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrim
    aryKeyBuilder();
    startPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKey
    Value.fromLong(baseStationNumber));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKey
    Value.fromLong(startTime));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKey
    Value.INF_MIN);
    rangeRowQueryCriteria.setInclusiveStartPrimaryKey(startPrimaryKeyBuilder.build());

    // Specify the primary key to end with.
    PrimaryKeyBuilder endPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBui
    lder();
    endPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKey
    Value.fromLong(baseStationNumber));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKey
    Value.fromLong(endTime));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKey
    Value.INF_MAX);
    rangeRowQueryCriteria.setExclusiveEndPrimaryKey(endPrimaryKeyBuilder.build());

    // Specify the name of the column to be read.
    rangeRowQueryCriteria.addColumnsToGet(colName);

    rangeRowQueryCriteria.setMaxVersions(1);

    String strBaseStationNum = String.format("%d", baseStationNumber);
    String strStartTime = String.format("%d", startTime);
    String strEndTime = String.format("%d", endTime);

    System.out.println("The duration of calls forwarded by the base station" +
        strBaseStationNum + "from" + strStartTime + "to" + strEndTime + "is listed as follows:");
    while (true) {
        GetRangeResponse getRangeResponse = client.getRange(new GetRangeRequest(
            rangeRowQueryCriteria));
        for (Row row : getRangeResponse.getRows()) {
            System.out.println(row);
        }

        // If the nextStartPrimaryKey value is not null, continue to read data from the base
        table.
        if (getRangeResponse.getNextStartPrimaryKey() != null) {
            rangeRowQueryCriteria.setInclusiveStartPrimaryKey(getRangeResponse.
            getNextStartPrimaryKey());
        } else {

```

```
        break;
    }
}
```

If you do not specify `Duration` as an attribute column for an index table, you have to retrieve `Duration` by querying the base table. However, when you specify `Duration` as an attribute column for an index table, this column is stored in both the base table and the index table. The configuration improves query performance at the cost of storage space consumption.

- You want to obtain the total call duration, the average call duration, the maximum call duration, and the minimum call duration of all calls forwarded by the base station 003 and whose call start time ranges from 1532574861 to 1532584054.

In this query, you want to obtain the statistics about the duration of all phone calls instead of the duration of each call queried in the previous scenario. You can obtain results by using the same method as in the previous query. Then you can perform calculations on the `Duration` column to obtain the required result. You can also use SQL-on-OTS to directly return the final statistical results without the need for client computing. You can use most MySQL syntax in SQL-on-OTS. Additionally, with SQL-on-OTS, you can easily process complicated calculations that are applicable to your business.

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.

ApsaraDB RDS for MySQL

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 the following advanced features:

- **Read-only instance:** 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.
- **Read/write splitting:** The read/write splitting feature provides an extra read/write splitting endpoint. This endpoint enables an automatic link for the primary instance and all its read-only instances. An application can use this method to read and write data by connecting to the read/write splitting endpoint. Write requests are automatically distributed to the primary instance while read requests are distributed to read-only instances based on their weights. To scale up the reading capacity of the system, you can add more read-only instances.

6.2 Benefits

6.2.1 Ease of use

ApsaraDB for RDS is a ready-to-use service that features on-demand upgrades, simple management, high transparency, and high compatibility.

Ready-to-use

You can use the API to create instances of any RDS instance type.

On-demand upgrade

When the database load or data storage capacity changes, you can upgrade an RDS instance by changing its type. The upgrades do not interrupt the data link service.

Transparency and compatibility

You can use ApsaraDB for RDS in the same way as native database engines. ApsaraDB for RDS is easily adopted, and no new knowledge is required. ApsaraDB for RDS is compatible with existing programs and tools. Data can be migrated to ApsaraDB for RDS through ordinary import and export tools.

Easy management

Alibaba Cloud is responsible for the routine maintenance and management tasks for ApsaraDB for RDS such as troubleshooting hardware and software issues or issuing database patches and updates. You can also manually add, delete, restart, back up, and restore databases by using the Apsara Stack Cloud Management (ASCM) console.

6.2.2 High performance

ApsaraDB for RDS implements parameter optimization, SQL optimization, and high-end back-end hardware to achieve high performance.

Parameter optimization

All RDS instance parameters have been optimized over their several years of production . Professional database administrators continue to optimize RDS instances over their lifecycles to ensure that ApsaraDB for RDS runs at peak efficiency.

SQL optimization

ApsaraDB for RDS locks inefficient SQL statements and provides recommendations to optimize code.

High-end back-end hardware

All servers used by ApsaraDB for RDS are evaluated by multiple parties to ensure stability.

6.2.3 High security

ApsaraDB for RDS implements anti-DDoS protection, access control, system security, and transparent data encryption (TDE) to guarantee the security of your databases.

DDoS attack prevention

**Note:**

You must activate Alibaba Cloud security services to use this feature.

When you access an ApsaraDB for RDS instance from the Internet, the instance is vulnerable to DDoS attacks. When a DDoS attack is detected, the RDS security system first scrubs the inbound traffic. If traffic scrubbing is not sufficient or if the traffic exceeds a specified threshold, black hole filtering is triggered.

Access control

You can configure an IP address whitelist for ApsaraDB for RDS to allow access for specified IP addresses and deny access for all others.

Each account can only view and operate their own respective database.

System security

ApsaraDB for RDS is protected by several layers of firewalls capable of blocking a variety of attacks to secure data.

ApsaraDB for RDS servers cannot be logged onto directly. Only the ports required for specific database services are provided.

ApsaraDB for RDS servers cannot initiate an external connection. They can only receive access requests.

TDE

Transparent Data Encryption (TDE) can be used to perform real-time I/O encryption and decryption on instance data files. Data is encrypted before it is written to disks and decrypted before it is read from disks to the memory. TDE will not increase the size of data files. Developers do not need to modify their applications before using the TDE feature.

6.2.4 High reliability

ApsaraDB for RDS provides hot standby, multi-copy redundancy, data backup, and data recovery to achieve high reliability.

Hot standby

ApsaraDB for RDS adopts a hot standby architecture. If the primary server fails, services will fail over to the secondary server within seconds. Applications running on the servers are not affected by the failover process and will continue to run normally.

Multi-copy redundancy

ApsaraDB for RDS servers implement a RAID architecture to store data. Data backup files are stored on OSS.

Data backup

ApsaraDB for RDS provides an automatic backup mechanism. You can schedule backups to be performed periodically, or manually initiate temporary backups as necessary to meet your business needs.

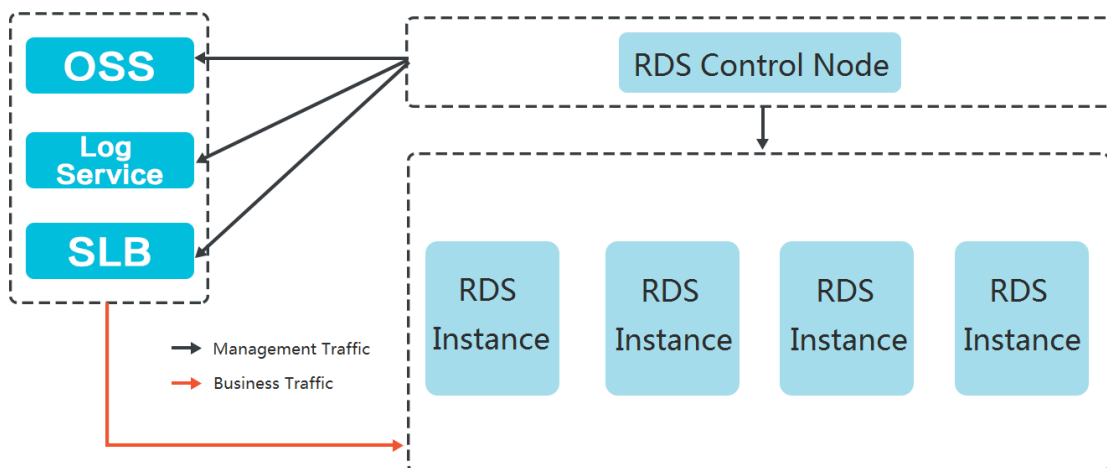
Data recovery

Data can be restored from backup sets or cloned instances created at previous points in time. After data is verified, the data can be migrated back to the primary RDS instance.

6.3 Architecture

The following figure shows the system architecture of ApsaraDB for RDS.

Figure 6-1: RDS system architecture



6.4 Features

6.4.1 Data link service

ApsaraDB for RDS provides all data link services, including DNS, Server Load Balancer (SLB).

ApsaraDB for RDS uses native database engines with similar database operations to minimize learning costs and facilitate database access.

DNS The DNS module can dynamically resolve domain names to IP addresses. Therefore, IP address changes do not affect the performance of RDS instances. After the domain name of an RDS instance is configured in the connection pool, the RDS instance can be accessed even if its corresponding IP address changes.

For example, the domain name of an RDS instance is **test.rds.aliyun.com**, and its corresponding IP address is **10.10.10.1**. The instance can be accessed when either **test.rds.aliyun.com** or **10.10.10.1** is configured in the connection pool of a program.

After a zone migration or version upgrade is performed for this RDS instance, the IP address may change to **10.10.10.2**. If the domain name **test.rds.aliyun.com** is configured in the connection pool, the instance can still be accessed. However, if the IP address **10.10.10.1** is configured in the connection pool, the instance will no longer be accessible.

SLB The SLB module provides both the internal IP address and public IP address of an RDS instance. Therefore, server changes do not affect the performance of the instance.

For example, the internal IP address of an RDS instance is **10.1.1.1**, and the corresponding Proxy or DB Engine runs on **192.168.0.1**. The SLB module typically redirects all traffic destined for **10.1.1.1** to **192.168.0.1**. If **192.168.0.1** fails, another server in the hot standby state with the IP address **192.168.0.2** will take over for the initial server. In this case, the SLB module will redirect all traffic destined for **10.1.1.1** to **192.168.0.2**, and the RDS instance will continue to provide services normally.

6.4.2 High-availability service

The high-availability (HA) service consists of modules such as the Detection, Repair, and Notice.

The HA service guarantees the availability of data link services and processes internal database exceptions.

Detection

The Detection module checks whether the primary and secondary nodes of the DB Engine are providing services normally. The HA node uses heartbeat information taken at 8 to 10 second intervals to determine the health status of the primary node. This information, along with the health status of the secondary node and heartbeat information from other HA nodes, provides a reference for the Detection module. All this information helps the module avoid misjudgment caused by exceptions such as network jitter. Failover can be completed quickly.

Repair

The Repair module maintains the replication relationship between the primary and secondary nodes of the DB Engine. It can also correct errors that occur on either node during normal operations.

For example:

- It can automatically restore primary/secondary replication after a disconnection.
- It can automatically repair table-level damage to the primary or secondary node.
- It can save and automatically repair the primary or secondary node in case of crashes.

Notice

The Notice module informs the SLB or Proxy module of status changes to the primary and secondary nodes to ensure that you always access the correct node.

For example, the Detection module discovers problems with the primary node and instructs the Repair module to resolve these problems. If the Repair module fails to resolve a

problem, it instructs the Notice module to perform traffic switchover. The Notice module forwards the switching request to the SLB or Proxy module, and then all traffic is redirected to the secondary node. Meanwhile, the Repair module creates a new secondary node on a different physical server and synchronizes this change back to the Detection module. The Detection module rechecks the health status of the instance.

HA policies

Each HA policy defines a combination of service priorities and data replication modes defined to meet the needs of your business.

There are two service priorities:

- **Recovery time objective (RTO):** The database preferentially restores services to maximize the availability time. Use the RTO policy if you require longer database uptime.
- **Recovery point objective (RPO):** The database preferentially ensures data reliability to minimize data loss. Use the RPO policy if you require high data consistency.

There are three data replication modes:

- **Asynchronous replication (Async):** When an application initiates an update request such as add, delete, or modify operations, the primary node responds to the application immediately after the primary node completes the operation. The primary node then replicates data to the secondary node asynchronously. This means that the operation of the primary database is not affected if the secondary node is unavailable. Data inconsistencies may occur if the primary node is unavailable.
- **Forced synchronous replication (Sync):** When an application initiates an update request such as add, delete, or modify operations, the primary node replicates data to the secondary node immediately after the primary node completes the operation. The primary node then waits for the secondary node to return a success message before the primary node responds to the application. The primary node replicates data to the secondary node synchronously. Unavailability of the secondary node will affect the operation on the primary node. Data will remain consistent even when the primary node is unavailable.
- **Semi-synchronous replication (Semi-Sync):** Data is typically replicated in Sync mode. When trying to replicate data to the secondary node, if an exception occurs causing the primary and secondary nodes to be unable to communicate with each other, the primary node will suspend response to the application. If the connection cannot be restored, the primary node will degrade to Async mode and restore response to the

application after the Sync replication times out. In a situation such as this, the primary node becoming unavailable will lead to data inconsistency. After the secondary node or network connection is recovered, data replication between the two nodes is resumed, and the data replication mode will change from Async to Sync.

You can select different combinations of service priorities and data replication modes to improve availability based on the business features.

6.4.3 Backup and recovery service

This service supports data backup, storage, and recovery functions.

ApsaraDB for RDS can back up databases at any time and restore them to any point in time based on the backup policy, which makes the data more traceable.

Backup

The Backup module compresses and uploads data and logs on both the primary and secondary nodes. ApsaraDB for RDS uploads backup files to OSS by default. When the secondary node is operating properly, backups are always created on the secondary node. This way, the services on the primary node are not affected. When the secondary node is unavailable or damaged, the Backup module creates backups on the primary node.

Recovery

The Recovery module restores backup files from OSS to a destination node.

- Primary node rollback: rolls back the primary node to a specified point in time if an operation error occurs.
- Secondary node repair: creates a new secondary node to reduce risks when an irreparable fault occurs on the secondary node.
- Read-only instance creation: creates a read-only instance from backup files.

Storage

The Storage module uploads, dumps, and downloads backup files. All backup data is uploaded to OSS for storage. You can obtain temporary links to download backups as necessary. In certain scenarios, the Storage module allows you to dump backup files from OSS to Archive Storage for more cost-effective and longer-term offline storage.

6.4.4 Monitoring service

ApsaraDB for RDS provides multilevel monitoring services across the physical, network, and application layers to ensure service availability.

Service

The Service module tracks the status of services that RDS depends on, such as SLB, OSS, log service, and Archive Storage, to ensure they are operating properly. Monitored metrics include functionality and response time. The Service module also uses logs to determine whether the internal RDS services are operating properly.

Network

The Network module tracks statuses at the network layer. It monitors the connectivity between ECS and RDS and between physical RDS servers. It also monitors the rates of packet loss on the VRouter and VSwitch.

OS

The OS module tracks the status of hardware and the OS kernel. The monitored items include:

- **Hardware maintenance:** The OS module constantly checks the operating status of the CPU, memory, motherboard, and storage device. It can predict faults in advance and automatically submit repair reports when it determines a fault is likely to occur.
- **OS kernel monitoring:** The OS module tracks all database calls and analyzes the causes of slow calls or call errors based on the kernel status.

Instance

The Instance module collects the following information on RDS instances:

- Instance availability information
- Instance capacity and performance metrics
- Instance SQL execution records

6.4.5 Scheduling service

The Resource module implements the scheduling of resources and services.

Resource

The Resource module allocates and integrates underlying RDS resources when you activate and migrate instances. When you use the RDS console or API to create an instance, the Resource module calculates the most suitable host to carry the traffic to and from the

instance. This module also allocates and integrates the underlying resources required to migrate RDS instances. After repeated instance creation, deletion, and migration operations, the Resource module calculates the degree of resource fragmentation. It also regularly integrates resources to improve the service carrying capacity.

6.4.6 Migration service

ApsaraDB for RDS provides Data Transmission Service (DTS) to help you migrate databases quickly.

The migration service helps you migrate data from the on-premises database to ApsaraDB for RDS, or migrate data between RDS instances.

DTS

DTS enables data migration from on-premises databases to RDS instances or between RDS instances.

DTS provides three migration methods: schema migration, full migration, and incremental migration.

- Schema migration

DTS migrates the schema definitions of migration objects to the destination instance.

Tables, views, triggers, stored procedures, and stored functions can be migrated in this mode.

- Full migration

DTS migrates all data of migration objects from the source database to the destination instance.

**Notice:**

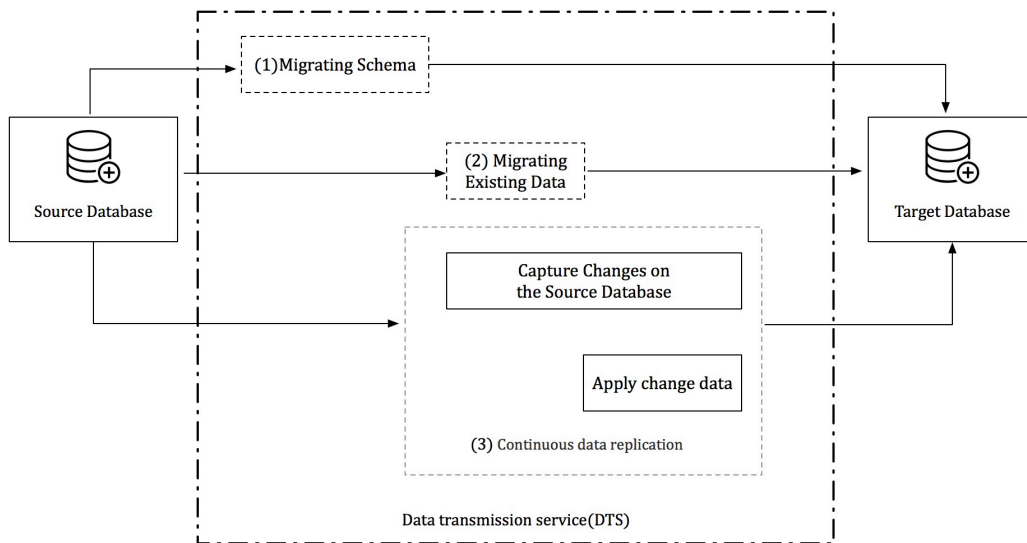
To ensure data consistency, non-transaction tables that do not have primary keys will be locked when performing a full migration. You cannot write data to locked tables. The lock duration depends on the amount of data in the tables. The tables are unlocked only after they are fully migrated.

- Incremental migration

DTS synchronizes data changes made in the migration process to the destination instance.

**Notice:**

If a DDL operation is performed during data migration, schema changes will not be synchronized to the destination instance.



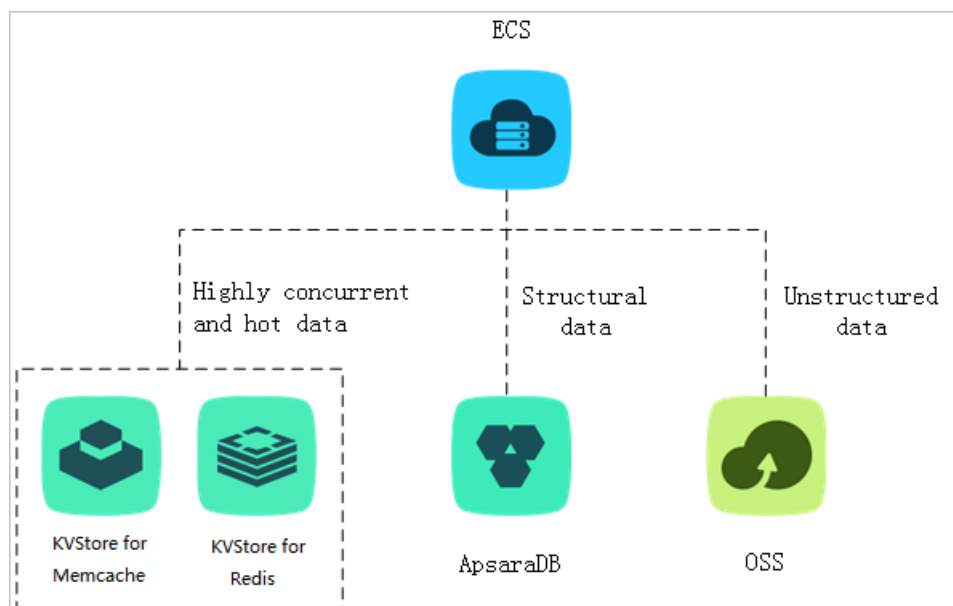
6.5 Scenarios

6.5.1 Diversified data storage

ApsaraDB for RDS provides cache data persistence and multi-structure data storage.

You can diversify the storage capabilities of ApsaraDB for RDS by using services such as KVStore for Redis and OSS, as shown in [Figure 6-2: Diversified data storage](#).

Figure 6-2: Diversified data storage



Cache data persistence

ApsaraDB for RDS can be used with KVStore for Memcache and KVStore for Redis to form a high-throughput and low-latency storage solution. These cache services have the following benefits over ApsaraDB for RDS:

- High response speed: The request latency of KVStore for Memcache and KVStore for Redis is only a few milliseconds.
- The cache area supports a higher number of queries per second (QPS) than ApsaraDB for RDS.

Multi-structure data storage

OSS is a secure, reliable, low-cost, and high-capacity storage service offered by Alibaba Cloud. ApsaraDB for RDS can be used with OSS to implement a multi-type data storage solution. For example, imagine ApsaraDB for RDS and OSS are used together to set up an online forum. Resources such as the posts and images uploaded to the forum can be stored in OSS to reduce storage needs on ApsaraDB for RDS.

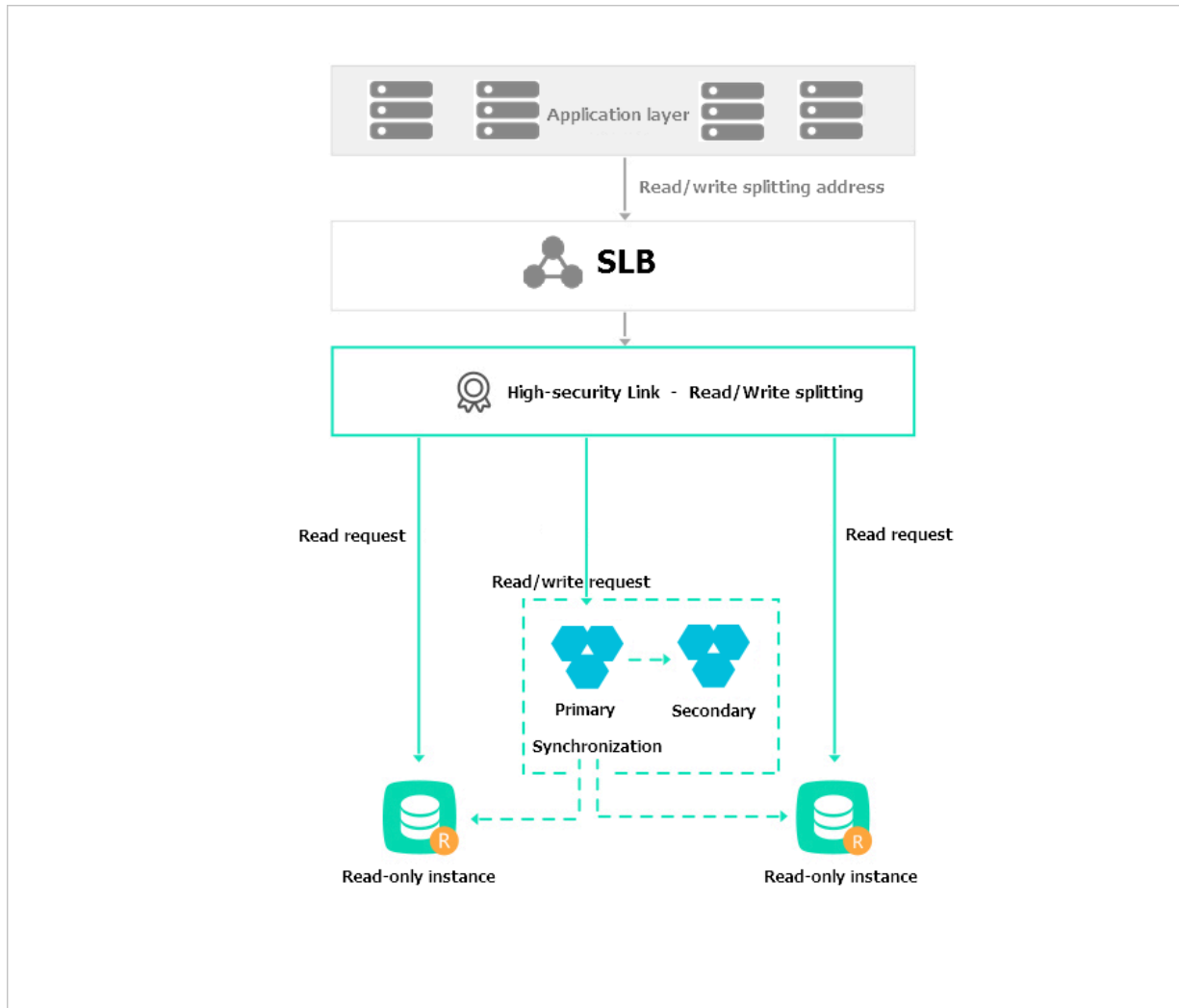
6.5.2 Read/write splitting

This feature allows you to split read requests and write requests across different instances to expand the processing capability of the system.

ApsaraDB RDS for MySQL allows you to directly attach read-only instances to ApsaraDB for RDS to reduce read pressure on the primary instance. The primary instance and read-only instances of ApsaraDB RDS for MySQL each have their own connection addresses. The system also offers an extra read/write splitting address after read/write splitting is enabled. This address associates the primary instance with all of its read-only instances for automatic read/write splitting, allowing applications to send all read and write requests to a single address. Write requests are automatically routed to the primary instance, and read requests are routed to each read-only instance based on their weights. You can scale

out the processing capability of the system by adding more read-only instances. There is no need to modify applications, as shown in [Read/write splitting](#).

Figure 6-3: Read/write splitting

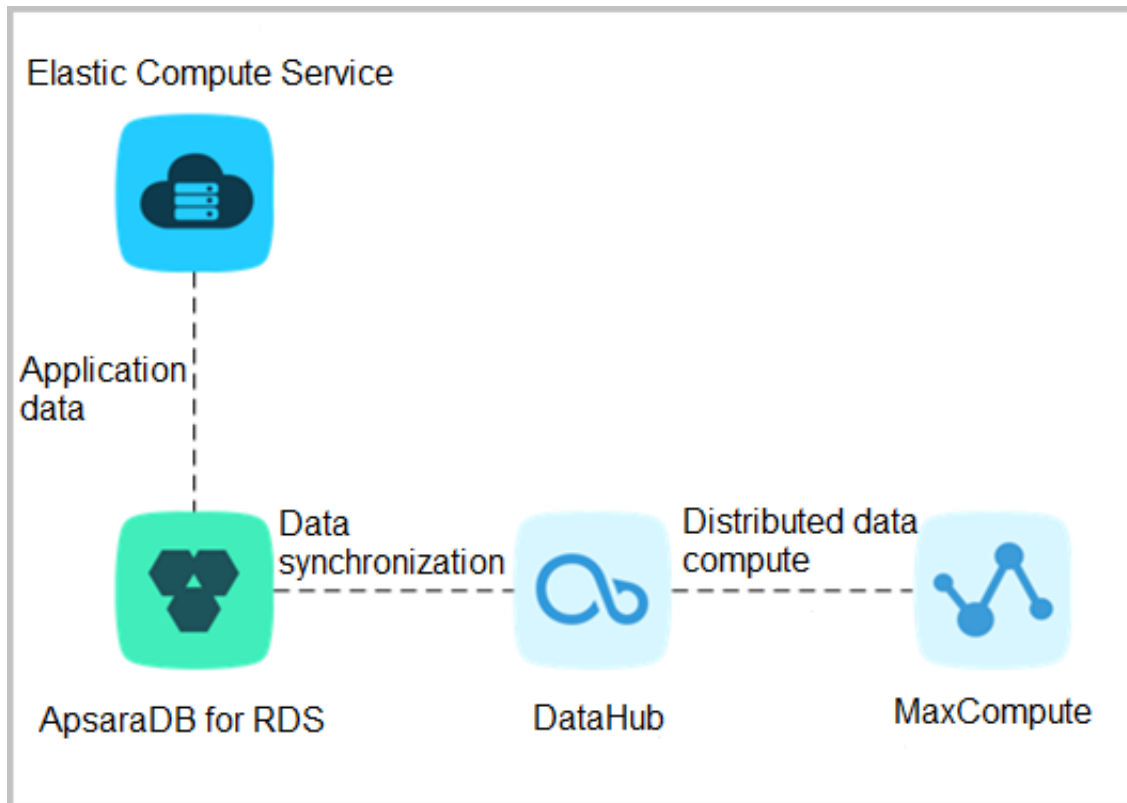


6.5.3 Big data analysis

You can import data from RDS to MaxCompute to enable large-scale data computing.

MaxCompute is used to store and compute batches of structured data. It provides various data warehouse solutions as well as big data analysis and modeling services, as shown in [Big data analysis diagram](#).

Figure 6-4: Big data analysis diagram



6.6 Limits

6.6.1 Limits on ApsaraDB RDS for MySQL

Before you use ApsaraDB RDS for MySQL, you must understand its limits and take the necessary precautions.

To guarantee instance stability and security, ApsaraDB RDS for MySQL has some service limits, as listed in [Table 6-1: Limits on ApsaraDB RDS for MySQL](#).

Table 6-1: Limits on ApsaraDB RDS for MySQL

Operation	Description
Database parameter modification	Database parameters can only 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> Logical restoration can be performed through the CLI or GUI. Physical restoration can only be performed through the RDS console or API operations.
Data import	<ul style="list-style-type: none"> Logical import can be performed through the CLI or GUI. Data can be imported through the MySQL CLI or DTS.
ApsaraDB RDS for MySQL storage engine	<ul style="list-style-type: none"> Only InnoDB and TokuDB are supported. Due to the inherent shortcomings of the MyISAM engine, some data may be lost. Only some existing instances use the MyISAM engine. MyISAM engine tables in newly created instances will be automatically converted to InnoDB engine tables. For safety performance and security considerations, we recommend that you use the InnoDB storage engine. The Memory engine is not supported. Newly created Memory tables will be automatically converted into InnoDB tables.
Database replication	ApsaraDB RDS for MySQL provides dual-node clusters based on a primary/secondary replication architecture. The secondary instances in this replication architecture are hidden and cannot be accessed directly.
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.

Operation	Description
Standard account	<ul style="list-style-type: none"> Custom authorization is not supported. The RDS console allows you to manage accounts and databases. Instances that support standard accounts also support privileged accounts.
Privileged account	<ul style="list-style-type: none"> Custom authorization is supported. 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.

6.7 Terms

Term	Description
Region	The geographical location where the server of your RDS instance resides . You must specify a region when you create an RDS instance. The region of an instance cannot be changed after instance creation. RDS must be used together with ECS and only supports intranet access. Because of this , RDS instances must be located in the same region as their corresponding ECS instances.
Zone	The physical area with an independent power supply and network in a region. Zones in a region can communicate through the intranet. Network latency for resources within the same zone is lower than for those across zones. Faults are isolated between zones. Single zone refers to the case where the three nodes in the RDS instance replica set are all located in the same zone. Network latency is reduced if an ECS instance and its corresponding RDS instance are both deployed in the same zone.
Instance	The most basic unit of RDS. An instance is the operating environment of ApsaraDB for RDS and works as an independent process on a host. You can create, modify, or delete an RDS instance from the RDS console . Instances are mutually independent and their resources are isolated. They do not compete for resources such as CPU, memory, or I/O. Each instance has its own features, such as database type and version. RDS controls instance behavior by using corresponding parameters.
Memory	The maximum amount of memory that can be used by an ApsaraDB for RDS instance.

Term	Description
Disk capacity	The amount of disk space selected when creating an ApsaraDB for RDS instance. Instance data that occupies disk space includes aggregated data as well as data required for normal instance operations such as system databases, database rollback logs, redo logs, and indexing. Ensure that the disk capacity is sufficient for the RDS instance to store data. Otherwise, the RDS instance may be locked. If the instance is locked due to insufficient disk capacity, you can unlock the instance by expanding the disk capacity.
IOPS	The maximum number of read/write operations performed per second on block devices at a granularity of 4 KB.
CPU core	The maximum computing capability of the instance. A single Intel Xeon series CPU core has at least 2.3 GHz of computational power with hyper-threading capabilities.
Number of connections	The number of TCP connections between a client and an RDS instance. If the client uses a connection pool, the connection between the client and RDS instance is a persistent connection. Otherwise, it is a transient connection.

6.8 Instance types

Instances of different editions, versions, and types each perform differently from one another.

Table 6-2: ApsaraDB RDS for MySQL instance types

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
High-availability Edition	5.6 or 5.7	Dedicated instance (X8)	mysql.x8.medium.2	2 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	Single-data center deployment and dual-data center deployment
			mysql.x8.large.2	4 cores, 32 GB	5,000	9,000	50 GB to 1,000 GB (in 5 GB increments)	
			mysql.x8.xlarge.2	8 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x8.2xlarge.2	16 cores, 128 GB	20,000	36,000	500 GB to 3,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
		Dedicated instance (X4)	mysql.x4.large.2	4 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	
			mysql.x4.xlarge.2	8 cores, 32 GB	5,000	9,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x4.2xlarge.2	16 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x4.4xlarge.2	32 cores, 128 GB	20,000	36,000	1,000 GB to 3,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
		Dedicated host	dds.mysql.st.d13	30 cores, 220 GB	64,000	20,000	1,000 GB to 3,000 GB (in 5 GB increments)	
Enterprise Edition	5.6	Dedicated instance (with high memory)	mysql.x8.medium.4	2 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	Dual-data center deployment
			mysql.x8.large.4	4 cores, 32 GB	5,000	9,000	50 GB to 1,000 GB (in 5 GB increments)	
			mysql.x8.xlarge.4	8 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
			mysql.x8.2xlarge.4	16 cores, 128 GB	20,000	36,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x8.4xlarge.4	32 cores, 256 GB	40,000	72,000	1,000 GB to 3,000 GB (in 5 GB increments)	
	5.7	Dedicated instance (with high CPU)	mysql.x4.large.25	4 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	Single-data center deployment
			mysql.x4.xlarge.25	8 cores, 32 GB	5,000	9,000	500 GB to 3,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
			mysql.x4.2xlarge.25	16 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x4.4xlarge.25	32 cores, 128 GB	20,000	36,000	1,000 GB to 3,000 GB (in 5 GB increments)	
		Dedicated instance (with high memory)	mysql.x8.medium.25	2 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	
			mysql.x8.large.25	4 cores, 32 GB	5,000	9,000	50 GB to 1,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
			mysql.x8.xlarge.25	8 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x8.2xlarge.25	16 cores, 128 GB	20,000	36,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x8.4xlarge.25	32 cores, 256 GB	40,000	72,000	1,000 GB to 3,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
Read-only instance	5.6 or 5.7	Dedicated instance (X8)	mysqlro.x8.medium.1	2 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	Single-data center deployment and dual-data center deployment
			mysqlro.x8.large.1	4 cores, 32 GB	5,000	9,000	50 GB to 1,000 GB (in 5 GB increments)	
			mysqlro.x8.xlarge.1	8 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysqlro.x8.2xlarge.1	16 cores, 128 GB	20,000	36,000	500 GB to 3,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
		Dedicated instance (X4)	mysqlro.x4.large.1	4 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	
			mysqlro.x4.xlarge.1	8 cores, 32 GB	5,000	9,000	50 GB to 1,000 GB (in 5 GB increments)	
			mysqlro.x4.2xlarge.1	16 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysqlro.x4.4xlarge.1	32 cores, 128 GB	20,000	36,000	500 GB to 3,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
		Dedicated host	dds.mysql.st.d13	30 cores, 220 GB	64,000	20,000	1,000 GB to 3,000 GB (in 5 GB increments)	

Table 6-3: ApsaraDB RDS for SQL Server instance types

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
High-availability Edition	2008 R2	Dedicated instance	mssql.x8.medium.2	2 cores, 16 GB	2,500	4,500	250 GB	Single-data center deployment and dual-data center deployment. This instance type is not available in new data centers, and does not support V5 servers.
			mssql.x8.large.2	4 cores, 32 GB	5,000	9,000	500 GB	
			mssql.x8.xlarge.2	8 cores, 64 GB	10,000	18,000	1,000 GB	
			mssql.x8.2xlarge.2	16 cores, 128 GB	20,000	36,000	2,000 GB	
		Dedicated host	dds.mssql.st.d13	30 cores, 220 GB	64,000	20,000	2 TB	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
	2012 EE: 2012_ent_ha 2016 EE: 2016_ent_ha	Dedicated instance family	mssql.x4.medium.e2	2 cores, 8 GB	No limit	Depends on the performance of cloud disks	20 GB to 4,000 GB	Single-data center deployment
			mssql.x8.medium.e2	2 cores, 16 GB				
			mssql.x4.large.e2	4 cores, 16 GB				
			mssql.x8.large.e2	4 cores, 32 GB				
			mssql.x4.xlarge.e2	8 cores, 32 GB				
			mssql.x8.xlarge.e2	8 cores, 64 GB				
			mssql.x4.2xlarge.e2	16 cores, 64 GB				
			mssql.x8.2xlarge.e2	16 cores, 128 GB				
			mssql.x4.3xlarge.e2	24 cores, 96 GB				
			mssql.x4.4xlarge.e2	32 cores, 128 GB				
			mssql.x8.4xlarge.e2	32 cores, 256 GB				
			mssql.x8.7xlarge.e2	56 cores, 480 GB				
			mssql.x4.8xlarge.e2	64 cores, 256 GB				
			mssql.x8.8xlarge.e2	64 cores, 512 GB				

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk capacity	Zone-disaster recovery deployment
	2012 SE: 2012_std_ha	Dedicated instance	mssql.x4.medium.s2	2 cores, 8 GB				
	mssql.x8.medium.s2		2 cores, 16 GB					
	mssql.x4.large.s2		4 cores, 16 GB					
	mssql.x8.large.s2		4 cores, 32 GB					
	mssql.x4.xlarge.s2		8 cores, 32 GB					
	mssql.x8.xlarge.s2		8 cores, 64 GB					
	mssql.x4.2xlarge.s2		16 cores, 64 GB					
	mssql.x8.2xlarge.s2		16 cores, 128 GB					
	mssql.x4.3xlarge.s2		24 cores, 96 GB					

7 AnalyticDB for PostgreSQL


7.1 What is AnalyticDB for PostgreSQL?





AnalyticDB for PostgreSQL (formerly known as HybridDB for PostgreSQL) is a distributed analytic database that adopts a massive parallel process (MPP) architecture and consists of multiple compute nodes. AnalyticDB for PostgreSQL provides MPP warehousing services and supports horizontal scaling of storage and compute capabilities, online analysis for petabyte levels of data, and offline extract, transform, and load (ETL) task processing.

AnalyticDB for PostgreSQL is developed based on the PostgreSQL kernel and has the following features:

- Supports the SQL:2003 standard, OLAP aggregate functions, views, Procedural Language for SQL (PL/SQL), user-defined functions (UDFs), and triggers. AnalyticDB for PostgreSQL is partially compatible with the Oracle syntax.
- Uses the horizontally scalable MPP architecture and supports range and list partitioning.
- Supports row store, column store, and multiple indexes. It also supports multiple compression methods based on column store to reduce storage costs.
- Supports standard database isolation levels and distributed transactions to ensure data consistency.
- Provides the vector computing engine and the CASCADE-based SQL optimizer to ensure high-performance SQL analysis capabilities.
- Supports the primary/secondary architecture to ensure dual-copy data storage.
- Provides online scaling, monitoring, and disaster recovery to reduce O&M costs.

7.2 Benefits

 Real-time analysis	Built on the MPP architecture for horizontal scaling and PB/s data processing. AnalyticDB for PostgreSQL supports the leading vector computing feature and intelligent indexes of column store. It also supports the CASCADE-based SQL optimizer to make complex queries without the need for tuning.
--	---

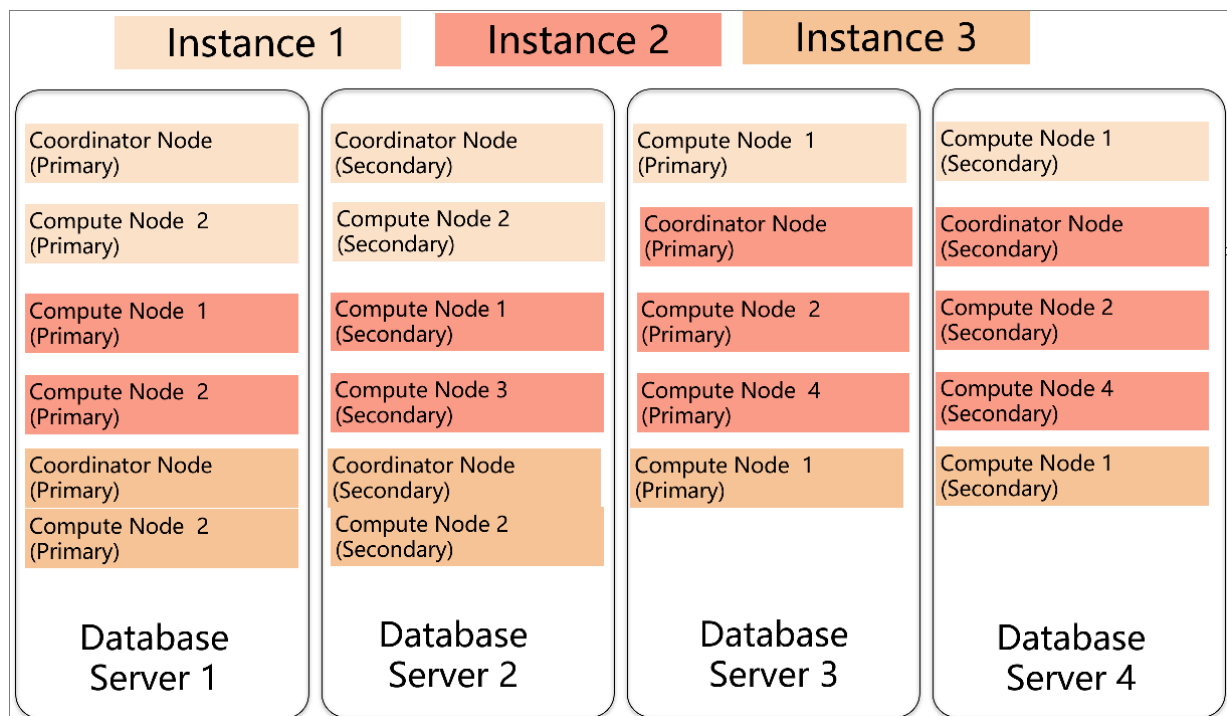
 Stability and reliability	Provides ACID properties for distributed transactions. Transactions are consistent across nodes and all data is synchronized between primary and secondary nodes. AnalyticDB for PostgreSQL supports distributed deployment and provides transparent monitoring, switching, and restoration to secure your data infrastructure.
 Easy to use	Supports a large number of SQL syntax and functions, Oracle functions, stored procedures, user-defined functions (UDFs), and isolation levels of transactions and databases. You can use popular BI software and ETL tools online.
 Ultra-high performance	Supports row store, column store, and multiple indexes. The vector engine provides high-performance analysis and computing capabilities. The CASCADE-based SQL optimizer enables complex queries without the need for tuning. It supports high-performance parallel import of data from OSS.
 Scalability	<p>Enables you to scale up compute nodes, CPU, memory, and storage resources on demand to improve OLAP performance.</p> <p>Supports transparent OSS operations. OSS offers a larger storage capacity for cold data that does not require online analysis.</p>

7.3 Architecture

Physical cluster architecture

The following figure shows the physical cluster architecture of AnalyticDB for PostgreSQL.

Figure 7-1: Physical cluster architecture



You can create multiple instances within a physical cluster of AnalyticDB for PostgreSQL.

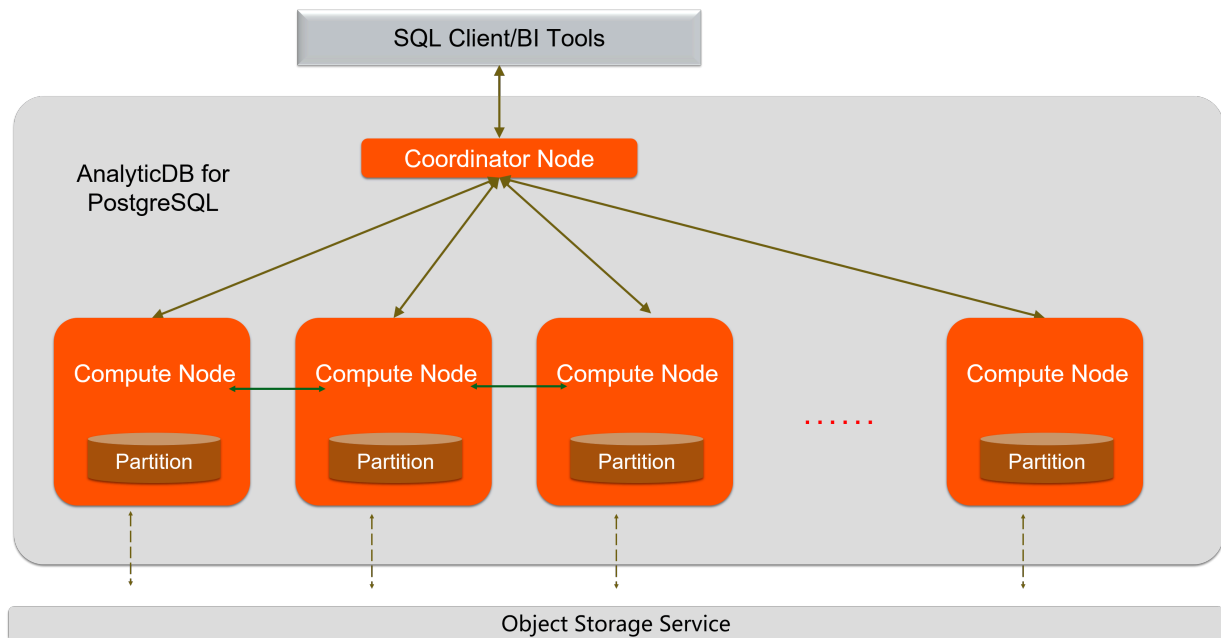
Each cluster includes two components: the coordinator node and the compute node.

- The coordinator node is used for access from applications. It receives connection requests and SQL query requests from clients and dispatches computing tasks to compute nodes. The cluster deploys a secondary node of the coordinator node on an independent physical server and replicates data from the primary node to the secondary node for failover. The secondary node does not accept external connections.
- Compute nodes are independent instances in AnalyticDB for PostgreSQL. Data is evenly distributed across compute nodes by hash value or RANDOM function, and is analyzed and computed in parallel. Each compute node consists of a primary node and a secondary node for automatic failover.

Logical architecture of an instance

You can create multiple instances within a cluster of AnalyticDB for PostgreSQL. The following figure shows the logical architecture of an instance.

Figure 7-2: Logical architecture of an instance



Data is distributed across compute nodes by hash value or RANDOM function of a specified distributed column. Each compute node consists of a primary node and a secondary node to ensure dual-copy storage. High-performance network communication is supported across nodes. When the coordinator node receives a request from the application, the coordinator node parses and optimizes SQL statements to generate a distributed execution plan. After the coordinator node sends the execution plan to the compute nodes, the compute nodes will perform an MPP execution of the plan.

7.4 Specification description

AnalyticDB for PostgreSQL supports two storage types: SSD storage and HDD storage. These storage types provide different features that are better suited for different scenarios.

- SSD storage: provides better I/O capabilities and higher analysis performance.
- HDD storage: provides larger and more affordable space to meet higher storage requirements.

Specifications

AnalyticDB for PostgreSQL supports the following compute node specifications:

Storage type	Core	Memory	Valid storage space	Total dual-copy space	Description
High-performance SSD	1	8 GB	80 GB	160 GB	This storage type is applicable to scenarios where the number of concurrent queries is less than five and the number of nodes of an instance is less than 32. This SSD configuration allows you to create an instance consisting of 4 to 32 nodes.
High-performance SSD	4	32 GB	320 GB	640 GB	The recommended type of high-performance SSD storage. This SSD configuration allows you to create an instance consisting of 8 to 2,048 nodes.
High-capacity HDD	2	16 GB	1 TB	2 TB	This storage type is applicable to scenarios where less than five concurrent queries are performed and the number of nodes of an instance is less than 8. This HDD configuration allows you to create an instance consisting of 4 to 32 nodes.

Storage type	Core	Memory	Valid storage space	Total dual-copy space	Description
High-capacity HDD	4	32 GB	2 TB	4 TB	The recommended type of high-capacity HDD storage. This HDD configuration allows you to create an instance consisting of 8 to 2,048 nodes.

7.5 Features

7.5.1 Distributed architecture

AnalyticDB for PostgreSQL is built on MPP architecture. Data is distributed evenly across nodes by hash value or RANDOM function, and is analyzed and computed in parallel. Storage and computing capacities are scaled horizontally as more nodes are added to ensure a quick response as the data volume increases.

AnalyticDB for PostgreSQL supports distributed transactions to ensure data consistency among nodes. It supports three transaction isolation levels: SERIALIZABLE, READ COMMITTED, and READ UNCOMMITTED.

7.5.2 High-performance data analysis

AnalyticDB for PostgreSQL supports column store and row store for tables. Row store provides high update performance and column store provides high OLAP aggregate analysis performance for tables. AnalyticDB for PostgreSQL supports the B-tree index, bitmap index, and hash index that enable high-performance analysis, filtering, and query.

AnalyticDB for PostgreSQL adopts the CASCADE-based SQL optimizer. AnalyticDB for PostgreSQL combines the cost-based optimizer (CBO) with the rule-based optimizer (RBO) to provide SQL optimization features such as automatic subquery decorrelation. These features enable complex queries without the need for tuning.

7.5.3 High-availability service

AnalyticDB for PostgreSQL builds a system based on the Apsara system of Alibaba Cloud for automatic monitoring, diagnostics, and error handling to reduce O&M costs.

The coordinator node compiles and optimizes SQL statements by storing database metadata and receiving query requests from clients. The coordinator node adopts a primary/secondary architecture to ensure strong consistency of metadata. If the primary coordinator node fails, the service will be automatically switched to the secondary coordinator node.

All compute nodes adopt a primary/secondary architecture to ensure strong data consistency between primary and secondary nodes when data is written into or updated. If the primary compute node fails, the service will be automatically switched to the secondary compute node.

7.5.4 Data synchronization and tools

You can use Data Transmission Service (DTS) or DataWorks to synchronize data from MySQL or PostgreSQL databases to AnalyticDB for PostgreSQL. Popular extract, transform, and load (ETL) tools can import ETL data and schedule jobs on AnalyticDB for PostgreSQL databases. You can also use standard SQL syntax to query data from formatted files stored in OSS by using external tables in real time.

AnalyticDB for PostgreSQL supports Business Intelligence (BI) reporting tools including Quick BI, DataV, Tableau, and FineReport. It also supports ETL tools, including Informatica and Kettle.

7.5.5 Data security

AnalyticDB for PostgreSQL supports IP whitelist configuration. You can add up to 1,000 IP addresses of servers to the whitelist to allow access to your instance and control risks from access sources. AnalyticDB for PostgreSQL also supports Anti-DDoS that monitors inbound traffic in real time. When large amounts of malicious traffic is identified, the traffic is scrubbed through IP filtering. If traffic scrubbing is not sufficient, the black hole process will be triggered.

7.5.6 Supported SQL features

- Supports row store and column store.
- Supports multiple indexes including the B-tree index, bitmap index, and hash index.
- Supports distributed transactions and standard isolation levels to ensure data consistency among nodes.
- Supports character, date, and arithmetic functions.

- Supports stored procedures, user-defined functions (UDFs), and triggers.
- Supports views.
- Supports range partitioning, list partitioning, and the definition of multi-level partitions.
- Supports multiple data types. The following table provides a list of data types and their information.

Parameter	Alias	Storage size	Range	Description
bigint	int8	8 bytes	-9223372036854775808 to 9223372036854775807	Large-range integer
bigserial	serial8	8 bytes	1 to 9223372036854775807	Large auto-increment integer
bit [(n)]	None	n bits	Bit string constant	Fixed-length bit string
bit varying [(n)]	varbit	Variable-length bit string	Bit string constant	Variable-length bit string
boolean	bool	1 byte	true/false, t/f, yes/no, y/n, 1/0	Boolean value (true/false)
box	None	32 bytes	((x1,y1),(x2,y2))	A rectangular box on a plane, not allowed in distribution key columns
bytea	None	1 byte + binary string	Sequence of octets	Variable-length binary string
character [(n)]	char [(n)]	1 byte + n	String up to n characters in length	Fixed-length, blank-padded string
character varying [(n)]	varchar [(n)]	1 byte + string size	String up to n characters in length	Variable length with limit
cidr	None	12 or 24 bytes	None	IPv4 and IPv6 networks
circle	None	24 bytes	<(x,y),r> (center and radius)	A circle on a plane, not allowed in distribution key columns
date	None	4 bytes	4,713 BC to 294,277 AD	Calendar date (year, month, day)
decimal [(p, s)]	numeric [(p, s)]	variable	No limit	User-specified precision, exact

Parameter	Alias	Storage size	Range	Description
double precision	float8	8 bytes	Precise to 15 decimal digits	Variable precision, inexact
	float			
inet	None	12 or 24 bytes	None	IPv4 and IPv6 hosts and networks
Integer	int or int4	4 bytes	-2.1E+09 to +2147483647	Typical choice for integer
interval [(p)]	None	12 bytes	-178000000 years to 178000000 years	Time span
json	None	1 byte + JSON size	JSON string	Unlimited variable length
lseg	None	32 bytes	((x1,y1),(x2,y2))	A line segment on a plane, not allowed in distribution key columns
macaddr	None	6 bytes	None	Media Access Control (MAC) addresses
money	None	8 bytes	-92233720368547758.08 to +92233720368547758.07	Currency amount
path	None	16+16n bytes	[(x1,y1),...]	A geometric path on a plane, not allowed in distribution key columns
point	None	16 bytes	(x,y)	A geometric point on a plane, not allowed in distribution key columns
polygon	None	40+16n bytes	((x1,y1),...)	A closed geometric path on a plane, not allowed in distribution key columns
real	float4	4 bytes	Precise to 6 decimal digits	Variable precision, inexact
serial	serial4	4 bytes	1 to 2147483647	Auto-increment integer
smallint	int2	2 bytes	-32768 to 32767	Small-range integer
text	None	1 byte + string size	Variable-length string	Unlimited variable length

Parameter	Alias	Storage size	Range	Description
time [(p)] [without time zone]	None	8 bytes	00:00:00[.000000] to 24:00:00[.000000]	Time of day (without time zone)
time [(p)] with time zone	timetz	12 bytes	00:00:00+1359 to 24:00:00-1359	Time of day (with time zone)
timestamp [(p)] [without time zone]	None	8 bytes	4,713 BC to 294,277 AD	Date and time
timestamp [(p)] with time zone	timestampz	8 bytes	4,713 BC to 294,277 AD	Date and time (with time zone)
xml	None	1 byte + XML size	Variable-length XML string	Unlimited variable length

- For more information about the supported standard SQL syntax, see [SQL syntax](#).

7.6 Scenarios

AnalyticDB for PostgreSQL is applicable to the following OLAP data analysis services.

- ETL for offline data processing

AnalyticDB for PostgreSQL has the following features that make it ideal for optimizing complex SQL queries and aggregating and analyzing large amounts of data:

- Supports standard SQL, OLAP window functions, and stored procedures.
- Provides the CASCADE-based SQL optimizer to make complex queries without the need for tuning.
- Built on the MPP architecture for horizontal scaling and PB/s data processing.
- Provides high performance, column store-based storage and aggregation of tables at a high compression ratio to save storage space.

- Online high-performance query

AnalyticDB for PostgreSQL provides the following benefits for real-time exploration, warehousing, and updating of data:

- Allows you to write and update high-throughput data through INSERT, UPDATE, and DELETE operations.
- Allows you to query data based on row store and multiple indexes (B-tree, bitmap, and hash) to obtain results in milliseconds.
- Supports distributed transactions, standard database isolation levels, and HTAP.

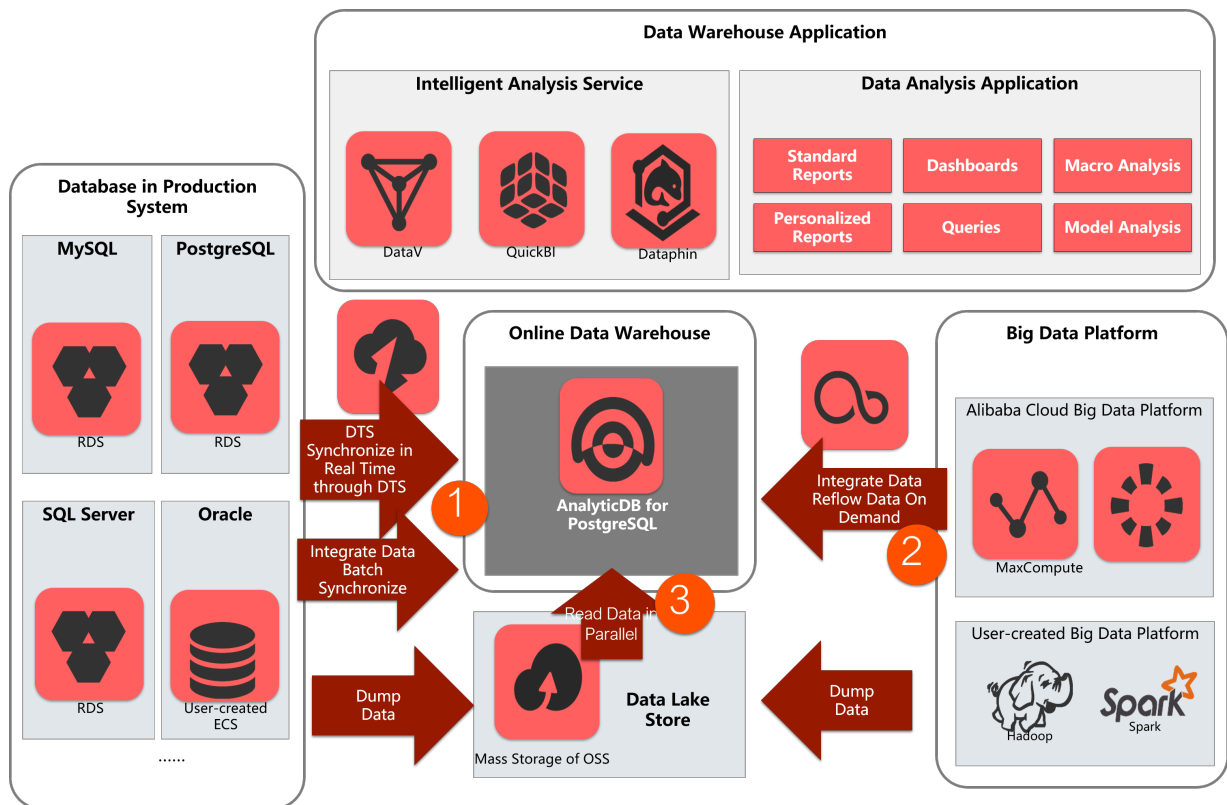
- Multi-model data analysis

AnalyticDB for PostgreSQL provides the following benefits for processing of a variety of unstructured data sources:

- Supports the PostGIS extension for geographic data analysis and processing.
- Uses the MADlib library of in-database machine learning algorithms to implement AI-native databases.
- Provides high-performance retrieval and analysis of unstructured data such as images, speech, and text through vector retrieval.
- Supports formats such as JSON and can process and analyze semi-structured data such as logs.

Typical scenarios

AnalyticDB for PostgreSQL is applicable to the following three scenarios:



- Data warehousing service

Data Transmission Service (DTS) can synchronize data in real time in production system databases such as ApsaraDB RDS for MySQL, ApsaraDB RDS for PostgreSQL, Apsara PolarDB, and traditional databases such as Oracle and SQL Server. Data can also be synchronized in batches to AnalyticDB for PostgreSQL through the data integration service (DataX). AnalyticDB for PostgreSQL supports complex extract, transform, and load (ETL) operations on large amounts of data. These tasks can also be scheduled by Dataworks. AnalyticDB for PostgreSQL also provides high-performance online analysis capabilities and can use Quick BI, DataV, Tableau, and FineReport for report presentation and real-time query.

- Big data analytics platform

You can import huge amounts of data from MaxCompute, Hadoop, and Spark to AnalyticDB for PostgreSQL through DataX or OSS for high-performance analysis, processing, and exploration.

- Data lake analytics

AnalyticDB for PostgreSQL can use an external table mechanism to access the huge amounts of data stored in OSS in parallel and build an Alibaba Cloud data lake analytics platform.

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 types of data, such as strings, lists, sets, sorted sets, and hash tables. The service also supports advanced features, such as transactions, message subscription, and message publishing. Based on the hybrid storage of memory and hard disks, KVStore for Redis can provide high-speed data read/write capability and support data persistence.

As a cloud computing service, KVStore for Redis works with hardware and data deployed in the cloud, and provides comprehensive infrastructure planning, network security protections, and system maintenance services. This service allows you to focus on business innovation.

8.2 Benefits

High performance

- Supports cluster features and provides cluster instances of 128 GB or higher to meet large capacity and high performance requirements.
- Provides primary/secondary instances of 32 GB or smaller to meet general capacity and performance requirements.

Elastic scaling

- Easy scaling of storage capacity: you can scale instance storage capacity in the KVStore for Redis console based on business requirements.
- Online scaling without interrupting services: you can scale instance storage capacity on the fly. This does not affect your business.

Resource isolation

Instance-level resource isolation provides enhanced stability for individual services.

Data security

- Persistent data storage: based on the hybrid storage of memory and hard disks, KVStore for Redis can provide high-speed data read/write capability and support data persistence.
- Dual-copy backup and failover: KVStore for Redis backs up data on both a primary node and a secondary node and supports the failover feature to prevent data loss.
- Access control: KVStore for Redis requires password authentication to ensure secure and reliable access.
- Data transmission encryption: KVStore for Redis supports encryption based on Secure Sockets Layer (SSL) and Secure Transport Layer (TLS) to secure data transmission.

High availability

- Primary/secondary structure: each instance runs in this structure to eliminate the possibility of single points of failure (SPOFs) and guarantee high availability.
- Automatic detection and recovery of hardware faults: the system automatically detects hardware faults and performs the failover operation within several seconds. This can minimize your business losses caused by unexpected hardware faults.

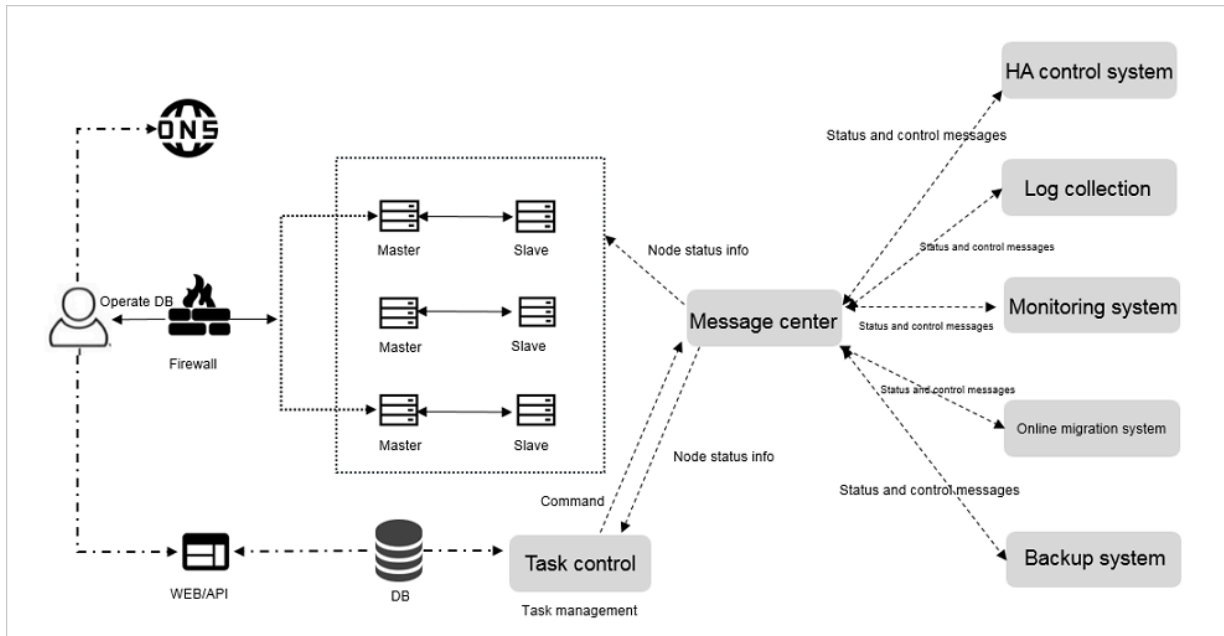
Easy to use

- Out-of-the-box service: KVStore for Redis requires no setup or installation. You can use the service immediately after purchase to ensure efficient business deployment.
- Compatible with open-source Redis: KVStore for Redis is compatible with Redis commands. You can use any Redis clients to easily connect to KVStore for Redis and perform data operations.

8.3 Architecture

The architecture of KVStore for Redis is as shown in [Figure 8-1: Architecture diagram](#).

Figure 8-1: Architecture diagram



KVStore for Redis automatically builds a primary/secondary structure. You can use this structure directly.

- **HA control system**

A high-availability (HA) detection module is used to detect and monitor the operating status of KVStore for Redis instances. If this module determines that a primary node is unavailable, the module automatically performs the failover operation to ensure high availability of KVStore for Redis instances.

- **Log collection**

This module collects instance operation logs, including slow query logs and access control logs.

- **Monitoring system**

This module collects performance monitoring information of KVStore for Redis instances , including basic group monitoring, key group monitoring, and string group monitoring.

- **Online migration system**

When an error occurs on the physical server that hosts a KVStore for Redis instance, this module recreates an instance on the fly based on the backup files stored in the backup system. This ensures high availability of your business.

- **Backup system**

This module generates backup files of KVStore for Redis instances, and stores the backup files in Object Storage Service (OSS). The backup system allows you to customize the backup settings, and retains backup files for up to seven days.

- **Task Control**

KVStore for Redis instances support various management and control tasks, including instance creation, specifications changes, and instance backups. The task system flexibly controls and tracks tasks and manages errors according to your instructions.

8.4 Features

- **High-availability technology ensures service stability**

The system synchronizes data between the primary node and the secondary node in real time. If the primary node fails, the system automatically performs the failover operation and restores services within a few seconds. The secondary node takes over services. This process does not affect your business, and ensures high availability of system services.

Cluster instances run in a distributed architecture. Each node uses a primary/secondary high-availability structure to automatically perform failover and disaster recovery and ensure high availability of system services.

- **Easy backup and recovery support custom backup policies**

You can back up data in the console and customize automatic backup policies. The system automatically retains backup data for seven days. You can easily restore data in the case of accidental data operations to minimize your business losses.

- **Multiple network security protections secure your data**

A Virtual Private Cloud (VPC) isolates network transmission at the transport layer. The Anti-Distributed-Denial-of-Service (DDoS) protection service monitors and protects against DDoS attacks. The system supports a whitelist that contains a maximum of 1,000 IP addresses or CIDR blocks to prevent malicious login attempts.

- **Kernel optimization avoids vulnerability exploits**

The experts of Alibaba Cloud have performed in-depth kernel optimization for the Redis source code to effectively prevent running out of memory, fix security vulnerabilities, and protect your business.

- **Elastic scaling eliminates capacity and performance bottlenecks**

KVStore for Redis supports multiple memory types. You can upgrade the memory type based on your service requirements.

The cluster architecture allows you to elastically scale the storage space and throughput performance of the database system. This eliminates the performance bottlenecks.

- **Multiple instance types support flexible specifications changes**

The single-node cache architecture and two-node storage architecture are applicable to various service scenarios. You can flexibly change instance specifications.

- **Monitoring and alerts allow you to check instance status in real time**

KVStore for Redis provides monitoring and alerts of instance information, such as CPU usage, connections, and disk utilization. You can check instance status anywhere and at any time.

- **Visual management simplifies operations and maintenance**

The KVStore for Redis console, a visual management platform, allows you to easily perform frequent and risky operations, such as instance cloning, backup, and data restoration.

- **Automatic engine version upgrades prevent software flaws**

The system automatically upgrades engine versions and efficiently fixes flaws so that you can easily manage database versions.

- **Custom parameters support individual requirements**

You can set parameters in the KVStore for Redis console to make full use of system resources.

8.5 Scenarios

Game industry applications

KVStore for Redis can be an important part of the business architecture for deploying a game application.

Scenario 1: KVStore for Redis works as a storage database

The architecture for deploying a game application is simple. You can deploy a main program on an ECS instance and all business data on a KVStore for Redis instance. The KVStore for Redis instance works as a persistent storage database. KVStore for Redis supports data persistence, and stores redundant data on primary and secondary nodes.

Scenario 2: KVStore for Redis works as a cache to accelerate connections to applications

KVStore for Redis can work as a cache to accelerate connections to applications. You can store data in a Relational Database Service (RDS) database that works as a backend database.

Reliability of the KVStore for Redis service is vital to your business. If the KVStore for Redis service is unavailable, the backend database is overloaded when processing connections to your application. KVStore for Redis provides a two-node hot standby architecture to ensure high availability and reliability of services. The primary node provides services for your business. If this node fails, the system automatically switches services to the secondary node. The complete failover process is transparent.

Live video applications

In live video services, KVStore for Redis works as an important measure to store user data and relationship information.

Two-node hot standby ensures high availability

KVStore for Redis uses the two-node hot standby method to maximize service availability.

Cluster editions eliminate the performance bottleneck

KVStore for Redis provides cluster instances to eliminate the performance bottleneck that is caused by Redis single-thread mechanism. Cluster instances can effectively handle traffic bursts during live video streaming and support high-performance requirements.

Easy scaling relieves pressure at peak hours

KVStore for Redis allows you to easily perform scaling. The complete upgrade process is transparent. Therefore, you can easily handle traffic bursts at peak hours.

E-commerce industry applications

In the e-commerce industry, the KVStore for Redis service is widely used in the modules such as commodity display and shopping recommendation.

Scenario 1: rapid online sales promotion systems

During a large-scale rapid online sales promotion, a shopping system is overwhelmed by traffic. A common database cannot properly handle so many read operations.

However, KVStore for Redis supports data persistence, and can work as a database system.

Scenario 2: counter-based inventory management systems

In this scenario, you can store inventory data in an RDS database and save count data to corresponding fields in the database. In this way, the KVStore for Redis instance reads count data, and the RDS database stores count data. KVStore for Redis is deployed on a physical server. Based on solid-state drive (SSD) high-performance storage, the system can provide a high-level data storage capacity.

8.6 Limits

Item	Description
List data type	The number of lists is not limited. The size of each element is 512 MB or less. We recommend that the number of elements in a list is less than 8,192. The value length is 1 MB or less.
Set data type	The number of sets is not limited. The size of each element is 512 MB or less. We recommend that the number of elements in a set is less than 8,192. The value length is 1 MB or less.
Sorted set data type	The number of sorted sets is not limited. The size of each element is 512 MB or less. We recommend that the number of elements in a sorted set is less than 8,192. The value length is 1 MB or less.
Hash data type	The number of fields is not limited. The size of each element in a hash table is 512 MB or less. We recommend that the number of elements in a hash table is less than 8,192. The value length is 1 MB or less.
Number of databases (DBs)	Each instance supports 256 DBs.
Supported Redis commands	For more information, see the " Supported Redis commands " topic of KVStore for Redis User Guide .
Monitoring and alerts	KVStore for Redis does not provide capacity alerts. You have to configure this feature in CloudMonitor. We recommend that you set alerts for the following metrics: instance faults, instance failover, connection usage, failed operations, capacity usage, write bandwidth usage, and read bandwidth usage.

Item	Description
Expired data deletion policies	<ul style="list-style-type: none"> Active expiration: the system periodically detects and deletes expired keys in the background. Passive expiration: the system deletes expired keys when you access these keys.
Idle connection recycling mechanism	KVStore for Redis does not actively recycle idle connections to KVStore for Redis. You can manage the connections.
Data persistence policy	KVStore for Redis uses the AOF_FSYNC_EVERYSEC policy, and runs the fsync command at a one-second interval.

8.7 Terms

Redis

A high-performance key-value storage system that works as a cache and store and that is compatible with BSD open-source protocols.

Instance ID

An instance corresponds to a user space, and serves as the basic unit of using Redis.

Redis has limits on instance configurations, such as connections, bandwidth, and CPU processing capacity. These limits vary according to different instance types. You can view the list of instance identifiers that you have purchased in the console. KVStore for Redis instances are classified into master-replica instances and high-performance cluster instances.

Master-replica instance

The KVStore for Redis instance that contains a master-replica structure. The master-replica instance provides limited capacity and performance.

High-performance cluster instance

The KVStore for Redis instance that runs in a scalable cluster architecture. Cluster instances provide better scalability and performance, but they still have limited features.

Connection address

The host address for connecting to KVStore for Redis. The connection address is displayed as a domain name. To obtain the connection address, go to the **Instance Information** tab page, and check the address in the **Connection Information** field.

Eviction policy

The policy that KVStore for Redis uses to delete earlier data when the memory of KVStore for Redis reaches the upper limit as specified in **maxmemory**. Eviction policies of KVStore for Redis are consistent with Redis eviction policies. For more information, see [Using Redis as an LRU cache](#).

DB

The abbreviation of the word "database" to indicate a database in KVStore for Redis. Each KVStore for Redis instance supports 256 databases numbered DB 0 to DB 255.

8.8 Instance types



Note:

The maximum bandwidth includes the maximum upstream bandwidth and the maximum downstream bandwidth.

Standard dual-replica edition

Table 8-1: Standard plan

Type	Service code	Maximum number of connections	Maximum bandwidth (MB)	CPU	Description	Zone-disaster recovery deployment
1 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb1g.1db.0rodb.4proxy.default	10,000	10	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region

Type	Service code	Maximum number of connections	Maximum bandwidth (MB)	CPU	Description	Zone-disaster recovery deployment
2 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb2g.1db.0rodb.4proxy.default	10,000	16	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
4 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb4g.1db.0rodb.4proxy.default	10,000	24	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
8 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb8g.1db.0rodb.4proxy.default	10,000	24	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
16 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb16g.1db.0rodb.4proxy.default	10,000	32	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region

Type	Service code	Maximum number of connections	Maximum bandwidth (MB)	CPU	Description	Zone-disaster recovery deployment
32 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredisdb32g.1db.0rodb.4proxy.default	10,000	32	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region

Table 8-2: Premium plan

Type	Service code	Maximum number of connections	Maximum bandwidth (MB)	CPU	Description	Zone-disaster recovery deployment
1 GB advanced primary/secondary edition	redis.master.small.special2x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone
2 GB advanced primary/secondary edition	redis.master.mid.special2x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone
4 GB advanced primary/secondary edition	redis.master.stand.special2x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone

Type	Service code	Maximum number of connections	Maximum bandwidth (MB)	CPU	Description	Zone-disaster recovery deployment
8 GB advanced primary/secondary edition	redis.master.large.special1x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone
16 GB advanced primary/secondary edition	redis.master.2xlarge.special1x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone
32 GB advanced primary/secondary edition	redis.master.4xlarge.special1x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone

Table 8-3: Cluster edition

Type	Service code	Maximum number of connections	Maximum bandwidth (MB)	CPU	Description
16 GB cluster edition	redis.sharding.small.default	80,000	384	4-core	High-performance cluster instance
32 GB cluster edition	redis.sharding.mid.default	80,000	384	8-core	High-performance cluster instance
64 GB cluster edition	redis.sharding.large.default	80,000	384	8-core	High-performance cluster instance

Type	Service code	Maximum number of connections	Maximum bandwidth (MB)	CPU	Description
128 GB cluster edition	redis.sharding.2xlarge.default	160,000	768	16-core	High-performance cluster instance
256 GB cluster edition	redis.sharding.4xlarge.default	160,000	768	16-core	High-performance cluster instance
512 GB cluster edition	redis.logic.sharding.16g.32db.0rodb.32proxy.default	320,000	1,536	8-core	
1 TB cluster edition	redis.sharding.16xlarge.default	640,000	3,072	8-core	
2 TB cluster edition	redis.sharding.32xlarge.default	1,280,000	6,144	8-core	
4 TB cluster edition	redis.logic.sharding.16g.256db.0rodb.256proxy.default	2,560,000	12,288	16-core	

Table 8-4: Cluster edition for zone-disaster recovery

Type	Service code	Maximum number of connections	Maximum bandwidth (MB)	CPU	Description
16 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredisdb16g.8db.0rodb.8proxy.default	80,000	384	8-core	Cluster instance for zone-disaster recovery

Type	Service code	Maximum number of connections	Maximum bandwidth (MB)	CPU	Description
32 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredismdb32g.8db.0rodb.8proxy.default	80,000	384	8-core	Cluster instance for zone-disaster recovery
64 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredismdb64g.8db.0rodb.8proxy.default	80,000	384	8-core	Cluster instance for zone-disaster recovery
128 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredismdb128g.16db.0rodb.16proxy.default	160,000	768	16-core	Cluster instance for zone-disaster recovery
256 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredismdb256g.16db.0rodb.16proxy.default	160,000	768	16-core	Cluster instance for zone-disaster recovery
256 GB cluster edition for zone-disaster recovery	redis.logic.sharding.16g.32db.0rodb.32proxy.default	320,000	1,536	8-core	Cluster instance for zone-disaster recovery
1 TB cluster edition for zone-disaster recovery	redis.sharding.16xlarge.default	640,000	3,072	8-core	Cluster instance for zone-disaster recovery

Type	Service code	Maximum number of connections	Maximum bandwidth (MB)	CPU	Description
2 TB cluster edition for zone-disaster recovery	redis.sharding.32xlarge.default	1,280,000	6,144	8-core	Cluster instance for zone-disaster recovery
4 TB cluster edition for zone-disaster recovery	redis.logic.sharding.16g.256db.0rodb.256proxy.default	2,560,000	12,288	16-core	Cluster instance for zone-disaster recovery

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 Benefits

DTS supports transmitting data between data sources such as relational databases and OLAP databases. DTS provides you with multiple data transmission methods such as data migration, real-time data subscription, and real-time data synchronization. Compared with other third-party data migration and synchronization tools, DTS provides multiple transmission channels with high performance, security, and reliability. DTS also makes it easy to create and manage transmission channels.

Diverse transmission methods

DTS supports multiple data transmission features, including data migration, data subscription, and data synchronization. In data subscription and data synchronization, data is transmitted in real time.

Data migration enables you to migrate data between databases without interrupting application operations. The application service downtime during data migration is reduced to minutes.

High performance

DTS uses servers with high specifications to ensure high data transmission performance for each synchronization or migration channel.

At the underlying layer, multiple measures are taken to improve DTS performance.

Compared with traditional data synchronization tools, the real-time synchronization feature of DTS enables you to concurrently transmit transactions. It also allows you to synchronize table data you want to update at a time. This greatly improves synchronization performance.

High security and reliability

DTS is implemented using clusters. If a node in a cluster is down or faulty, the control center quickly moves all tasks from this node to another healthy node in the cluster.

DTS provides a 24 x 7 mechanism for validating data accuracy in some transmission channels to quickly locate and correct incorrect data. This helps ensure reliable data transmission.

Secure transmission protocols and tokens are used for authentication across DTS modules to ensure reliable data transmission.

Easy-to-use

The DTS console is a visual management interface that provides a wizard-like process to assist you in creating data transmission channels.

You can also view data transmission information in the DTS console, including the transmission status, progress, and performance, to better manage the transmission channels.

DTS supports resumable transmission, and regularly monitors channel status to avoid interruptions resulting from network or system exceptions. When DTS detects a channel exception, it automatically repairs or restarts the channel. In cases where manual operations are needed, you can directly repair the channel and restart it in the DTS console.

9.3 Environment requirements

Use Data Transmission Service (DTS) on hosts of the following models:

- PF51.*
- PV52P2M1.*

- DTS_E. *
- PF61. *
- PF61P1. *
- PV62P2M1. *
- PV52P1. *
- Q5F53M1. *
- PF52M2. *
- Q41. *
- Q5N1.22
- Q5N1.2B
- Q46.22
- Q46.2B
- W41.22
- W41.2B
- W1.22
- W1.2B
- W1.2C
- D13.12

Use the following operating system:

AliOS7U2-x86-64

**Notice:**

- Do not use DTS on hosts whose models are excluded from the preceding list.
- The /apsara directory used by DTS resides on only one hard disk. Make sure that the space of the hard disk is larger than 2 TB.

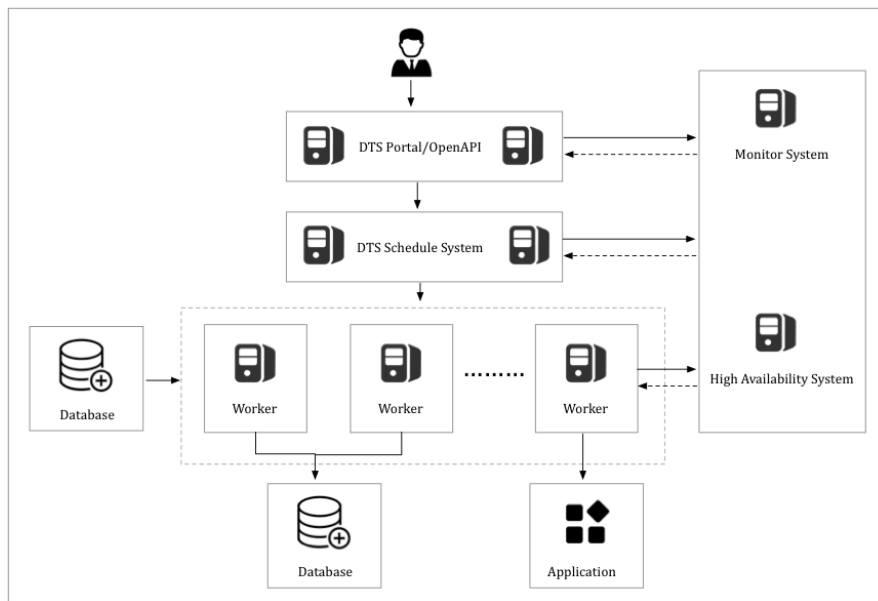
If the space of the hard disk where the /apsara directory resides is smaller than 2 TB, tasks may fail to run and errors may occur. In this case, DTS cannot restore failed tasks or pull data properly.

9.4 Architecture

System architecture

Figure 9-1: System architecture shows the system architecture of DTS.

Figure 9-1: System architecture



- **High availability**

Each DTS module comes with a primary-secondary architecture to ensure high availability of the system. The disaster recovery module runs a health check on each node in real time. Once a node exception is detected, the module switches the channel to another healthy node within seconds.

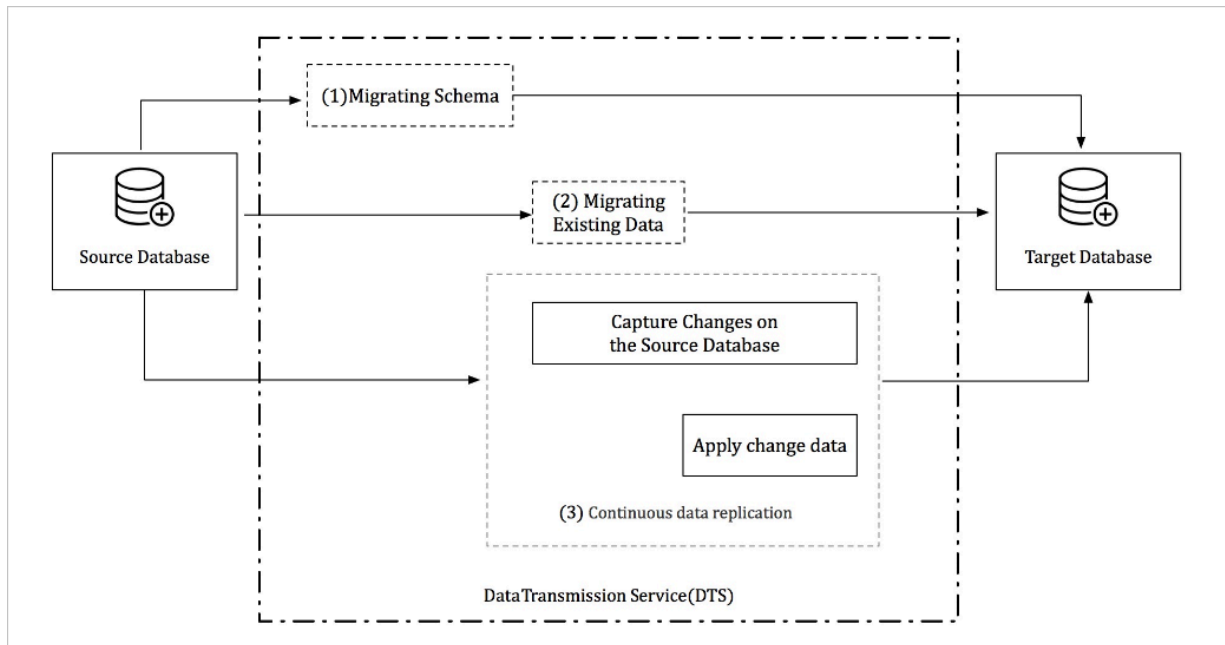
- **Monitor changes in the data source IP address**

For data subscription and synchronization channels, the disaster recovery module checks for any changes. For example, once it detects a change in the data source address, the module dynamically changes the method for connecting to the data source to ensure channel stability.

Data migration process

Figure 9-2: Data migration workflow shows how data migration works.

Figure 9-2: Data migration workflow



Data migration supports schema migration, real-time full data migration, and real-time incremental data migration. To implement migration without service interruption, follow these steps:

1. Schema migration
2. Full data migration
3. Incremental data migration

For migration between heterogeneous databases, DTS reads the schema using the syntax of the source database, translates the schema into the syntax of the destination database, and then imports the schema to the destination instance.

Full data migration takes a longer time. In this process, new data is continuously written into the source instance. To ensure data consistency, DTS starts the incremental data pulling module before full data migration. This module pulls the incremental data from the source instance and then parse, encapsulate, and store the data locally.

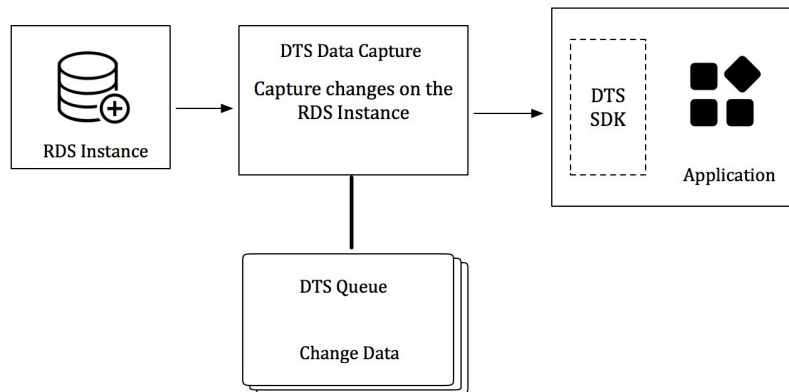
When full data migration is complete, DTS starts the incremental data playback module. The module retrieves the incremental data from the incremental data pulling module. After reverse parsing, filtering, and encapsulation, the data is synchronized to the destination

instance. Eventually, data is synchronized between the source and destination instances in real time.

Data subscription process

[Figure 9-3: Data subscription workflow](#) shows how data subscription works.

Figure 9-3: Data subscription workflow



Data subscription supports pulling incremental data from the RDS instance in real time. You can subscribe to the incremental data on the data subscription server using the DTS SDK. You can also customize data consumption based on business requirements.

The data pulling module of the DTS server captures raw data from the data source, and makes the incremental data locally persistent by parsing, filtering, and formatting it.

The data capturing module connects to the source instance using the database protocol and pulls the incremental data from the source instance in real time. For example, the data capturing module connects to an RDS for MySQL instance using the binlog dump command

.

DTS guarantees the high availability of the data pulling module and downstream consumption SDKs.

To ensure the high availability of the data pulling module, the DTS disaster recovery module restarts the data pulling module on a healthy service node once an exception is detected in the data pulling module.

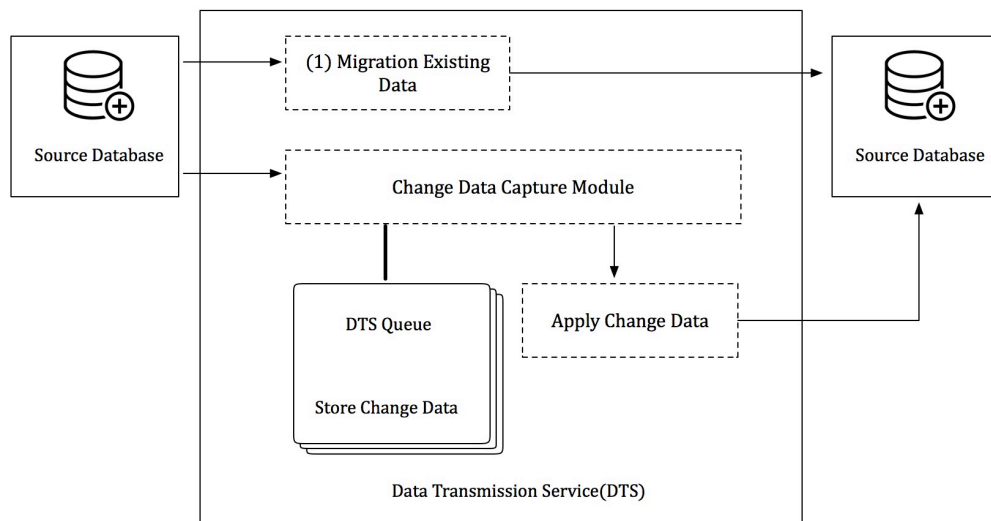
The DTS server ensures the high availability of downstream consumption SDKs. If you start multiple consumption SDKs for the same subscription channel, the server pushes the incremental data to only one SDK at a time. If the consumption encounters an exception

, the service end selects another consumption process from other healthy downstream nodes to push data to that consumption process. In this way, the high availability of downstream consumption processes can be guaranteed.

Real-time synchronization workflow

Figure 9-4: Real-time synchronization workflow shows how real-time synchronization works.

Figure 9-4: Real-time synchronization workflow



The data synchronization feature in DTS enables real-time synchronization of incremental data between any two RDS instances.

To create a synchronization channel, follow these steps:

- Initial synchronization: The existing data in the source instance is synchronized to the destination instance.
- Incremental data synchronization: After initial synchronization, the incremental data starts to be synchronized between the source instance and destination instance in real time. During this phase, data is eventually synchronized between the source and destination instances.

DTS provides the following underlying modules for real-time incremental data synchronization:

- Data reading module

The data reading module reads raw data from the source instance and makes the data locally persistent by parsing, filtering, and formatting it. The data reading module

connects to the source instance using the database protocol and reads the incremental data from the source instance. For example, the data reading module connects to an RDS for MySQL instance using the binlog dump command.

- Data playback module

The data playback module requests incremental data from the data reading module, filters data based on the objects to be synchronized, and then synchronizes the data to the destination instance without compromising the transaction sequence and consistency.

DTS ensures the high availability of the data reading module and data playback module. When a channel exception is detected, the disaster recovery module restarts the channel on a healthy service node. In this way, the high availability of the synchronization channels is guaranteed.

9.5 Features

9.5.1 Data migration

Data migration allows you to migrate data between multiple data sources. Typical scenarios include data migration to the cloud, data migration between instances within Apsara Stack, and database sharding and scaling. DTS supports data migration between homogeneous and heterogeneous data sources. It also supports ETL features such as data mapping at three levels (databases, tables, and columns) and data filtering.

Data sources

DTS supports migrating data between multiple data sources. [Table 9-1: Data sources supported by each data migration type](#) lists the data sources supported by each data migration type.

Table 9-1: Data sources supported by each data migration type

Data source	Schema migration	Full data migration	Incremental data migration
MySQL > RDS for MySQL	Supported	Supported	Supported
MySQL > Oracle	Not supported	Supported	Supported

Data source	Schema migration	Full data migration	Incremental data migration
Oracle > RDS for MySQL	Supported	Supported	Supported

DTS supports migrating data from the following data sources:

- RDS instances
- Oracle databases
- On-premises databases

DTS supports migrating data to the following data sources:

- RDS instances

Online migration

DTS uses online migration. You only need to configure the source instance, destination instance, and objects to be migrated. DTS automatically completes the entire data migration process. To minimize the impact of online data migration on your services, you must ensure that the DTS server can connect to both the source and destination instances.

Data migration types

DTS supports schema migration, full data migration, and incremental data migration.

- Schema migration: migrates schemas from the source instance to the destination instance.
- Full data migration: migrates historical data from the source instance to the destination instance.
- Incremental data migration: migrates incremental data generated during migration from the source instance to the destination instance in real time. You can select schema migration, full data migration, and incremental migration to migrate data with minimal downtime.

ETL features

Data migration supports the following ETL features:

- Object name mappings of databases, tables, and columns. With this feature, you can migrate data between two databases, tables, or columns that have different names.

- Data filtering. With this feature, you can use SQL conditions to filter the required data in a specific table. For example, you can specify a time range to migrate only the latest data.

Alerts

Data migration supports sending alerts when errors occur. When an error occurs, DTS immediately sends an SMS alert to the task owner, allowing the owner to promptly handle the exception.

Migration task

A migration task is a basic unit of data migration. To migrate data, you must create a data migration task in the DTS console. To create a migration task, you must configure the required information such as the connection type of the source and destination instances, the migration type, and the objects to be migrated. You can create, manage, stop, and delete data migration tasks in the DTS console.

[Table 9-2: Task statuses](#) describes possible statuses of a migration task.

Table 9-2: Task statuses

Task status	Description	Available operations
Not started	The migration task configurations have been completed, and the precheck is not performed.	<ul style="list-style-type: none"> • Perform the precheck. • Delete the migration task.
Prechecking	A precheck is being performed before the migration task starts.	Delete the migration task.

Task status	Description	Available operations
Precheck passed	The migration task has passed the precheck and has not started.	<ul style="list-style-type: none"> Start the migration task. Delete the migration task.
Migrating	Data is being migrated.	<ul style="list-style-type: none"> Pause the migration task. Stop the migration task. Delete the migration task.
Migration failed	An error occurred during migration. You can determine the specific phase in which the migration failed based on the progress of the migration task.	Delete the migration task.
Paused	The migration task is paused.	<ul style="list-style-type: none"> Start the migration task. Delete the migration task.
Completed	The migration task is complete, or you have clicked End to stop data migration.	Delete the migration task.

9.5.2 Data synchronization

You can use DTS to synchronize data between data sources in real time.

This feature applies to the following scenarios: active geo-redundancy, geo-disaster recovery, zone-disaster recovery, cross-border data synchronization, query load balancing , cloud BI systems, and real-time data warehousing.

Synchronization modes

The [Table 9-3: Synchronization modes](#) table lists the synchronization modes that are supported by DTS.

Table 9-3: Synchronization modes

Source instance	Destination instance	One-way/two-way synchronization
MySQL	MySQL	<ul style="list-style-type: none">One-way synchronizationTwo-way synchronization
MySQL	MaxCompute	One-way synchronization
MySQL	AnalyticDB	One-way synchronization
MySQL	DataHub	One-way synchronization

Objects to be synchronized

- You can select columns, tables, or databases as the objects to be synchronized. You can specify one or more tables that you want to synchronize.
- DTS allows you to synchronize data between tables that have different names or databases that have different names. You can use the object name mapping feature to specify the names of destination columns, tables, and databases.
- You can specify one or more columns that you want to synchronize.

Synchronization tasks

A synchronization task is used to synchronize data in real time. To synchronize data between two instances, you must create a synchronization task in the DTS console.

The [Table 9-4: Task statuses](#) table describes the statuses of a synchronization task.

Table 9-4: Task statuses

Task status	Description	Available operations
Prechecking	A precheck is being performed before the synchronization task is started.	<ul style="list-style-type: none"> • View the configurations of the synchronization task • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts
Precheck Failed	The synchronization task has failed the precheck.	<ul style="list-style-type: none"> • Perform the precheck • View the configurations of the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts
Not Started	The synchronization task has passed the precheck but has not been started.	<ul style="list-style-type: none"> • Perform the precheck • Start the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts

Task status	Description	Available operations
Performing Initial Synchronization	The initial synchronization is in progress.	<ul style="list-style-type: none"> • View the configurations of the synchronization task • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts.
Initial Synchronization Failed	The task has failed during initial synchronization.	<ul style="list-style-type: none"> • View the configurations of the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts
Synchronizing	Data is being synchronized.	<ul style="list-style-type: none"> • View the configurations of the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Pause the synchronization task • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts

Task status	Description	Available operations
Synchronization Failed	An error occurred during synchronization.	<ul style="list-style-type: none"> • View the configurations of the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Start the synchronization task • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts
Paused	The synchronization task is paused.	<ul style="list-style-type: none"> • View the configurations of the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Start the synchronization task • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts

Advanced features

You can use the following advanced features to facilitate data synchronization:

- Add and remove the objects to be synchronized

You can add and remove the objects to be synchronized during data synchronization.

- View and analyze synchronization performance

DTS provides trend charts that allow you to view and analyze the performance of your synchronization tasks. The synchronization performance is measured based on bandwidth, synchronization speed (TPS), and synchronization delay.

- Monitor synchronization tasks

DTS monitors the statuses and delays of synchronization tasks and generates alerts if the predefined threshold is reached. You can set the threshold for synchronization delay alerts based on the sensitivity of your businesses to synchronization delays.

9.5.3 Data subscription

Real-time data subscription is designed to help users retrieve incremental data from Relational Database Service (RDS) in real time. In business scenarios such as cache update, asynchronous decoupling, real-time data synchronization between heterogeneous data sources, and real-time data synchronization with complex ETL, you may choose to consume the incremental data as required.

Features

- Supports data subscription for RDS for MySQL instances in classic and VPC networks.

Data source types

Real-time data subscription supports the following data sources:

- RDS for MySQL

Objects to be subscribed

Objects to be subscribed include databases and tables. You can subscribe to the incremental data of specified tables as needed.

In data subscription, incremental data is further divided into data updates (DML) and schema updates (DDL). When you configure data subscription, you can select a data change type as needed.

Subscription channels

Subscription channels are used for incremental data subscription and consumption. To subscribe to incremental data of an RDS instance, you must create a subscription channel in the DTS console for this instance. The subscription channel reads incremental data in the RDS instance in real time and stores the most recent increments. You can use the SDK

provided by DTS to subscribe to and consume the incremental data in the channel. You can also create, manage, and delete subscription channels in the DTS console.

Data in a subscription channel can only be subscribed and consumed using one SDK. To subscribe to an RDS instance for multiple downstream SDKs, you must create an equivalent number of subscription channels. RDS instances subscribed by these channels share the same instance ID.

[Table 9-5: Subscription channel status and descriptions](#) shows the different status of a subscription channel during its lifecycle.

Table 9-5: Subscription channel status and descriptions

Channel status	Description	Available actions
Prechecking	The subscription channel has been configured and a preliminary check is being performed.	Delete the subscription channel.
Not started	The migration task has passed the precheck, but is not started.	<ul style="list-style-type: none"> Start the subscription channel. Delete the subscription channel.
Initial subscription	Initial subscription is being enabled and takes about one minute.	Delete the subscription channel.
Running	The subscription channel is reading incremental data from an RDS instance.	<ul style="list-style-type: none"> View sample code. View the subscribed data. Delete the subscription channel.
Error	An exception occurs when the subscription channel reads incremental data from an RDS instance.	<ul style="list-style-type: none"> View sample code. Delete the subscription channel.

Advanced features

The following features are used in data subscription:

- Add and remove objects to be subscribed as required

You can add or remove the objects to be subscribed during a data subscription.

- View the subscribed data online

You can view the subscribed incremental data in the DTS console.

- Modify the data consumption checkpoint

You can modify the time for data consumption at any time.

- Complete monitoring system

Data subscription monitors the subscription channel status and reports an alert when the threshold for downstream consumption delays is reached. You can set the alert threshold according to business sensitivity.

9.6 Scenarios

DTS supports multiple features including data migration, real-time data subscription, and real-time data synchronization to meet the following scenarios.

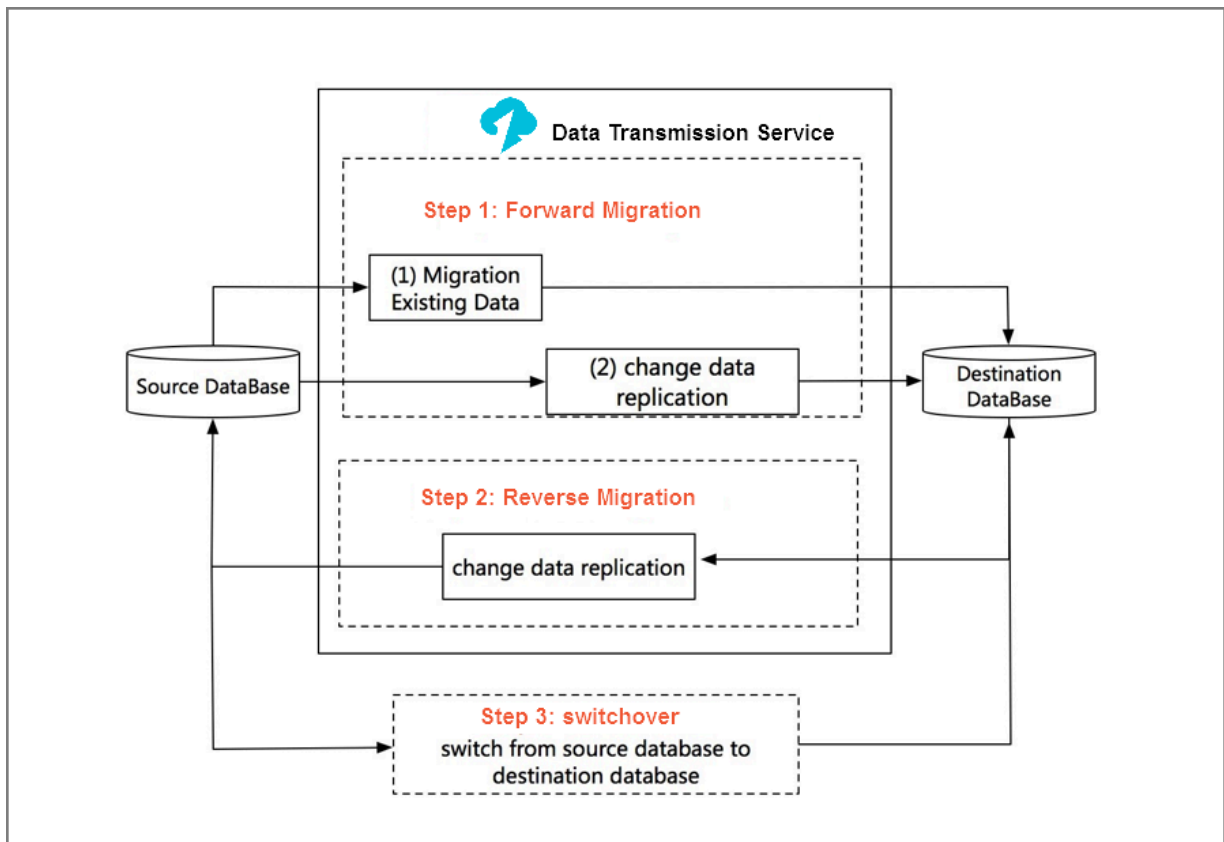
Migration with service downtime reduced to minutes

Many users seek for a way to migrate systems without affecting their services. However, data changes if services are not suspended during the migration. To ensure data consistency, many third-party migration tools require that the service be suspended during data migration. It may take hours or even days throughout the migration and result in a significant loss in service availability.

To reduce the barrier of database migration, DTS provides an interruption-free migration solution that minimizes the service downtime to minutes.

Figure 9-5: Interruption-free migration shows how interruption-free migration works.

Figure 9-5: Interruption-free migration



The interruption-free data migration process involves schema migration, full data migration, and incremental data migration. In the incremental data migration phase, data is synchronized between the source and destination instances in real time. You can validate the service in the destination database. After the validation is complete, the service is migrated to the destination database. The entire system is then eventually migrated.

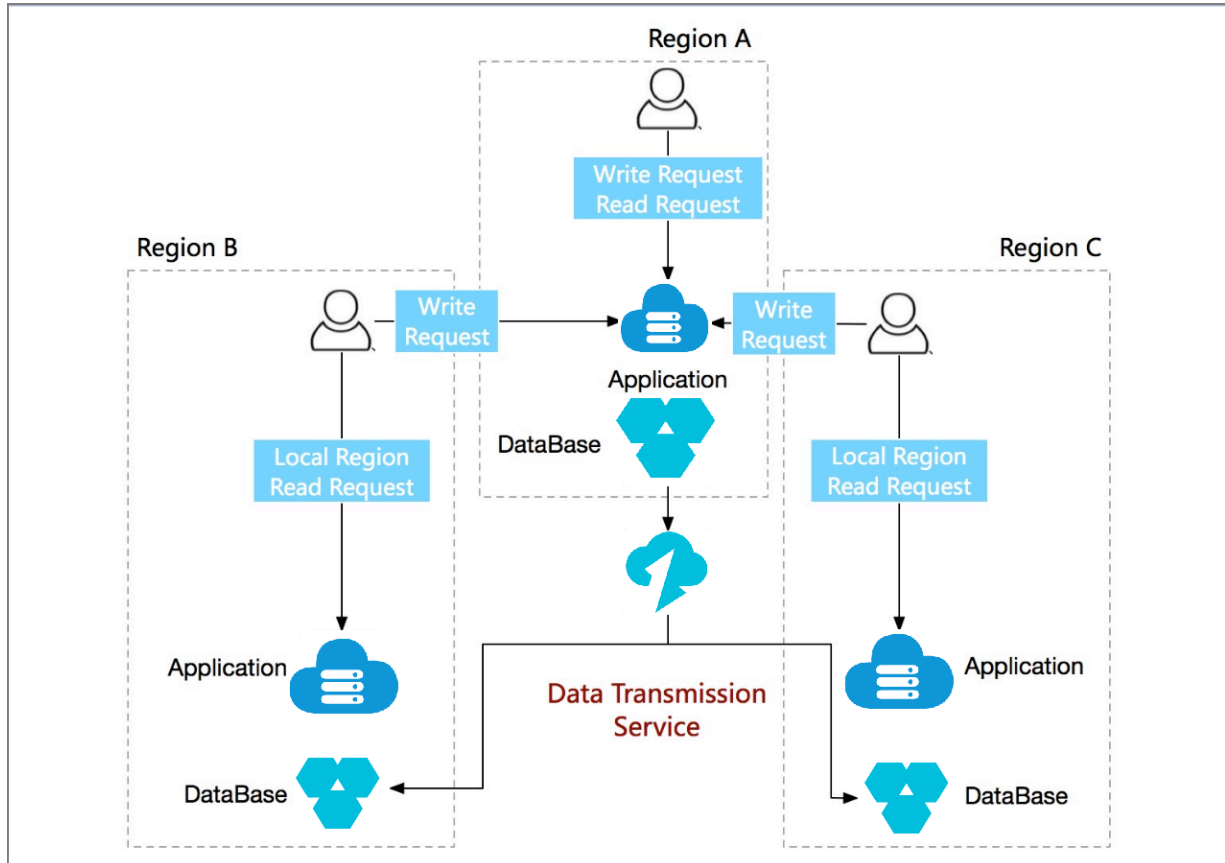
Throughout the migration process, the service experiences interruptions only when it is switched from the source instance to the destination instance.

Accelerated access to global services to empower cross-border businesses

If services with widely distributed users, such as global services, are deployed only in one region, users in other regions have to access them remotely, resulting in high access latency and poor user experience. To accelerate the access to global services and improve

access experience, you can adjust the architecture, as shown in [Figure 9-6: Reduced cross-region access latency](#).

Figure 9-6: Reduced cross-region access latency



This architecture consists of one center and multiple units. Write requests of users in all regions are routed back to the center. DTS synchronizes data in the center to all units. Read requests of users in different regions can be routed to nearby units to avoid remote access and reduce access latency. In this way, access to global services is accelerated.

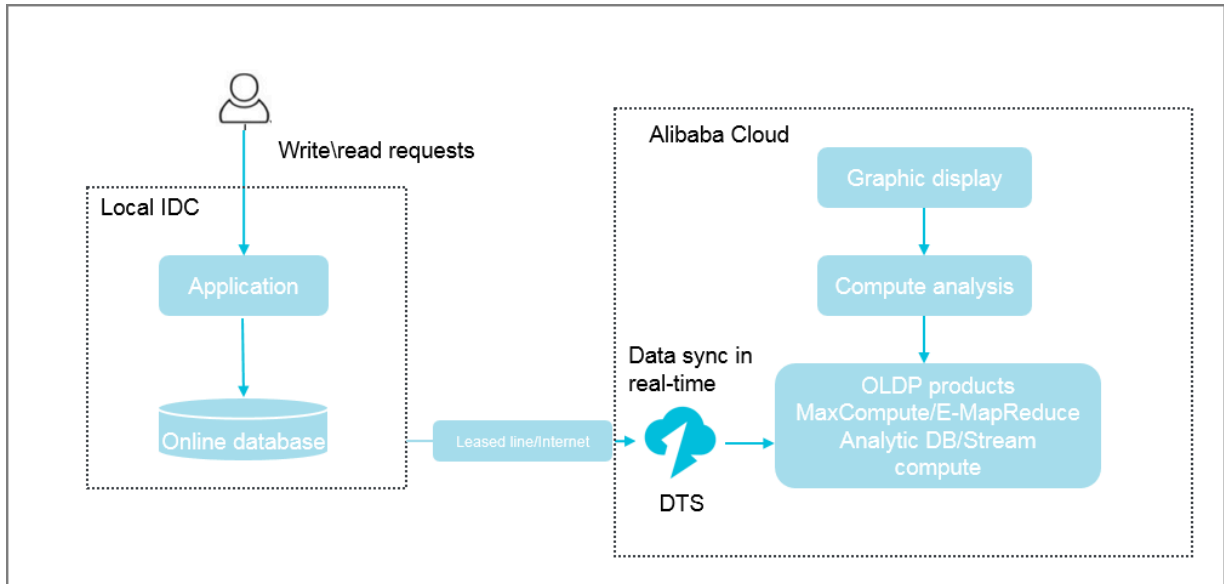
Custom cloud BI system built with more efficiency

User-created business intelligence (BI) systems cannot meet the increasing demand for real-time performance and are difficult to manipulate. With the Apsara Stack BI architecture, you can quickly build a BI system without affecting the current architecture. For this reason, more and more users choose to build BI systems that meet their own business requirements on Apsara Stack.

DTS can help you synchronize data stored in local databases to an Apsara Stack BI system (such as MaxCompute or StreamCompute) in real time. You can then perform subsequent data analysis with various compute engines while viewing the computing results in real

time with a visualization tool. You can also synchronize those results back to the local IDC with a migration tool. [Figure 9-7: Cloud BI architecture](#) shows the implementation architecture.

Figure 9-7: Cloud BI architecture

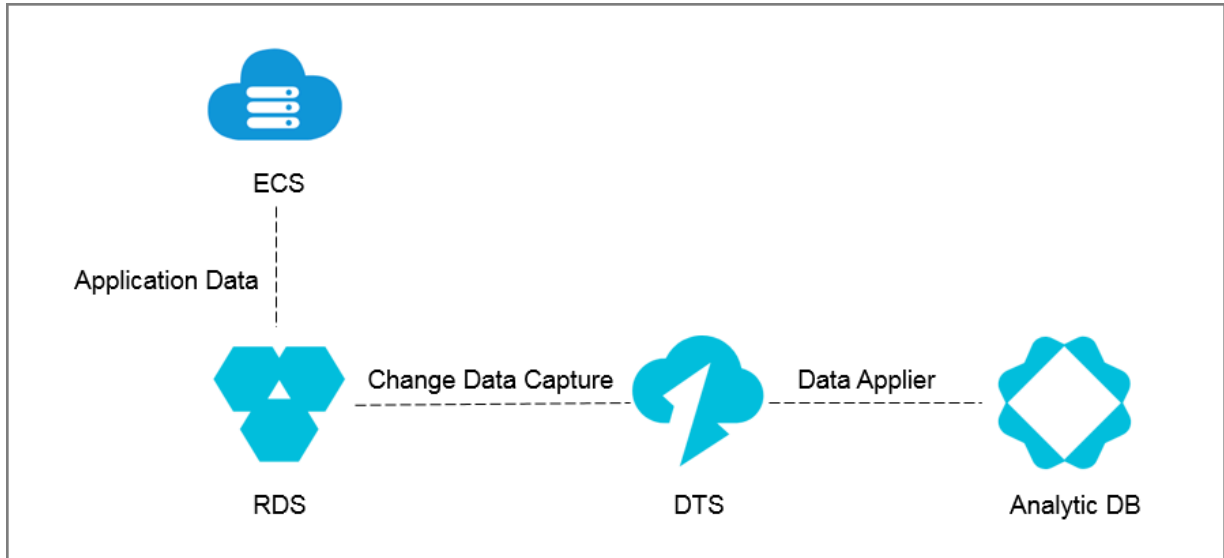


Real-time data analysis to rapidly respond to market conditions

Data analysis is essential in improving enterprise insights and user experience. Real-time data analysis enables enterprises to adjust marketing strategies more quickly and flexibly so that they can adapt to the rapidly changing marketing conditions and demands for higher user experience. To implement real-time data analysis without affecting online services, service data needs to be synchronized to the analysis system in real time. For this reason, acquiring service data in real time becomes essential. In DTS, the data subscription feature can help you acquire real-time incremental data without affecting online services

and synchronize the data to the analysis system using the SDK for real-time data analysis, as shown in [Figure 9-8: Real-time data analysis](#).

Figure 9-8: Real-time data analysis



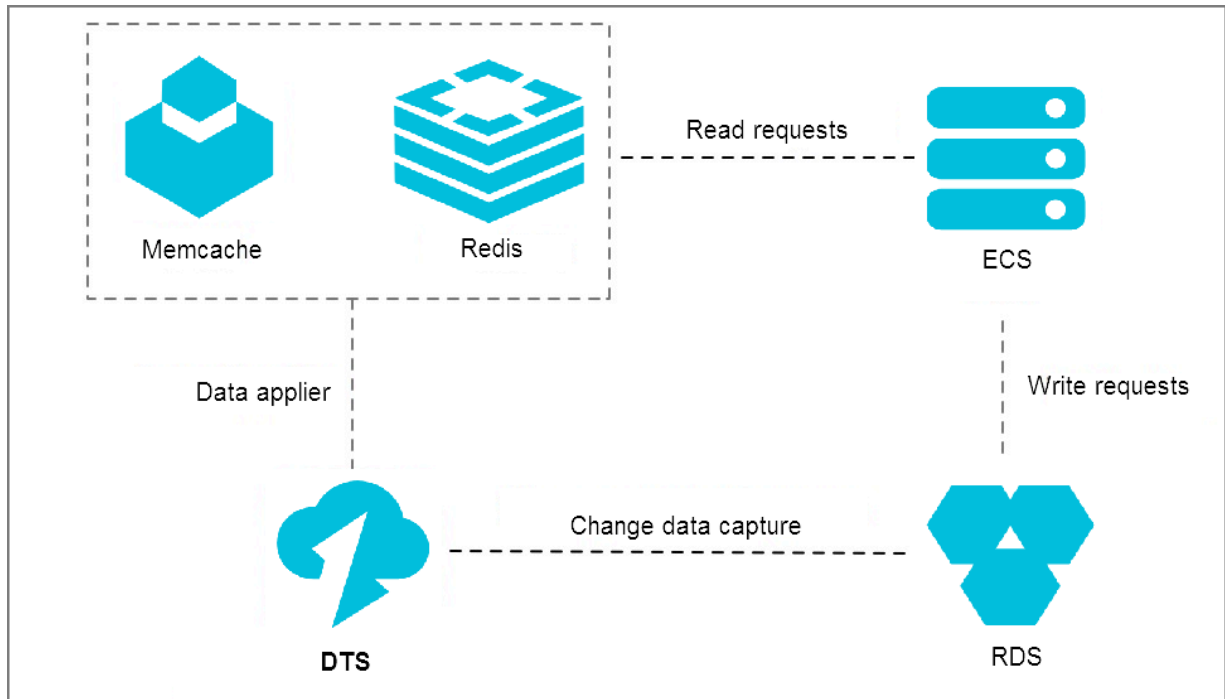
Lightweight cache update policies to make core services more simple and reliable

To accelerate service access and improve concurrent read performance, many enterprises introduce the caching layer to the service architecture. In this architecture, all the read requests are routed to the caching layer, and the memory reading mechanism greatly improves read performance. Cached data cannot persist. If caching ends abnormally, data in the cache memory is lost. To ensure data integrity, the updated service data is kept in a persistent storage medium, such as a database.

In this condition, the service data is inconsistent between the cache and the persistent databases. The data subscription feature can help asynchronously subscribe to the incremental data in those databases and update the cached data to implement lightweight

cache update policies. [Figure 9-9: Cache update policies](#) shows the architecture of these policies.

Figure 9-9: Cache update policies



Cache update policies offer the following benefits:

- Quick update with low latency

Cache invalidation is an asynchronous process, and the service returns data directly after the database update is complete. For this reason, you do not need to consider the cache invalidation process, and the entire update path is short with low latency.

- Simple and reliable applications

The complex doublewrite logic is not required for the application. You only need to start the asynchronous thread to monitor the incremental data and update the cached data.

- Application updates without extra performance consumption

Because data subscription acquires incremental data by parsing incremental logs in the database, the acquisition process does not damage the performance of services and databases.

Asynchronous service decoupling to make core services simpler and more reliable

Data subscription optimizes intensive coupling to asynchronous coupling by using real-time message notifications. This makes the core service logic simpler and more

reliable. This application has been widely implemented in Alibaba. Tens of thousands of downstream services in the Taobao ordering system acquire real-time data updates through data subscription to trigger the business logic every day.

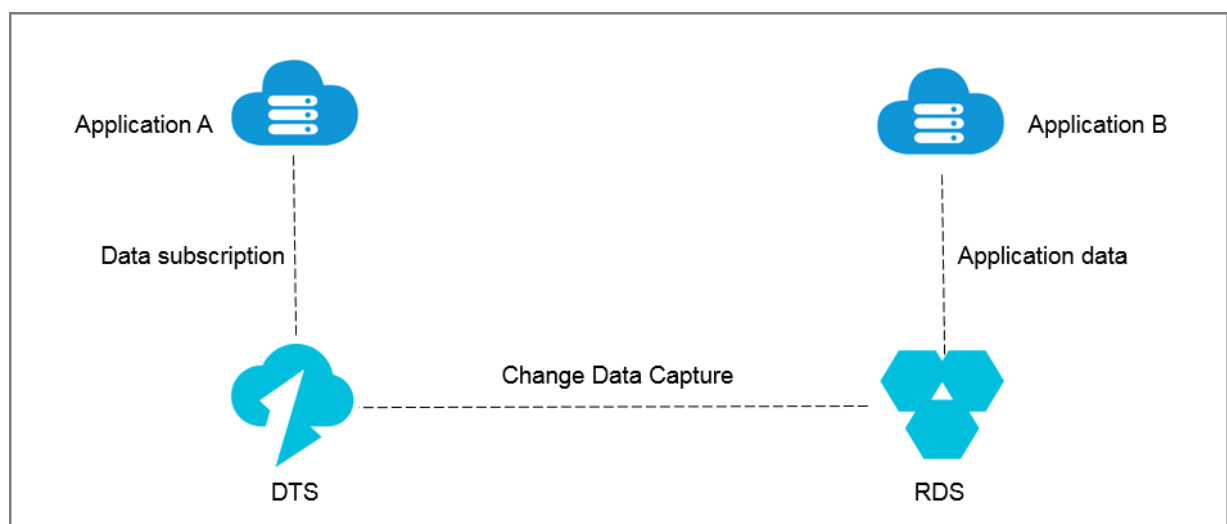
The following uses a simple example to describe the benefits of implementing data subscription in this scenario.

The e-commerce industry involves multiple services including the order management system, inventory management, and the shipping of goods. An ordering process with all of those services included is as follows: After a user places an order, downstream services including seller inventory notification and goods shipping are modified. When all logic modifications are complete, the order result is returned to the user. However, this ordering logic has the following issues:

- The lengthy ordering process results in poor user experience.
- The system is unstable and any downstream fault directly affects the availability of the ordering system.

To improve user experience of core applications, you can decouple the core applications and the dependent downstream services so that they can work asynchronously. In this way, the core applications become more stable and reliable. [Figure 9-10: Asynchronous service decoupling](#) shows how to adjust the logic.

Figure 9-10: Asynchronous service decoupling



The ordering system returns the order result directly after order placement. With DTS, the underlying layer acquires the updated data from the ordering system in real time. Then, the downstream service subscribes to the modified data using the SDK and triggers the service

logic such as inventory and shipping. In this way, the ordering system becomes simpler and more reliable.

Horizontal scaling to improve read performance and quickly adapt to business growth

A single RDS instance may not be able to support a large number of read requests, which may affect the main service process. To elastically improve the read performance and reduce database workload, you can create read-only instances using the real-time synchronization feature of DTS. These read-only instances take on large amounts of the database reading workload and expand the throughput of applications.

9.7 Concepts

Precheck

Precheck is an essential stage before a migration task starts. It mainly checks the prerequisites that may affect a successful migration, such as the connectivity of the source and destination instances and the permissions of the migration accounts. If the precheck fails, you can fix the problems as instructed and run the precheck again.

Schema migration

Schema migration is a type of migration tasks. In database migration, it refers to migrating the schema syntax, including tables, views, triggers, stored procedures, stored functions, and synonyms. For migration between heterogeneous databases, data types are mapped during schema migration, and the schema syntax is adjusted according to the schema syntax of the source and destination instances.

Full data migration

Full data migration is a type of migration task. It refers to migrating all the data except the schema syntax from the source instance to the destination instance. If you select Full Data Migration only and leave Schema Migration unselected, new data generated in the source instance will not be migrated to the destination instance.

Incremental data migration

Incremental data migration is a type of migration tasks. It refers to synchronizing the new data written to the source instance to the destination instance during the migration . When creating a migration task, if you select both Full Data Migration and Incremental Data Migration, DTS will first perform a static snapshot on the source instance, migrate the snapshot data to the destination instance, and then synchronize the new data from

the source instance to the destination instance during the migration. Incremental data migration is a process of synchronizing data between the source and destination instances in real time. This process does not automatically end. If you want to stop migrating data, you must manually disable the task in the console.

Initial synchronization

Initial synchronization refers to synchronizing the historical data of the objects to be synchronized to the destination instance before synchronizing the incremental data through the synchronization channel.

Initial synchronization includes initial schema synchronization and initial full data synchronization. Initial schema synchronization refers to synchronizing the required schema syntax in the initial stage. Initial full data synchronization refers to synchronizing the data of the objects for the first time.

Synchronization performance

Synchronization performance is measured based on the number of records that are synchronized to the destination instance per second. The measurement unit is records per second (RPS).

Synchronization delay

Synchronization delay refers to the duration between the timestamp when the latest data in the destination instance is starting to be synchronized from the source instance and the current timestamp of the source instance. It reflects the time difference between the data in the source and destination instances. If the synchronization delay is zero, data in the source instance is in sync with that in the destination instance.

Subscription channel ID

The subscription channel ID is a unique identifier of a subscription channel. After you purchase a subscription channel, DTS automatically generates a subscription channel ID. To consume the incremental data using the SDK, you must configure a correct subscription channel ID. You can find the ID that corresponds to each subscription channel in the subscription list of the DTS console.

Data update

In DTS, you can update data or its schema. A data update only modifies the data. The schema syntax is not changed. Operations including INSERT, UPDATE, and DELETE fall into this category.

Schema update

In DTS, you can update data or its schema. Schema update modifies the schema syntax. Operations including CREATE TABLE, ALTER TABLE, and DROP VIEW fall into this category. You can choose whether to subscribe to schema update when you create a subscription channel.

Data range

Data range refers to the range of timestamps of incremental data stored in the subscription channel. The timestamp of a piece of incremental data is the time when the incremental data is applied and written to the transaction log in the database instance. By default, only data generated on the most recent day is retained in the subscription channel. DTS regularly cleans the expired incremental data and updates the data range of the subscription channel.

Consumption checkpoint

The consumption checkpoint is the timestamp of the latest consumed incremental data that is subscribed using the downstream SDK. The SDK sends an ACK message to DTS for every piece of data that is consumed. The server updates and saves the consumption checkpoint corresponding to the SDK. When the SDK encounters an exception, the server restarts and automatically pushes the data at the latest consumption checkpoint.

10 Data Management (DMS)

10.1 What is DMS?

Data Management (DMS) is a web-based data management software used to manage relational databases such as MySQL, as well as OLAP databases such as AnalyticDB.

It integrates data management with schema management.

10.2 Advantages

Extensive options for data sources

- Relational databases such as MySQL

Visualized data analysis

Visualized analysis of the numbers of read, inserted, deleted, and updated rows in business tables

Efficient R&D

- Schema comparison
- Smart SQL completion
- Convenient reuse of custom SQL statements and SQL templates
- Automatic recovery of working environments
- Export of dictionary files

10.3 Architecture

Data Management (DMS) consists of the business layer, scheduling layer, and connection layer. It processes real-time data access and schedules data-related background tasks for relational databases.

Business layer

- The business layer supports online GUI-based database operations, and can be scaled to improve the general service capabilities of DMS.
- DMS supports stateless failover to ensure 24/7 availability.

Scheduling layer

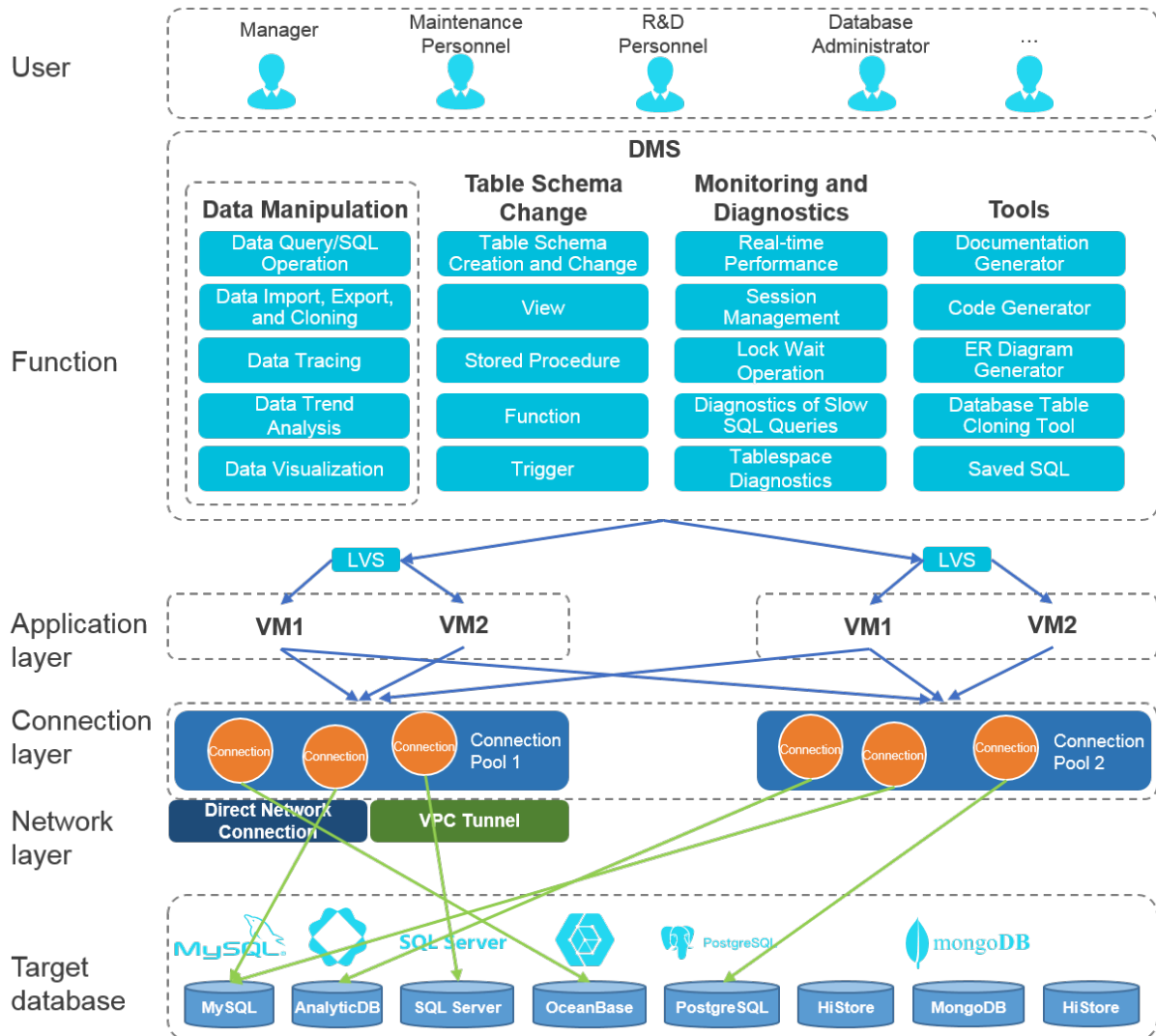
- The scheduling layer allows you to import and export tables and compare schemas. This layer schedules tasks by using the thread pool in real-time scheduling or background periodic scheduling mode.
- Real-time scheduling allows you to schedule and execute a task on the frontend. After you submit a task, DMS automatically executes the task in the background. After the task is completed, you can download or view the execution result.
- Background periodic scheduling allows you to periodically obtain specified data, such as data trends. DMS collects business data in the background based on scheduled tasks for your reference and analysis.

Connection layer

The connection layer is the core component to access data in DMS. It has the following characteristics:

- Processes requests from MySQL.
- Supports session isolation and persistence. SQL windows opened in DMS are isolated from each other and the sessions in each SQL window are persistent to simulate the client experience.
- Controls the number of instance sessions to avoid establishing a large number of connections to a single instance.

- Provides different connection release policies for different features. This improves user experience and reduces the number of connections to the databases.

Figure 10-1: DMS system architecture

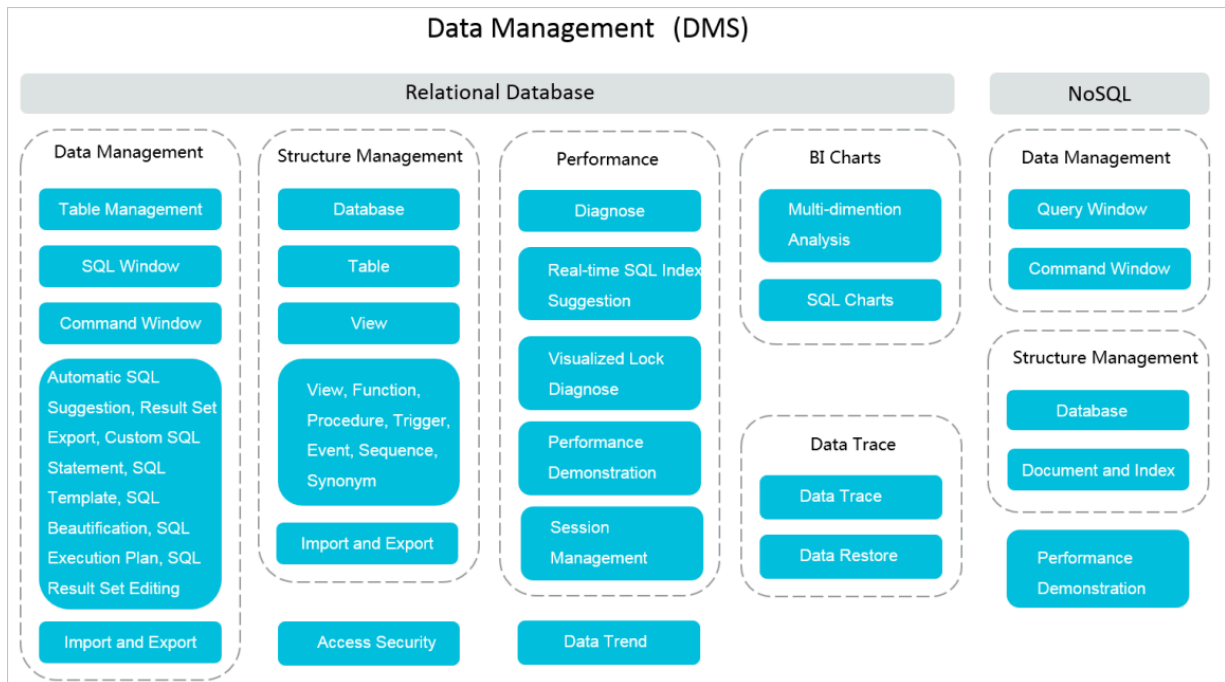
10.4 Features

Relational database management

- **Data management:** includes functions such as SQL windows, SQL command lines, table data, intelligent SQL prompts, SQL formatting, custom SQL statements, SQL templates, SQL execution plans, and import and export operations.
- **Structure management:** includes functions such as table structure comparison, and management of objects (databases, tables, views, functions, storage procedures, triggers, events, series, and synonyms).

Feature diagram

Figure 10-2: Feature diagram



10.5 Scenarios

10.5.1 Convenient data operations

Pain point

You need a lightweight product that features full functionality to create SQL statements, save frequently used SQL statements, and use these statements in your business.

Solution

- You can open a table in DMS and perform operations on table data as you would in an Excel worksheet. You can add, delete, change, query, and make statistical analysis of table data without understanding SQL.
- You can customize SQL statements, save frequently used SQL statements, and apply these SQL statements to databases or instances.

10.5.2 Prohibiting data export

Pain point

When cooperating with a partner, an enterprise manages data and its partner develops functions. The partner needs to have access to view the enterprise's data but cannot have the ability to export data to ensure data security.

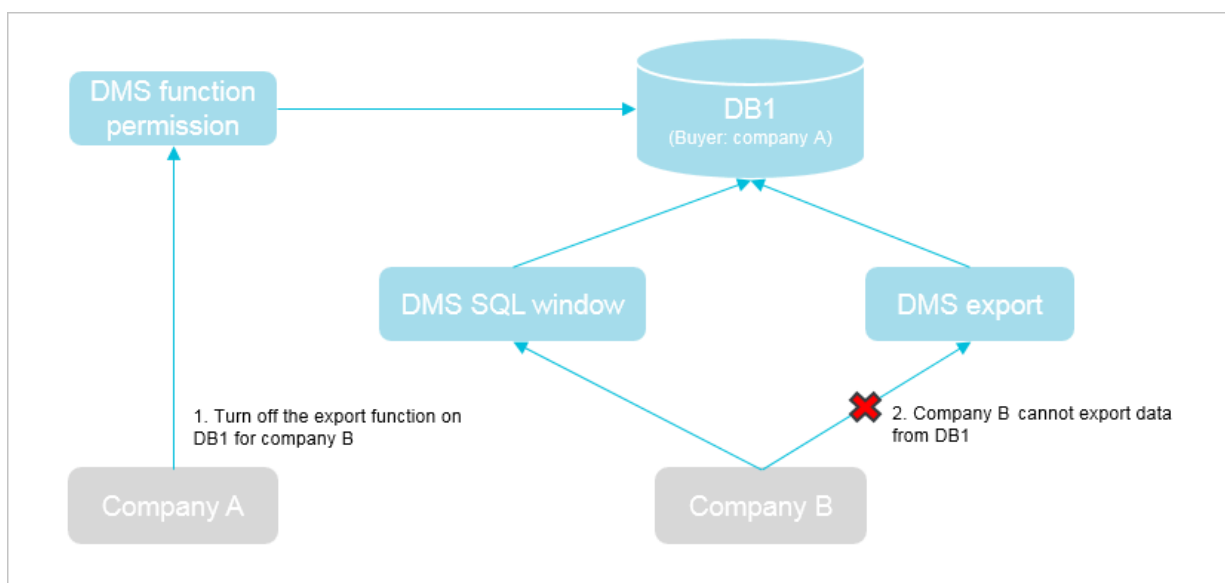
Solution

Enterprise users can log on to the DMS console to grant their partners access permissions on the corresponding database instances, disabling data exporting to protect their data.

Partners are permitted only to query and view data, eliminating the risk of data leakage.

[Figure 10-3: DMS # Function-based authorization](#) shows how to use the function-based authorization feature to prohibit partners from exporting data.

Figure 10-3: DMS # Function-based authorization



10.5.3 SQL statement reuse

Pain point

SQL statements are used when you access a database. While simple queries are easy to use, rewriting SQL queries for complex data analysis or SQL queries that contain service logic is time-consuming. Even if you save these SQL queries to files, you have to maintain the files and you cannot use them without access to the files.

Solution

You can use the **My SQL** function provided to save frequently used SQL statements to DMS. As the SQL statements are not saved locally, they can be reused in any databases or instances.

10.6 Limits

Relational databases

Table 10-1: Support for relational databases

Module	Function	MySQL	PostgreSQL
Data management	Table data management	√	√
	SQL windows	√	√
	SQL command lines	√	√
	SQL templates	√	
	SQL formatting	√	√
	Custom SQL statements	√	
	Intelligent SQL prompts	√	
	SQL execution plans	√	√
Structure management	Database management	√	√
	Table management	√	√
	Management of objects such as indexes, views, stored processes, functions, triggers, and events	√	√
	Entity relationship diagram display	√	
	Data dictionaries	√	
Import and export	Basic import and export functions	√	√
	Export of large volumes of data	√	√

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

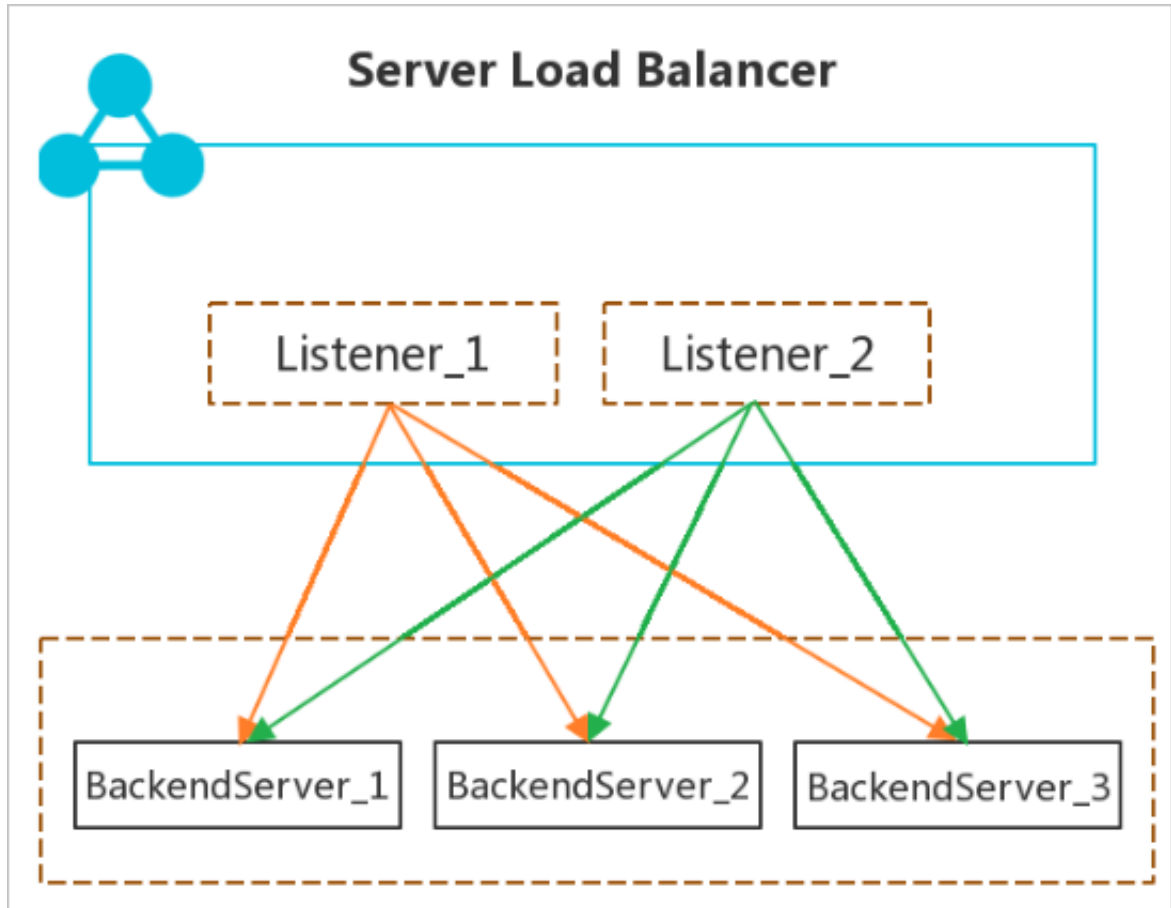
An SLB instance is a running load balancing service that receives inbound traffic and distributes the traffic to the backend servers. To use the SLB service, you must create an SLB instance with at least one listener and two backend servers.

- 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 High availability

This topic describes the high availability architecture of Server Load Balancer (SLB) in terms of system design and product configurations. You can use SLB in concert with Apsara Stack DNS to implement geo-disaster recovery to suit your business needs. SLB is designed to offer a multi-zone service availability of 99.99% and a single-zone service availability of 99.90%.

High availability of the SLB system

SLB instances are deployed in clusters to synchronize sessions and protect backend servers from single points of failure (SPOFs), improving redundancy and ensuring service stability . Layer-4 SLB uses the open source Linux Virtual Server (LVS) and Keepalived software to balance loads, whereas Layer-7 SLB uses Tengine. Tengine, a web server project launched by Taobao, is based on NGINX and adds advanced features dedicated for high-traffic websites.

Requests from the public network reach an LVS cluster along Equal-Cost Multi Path (ECMP) routes. In the LVS cluster, each machine uses multicast packets to synchronize sessions with the other machines. At the same time, the LVS cluster performs health checks on the Tengine cluster and removes unhealthy machines from the Tengine cluster to ensure the availability of Layer-7 SLB.

Best practice:

You can use session synchronization to prevent persistent connections from being affected by server failures within a cluster. However, for short-lived connections or if the session synchronization rule is not triggered by the connection (that is, the three-way handshake is not completed), server failures in the cluster may still affect user requests. To prevent session interruptions caused by server failures within the cluster, you can add a retry mechanism to the service logic to reduce the impact on user access.

High availability of a single SLB instance

To provide more stable and reliable load balancing services, Apsara Stack allows you to deploy SLB instances across multiple zones in most regions to achieve cross-data center disaster recovery. Specifically, you can deploy an SLB instance in two zones within the same region. One zone is the primary zone, while the other zone is the secondary zone. If the primary zone fails or becomes unavailable, the SLB instance will fail over to the secondary zone in about 30 seconds. When the primary zone recovers, the SLB instance will automatically switch back to the primary zone.

**Note:**

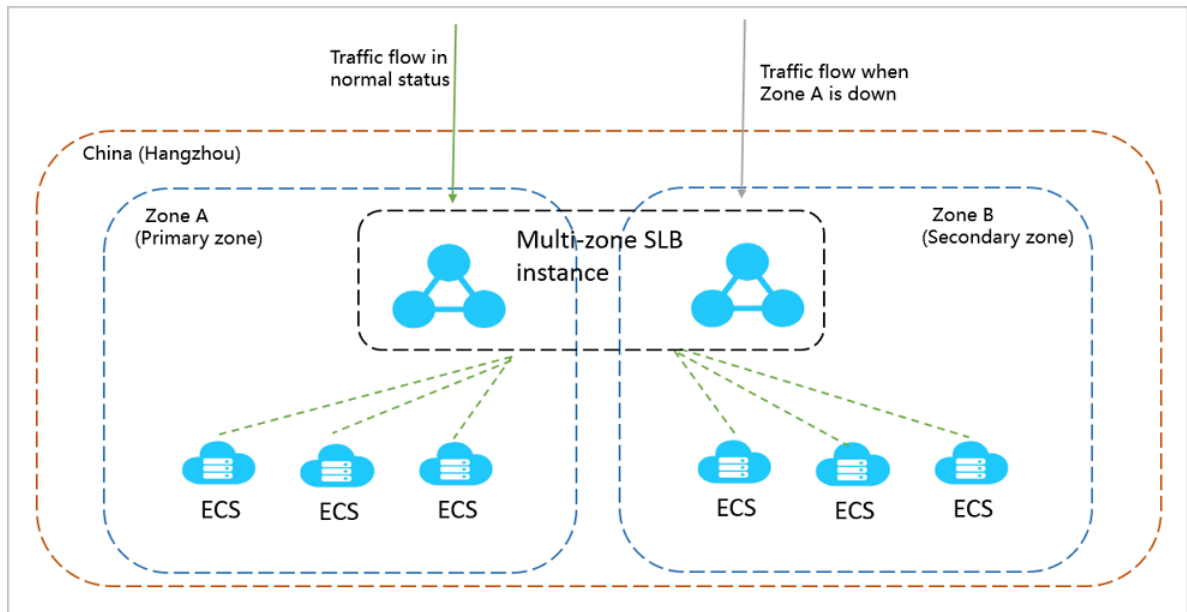
Zone-disaster recovery is implemented between the primary and secondary zones. The SLB instance fails over to the secondary zone only when the primary zone is unavailable due to power outage or optical cable failures, not simply when an ECS instance in the primary zone fails.

Best practice:

1. We recommend that you create an SLB instance in a region that has multiple zones for zone-disaster recovery.

2. You can determine the primary and secondary zones for an SLB instance based on the zone distribution of ECS instances. Select the zone where most of the ECS instances are located as the primary zone to minimize access latency.

We recommended that you do not deploy all ECS instances in the primary zone, but instead deploy a small number of ECS instances to the secondary zone to implement failover in extreme cases, such as when the entire primary zone becomes unavailable.

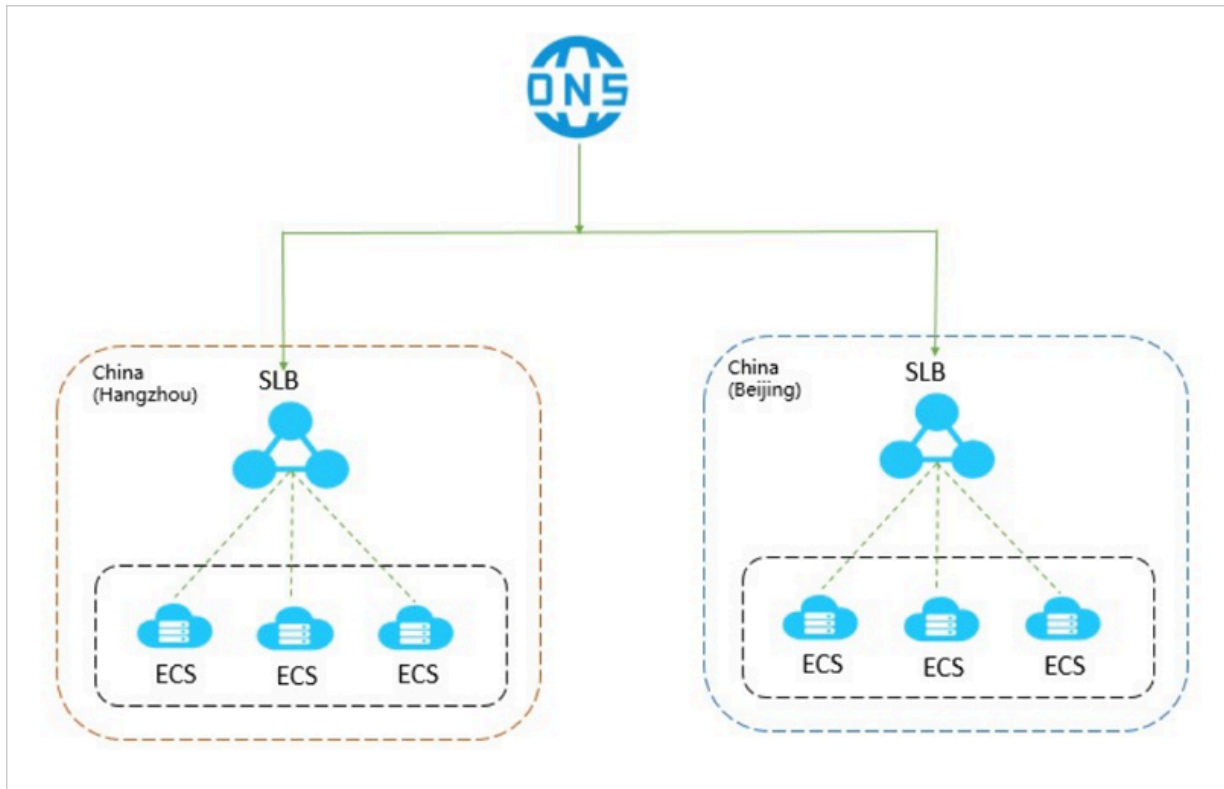


High availability of multiple SLB instances

You can create multiple SLB instances if a single SLB instance cannot meet your availability requirements. You can use Apsara Stack DNS to schedule access requests or use the global SLB solution to achieve cross-region backup and disaster recovery.

Best practice:

You can deploy SLB instances and ECS instances in multiple zones within a region or across multiple regions and schedule access requests by using Apsara Stack DNS.



High availability of backend ECS instances

SLB performs health checks to check the availability of backend ECS instances. The health check function can improve the overall availability of the frontend service and mitigate the impact of exceptions in backend ECS instances on services.

If a check result shows that a backend ECS instance is unhealthy, the corresponding SLB instance automatically distributes new requests to other healthy ECS instances. When the unhealthy ECS instance recovers, the SLB instance automatically resumes distributing requests to the ECS instance. For more information, see [Health check overview](#) in the SLB User Guide.

Best practice:

Enable and correctly configure the health check function. For more information, see [Configure health checks](#) in the SLB User Guide.

11.3 Architecture

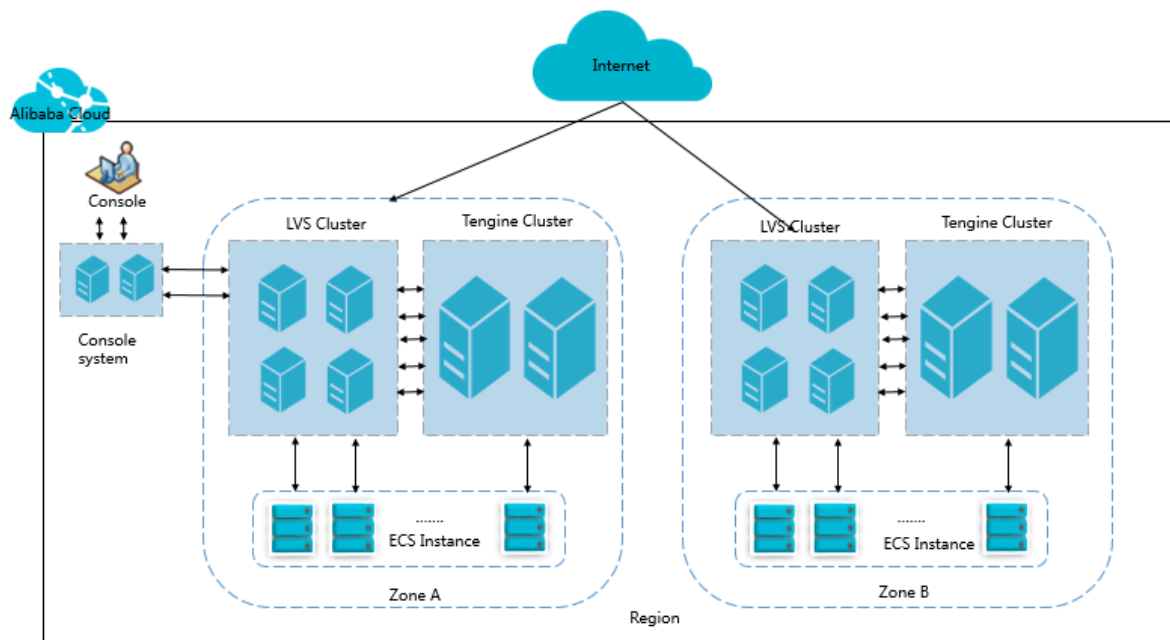
This topic describes the architecture of Server Load Balancer (SLB). SLB is deployed in clusters, which enables session synchronization and mitigates the creation of SPOFs in backend servers. As a traffic forwarding service, SLB forwards requests from clients to

backend servers through SLB clusters, and backend servers return the responses to SLB through the internal network.

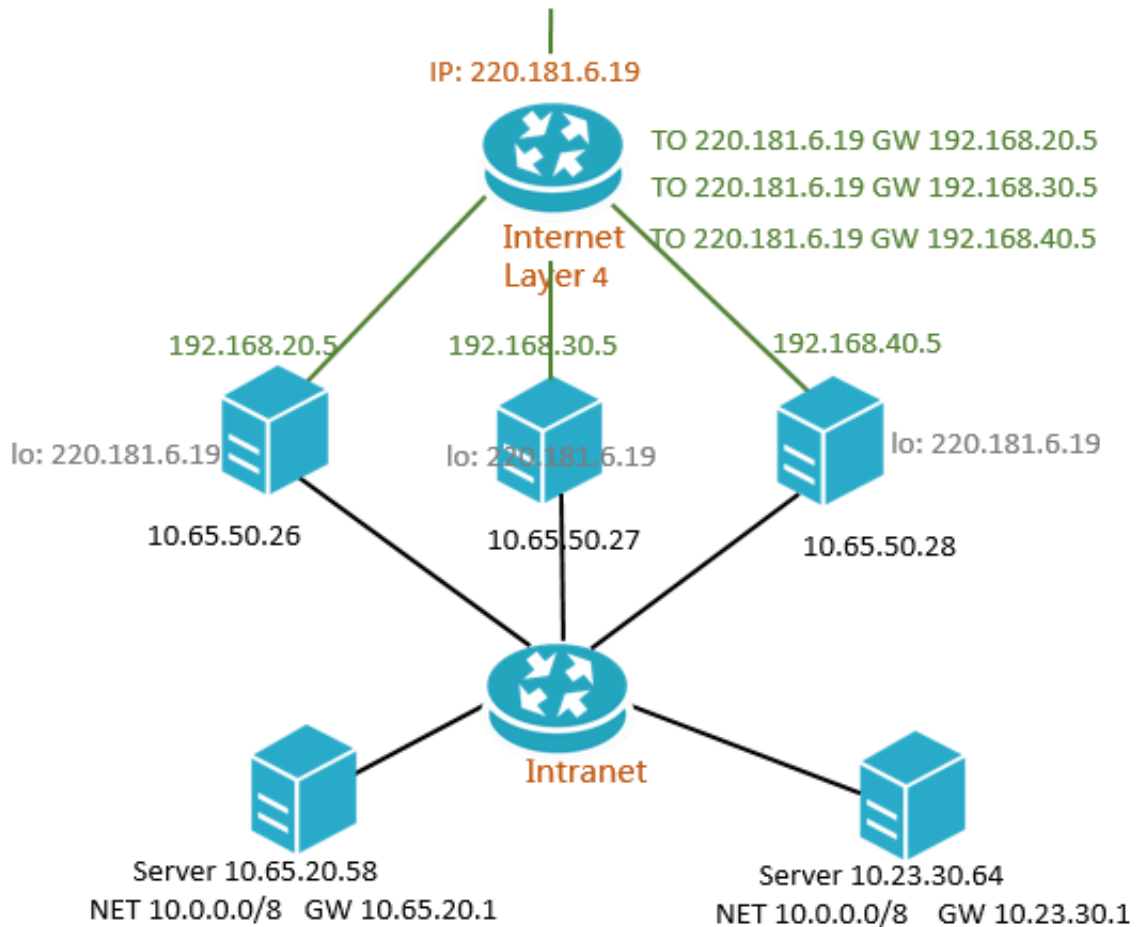
Architecture

Currently, Alibaba Cloud provides layer-4 (TCP protocol and UDP protocol) and layer-7 (HTTP protocol and HTTPS protocol) load balancing services.

- Layer-4 SLB uses the open source software Linux Virtual Server (LVS) and Keepalived to achieve load balancing. It also customizes the software to adapt to cloud computing requirements.
- Layer-7 SLB uses Tengine to achieve load balancing. Tengine is a web server project launched by another branch of the Alibaba Cloud Group (Taobao). Based on Nginx, Tengine provides a wide range of advanced features to support high-traffic websites.



As shown in the following figure, layer-4 SLB in each region actually runs in a cluster of multiple LVS machines. The cluster deployment model strengthens the availability, stability, and scalability of the load balancing services in abnormal circumstances.

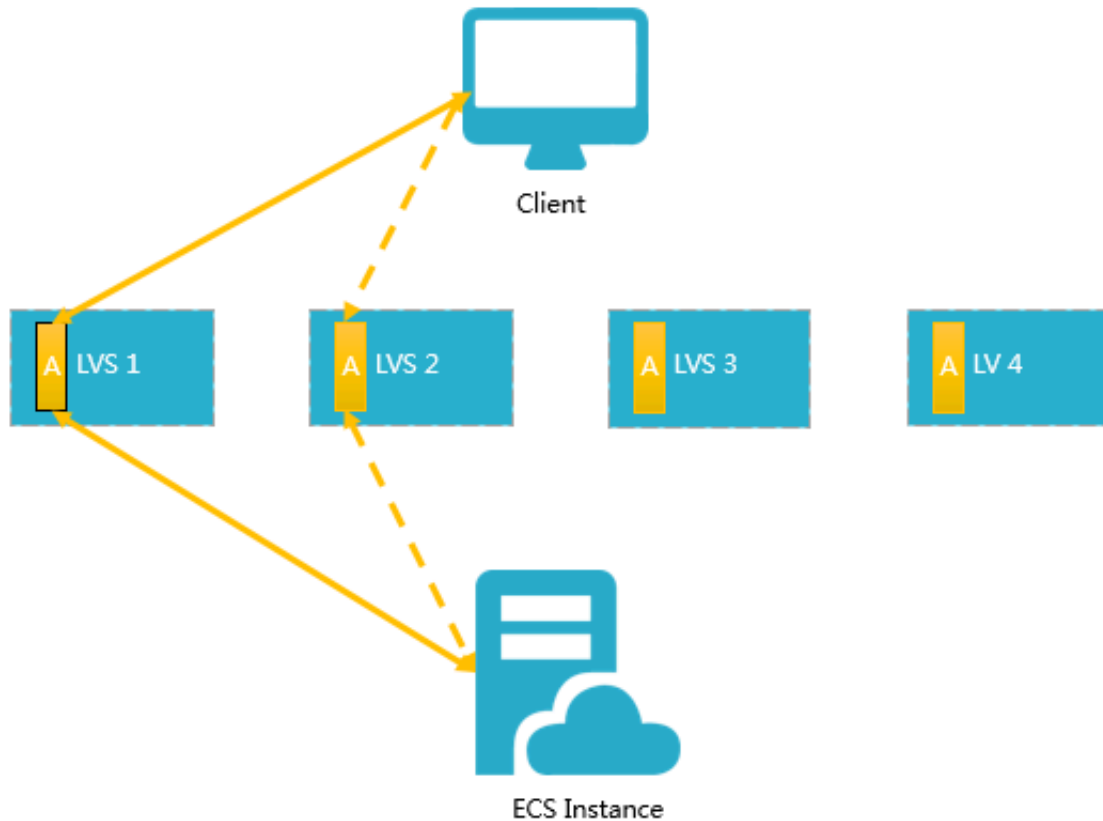


Additionally, each LVS machine in an LVS cluster uses multicast packets to synchronize sessions to other LVS machines. As shown in the following figure, after the client sends three packets to the server, session A is established on LVS1 and this session is synchronized to other LVS machines. In normal situations, the session request is sent to LVS1 as the solid line shows. If LVS1 fails or is being maintained, the session request is sent to other normally working machines, such as LVS2, as the dotted line shows. Therefore, SLB clusters support hot upgrades, and machine failures and system maintenance do not affect your business.




Note:

If a connection is not established (that is, a three-way handshake is not completed), or a connection has been established but the session synchronization is not triggered during a hot upgrade, your service may be interrupted and the client needs to re-initiate the connection.



11.4 Features

This topic describes the key features of Apsara Stack SLB and whether Layer-4 SLB and Layer-7 SLB support the features.

Feature	Layer-4 SLB	Layer-7 SLB
Scheduling algorithms SLB supports the following scheduling algorithms: round-robin, weighted round-robin (WRR), weighted least connections (WLC), and consistent hash.	√	√ <div>  Note: Currently, Layer-7 SLB does not support the scheduling algorithm of consistent hash. </div>

Feature	Layer-4 SLB	Layer-7 SLB
Health check SLB checks the health status of backend servers. If a backend server is declared as unhealthy, the corresponding SLB instance stops distributing inbound traffic to it and distributes the traffic to other healthy backend servers.	√	√
Session persistence SLB provides session persistence. In a session, the SLB instance can distribute all of the requests from the same client to the same backend server.	√	√
Access control SLB uses whitelists and blacklists to control access to your applications.	√	√
High availability SLB can forward inbound traffic to backend servers in different zones. In addition, SLB is deployed in primary/secondary mode in most regions, meaning that an SLB instance is deployed across two zones within the same region. One zone is the primary zone and the other zone is the secondary zone. If the primary zone fails, the SLB instance will automatically fail over to the secondary zone, ensuring the high availability of the SLB service.	√	√
Security SLB can be used with Apsara Stack Security to provide a DDoS protection capability of 5 Gbit/s.	√	√

Feature	Layer-4 SLB	Layer-7 SLB
<p>Public and internal network load balancing</p> <p>SLB can load-balance both public network traffic and internal network traffic. You can create an internal SLB instance to balance traffic within a VPC, or create a public SLB instance to balance traffic coming from the public network.</p>	√	√
<p>Monitoring</p> <p>You can use CloudMonitor to view the monitoring data of SLB instances such as the number of connections and inbound and outbound traffic volumes.</p>	√	√
<p>IPv6 support</p> <p>SLB can forward requests from IPv6 clients.</p>	√	√
<p>Domain name forwarding and URL forwarding</p> <p>Layer-7 SLB supports configuration of domain name forwarding rules and URL forwarding rules to forward requests from different domain names or URLs to different backend servers.</p>	-	√
<p>Certificate management</p> <p>SLB provides centralized certificate management for HTTPS listeners. You do not need to upload certificates to backend servers. Deciphering is performed on SLB instances to reduce the CPU utilization of backend servers.</p>	-	√
<p>SNI support</p> <p>SLB allows you to associate multiple certificates to an HTTPS listener to distribute requests from different domain names to different backend servers.</p>	-	√
<p>Redirection</p> <p>SLB supports redirecting HTTP requests to HTTPS.</p>	-	√

Feature	Layer-4 SLB	Layer-7 SLB
WS/WSS support WebSocket is a new HTML5 protocol. It provides bidirectional communication channels between clients and servers, saving server resources and bandwidth, and achieving real-time communication.	-	√
HTTP/2 support HTTP/2 is the second version of HTTP. It is backward compatible with HTTP1.X and significantly improves performance.	-	√

11.5 Scenarios

Server Load Balancer (SLB) is suitable for applications with high access traffic. It improves the availability and reliability of your applications.

Balance the loads of your applications

If your applications experience high access traffic, SLB can help you distribute inbound traffic among multiple backend servers (ECS instances) according to the listening rules you set. Additionally, you can configure session persistence to distribute the requests from the same client to the same backend server.

Scale your applications

You can add or remove backend servers at any time to scale the serving capacity of your applications. SLB supports various web servers and application servers.

Protect your applications from single points of failure

You can add multiple ECS instances to an SLB instance. If some ECS instances are faulty, SLB automatically shields faulty ECS instances and distributes requests to healthy ECS instances, guaranteeing that your applications can work normally.

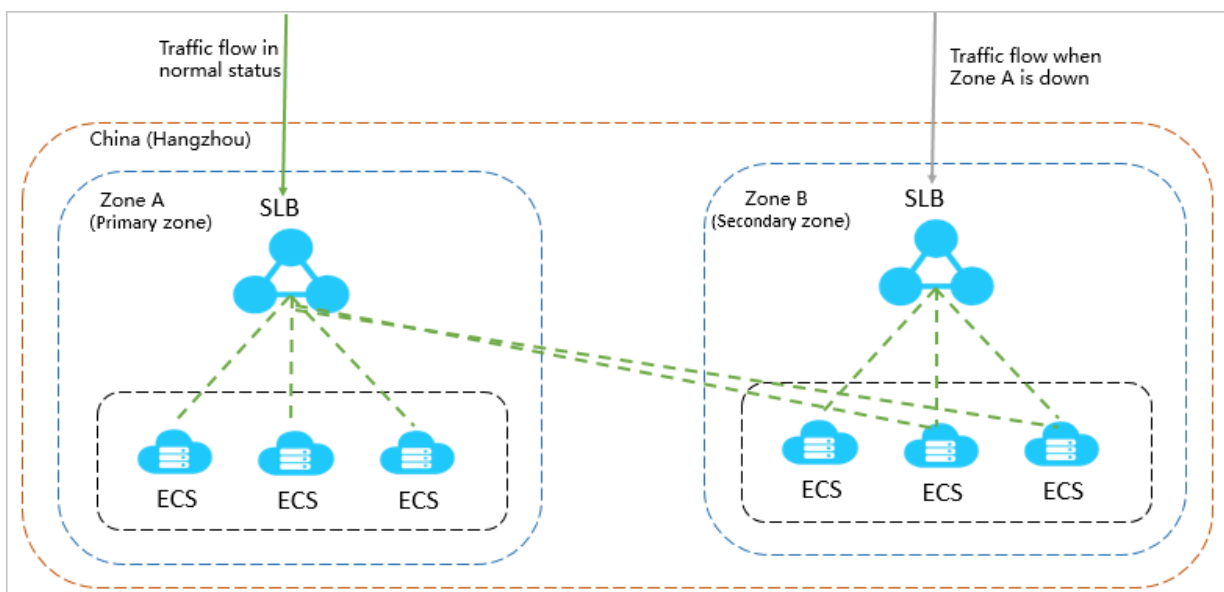
Realize disaster tolerance across zones

To deliver more stable and reliable services, SLB provides multiple zones in most regions to achieve same-region disaster tolerance. If the primary zone becomes unavailable, SLB switches to the secondary zone in about 30 seconds without affecting the load balancing

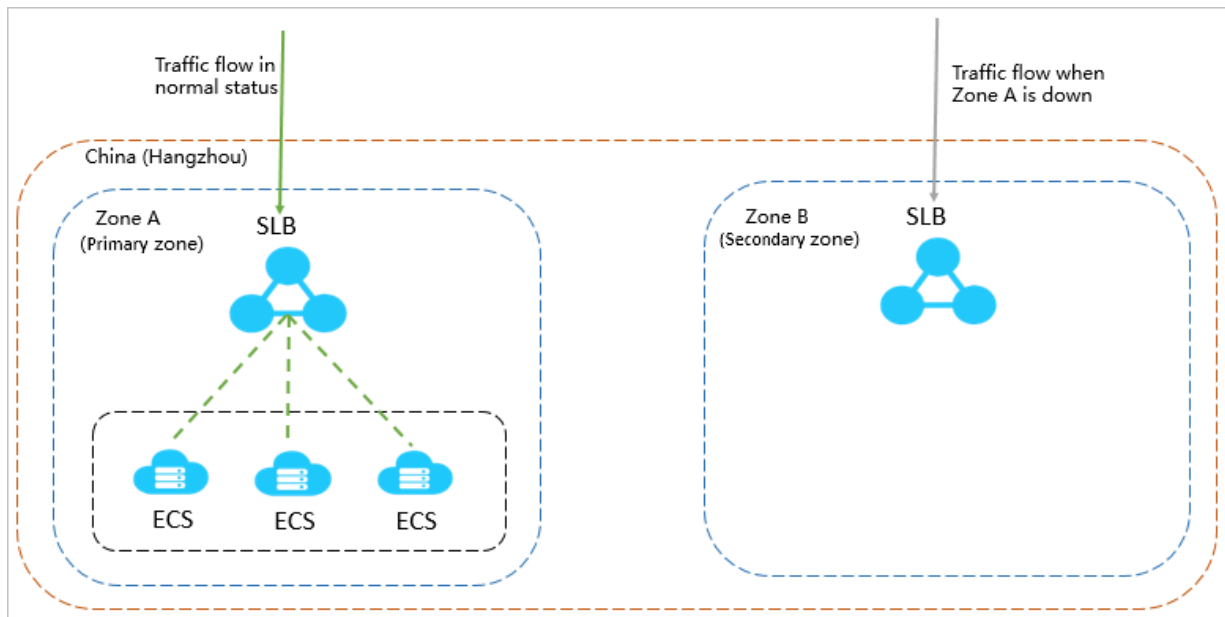
service. After the primary zone becomes available, SLB automatically switches back to the primary zone.

SLB is most effective when you create an SLB instance in a region with multiple zones and add at least one backend server for each zone.

As shown in the following figure, an SLB instance is configured with ECS instances in different zones. In normal situation, SLB distributes traffic to ECS instances both in the primary zone (Zone A) and in the secondary zone (Zone B). If the primary zone is unavailable, the traffic is distributed to the ECS instances in the secondary zone. This avoids service interruptions caused by the failure of a single zone, and also reduces latency.

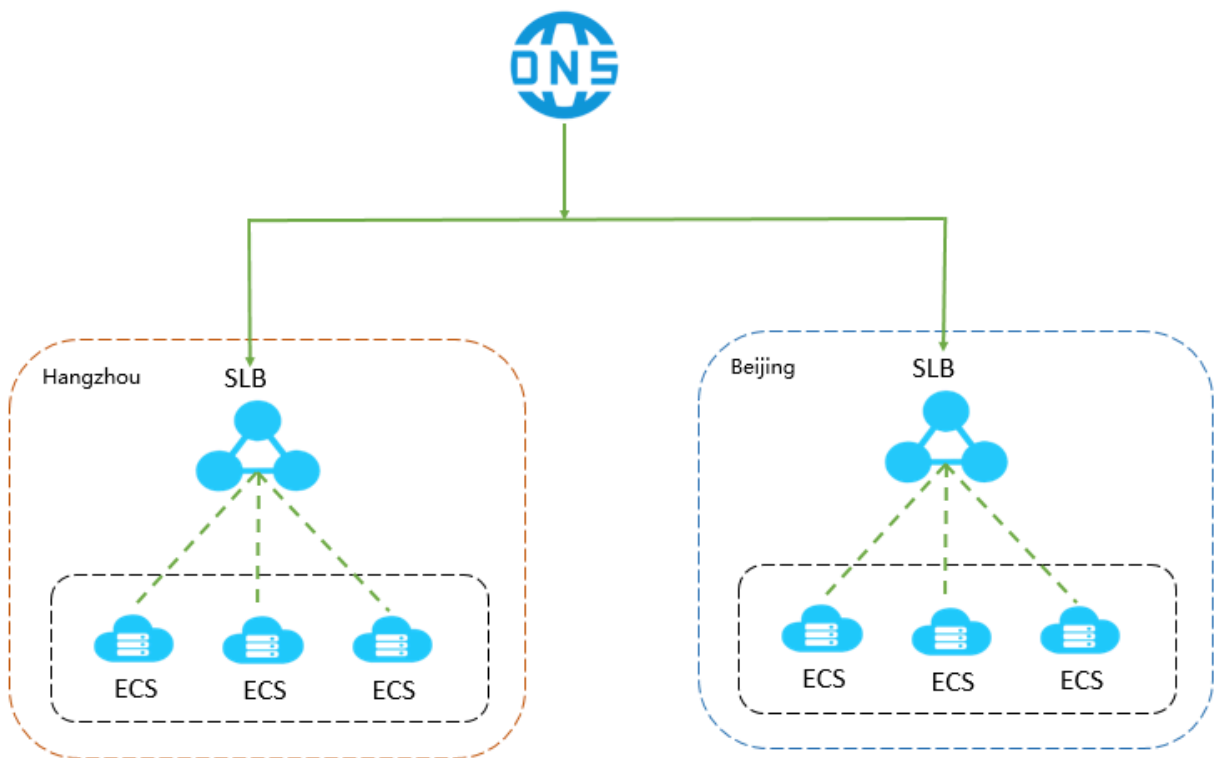


However, if you deploy all ECS instances in the primary zone, your service will be interrupted when the primary zone is unavailable, because no ECS instances are available in the secondary zone to handle the distributed requests. This deployment mode achieves low latency at the expense of high availability.



Achieve disaster tolerance across regions

You can deploy SLB instances in different regions, and add ECS instances of different zones in the same region to an SLB instance. Then, you can use Alibaba Cloud DNS to resolve domain names to service addresses of SLB instances in different regions, thus implementing global load balancing. When a region becomes unavailable, you can temporarily stop DNS in that region, and no user access is affected.



11.6 Limits

This topic lists the resource quotas of Server Load Balancer (SLB).

SLB provides an API to query the default limits of an SLB instance.

Resource	Default limit
Limits on SLB instances	
The number of SLB instances that can be created under each account	6030
The number of times that an ECS instance can be added to SLB instances	50
The number of backend servers that can be added to an SLB instance	200
The number of listeners that can be added to an SLB instance	50
The number of domain name-based and URL-based forwarding rules that can be added to a layer-7 listener	40
The number of domain name extensions that an HTTPS listener can create	3
The range of frontend and backend port that can be used by a listener (public cloud)	1 to 65535
The range of frontend and backend port that can be used by a listener (Ant Financial Cloud)	80, 443, 2800 to 3300, 5000 to 10000, 13000 to 14000
Limits on certificates	
The number of server certificates that can be uploaded in a region	100
The number of CA certificates that can be uploaded in a region	100
Limits on access control lists	
The number of access control lists that can be created in a region	50
The number of IP entries that can be added to an access control list	300

Resource	Default limit
The number of IP entries that can be added to an access control list in an API call	50
The number of times that an access control list can be added to listeners	50

11.7 Terms

This topic describes the terms commonly used in Alibaba Cloud Server Load Balancer.

Term	Description
Server Load Balancer	Alibaba Cloud Server Load Balancer (SLB) is a traffic distribution and control service that distributes the incoming traffic among multiple Elastic Compute Service (ECS) instances according to configured forwarding rules.
SLB instance	An SLB instance is a running entity of the SLB service. To use SLB, you must first create an SLB instance.
Endpoint	An IP address allocated to an SLB instance. According to the instance type, the IP address is either a public IP address or a private IP address. You can resolve a domain name to a public IP address to provide external services.
Listener	A listener defines how incoming requests are distributed. An SLB instance must contain at least one listener.
Backend server	The ECS instances that are added to an SLB instance to process distributed requests.
Default server group	A group of ECS instances that process the distributed requests. If a listener is not associated with a VServer group or an active/standby server group, the default server group is used. Incoming traffic is distributed to ECS instances in the default server group.
VServer group	A group of ECS instances that process the distributed requests. Different listeners can be associated with different VServer groups to make sure that requests are distributed only to desired VServer groups.

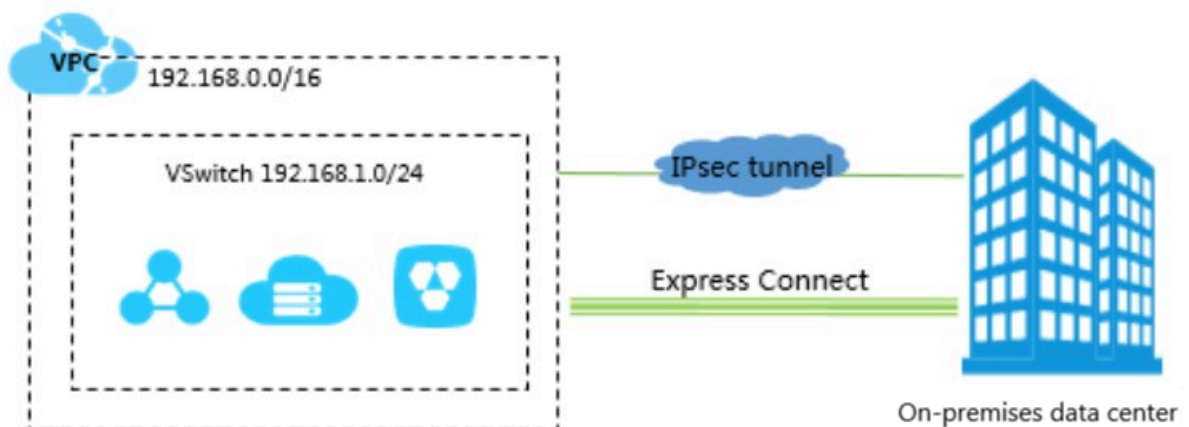
Term	Description
Active/standby server group	An active/standby server group contains only two ECS instances . One is the active server and the other is the standby server. When the health check of the active server fails, SLB automatically forwards traffic to the standby server.

12 Virtual Private Cloud (VPC)

12.1 What is a VPC?

A Virtual Private Cloud (VPC) is a private network dedicated to your Alibaba Cloud account. You have full control over your VPC, which you can define and customize by specifying the IP address range and configuring route tables and network gateways. You can also launch Alibaba Cloud resources such as Elastic Compute Service (ECS), Relational Database Service (RDS), and Server Load Balancer (SLB) in your own VPC.

Furthermore, you can connect your VPC to other VPCs or local networks to create a custom network environment. In this way, you can smoothly migrate applications to the cloud and extend on-premises data centers.



Components

Each VPC consists of one VRouter, at least one private CIDR block, and one or more VSwitches.

- Private CIDR block

When you create a VPC or a VSwitch, you must specify the private IP address range in the form of a CIDR block.

You can use the standard private CIDR blocks listed in the following table and their subnets as CIDR blocks of your VPCs. For more information, see "Plan a VPC network" chapter of the User Guide.

CIDR block	Number of available private IP addresses (excluding those reserved by the system)
192.168.0.0/16	65,532
172.16.0.0/12	1,048,572
10.0.0.0/8	16,777,212

- VRouter

A VRouter is a hub that connects all VSwitches in a VPC and serves as a gateway between the VPC and other networks. After a VPC is created, a VRouter is automatically created for the VPC. Each VRouter is associated with a route table.

For more information, see "Route table overview" chapter of the User Guide.

- VSwitch

A VSwitch is a basic network device that connects different cloud resources in a VPC. After you create a VPC, you can create one or more subnets in the VPC by creating VSwitches. The VSwitches within a VPC are interconnected. You can deploy your applications in VSwitches that belong to different zones to improve service availability.

For more information, see "Create a VSwitch" chapter of the User Guide.

12.2 Benefits

This topic describes the benefits of using VPCs.

High security

Each VPC has a unique tunnel ID, and each tunnel ID corresponds to a virtual network. Different VPCs are isolated by tunnel IDs:

- Similar to traditional networks, VPCs can also be divided into subnets. ECS instances in the same subnet use the same VSwitch to communicate with each other, while ECS instances in different subnets use VRouters to communicate with each other.

- VPCs are completely isolated from each other and can only be interconnected by mapping an EIP or a NAT IP address.
- ECS IP packets are encapsulated by using the tunneling technique. Therefore, information about the data link layer (layer-2 MAC address) of ECS does not go to the physical network. As a result, the layer-2 network between different ECS instances or between different VPCs is isolated.
- ECS instances in a VPC use security groups as firewalls to control traffic going to and from ECS instances. This is layer-3 isolation.

High flexibility

You can use security groups or whitelists to flexibly control traffic going to and from the cloud resources in a VPC.

Ease of use

You can quickly create and manage VPCs in the VPC console. After a VPC is created, the system automatically creates a VRouter and a route table for the VPC.

High scalability

You can create multiple subnets in a VPC to deploy different services. Additionally, you can connect a VPC to other VPCs or on-premises data centers to expand your network.

12.3 Architecture

Based on the tunneling technique, VPCs isolate virtual networks. Each VPC has a unique tunnel ID, and each tunnel ID corresponds to only one VPC.

Background information

With the development of cloud computing, a variety of network virtualization techniques have been developed to meet the increasing demands for virtual networks with higher scalability, security, reliability, privacy, and connectivity.

Earlier solutions combined the virtual network with the physical network to form a flat network, for example, the large layer-2 network. However, with the increase of virtual network scale, problems such as ARP spoofing, broadcast storms, and host scanning are becoming more serious. To resolve these problems, various network isolation techniques are developed to completely isolate the physical network from the virtual network. One of these techniques can isolate users with a VLAN. However, a VLAN only supports up to 4,096 nodes, which are insufficient for the large number of users in the public cloud.

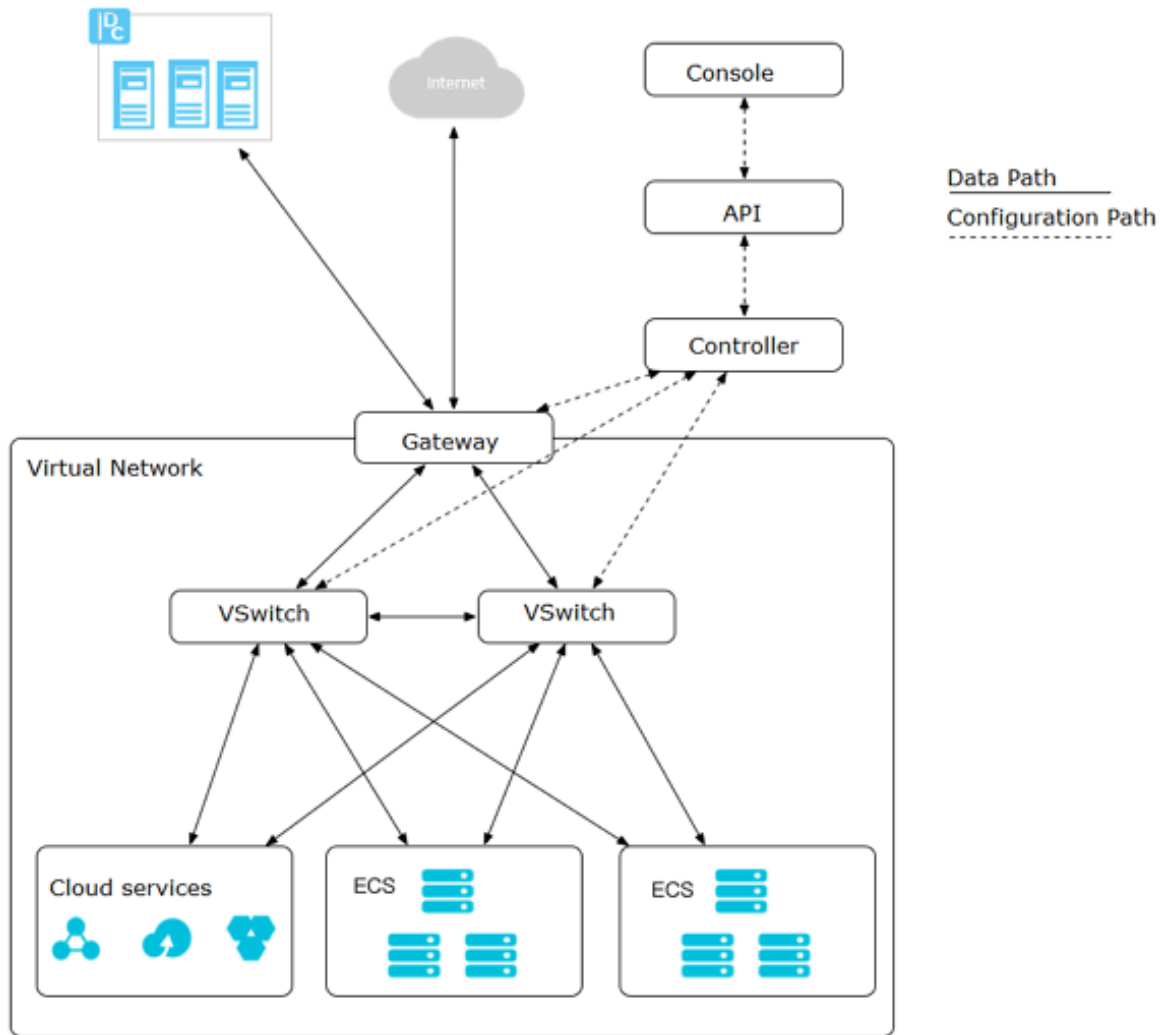
Principles

Based on the tunneling technique, VPCs isolate virtual networks. Each VPC has a unique tunnel ID, and each tunnel ID corresponds to only one VPC. A tunnel encapsulation carrying a unique tunnel ID is added to each data packet transmitted over the physical network between ECS instances in a VPC. In different VPCs, ECS instances with different tunnel IDs are located on two different routing planes. Therefore, these ECS instances cannot communicate with each other.

Based on the tunneling and Software Defined Network (SDN) techniques, Alibaba Cloud has developed VPCs that are integrated with gateways and VSwitches.

Logical architecture

As shown in the following figure, a VPC consists of a gateway, a controller, and one or more VSwitches. The VSwitches and gateway form a key data path. By using a protocol developed by Alibaba Cloud, the controller distributes the forwarding table to the gateway and VSwitches to provide a key configuration path. In the overall architecture, the configuration path and data path are separated from each other. The VSwitches are distributed nodes, the gateway and controller are deployed in clusters, and all links are equipped with disaster recovery. These features improve the availability of the VPC.



12.4 Features

Custom VPCs

You can customize VPCs. When you create VPCs or VSwitches, you can specify private CIDR blocks for them. You can divide a VPC into multiple subnets and deploy services in different subnets to improve service availability.

Custom routes

You can add custom routes to the VPC routing table to forward traffic to the specified next hop. The routing table uses the longest prefix matching rule for traffic routing. The routing entry with the longest subnet mask will be used because it is the most specific route.

Varied connection methods

A VPC provides you with varied connection methods. You can connect a VPC to the public network, a local IDC, or another VPC.

- Connect a VPC to the public network

You can connect a VPC to the public network by binding an EIP to the VPC or configuring NAT Gateway, so that cloud services in the VPC can communicate with the public network

- Connect a VPC to another VPC

You can connect a VPC to another VPC by creating a pair of router interfaces to build high speed and secure intranet communication.

- Connect a VPC to a local IDC

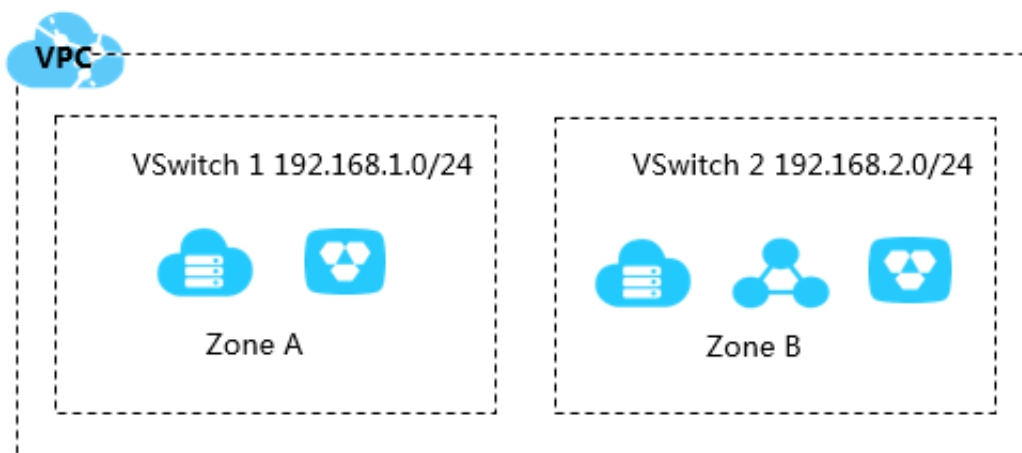
You can connect a VPC to a local IDC by using a leased line to smoothly migrate local applications to the cloud.

12.5 Scenarios

This topic describes the scenarios in which VPCs are used to guarantee a high level of data security and service availability.

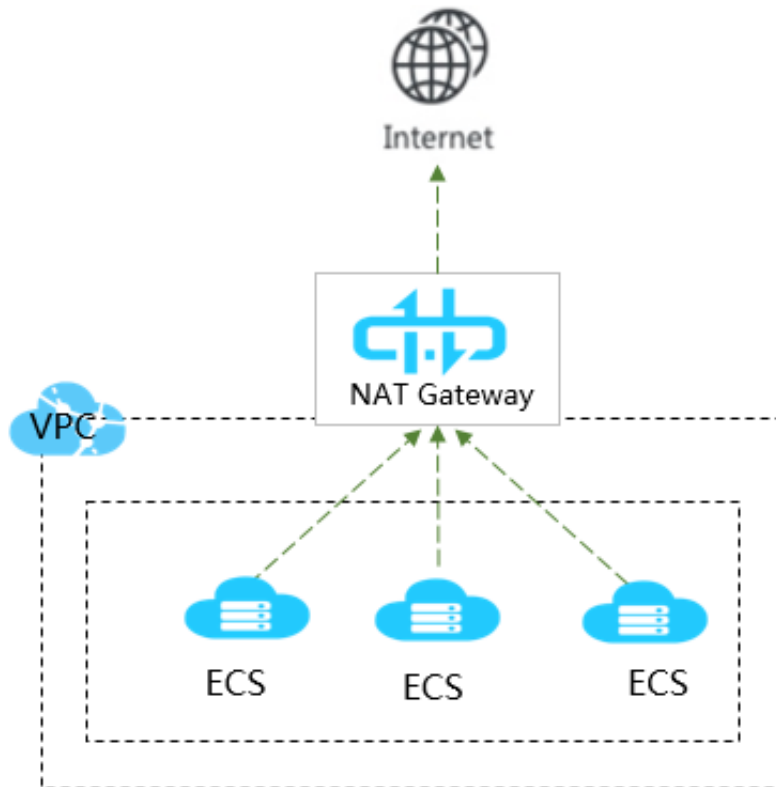
Host applications that provide external services

You can host applications that provide external services in a VPC and control access to these applications from the Internet by creating security group rules and access control whitelists. You can also isolate Internet-based mutual access between the application server and the database. For example, you can deploy the web server in a subnet that can access the Internet and deploy the application database in a subnet that cannot access the Internet.



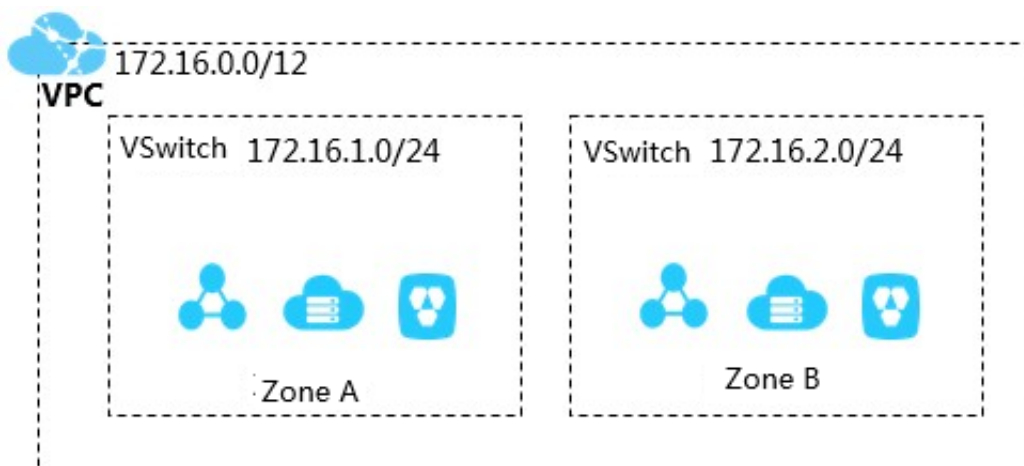
Host applications that require access to the Internet

You can host applications that require access to the Internet in a subnet of a VPC and route traffic through network address translation (NAT). After you configure SNAT rules, instances in the subnet can access the Internet without exposing their private IP addresses, which can be changed to public IP addresses any time to avoid external attacks.



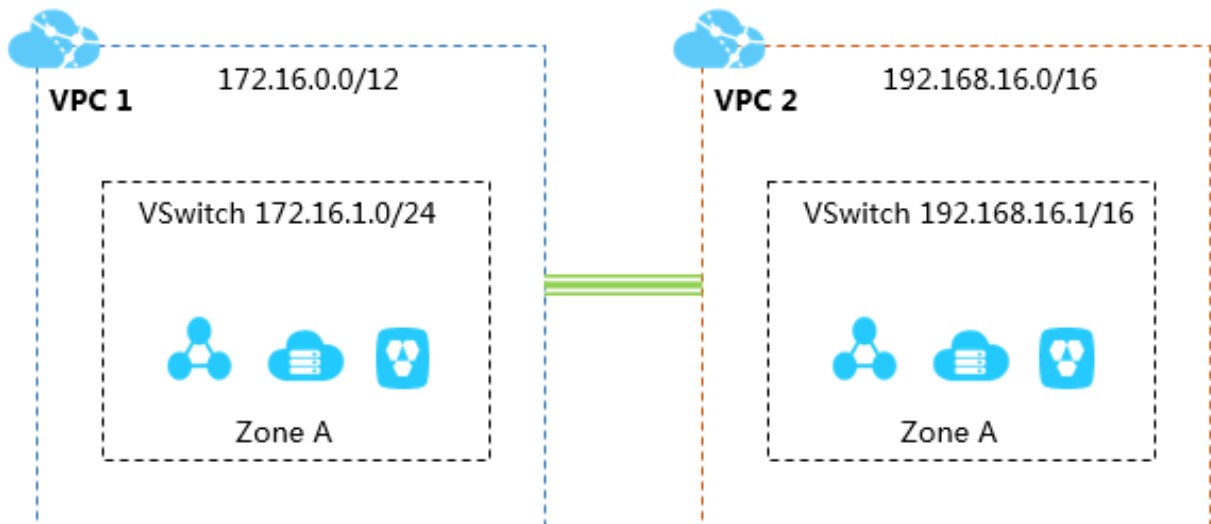
Implement disaster tolerance across zones

You can create one or multiple subnets in a VPC by creating VSwitches. VSwitches in a VPC can communicate with each other. You can deploy resources on VSwitches in different zones for disaster tolerance.



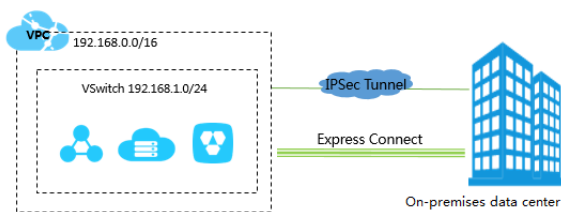
Isolate business systems

VPCs are logically isolated from each other. Therefore, you can create multiple VPCs to isolate multiple business systems, for example, isolate the production environment from the test environment. You can also create a peering connection between two VPCs if they need to communicate with each other.



Build a hybrid cloud

You can create a dedicated connection to connect your VPC to an on-premises data center to expand your local network. By doing so, you can seamlessly migrate your local applications to the cloud without changing the method of access to these applications.



12.6 Terms

This topic describes the terms about VPCs.

Term	Description
Virtual Private Cloud (VPC)	A VPC is a private network established in Alibaba Cloud. VPCs are logically isolated from each other. You can create and manage cloud resources in your VPC, such as ECS, SLB, and RDS.

Term	Description
VSwitch	A VSwitch is a basic network device that connect different cloud resources in a VPC. When you create a cloud resource in a VPC, you must specify the VSwitch to which the cloud resource is connected.
VRouter	A VRouter is a hub that connects all VSwitches in a VPC and serves as a gateway that connects the VPC to other networks. A VRouter also forwards network traffic according to the route entries in its route table.
Route table	A route table is a list of route entries in a VRouter.
Route entry	Each item in a route table is a route entry. A route entry specifies the next hop address for the network traffic directed to a destination CIDR block. Route entries are divided into system route entries and custom route entries.

12.7 Limits

The following table describes the limits that apply to Virtual Private Cloud (VPC) networks.

VPC

Resource	Limit
Maximum number of VRouters per VPC	1
Maximum number of route tables per VPC	1
Maximum number of VSwitches per VPC	24
Maximum number of route entries per route table	48

VRouter and VSwitch

Resource	Limit
VRouter	<ul style="list-style-type: none"> Each VPC has only one VRouter. Each VRouter has only one route table. BGP, OSPF, or other dynamic routing protocols are not supported.
VSwitch	Layer-2 broadcast or multicast is not supported.

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.
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.

Security domain	Service	Description
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 Benefits

As a pioneer of cloud security, Apsara Stack Security has received a variety of authoritative certifications. Through mature security systems and advanced security technologies, Apsara Stack Security can fully protect the security of the Apsara Stack environment.

Pioneer of cloud security

The Apsara Stack Security team has accumulated a wealth of security experience by protecting all internal business systems of Alibaba Group since 2005. Since its release in 2011, Apsara Stack Security has become a pioneer in providing comprehensive protection for cloud security.

Apsara Stack Security protects more than 40% of all websites in China. It prevents more than 50% of all distributed denial of service (DDoS) attacks and blocks up to 3.6 billion attacks every day. It has fixed over 6.13 million vulnerabilities over the last year.

Mature systems and advanced technologies

Apsara Stack Security is a service born from ten years of protection experience. After a decade of experience in providing security operations services for the internal business systems of Alibaba Group, Alibaba has obtained considerable security research achievements, security data, and security operations methods, and has built a professional cloud security team. Apsara Stack Security brings together the rich experience of these experts to develop the sophisticated systems that provide enhanced security for cloud computing platforms. This service can protect the cloud platform, cloud network environments, and cloud business systems of Apsara Stack users.

Comparison with traditional security products

Feature	Traditional security product	Apsara Stack Security
Comprehensive industry-leading security capabilities among Internet enterprises	A traditional security service provider only has limited products and features and cannot provide a comprehensive security protection system.	Alibaba has accumulated a large number of intelligence sources through years of attack prevention experience . This has allowed it to detect common Internet attacks including zero-day exploits, and provide comprehensive security capabilities.
Early risk detection	Traditional security service providers cannot detect risks due to a lack of complete monitoring systems.	Apsara Stack Security can detect and respond to critical vulnerabilities and security events quickly to prevent security issues.
Security big data modeling analysis	Traditional security service providers cannot detect threats through signature scanning. The traditional log analysis feature only provides data collection and reporting. It does not provide data modeling analysis.	Big data modeling analysis enables Apsara Stack Security to detect threats in the entire network and display the security data. More than 30 algorithmic models are used to analyze the historical data, network data, and server data. This enables security situation awareness.

Feature	Traditional security product	Apsara Stack Security
Scalability and decoupling with hardware	Traditional security products are developed based on the existing hardware devices. Security product software relies on the virtual machines created on virtualization platforms.	<ul style="list-style-type: none"> Hardware and software decoupling: All modules are developed based on the cloud computing architecture and the common x86 hardware platform, and therefore do not rely on specific hardware. Scalability: You can increase the amount of hardware for higher performance without the need to change the network architecture.
Collaboration between the network and servers	Traditional security service providers increase security features by adding devices. The devices can only collect device logs and status data and display the data on the management platform. They cannot collaborate to provide more features.	Apsara Stack Security provides complete Internet protection to guarantee the security of networks, applications, and servers. The security modules interact with each other to form a comprehensive protection system that blocks attacks effectively.
Compatibility with all data center environments and decoupling with specific cloud platforms	Most traditional security products are provided as hardware appliances. This makes the product incompatible with the cloud platforms based on Software Defined Network (SDN) technology.	Based on the interactions between servers and the operating system, Apsara Stack Security detects threats at the network perimeter through data analysis. This enables the service compatibility with all data center environments by avoiding the complex network topology inside the data centers.

13.3 Architecture

Apsara Stack Security consists of Apsara Stack Security Standard Edition and optional security services.

Apsara Stack Security Standard Edition

- Threat Detection Service

This module collects network traffic and server information, and detects possible vulnerability exploits, intrusions, and virus attacks through machine learning and data modeling. It also provides you with up-to-date information about ongoing attacks to help you monitor the security status of your businesses.

- Network Traffic Monitoring System

This module is deployed on the network perimeter of Apsara Stack. It allows you to inspect and analyze each inbound or outbound packet of an Apsara Stack network through traffic mirroring. The analysis results are used by other Apsara Stack Security modules.

- Server Guard

This module safeguards ECS instances by providing security features such as vulnerability management, baseline check, intrusion detection, and asset management. To do this, the module performs operations such as log monitoring, file analysis, and signature scanning.

- Web Application Firewall

This module protects web applications against common web attacks reported by Open Web Application Security Project (OWASP), such as Structured Query Language (SQL) injections, cross-site scripting (XSS), exploitation of vulnerabilities in web server plug-ins, Trojan uploads, and unauthorized access. It also blocks a large number of malicious visits to avoid data leaks and ensure both the security and availability of your websites.

Apsara Stack Security Standard Edition also provides on-premises security services. These services help you better use the features of Apsara Stack products such as Apsara Stack Security to secure your applications.

On-premises security services include pre-release security assessment, access control policy management, Apsara Stack Security configuration, periodic security check, routine security inspection, and urgent event handling. These services cover the entire lifecycle of your businesses in Apsara Stack and help you create a security operations system. This

system enhances the security of your application systems and ensures both the security and stability of your businesses.

Optional security services

You can also choose the following service modules to enhance your system security.

- DDoS Traffic Scrubbing

This module detects and blocks distributed denial of service (DDoS) attacks.

- Sensitive Data Discovery and Protection

This module uses Alibaba Cloud's big data analytics capabilities and artificial intelligence (AI) technologies to detect and classify sensitive data based on your business requirements. It masks sensitive data both in transit and at rest, monitors dataflows, and detects abnormal activities. This module provides visible, controllable, and industry-compliant security protection for your sensitive data by means of precise detection and analysis.

13.4 Features

13.4.1 Apsara Stack Security Standard Edition

Apsara Stack Security is developed based on the Apsara Stack environment and adopts a cloud security architecture that enables in-depth defense and multi-module collaboration. Unlike traditional software and hardware security products, Apsara Stack Security provides comprehensive and integrated cloud security protection at multiple layers, such as the network layer, application layer, and server layer.

The following table describes the features of Apsara Stack Security Standard Edition.

Table 13-1: Features of Apsara Stack Security Standard Edition

Module	Feature	Description
Network Traffic Monitoring System	Traffic collection and analysis	Uses a bypass in traffic mirroring mode to collect inbound and outbound traffic that passes through the interconnection switch (ISW), and generates a traffic diagram.
	Malicious server identification	Detects attacks launched by internal servers and identifies the internal servers.

Module	Feature	Description
	Abnormal traffic detection	Uses a bypass in traffic mirroring mode to detect abnormal traffic that has exceeded a specific threshold.
	Web application protection	Uses a bypass to block common attacks on web applications at the network layer based on default web attack detection rules.
Server Guard	Baseline check	Performs security baseline checks for Elastic Compute Service (ECS) instances. The check items include accounts, weak passwords, and at-risk configuration items. The baseline checks ensure that the ECS instances comply with the security standards for enterprise servers.
	Vulnerability management	<ul style="list-style-type: none"> Scans ECS instances for software vulnerabilities, and provides suggestions on vulnerability fixes. Provides quick fixes for critical vulnerabilities in applications and operating systems on your ECS instances.
	Webshell detection and removal	Accurately detects and removes webshells based on specific rules, and allows you to manually isolate webshells.
	Brute-force attack blocking	Detects and blocks brute-force attacks in real time.
	Abnormal logon alerting	Detects abnormal logons based on the approved logon settings, and generates alerts.
	Suspicious server detection	Detects suspicious activities such as reverse shells, Java processes running CMD commands, and abnormal file downloads with Bash.
	Asset fingerprints	Collects up-to-date information about the servers, such as ports, accounts, processes, and applications, to perform event tracking.
	Log retrieval	Centrally manages server logs of processes, networks, and system logons. This helps you quickly locate the cause of an issue from the logs.

Module	Feature	Description
Web Application Firewall	Protection against common web attacks	<p>Detects Structured Query Language (SQL) injections, cross-site scripting (XSS), intelligence, cross-site request forgery (CSRF), server-side request forgery (SSRF), Hypertext Preprocessor (PHP) deserialization, Java deserialization, Active Server Pages (ASP) code injections, file inclusion attacks, file upload attacks, PHP code injections, command injections, crawlers, and server responses.</p> <p>Web Application Firewall (WAF) provides the following five built-in protection templates: template with default protection policies, monitoring mode template, anti-DDoS template, template for financial customers, and template for Internet customers. WAF allows you to customize the decoding algorithms in the templates, enable or disable each attack detection module separately, and set the detection granularity.</p>
	HTTP flood mitigation	<p>Allows you to set access frequency control rules for domain names and URLs. This restricts the access frequency of IP addresses or sessions that meet the criteria, or blocks these IP addresses or sessions.</p> <p>Restricts the access frequency of known IP addresses or sessions, or blocks these IP addresses or sessions.</p> <p>HTTP flood mitigation rules are not applicable to IP addresses or sessions in a whitelist.</p>
	Custom and precise access control	<p>Supports precise access control based on the following individual HTTP message components or their combinations: URI, GET parameters, decoded path, HOST header, complete cookie, POST parameters, complete body, HTTP status code, and response content.</p>
Threat Detection Service	Security situation overview	Provides a comprehensive security overview with statistics on protected assets, attack trends, and traffic trends.
	Screens	Provides a map-based traffic screen and a server security screen.

Module	Feature	Description
	Alert handling	Allows you to view and handle security events, such as abnormal logons, abnormal accounts, webshells , malicious processes, abnormal process behavior , sensitive file tampering, and abnormal network connections.
	Attack analysis	Protects against common attacks on web applications and brute-force attacks.

13.4.2 On-premises security operations services

Apsara Stack Security Standard Edition provides on-premises security operations services that guarantee the security of your business systems.

The following table describes the on-premises security operations services that Apsara Stack Security provides.

Table 13-2: On-premises security operations services

Category	Service	Description
Security operations	Asset research	Periodically researches your business systems in the cloud under your authorization and develops a business list containing information such as the business system name, Elastic Compute Service (ECS) information, Relational Database Service (RDS) information, IP address, domain name, and owner.
	New system assessment	<ul style="list-style-type: none"> • Detects system and application vulnerabilities in a new business system by using both automation tools and manual operations before you migrate the system to the cloud. • Provides suggestions and verification on vulnerability fixes.
	Periodic security assessment	<ul style="list-style-type: none"> • Periodically uses automation tools to detect system vulnerabilities, application vulnerabilities, and security risks in running business systems. • Provides suggestions on handling detected risks, including but not limited to security policy settings , patch updates, and application vulnerability handling.

Category	Service	Description
	Access control management	Provides inspection and guidance on applying access control policies when a new business system is migrated to the cloud.
	Access control routine inspection	Periodically checks for access control risks of your business systems.
	Security risk routine inspection	Monitors and inspects security events in Apsara Stack Security, informs you of the verified events, and provides suggestions on event handling.
Apsara Stack Security maintenance	Rule update	Periodically updates the rules repository of Apsara Stack Security.
	Service integration	<ul style="list-style-type: none"> Provides support for integrating Apsara Stack Security with your business systems. Helps you customize and optimize security policies.
Security event response	Event alerts	Synchronizes security events information from Alibaba Cloud, and helps you remove the risks.
	Event handling	Handles urgent events such as attacker intrusions.

13.4.3 Optional security services

The following table describes the optional security services that Apsara Stack Security provides.

Table 13-3: Optional security services

Module	Feature	Description
DDoS Traffic Scrubbing	Traffic scrubbing against distributed denial of service (DDoS) attacks	Detects and prevents attacks such as SYN flood, ACK flood, ICMP flood, UDP flood, NTP flood, DNS flood, and HTTP flood.
	DDoS attack display	Allows you to search for DDoS attack events by IP address, status, and event information.
	DDoS traffic analysis	Allows you to analyze the traffic of a DDoS attack by viewing its protocol and the top 10 server IP addresses from which attacks are launched.

Module	Feature	Description
Sensitive Data Discovery and Protection	Security situation overview	Allows you to view the overall security status of sensitive data.
	Detection and processing of abnormal activities	Detects abnormal activities related to sensitive data and allows you to manually confirm or exclude the activities.
	Sensitive data detection	Detects sensitive data in services such as MaxCompute, Object Storage Service (OSS), AnalyticDB, ApsaraDB for RDS, and Table Store.
	Static data masking	Uses data masking algorithms to mask sensitive data at rest.
	Intelligent audit	Uses audit rules to intelligently audit services such as OSS, MaxCompute, and ApsaraDB for RDS.
	Data permission management	Displays departments and accounts hierarchically, and allows you to query and manage detailed permissions of accounts.
	Dataflow monitoring	Allows you to view the dataflow details of DataHub and Cloud Data Pipeline (CDP).
	Rule configuration	Allows you to configure detection rules, risk levels, and abnormal output rules to detect sensitive data.
	Access authorization	Supports department-based authorization and protects the data assets of authorized departments.

13.5 Restrictions

None

13.6 Terms

DDoS attacks

An attacker combines multiple computers by using the client-server model to form an attack platform and initiates a large number of valid requests to one or more targets from this platform to cause network failures. Distributed denial of service (DDoS) attacks are much stronger than common denial of service (DoS) attacks.

SQL injections

An attacker makes the server run malicious Structured Query Language (SQL) commands by inserting these commands in Web tables or inserting malicious strings in URL requests.

Traffic scrubbing

The traffic scrubbing service monitors the inbound traffic of a data center in real time and detects unusual traffic that may be from DDoS attacks and other attacks. This service scrubs the unusual traffic without affecting businesses.

Brute-force attacks

Brute-force attacks work by iterating through all possible combinations that can make up a password.

Webshells

A webshell is a script written in languages such as Active Server Pages (ASP) and Hypertext Preprocessor (PHP). Attackers can run a webshell on a Web server to perform risky operations. This enables attackers to obtain sensitive information or control the server through server penetration or privilege escalation.

Server intrusion detection

By analyzing server logs, Apsara Stack Security can detect attacks, such as system password cracking and logons from unusual IP addresses, and generate real-time alerts.

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 in 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 Apsara Stack DNS and a user-created DNS over a leased line.

14.2 Benefits

Management of enterprise domain names

Apsara Stack DNS manages and resolves enterprise domain names. Apsara Stack DNS provides the following functions:

- DNS resolution and reverse DNS resolution for domain names of cloud service instances, such as ECS instances.
- DNS resolution and reverse DNS resolution for your internal domain names.
- Addition, modification, and deletion of DNS records of the following types: A, AAAA, CNAME, NS, MX, TXT, SRV, and PTR.
- Addition of multiple DNS records of the A, AAAA, and PTR types on one host. By default, the resolution results include all matching records. The records can be randomly rotated for load balancing.

Flexible network creation and merging

Apsara Stack DNS forwards queries for enterprise domain names to allow you to flexibly create or merge networks. Apsara Stack DNS offers the following forwarding modes:

- Forward queries for all domain names.
- Forward queries for specific domain names.

Internet access from enterprise servers

When the Internet is accessible, Apsara Stack DNS supports recursive DNS for Internet domain names. You can access Internet services by using enterprise servers.

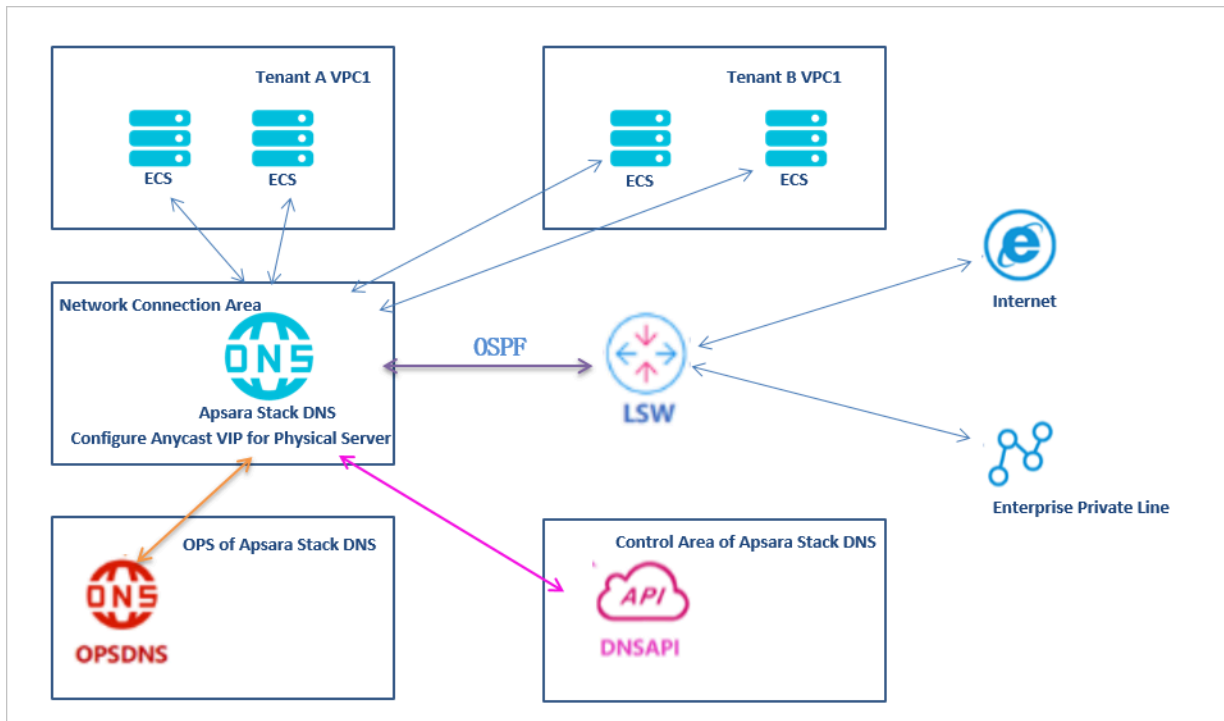
Unified management platform

The Apsara Stack DNS console is integrated into the Apsara Stack console. You can use the same account to manage both Apsara Stack DNS and other Apsara Stack services. In terms of control and management, Apsara Stack DNS has the following advantages:

- Easy-to-use web-based data and service management.
- Cluster-based deployment and high scalability. You can add DNSs based on your needs.
- Multi-zone deployment, zone active-active recovery, and zone-disaster recovery.
- Anycast-based deployment, which delivers high availability and disaster recovery.

14.3 Architecture

Figure 14-1: Apsara Stack DNS architecture



Apsara Stack DNS architecture description

- Two or more physical servers are deployed in the network access area.
- Two control interfaces are bound and uplinked to an access switch (ASW). The gateway is the default gateway of the internal network.
- Two service interfaces are uplinked to a LAN switch (LSW) by using Equal-Cost Multi-Path (ECMP) routing. These interfaces advertise anycast VIP routes in compliance with Open Shortest Path First (OSPF), and are connected to the Internet.
- The control system is deployed in a container in the control area.

14.4 Features

Management of internal domain names

Apsara Stack DNS manages internal domain names. It allows you to register, search, and delete internal domain names and add descriptions. You can also add, delete, and modify DNS records. Supported DNS record types include A, AAAA, CNAME, NS, MX, TXT, SRV, and PTR.

The management of internal domain names resolves internal domain names for servers deployed in a VPC. DNS server addresses are deployed in anycast mode to ensure the continuity of services when errors occur in data centers.

Management of domain name forwarding

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 the target DNS server is used for domain name resolution. If the target DNS server cannot resolve the domain names or the request times out, a message is returned to the DNS client indicating that the query failed.
- Forward all requests with recursion: The target DNS server is preferentially used for domain name resolution. If the target DNS server cannot resolve the domain names, the local DNS server is used instead.

Recursive DNS management

Apsara Stack DNS supports recursive DNS for Internet domain names, which enables your servers to access the Internet.

Option configuration

Apsara Stack DNS allows you to enable or disable global default forwarding, and modify the settings for global default forwarding.

14.5 Scenarios

Scenario 1: Accessing cloud resources from a VPC

Apsara Stack DNS allows Docker instances or VPC-connected ECS instances to access RDS, SLB, and OSS instances.

Scenario 2: Accessing hostnames from a VPC

If you have defined hostnames based on your own rules for your Docker instances or VPC-connected ECS instances, Apsara Stack DNS can be used to remotely access and control the ECS or Docker instances by using their hostnames.

Scenario 3: Accessing internal service domain names from a VPC

If you have developed your own SaaS on Apsara Stack and assigned a domain name that only allows internal access, you can use Apsara Stack DNS to provide access to your SaaS within a VPC network.

Scenario 4: Performing a round-robin traffic distribution for the internal network services provided by Apsara Stack

If you have developed your own SaaS on Apsara Stack and assigned a domain name that only allows internal access, and this service is deployed in multiple data centers or regions, Apsara Stack DNS helps you access your SaaS within a VPC network and evenly distributes traffic to different nodes.

Scenario 5: Accessing the Internet from a VPC

Apsara Stack DNS helps you access the Internet from a VPC.

Scenario 6: Connecting multiple networks on Apsara Stack

Apsara Stack DNS helps you establish connections between your internal network and Apsara Stack networks.

14.6 Limits

The following table describes the limits on Apsara Stack DNS clusters.

Table 14-1: Cluster limits

Cluster	Module	Server type	Configuration	Quantity
Access cluster	Basic edition - resolution module	Q46S1.2B	Minimum configuration: 16-core CPU + 96 GB of memory + two GE ports + two 10-GE ports + 600 GB of hard disk	2
Control cluster	Basic edition - resolution module	Base container	Minimum configuration: 4-core CPU + 8 GB of memory + 60 GB of hard disk + network connection support	2

14.7 Terms

DNS

Domain Name System (DNS) is a distributed database that is used for TCP/IP applications. It translates domain names into IP addresses, and selects paths for emails.

Domain name resolution

Domain name resolution maps domain names to IP addresses by using the DNS system. It includes both authoritative DNS and recursive DNS.

Recursive DNS

Recursive DNS queries domain names cached on the local DNS server or sends a request to the authoritative DNS system to obtain the corresponding IP addresses. You can use recursive DNS to resolve Internet domain names.

Authoritative DNS

Authoritative DNS resolves the names of root domains, top-level domains, and various other domains.

Authoritative domain names

Authoritative domain names are domain names resolved by the local DNS server. You can configure and manage the domain name resolution data on the local DNS server.

DNS forwarding

DNS forwarding uses two DNS servers to provide DNS resolution services. The local DNS server is used to configure and manage the domain name resolution data. The other DNS server is used to resolve domain names.

Default forwarding

If DNS queries for authoritative domain names are not resolved by the local DNS server, they are forwarded to another DNS server for resolution.

15 MaxCompute

15.1 What is MaxCompute?

MaxCompute is a highly efficient, highly available, and low-cost **EB-level** big data computing service independently developed by Alibaba Cloud. This service is used within Alibaba Group to process exabytes of data each day. MaxCompute is a distributed system oriented towards big data processing. As one of the core products in the Alibaba Cloud computing solution, the service is used to store and compute structured data.

MaxCompute is designed to support multiple tenants, and provide data security and horizontal scaling. Based on an abstract job processing framework, the service provides centralized programming interfaces for various data processing tasks of different users.

MaxCompute is used to store and compute large volumes of structured data. It provides various data warehouse solutions as well as big data analysis and modeling services. MaxCompute is designed to provide an easier approach to analyze and process large amounts of data. You can analyze big data without deep knowledge about distributed computing.

MaxCompute has the following features:

- Uses a distributed architecture that can be scaled as needed.
- Provides an automatic storage and fault tolerance mechanism to ensure high data reliability.
- Allows all computing tasks to run in sandboxes to ensure high data security.
- Uses RESTful APIs to provide services.
- Can upload or download high-concurrency, high-throughput data.
- Supports two service models: the offline computing model and the machine learning model.
- Supports data processing methods based on programming models such as SQL, MapReduce, Graph, and MPI.
- Supports multiple tenants, allowing multiple users to collaborate on data analysis.
- Provides user permission management based on ACLs and policies, allowing you to configure flexible data access control policies to prevent unauthorized access to data.
- Provides Elasticsearch on MaxCompute for enhanced applications.

- Provides Spark on MaxCompute for enhanced applications.
- Supports access and processing of unstructured data.
- Supports the deployment of multiple clusters in a single region.
- Can be deployed across different regions.

15.2 Integration with other Alibaba Cloud services

MaxCompute has been integrated with several other Alibaba Cloud services to quickly implement a variety of business scenarios.

MaxCompute and DataWorks

DataWorks uses MaxCompute as the core computing and storage engine to provide offline processing and analysis of large amounts of data. DataWorks offers fully hosted services for visual workflow development, scheduling, and O&M.

MaxCompute works with DataWorks to provide complete ETL and data warehouse management capabilities, as well as classic distributed computing models such as SQL, MapReduce, and Graph. These models enable you to process large amounts of data while reducing business costs and maintaining data security.

MaxCompute and Data Integration

You can use Data Integration to load data from different sources such as MySQL databases into MaxCompute, and export data from MaxCompute to various business databases.

Data Integration has been integrated into DataWorks and is configured and run as a data synchronization task. You can directly configure MaxCompute data sources in DataWorks , and then configure tasks to read or write data from or to MaxCompute tables. The entire process is completed on a single platform.

MaxCompute and Machine Learning Platform for AI

Machine Learning Platform for AI (PAI) is a machine learning algorithm platform based on MaxCompute. PAI provides an end-to-end machine learning platform for data processing, model training, service deployment, and prediction without data migration. After creating a MaxCompute project and activating PAI, you can use the algorithm components of the machine learning platform to perform operations such as model training on MaxCompute data.

MaxCompute and Quick BI

After processing data in MaxCompute, you can add the project as a Quick BI data source. Then, you can create reports based on MaxCompute table data in the Quick BI console for visual data analysis.

MaxCompute and AnalyticDB for MySQL

AnalyticDB for MySQL is a cloud computing service designed for online analytical processing (OLAP). It can process huge amounts of data in a highly concurrent and real-time manner. AnalyticDB for MySQL can be used in combination with MaxCompute to implement big data-driven business systems. You can use MaxCompute to compute and mine data offline and generate high-quality data. Then, you can import the data to AnalyticDB for MySQL for business systems to perform analysis.

You can use one of the following methods to import data from MaxCompute to AnalyticDB for MySQL:

- Use the import and export feature of DMS for AnalyticDB for MySQL.
- Use DataWorks to configure a data synchronization task and configure MaxCompute Reader and AnalyticDB Writer.

MaxCompute and Recommendation Engine

Recommendation Engine (RecEng) is a recommendation service framework established in the Alibaba Cloud computing environment. The recommendation service is typically composed of three parts: log collection, recommendation computing, and product connection. The input and output of offline recommendation computing are both MaxCompute tables.

On the Resources page of the RecEng console, you can add a MaxCompute project as a RecEng computing resource in the same way as you add a cloud computing resource.

MaxCompute and Table Store

Table Store is a distributed NoSQL data storage service built on the Apsara distributed operating system of Alibaba Cloud. MaxCompute 2.0 allows you to directly access and process table data in Table Store through external tables.

MaxCompute and Object Storage Service

Object Storage Service (OSS) is a secure, cost-efficient, and highly reliable cloud storage service that can store large amounts of data. MaxCompute 2.0 allows you to directly access and process table data in OSS through external tables.

MaxCompute and OpenSearch

Alibaba Cloud OpenSearch is a large-scale distributed search engine platform developed by Alibaba Cloud. After data is processed by MaxCompute, you can access MaxCompute data by adding data sources on the OpenSearch platform.

MaxCompute and Mobile Analytics

Mobile Analytics is a product launched by Alibaba Cloud to collect and analyze mobile application usage data, providing developers with an end-to-end digital operations service. When the basic analysis reports that come with Mobile Analytics cannot meet the personalized needs of app developers, app developers can synchronize data to MaxCompute with a single click. They can further process and analyze their data based on their business requirements.

MaxCompute and Log Service

Log Service allows you to quickly complete data collection, consumption, delivery, query , and analysis. After collecting log data, you need more personalized analysis and mining . You can synchronize Log Service data to MaxCompute through Data Integration in DataWorks, and then use MaxCompute to perform personalized and in-depth data analysis and mining on log data.

MaxCompute and RAM

Resource Access Management (RAM) is a service provided by Alibaba Cloud to manage user identities and resource access permissions.

Character sets supported by other Alibaba Cloud services

Service	Supported character set
Table Store	UTF-8
PAI	UTF-8
OSS	UTF-8
Quick BI	UTF-8

Service	Supported character set
DataWorks	UTF-8, GBK, CP936, and ISO-8859 are supported when data is uploaded in DataStudio. However, all data is encoded in UTF-8 in DataWorks. UTF-8 and GBK are supported when data is downloaded.

15.3 Benefits

China's only big data cloud service and real data sharing platform

- Warehousing, mining, analysis, and sharing of data can all be performed on the same platform.
- Alibaba Group implements this unified data processing platform in several of its own products such as Aliloan, Data Cube, DMP (Alimama), and Yu'e Bao.

Support for large numbers of clusters, users, and concurrent jobs

- A single cluster can contain more than 10,000 servers and maintain 80% linear scalability.
- A single MaxCompute instance can support more than 1 million servers in multiple clusters without restrictions (linear expansion is slightly affected). It supports the local multi-IDC mode.
- A single MaxCompute instance supports over 10,000 users, over 1,000 projects, and over 100 departments (of multi-tenants).
- A single MaxCompute instance supports more than 1 million jobs (daily submitted jobs on average) and more than 20,000 concurrent jobs.

Big data computing at your fingertips

You do not have to worry about the storage difficulties and the prolonged computing time caused by the increasing data volume. MaxCompute automatically expands the storage and computing capabilities of clusters based on the volume of data to process, allowing you to focus on maximizing the efficiency of data analysis and mining.

Out-of-the-box service

You do not have to worry about cluster construction, configuration, and O&M. Only a few simple steps are required to upload data, analyze data, and obtain analysis results in MaxCompute.

Secure and reliable data storage

User data is protected against loss, theft, and exposure by the multi-level data storage and access security mechanisms. These mechanisms include multi-replica technology, read/write request authentication, and application and system sandboxes.

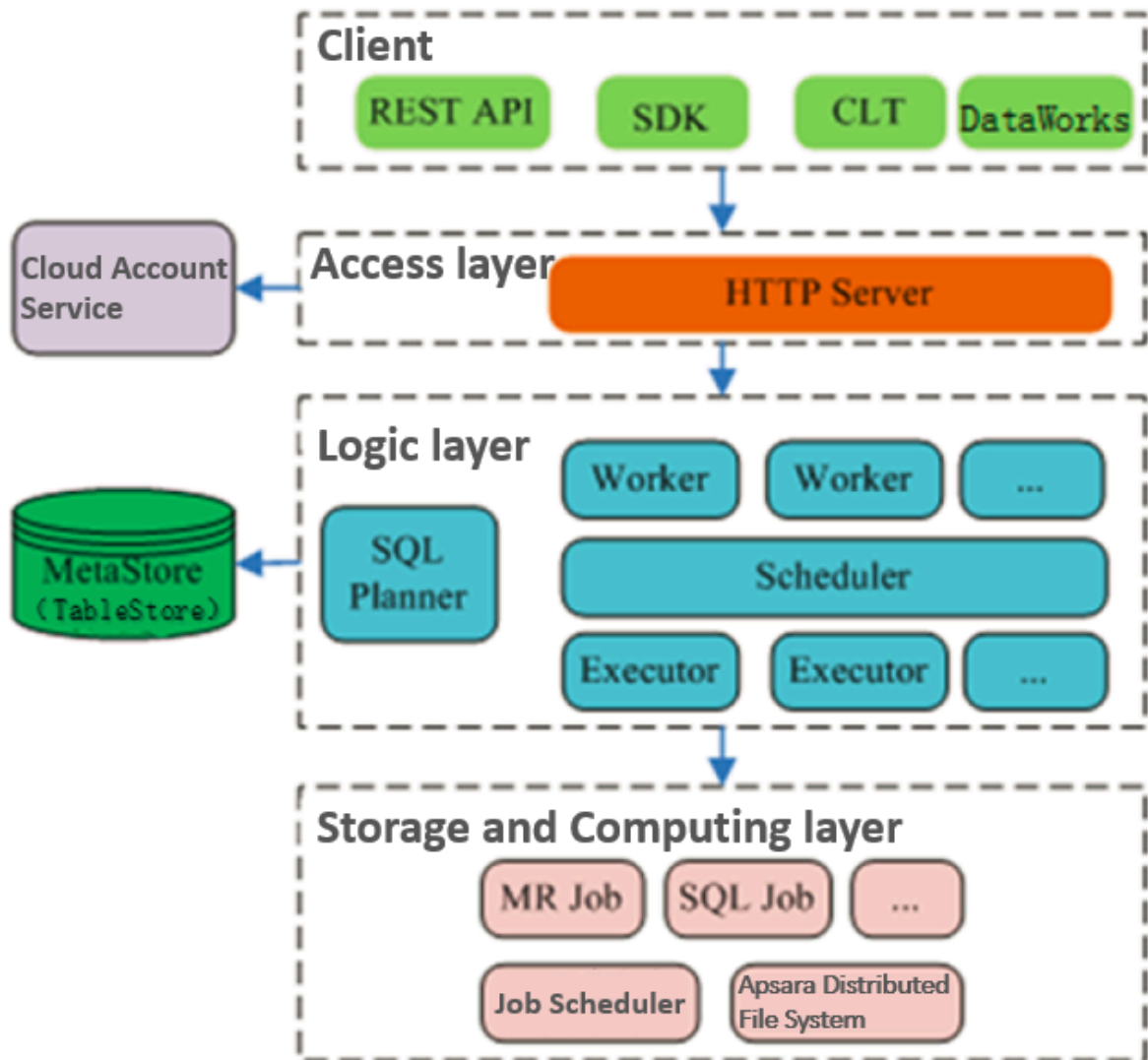
Multi-tenancy for multi-user collaboration

You can have multiple data analysts in your organization to work together by configuring different data access policies, while ensuring that each analyst can only access data within their own permissions. This maximizes work efficiency while ensuring data security.

15.4 Architecture

Figure 15-1: MaxCompute architecture shows the MaxCompute architecture.

Figure 15-1: MaxCompute architecture



The MaxCompute service is divided into four parts: **client**, **access layer**, **logic layer**, and **storage and computing layer**. Each layer can be scaled out.

The following methods can be used to implement the functions of a MaxCompute client:

- **API:** RESTful APIs are used to provide offline data processing services.
- **SDK:** RESTful APIs are encapsulated in SDKs. SDKs are currently available in programming languages such as Java.
- **Command line tool (CLT):** This client-side tool runs on Windows and Linux. CLT allows you to submit commands to manage projects and use DDL and DML.

- **DataWorks:** DataWorks provides upper-layer visual ETL and BI tools that allow you to synchronize data, schedule tasks, and create reports.

The access layer of MaxCompute supports HTTP, HTTPS, load balancing, user authentication, and service-level access control.

The logic layer is at the core of MaxCompute. It supports project and object management, command parsing and execution logic, and data object access control and authorization. The logic layer is divided into control and compute clusters. The control cluster manages projects and objects, parses queries and commands, and authorizes access to data objects. The compute cluster executes tasks. Both control and compute clusters can be scaled out as required. The control cluster is comprised of three different roles: Worker, Scheduler, and Executor. These roles are described as follows:

- **The Worker role processes all RESTful requests** and manages projects, resources, and jobs. Workers forward jobs that need to launch Fuxi tasks (such as SQL, MapReduce, and Graph jobs) to the Scheduler for further processing.
- **The Scheduler role schedules instances**, splits instances into multiple tasks, sorts tasks that are pending for submission, and queries resource usage from FuxiMaster in the compute cluster for throttling. If there are no idle slots in Job Scheduler, the Scheduler stops processing task requests from Executors.
- **The Executor role is responsible for launching SQL and MapReduce tasks.** Executors submit Fuxi tasks to FuxiMaster in the compute cluster and monitor the operating status of these tasks.

In summary, when you submit a job request, the Web server at the access layer queries the IP addresses of registered Workers and sends API requests to randomly selected Workers. The Workers then send these requests to the Scheduler for scheduling and throttling. Executors actively poll the Scheduler queue. If the necessary resources are available, the Executors start executing tasks and return the task execution status to the Scheduler.

The storage and computing layer of MaxCompute is a core component of the proprietary cloud computing platform developed by Alibaba Cloud. The architecture diagram illustrates only major modules.

15.5 Features

15.5.1 Tunnel

15.5.1.1 Terms

Tunnel is the data tunnel service provided by MaxCompute. You can use Tunnel to import data from various heterogeneous data sources into MaxCompute or export data from MaxCompute. As the unified channel for MaxCompute data transmission, Tunnel provides stable and high-throughput services.

Tunnel provides RESTful APIs and Java SDKs to facilitate programming. You can upload and download only table data (excluding view data) through Tunnel.

15.5.1.2 Tunnel features

- The channel through which data flows in to and out of MaxCompute
- Highly concurrent upload and download
- Horizontal expansion of service capabilities
- Tools based on MaxCompute Tunnel, such as TT, CDP, Flume, and Fluentd
- Support for reads and writes of tables (excluding views)
- Support for data writes in append mode
- Concurrency capabilities to improve total throughput
- Support for data upload only when target partitions exist
- Real-time upload mode

15.5.1.3 Data upload and download through Tunnel

Tunnel commands

```
odps@ > tunnel upload log.txt test_project.test_table/p1="b1",p2="b2 ";
```

```
odps@ > tunnel download test_project.test_table/p1="b1",p2="b2" log.txt;
```

Notes

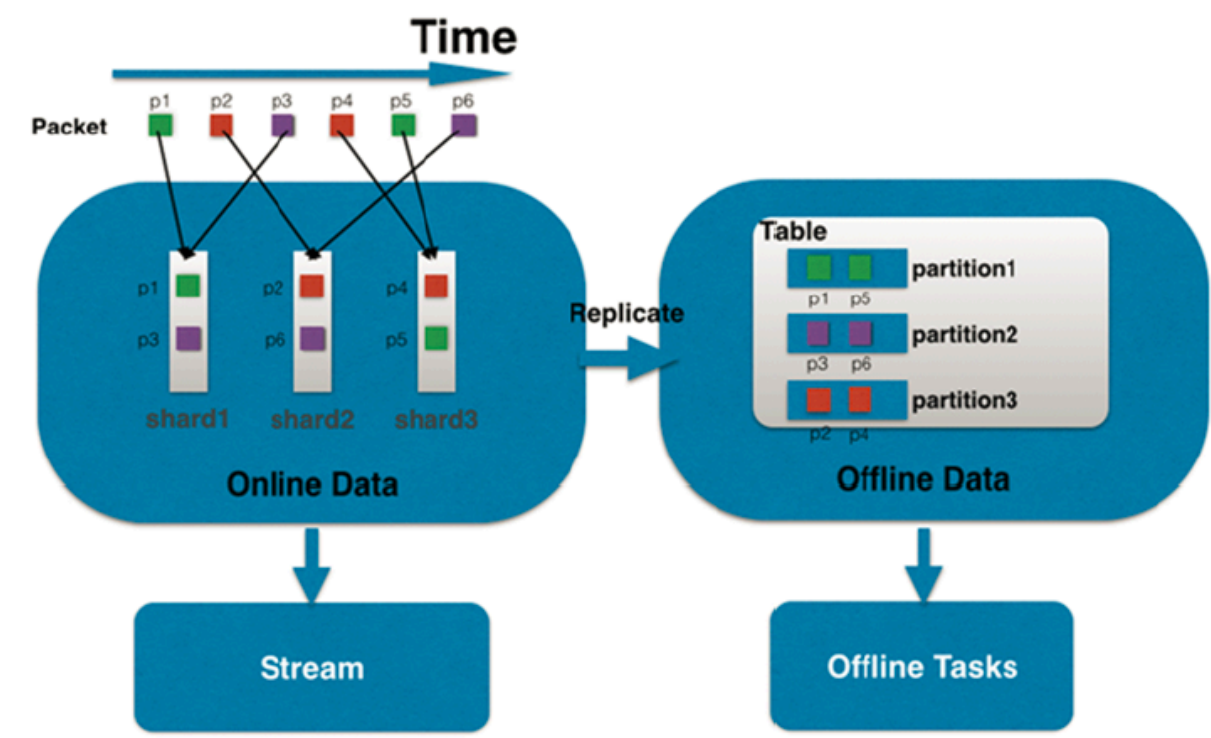
- Tunnel is a CLT based on the Tunnel SDK and can be used to upload local text files to MaxCompute or download data tables to your local device.
- You must create table partitions before using Tunnel.
- DataX, CDP, and TT provide enhanced Tunnel-based tools, which are used to exchange data between MaxCompute and relational databases.

- You can import log data by using the Flume and Fluentd tools.
- In some scenarios, you can develop custom tools based on Tunnel.

Real-time upload

- Upload in small batches
- High QPS performance
- Latency within milliseconds
- Subscription available

Figure 15-2: Real-time upload



15.5.2 SQL

15.5.2.1 Terms

The syntax of MaxCompute SQL is similar to SQL. It can be considered as a subset of standard SQL. However, MaxCompute SQL is not equivalent to a database, because it does not possess many characteristics that a database has, such as transactions, primary key constraints, and indexes. The maximum SQL statement size currently allowed in MaxCompute is 2 MB.

MaxCompute SQL offline computing is applicable to scenarios that have a large amount of data (measured in TBs) and that do not have high real-time processing requirements

. It takes a relatively long time to prepare and submit each job. Therefore, MaxCompute SQL is not optimal for services that need to process thousands of transactions per second. MaxCompute SQL online computing is applicable to scenarios that require near-real-time processing.

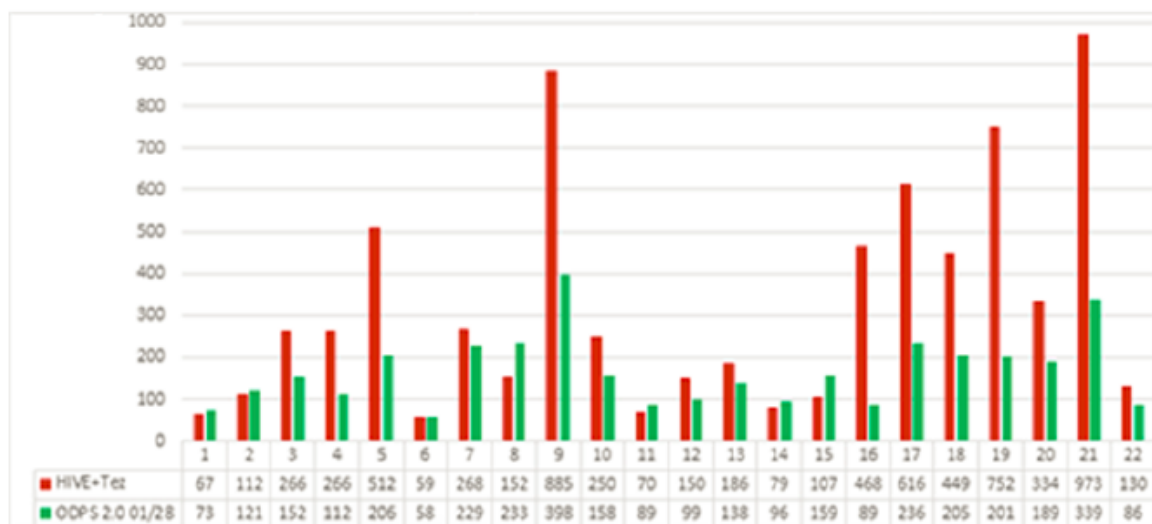
15.5.2.2 SQL characteristics

- It is suitable for processing large volumes of data (TBs or PBs).
- It has relatively high latency. The runtime of each SQL statement ranges from dozens of seconds to several hours.
- Its syntax is similar to that for Hive HQL. It is extended based on standard SQL syntax.
- It does not involve transactions or primary keys.
- It does not support UPDATE and DELETE operations.

15.5.2.3 Comparison with open source products

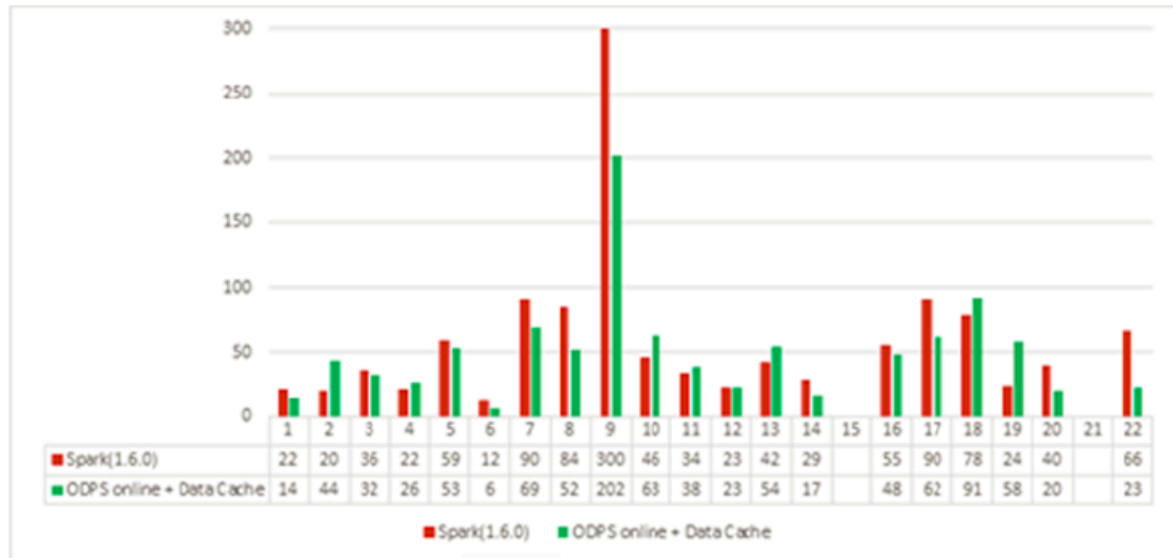
- TPC-H 1 TB data benchmark: Compared with Hive (Apache Hive-1.2.1-bin + Tez-UI-0.7.0 with CBO), MaxCompute has a 95.6% improvement in performance.

Figure 15-3: MaxCompute 2.0 VS Hive



- TPC-H 450 GB data benchmark: Compared with Spark SQL V1.6.0 (the latest release), MaxCompute has a 17.8% improvement in performance.

Figure 15-4: MaxCompute 2.0 VS Spark SQL



15.5.3 MapReduce

15.5.3.1 Terms

MapReduce is a programming model, which is basically equivalent to Hadoop MapReduce. The model is used for parallel MaxCompute operations on large-scale data sets (measured in TBs).

MaxCompute provides a MapReduce programming interface. You can use Java APIs, which is provided by MapReduce, to write MapReduce programs for processing data in MaxCompute.



Note:

All data in MaxCompute is stored as tables. The inputs and outputs of MaxCompute MapReduce can only be tables. Custom output formats are not supported, and no interface, such as a file system, is provided.

15.5.3.2 MapReduce characteristics

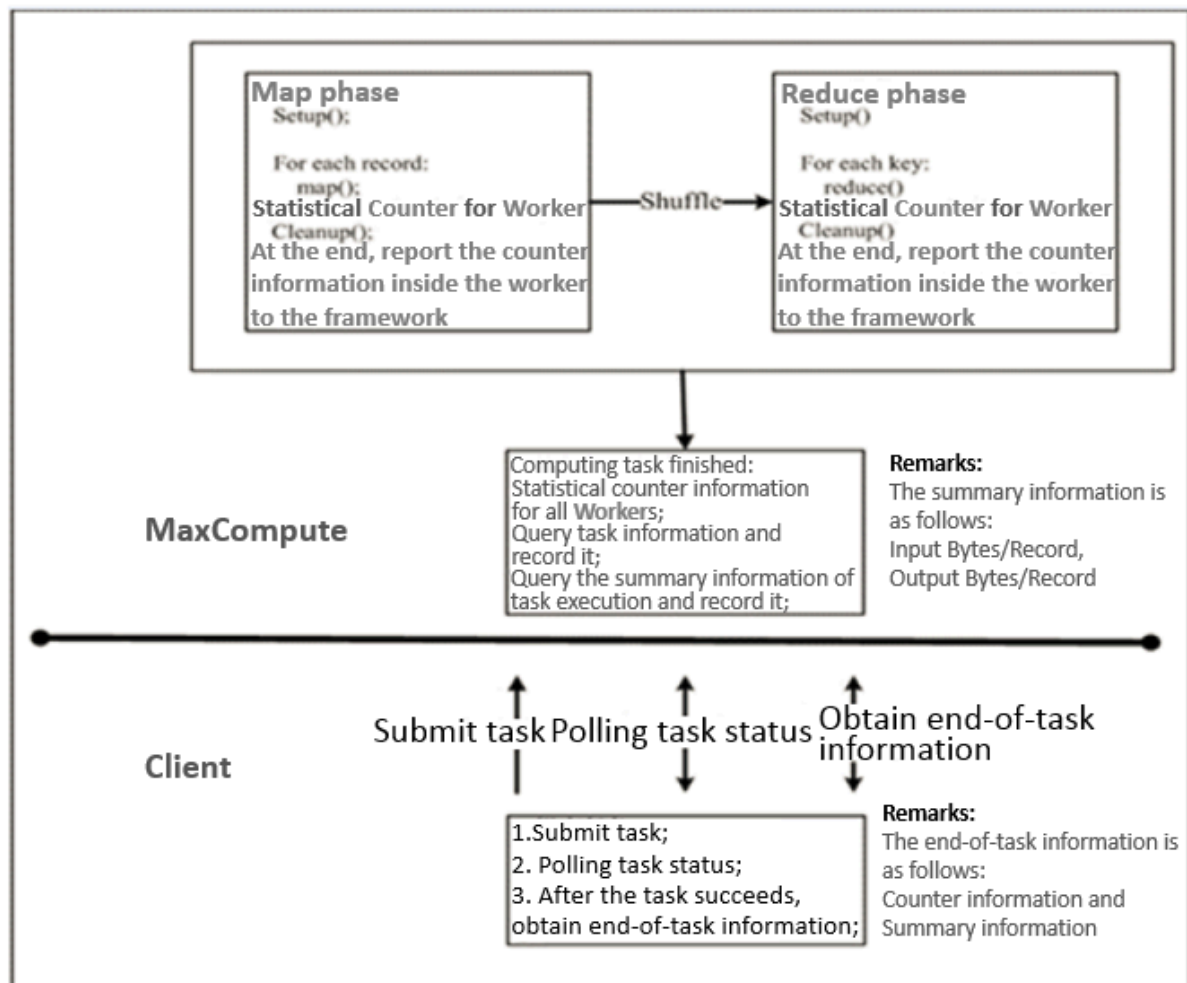
- It only supports the input and output of MaxCompute built-in data types.
- It supports the input and output of multiple tables to different partitions.
- It reads resources.

- It does not support using views as data inputs.
- It supports MapReduce programming only in the JDK 1.8 environment.
- It provides a limited sandbox security environment.

15.5.3.3 MaxCompute MapReduce process

The following figure shows the MapReduce process in MaxCompute:

Figure 15-5: MapReduce process



15.5.3.4 Hadoop MapReduce VS MaxCompute MapReduce

The following table describes the comparison between Hadoop MapReduce and MaxCompute MapReduce.

Table 15-1: Mapper/Reducer

Mapper/Reducer	
Hadoop MapReduce	MaxCompute MapReduce

Map (InKey key, InputValue value, OutputCollector<OutKey, OutValue> output, Reporter reporter)	Map (long key, Record record, TaskContext context)
Reduce (InKey key, Iterator<InValue> values, OutputCollector<OutKey, OutValue> output, Reporter reporter)	Reduce (lRecord key, Iterator<Record> values, TaskContext context)

Figure 15-6: MapReduce

```

@Override
public void map(long recordNum, Record record, TaskContext context)
    throws IOException {
    for (int i = 0; i < record.getColumnCount(); i++) {
        word.set(new Object[] { record.get(i).toString() });
        context.write(word, one);
    }
}

```

15.5.4 Graph

15.5.4.1 Terms

Graph is the computing framework of MaxCompute designed for iterative graph processing. It provides programming interfaces similar to Pregel, allowing you to use Java SDKs to develop efficient machine learning and data mining algorithms.

Graph jobs use graphs to build models. This process outputs a result after performing iterative graph editing and evolution.

15.5.4.2 Graph characteristics

- It is a graphic computing programming model (similar to Google Pregel).
- It loads data to the memory, which is superior in multiple iteration scenarios.
- It can be used to develop machine learning algorithms.
- It can support 10 billion vertices and 150 billion edges.
- Its typical applications include:
 - PageRank
 - K-means clustering
 - Level 1 and level 2 relationships and shortest path
- Graph jobs process graph data.

- The original data is stored in tables. The user-defined Graph Loader loads data in the table as vertexes and edges.
- It supports iterative computing.

15.5.4.3 Graph relational network models

A relational network engine provides a variety of business-oriented relational network models. It helps you quickly implement relational data mining at finer granularities.

Community discovery

- Input to the engine: relational data.
- Engine output: IDs and community IDs.
- Computing logic: locates N communities with the optimal global network connection. The communities are close enough internally, and sparse enough in between.

Semi-supervised category

- Input to the engine: problematic IDs.
- Engine output: potentially problematic IDs and weights.
- Computing logic: uses existing problematic IDs (of one or more categories) to determine potential problematic IDs of the same or multiple categories and corresponding weights based on the entire network connection relationships.

Isolated point detection

- Input to the engine: relational data.
- Engine output: isolated points and weights.
- Computing logic: determines whether there are relatively isolated nodes using the connection relationships in a relational network, and generates the result.

Key point mining

- Input to the engine: relational data.
- Engine output: key point IDs and categories.
- Computing logic: calculates the key type nodes in a computing network using the connection relationships (such as centrality, influence, and betweenness centrality) in a relational network.

Level N relationships

- Input to the engine: relational data.
- Engine output: retrievable relational networks.

- Computing logic: manages multi-dimensional relationships using the connection relationships in the relational network, and creates indexes to facilitate the query for specific associations of an ID.

15.5.5 Unstructured data processing (integrated computing scenarios)

Alibaba Cloud introduced the MaxCompute-based unstructured data processing framework so that MaxCompute SQL commands can directly process external user data, such as unstructured data from OSS. You are no longer required to first import data into MaxCompute tables.

You can run a simple DDL statement to create an external table in MaxCompute and associate MaxCompute tables with external data sources. This table can then act as an interface between MaxCompute and external data sources. The external table can be accessed in the same way as a MaxCompute table, and computed by MaxCompute SQL.

MaxCompute allows you to process the following data sources by creating external tables:

- Internal data sources: OSS, Table Store, AnalyticDB, ApsaraDB for RDS, HDFS (Alibaba Cloud), and TDDL.
- External data sources: HDFS (Open Source), ApsaraDB for MongoDB, and Hbase.

15.5.6 Unstructured data processing in MaxCompute

MaxCompute has the following problems when processing unstructured data:

MaxCompute stores data as volumes and must export generated unstructured data to an external system for processing.

To alleviate these problems, MaxCompute uses external tables to enable connections between MaxCompute and various data types. MaxCompute uses external tables to read and write data volumes as well as process unstructured data from external sources such as OSS.

15.5.7 Enhanced features

15.5.7.1 Spark on MaxCompute

15.5.7.1.1 Terms

Spark on MaxCompute is a solution developed by Alibaba Cloud to enable seamless use of Spark on the MaxCompute platform, extending the functions of MaxCompute.

Spark on MaxCompute provides a native Spark user experience with its native Spark components and APIs. It allows access to MaxCompute data sources and better security for multi-tenant scenarios. It also offers a management platform enabling Spark jobs to share resources, storage, and user systems with MaxCompute jobs. This guarantees high performance and low costs. Spark can work with MaxCompute to create better and more efficient data processing solutions. Spark Community applications can run seamlessly in **Spark on MaxCompute**.

Spark on MaxCompute has an independent data development node in DataWorks and supports data development in DataWorks.

15.5.7.1.2 Features of Spark on MaxCompute

Processing of data from MaxCompute and unstructured data sources

- Processes MaxCompute tables through APIs based on Scala, Python, Java, and R programming languages.
- Processes MaxCompute tables through components such as Spark SQL, Spark MLlib, GraphX, and Spark Streaming.
- Can process unstructured data from Alibaba Cloud OSS.

User-friendly experience and management functions

- Supports job submission in a way similar to Spark on YARN. **Spark on MaxCompute** is compatible with YARN and HDFS APIs.
- Supports components including Spark SQL, Spark MLlib, GraphX, and Spark Streaming.
- Can work with SQL and Graph components of MaxCompute to form optimized solutions.
- Can connect to the native Spark UI.
- Allows you to directly use the powerful management functions of MaxCompute.
- Supports not only Spark Community but also tools such as client, Livy, and Hue.

Scalability

Spark and MaxCompute share cluster resources. Spark resources can be scaled from large-scale MaxCompute clusters.

15.5.7.1.3 Spark features

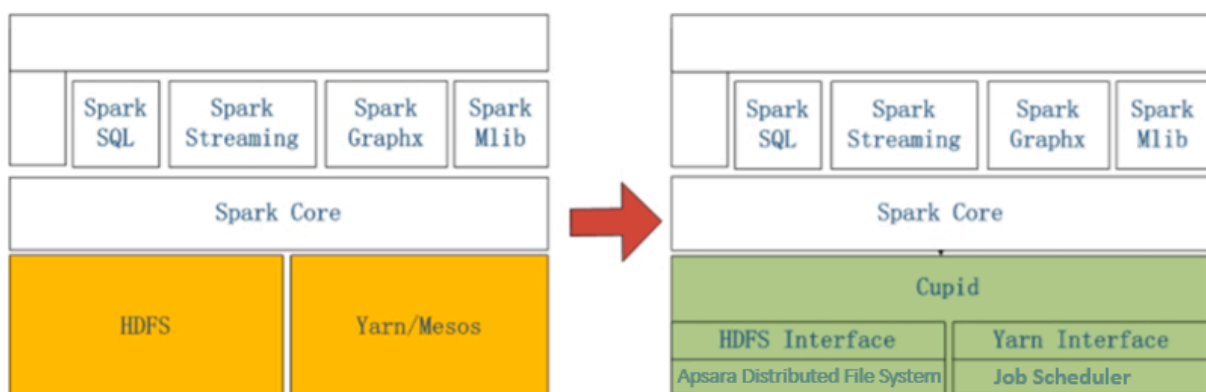
The following table describes **Spark on MaxCompute** features.

Table 15-2: Features

Type	Feature	Description
Distributed cluster	Cluster deployment Cluster monitoring	Provide an O&M platform to monitor clusters and nodes.
Data processing component	Support for components such as Spark SQL, Spark MLlib, GraphX, and Spark Streaming	Provide native Spark components.
Job management	Centralized resource management, life cycle management, and authentication	The features are available through compatible YARN APIs .
Data sources	Unstructured data Table data sources in MaxCompute	Provide data processing capabilities of SQL and MapReduce on MaxCompute.
Security management	User identification, data authentication, and multi-tenant job isolation	Harden Spark security through authentication and sandboxes.

15.5.7.1.4 Spark architecture

The following figure shows the architectural comparison between **Spark on MaxCompute** and **native Spark**.

Figure 15-7: Architectural comparison between Spark on MaxCompute and native Spark

**Note:**

On the left is the native Spark architecture and on the right is the **Spark on MaxCompute** architecture.

As shown in the figure, Spark on MaxCompute has the computing capabilities of native Spark and the functions related to management, O&M, scheduling, security, and data interconnection. The management function of Spark is implemented by starting a Cupid Task instance of MaxCompute. The resource application function is realized through layer-1 YARN APIs provided by MaxCompute. The security function is offered through the sandbox mechanism of MaxCompute. The processing of and interconnection between data and metadata are also made available. The module details are described as follows:

- The MaxCompute control cluster starts a Spark driver by using the Cupid Task instance . The Spark driver uses YARN APIs to apply for resources from FuxiMaster, the central resource manager.
- The MaxCompute control cluster manages user quota consumed by running Spark instances, life cycles of Spark instances, and permissions on accessible data sources.
- The MaxCompute computing cluster starts a Spark driver and Executor as parent and child processes and executes Spark code in the sandbox of MaxCompute, ensuring security in multi-tenant scenarios.
- MaxCompute allows you to use the native Spark UI through its Proxy Server and manage job information through its management components.

15.5.7.1.5 Benefits of Spark on MaxCompute

Support for the complete Spark ecosystem

Provides consistent user experience with that of open source Spark.

Full integration with MaxCompute

Implements centralized management of resources, data, and security features for both Spark and MaxCompute.

Combination of Spark and the Apsara system

Combines the flexibility and ease of use of Spark with the high availability, scalability, and stability of the Apsara system.

Support for multi-tenancy

Reduces costs by centrally scheduling resources in large-scale clusters and ensuring high performance of physical machines.

Support for cross-cluster scheduling

Maximizes the efficiency of cluster resources by effectively allocating clusters and scheduling resources across clusters.

Support for real-time scaling of Spark resources

Scales resources in Spark Community in real time to better utilize resources and avoid waste.

**Note:**

Real-time Spark resource scaling is not enabled on all MaxCompute clusters. To use this function, contact the MaxCompute team.

15.5.7.2 Elasticsearch on MaxCompute

15.5.7.2.1 Overview

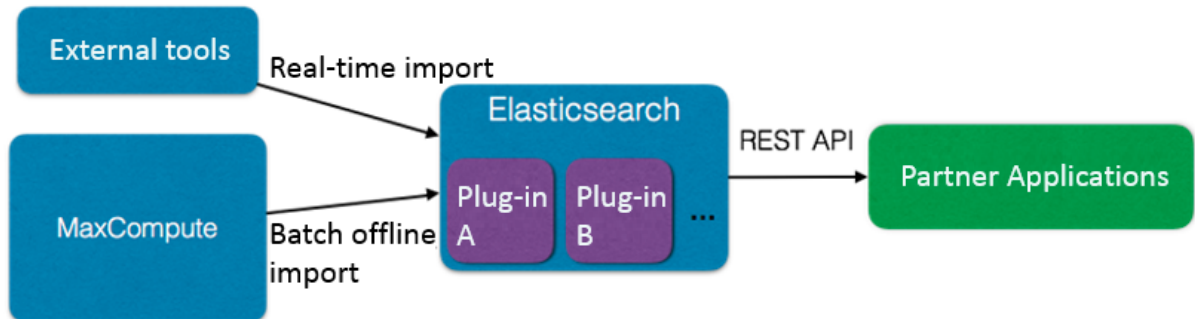
Elasticsearch on MaxCompute is an enterprise-class full-text retrieval system developed by Alibaba Cloud to retrieve large volumes of data with near-real-time search performance.

Elasticsearch on MaxCompute provides elastic full-text retrieval and supports native Elasticsearch APIs. You can import data from heterogeneous data sources and perform O&M for clusters and services. The centralized scheduling and management capabilities of MaxCompute allow Elasticsearch to provide more efficient core services for data retrieval at large volumes. **Elasticsearch on MaxCompute** can also work with plug-ins available from the Elasticsearch open source community to enhance retrieval functions.

Elasticsearch on MaxCompute allows you to use tools to import data from external sources in real time. You can also import offline data from MaxCompute. After the imported data is

indexed, Elasticsearch on MaxCompute provides retrieval services through RESTful APIs. The following figure shows its usage.

Figure 15-8: Elasticsearch on MaxCompute usage



15.5.7.2.2 Features of Elasticsearch on MaxCompute

Distributed cluster architecture

- Improves retrieval and reliability of data with a distributed architecture.
- Supports elastic scaling.
- Supports dynamic scaling.
- Supports service-level O&M and monitoring.

Robust full-text retrieval

- Performs full-text retrieval at the word, phrase, sentence, and section levels.
- Available in languages such as Chinese and English.
- Provides precise word segmentation with 100% recall for Chinese information retrieval.
- Supports complex searching methods, such as Boolean retrieval, proximity search, and fuzzy search.
- Sorts search results by relevance, field, and custom weight, and allows for secondary sorting.
- Performs statistical classification and analysis of search results.
- Allows real-time indexing and retrieval, so that inserted data can be retrieved immediately.
- Allows an index to be used multiple times after it is created.
- Allows you to modify the index structure in real time or rebuild the index to re-distribute data.

Support for multiple data sources

- Imports data from native Elasticsearch interfaces.
- Provides data import tools for MaxCompute.
- Supports full and incremental update.

Reliability

- Stores data in multiple copies, preventing user data from being lost during the downtime of machines.
- Implements a high availability architecture and comprehensive failover for nodes and services.
- Provides comprehensive O&M and monitoring functions.
- Authenticates access to protect data from malicious operations and ensure security.

15.5.7.2.3 Elasticsearch features

Elasticsearch on MaxCompute features are described as follows:

Table 15-3: Features

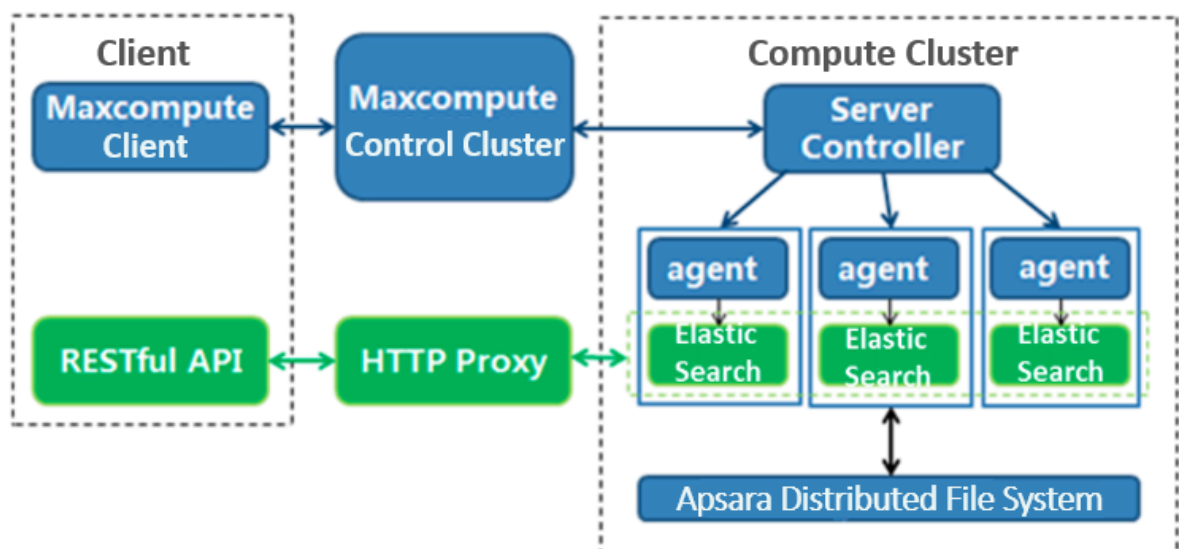
Type	Feature	Description
Distributed cluster	Cluster deployment Cluster monitoring	Provide an O&M platform to monitor clusters, nodes, and indexes.
Retrieval management	Index configuration management Structure definition and index rebuilding	Provide a retrieval management platform and support configuration.
Full-text retrieval	Retrieval Sorting Statistical analysis	The features are provided through RESTful APIs.
Data collection	Elasticsearch data import APIs MaxCompute data import tools Full and incremental collection	Support a variety of interfaces to collect native data. Provide integrated tools to import MaxCompute data.

Type	Feature	Description
Service authentication	Service-level user authentication	Allow you to configure user authentication in a centralized manner.

15.5.7.2.4 Elasticsearch architecture

Elasticsearch on MaxCompute provides core search engine services, management platforms for O&M and indexes, MaxCompute management system, and MaxCompute data import tools. It can work with universal data import interfaces and data retrieval SDKs of Elasticsearch, enabling you to retrieve applications and perform full-text retrieval of large volumes of data. The following figure shows the overall architecture.

Figure 15-9: Overall architecture



An Elasticsearch cluster corresponds to a MaxCompute Server Task instance in MaxCompute. You can quickly and flexibly deploy, operate, and expand Elasticsearch clusters on a MaxCompute client. In the overall architecture,

- The MaxCompute control cluster starts Server Controller and forwards control requests from a client.
- Server Controller is the core component for Elasticsearch cluster management. It applies for resources, starts each Elasticsearch node, and responds to the control requests that a client forwards through the control cluster. It also returns the running status of Elasticsearch clusters or adjusts the clusters.

- An agent starts Elasticsearch node processes, monitors node running status, handles failover events, and executes tasks distributed by Server Controller.
- **Elasticsearch on MaxCompute** stores its data in Apsara Distributed File System. Once a node is started successfully, Elasticsearch on MaxCompute can provide services through HTTP Proxy and allow users to use **its functions** through RESTful APIs.

15.5.7.2.5 Elasticsearch benefits

Integration of big data computation and data retrieval for resource sharing

It can access and import MaxCompute data to an Elasticsearch cluster for full-text retrieval, and manages and uses data in a centralized manner.

Centralized management of computing and storage resources

It solves storage problems or prolonged computing tasks caused by increased data volumes. It supports automatic scaling of your cluster storage and retrieval capacities based on the amount of data you have. In this way, you can focus on data analysis and mining to maximize your data value.

Provision of services such as Elasticsearch cluster deployment and O&M

It helps you deploy, configure, operate, and maintain clusters. You can upload data, analyze data, and obtain corresponding results by performing a few steps in the offline analysis service.

Secure and reliable data storage

It uses multi-replica technology to store user data at multiple layers, preventing data from being lost, leaked, or intercepted.

Open service interfaces

It provides Elasticsearch SDKs which are native and open, allowing data import, indexing, and data retrieval with **Elasticsearch on MaxCompute**.

15.6 Scenarios

15.6.1 Scenario 1: Migrate data to the cloud cost-effectively and quickly

Usage scenario: The customer is a data and information service provider focusing on the new energy power sector. The customer's target is to build a cloud platform for Internet big data application services of the new energy industry.

Results: The customer's entire business system has been migrated to the cloud within three months. The data processing time is decreased to less than one third when compared with the customer-built system. Cloud data security is ensured through multiple security mechanisms.

Customer benefits:

- **More focus on its core business:** The entire business system is migrated to the cloud within three months, which enables the customer to use a variety of cloud resources to improve the business.
- **Low investment and O&M costs:** The cloud platform helps to significantly lower the costs of infrastructure construction, O&M personnel, and R&D when compared with a customer-built big data platform.
- **Security and stability:** Alibaba Cloud's comprehensive service and stable performance guarantee data security on the cloud.

15.6.2 Scenario 2: Improve development efficiency and reduce storage and computing costs

Usage scenario: Massive log analysis services for weather query and advertising business are provided to meet the business needs of an emerging mobile Internet company aiming for an excellent weather service provider.

Results: After the Internet company's log analysis business is migrated to MaxCompute, the development efficiency is improved by more than five times, the storage and computing costs are reduced by 70%, and 2 TB of log data is processed and analyzed every day. This more efficiently empowers its personalized marketing strategies.

Customer benefits:

- **Improved work efficiency:** All log data is analyzed by using SQL, and the work efficiency is increased by more than 5 times.
- **Improved storage usage:** The overall storage and computing cost is reduced by 70%, and the performance and stability are also improved.
- **Personalized service:** Machine learning algorithms on MaxCompute are used to perform in-depth data mining and provide personalized services for users.
- **Easy use of big data:** MaxCompute provides plugins for a variety of open-source software to easily migrate data to the cloud.

15.6.3 Scenario 3: Use mass data to achieve precision marketing for millions of users

Usage scenario: To meet the business needs of a community-oriented vertical e-commerce app that focuses on the manicure industry, you can use MaxCompute to build a big data platform for the app. It is mainly used in four aspects: business monitoring, business analysis, precision marketing, and recommendation.

Results: This e-commerce app uses the big data platform built based on MaxCompute to achieve precision marketing for millions of users through the computing capability of MaxCompute, making e-commerce business more agile, intelligent, and insightful. The platform can quickly respond to the data and analysis needs of new business.

Customer benefits:

- **Improved business insights:** Through the computing capabilities of MaxCompute, precision marketing for millions of users is achieved.
- **Data-driven business:** The platform improves the business data analysis capability and effectively monitors business data to better empower businesses.
- **Fast response to business needs:** The MaxCompute ecosystem can quickly respond to changing business data analysis needs.

15.6.4 Scenario 4: Achieve precision marketing with big data

Usage scenario: MaxCompute is used to meet the business needs of an Internet company that focuses on precision marketing and advertising technologies and services. A core big data-based precision marketing platform will be built for the company.

Results: Based on MaxCompute, the company builds a core big data-based precision marketing platform. All log data is stored in MaxCompute, and offline scheduling and analysis are performed through DataWorks.

Customer benefits:

- **Efficient and low-cost analysis of massive data:** Statistical analysis of massive data can reduce expenditures by half to meet the same business needs, effectively saving costs and helping startup enterprises grow rapidly.
- **Real-time data query and analysis:** MaxCompute helps the enterprise establish technical advantages, overcoming the technical bottleneck of massive data processing and analysis, and real-time query and analysis. MaxCompute collects, analyzes, and stores more than 2 billion visitor activities every day. At the same time, it performs

millisecond-level queries in hundreds of millions of log tables based on user requirements.

- **Machine learning platform with low entry barrier:** As for a precision marketing and advertising provider, the quality of algorithm models is directly linked to its final revenue. Therefore, selecting the ease-of-use MaxCompute machine learning platform with low entry barrier can get twice the result with half the effort.

15.7 Limits

None.

15.8 Terms

project

The basic unit of operation in MaxCompute. A MaxCompute project is similar to a database or schema in a traditional database. MaxCompute projects set boundaries for isolation and access control between different users. A user can have permissions on multiple projects.



Note:

After being authorized, a user can access objects within a project, such as tables, resources, functions, and instances, from other projects.

table

The data storage unit in MaxCompute. A table is a two-dimensional data structure composed of rows and columns. Each row represents a record, and each column represents a field of the same data type. One record can contain one or more columns. The column names and data types comprise the schema of a table.



Note:

There are two types of MaxCompute tables: external tables and internal tables.

partitioned table

A logical structure used to divide a large table into smaller pieces called partitions. You can specify a partition when creating a table. Specifically, several fields in the table can be specified as partition columns. If you specify the name of a partition you want to access, the system only reads data from the specified partition instead of scanning the entire table, thus reducing costs and improving efficiency.

lifecycle

The validity period of a MaxCompute table or partition. The lifecycle of a MaxCompute table or partition is measured from the last update time. If the table or partition has not undergone any changes within a specified amount of time, MaxCompute will automatically recycle it. This amount of time is specified by the lifecycle.

- Lifecycle unit: days, positive integers only.
- When a lifecycle is specified for a non-partitioned table, the lifecycle is counted from the last time the table data was modified (LastDataModifiedTime). If the table has not been modified before the end of the lifecycle, MaxCompute will automatically recycle the table in a manner similar to the DROP TABLE operation.
- When a lifecycle is specified for a partitioned table, you can decide whether a partition should be recycled based on the LastDataModifiedTime value of the partition. Unlike non-partitioned tables, a partitioned table will not be deleted even if its last partition has been recycled.



Note:

Lifecycle scanning is performed at a scheduled time each day, and entire partitions are scanned. If a partition has not undergone any changes within its lifecycle, MaxCompute will automatically recycle it. Assume that the lifecycle of a partitioned table is one day and that the partition data was last modified at 15:00 on the 17th. If the table is scanned before 15:00 on the 18th, the aforementioned partition will not be recycled. During the lifecycle scanning scheduled on the 19th, if the last modification time of the partition exceeds the lifecycle period, the partition will be recycled.

- You can configure a lifecycle for tables, but not for partitions. You can specify a lifecycle when creating a table.
- If no lifecycle is specified, the table or partition cannot be automatically recycled by MaxCompute.

data type

A property of a field that defines the kinds of data the field can store. Columns in MaxCompute tables must be of one of the following data types: TINYINT, SMALLINT, INT, BIGINT, STRING, FLOAT, BOOLEAN, DOUBLE, DATETIME, DECIMAL, VARCHAR, BINARY, TIMESTAMP, ARRAY, MAP, and STRUCT.

resource

A unique concept in MaxCompute. To accomplish tasks by using user-defined functions (UDFs) or MapReduce features in MaxCompute, you must use resources.



Note:

Resource types in MaxCompute include file, MaxCompute table, JAR (compiled JAR package), and archive. Compressed files are identified by the extensions of resource names. Supported file types include .zip, .tgz, .tar.gz, .tar, and .jar.

function

A piece of code that operates as a single logical unit. MaxCompute provides SQL computing capabilities. In MaxCompute SQL, you can use built-in functions for computing and calculation. When the built-in functions are not sufficient to meet your requirements, you can use the Java programming interface provided by MaxCompute to develop UDFs.



Note:

UDFs can be further divided into scalar-valued functions, user-defined aggregate functions (UDAFs), and user-defined table functions (UDTFs).

task

The basic computing unit of MaxCompute. Computing jobs such as those involving SQL and MapReduce functions are completed by using tasks.

task instance

A snapshot of a task taken at a specified time. In MaxCompute, some tasks are converted into instances when being executed and subsequently exist as MaxCompute instances.

resource quota

A per-process limit on the use of system resources. There are two types of quotas: storage and computing. MaxCompute allows you to set an upper limit of storage for a project. When the storage space occupied approaches the upper limit, MaxCompute triggers an alert. The computing quota limits the use of memory and CPU resources. The memory usage and CPU utilization of running processes in a project cannot exceed the specified upper limit.

ACID semantics

This topic describes the ACID semantics of MaxCompute for concurrent jobs. ACID is an acronym that stands for Atomicity, Consistency, Isolation, Durability.

Terms

- Operation: a single job submitted in MaxCompute.
- Data object: an object that contains data, such as a non-partitioned table or partition.
- INTO jobs: such as INSERT INTO and DYNAMIC INSERT INTO.
- OVERWRITE jobs: such as INSERT OVERWRITE and DYNAMIC INSERT OVERWRITE.
- Data upload with Tunnel: an INTO or OVERWRITE job.

Description of ACID semantics

- Atomicity: An operation is either fully completed or not executed at all.
- Consistency: The integrity of data objects is not compromised during the entire period of an operation.
- Isolation: An operation is completed independent of other concurrent operations.
- Durability: After an operation is complete, data is available in its current state even in the event of a system failure.

Description of ACID semantics in MaxCompute

- Atomicity
 - At any time, MaxCompute ensures that only one job succeeds in the case of a conflict, and all other conflicting jobs fail.
 - The atomicity of the CREATE, OVERWRITE, and DROP operations on a single table or partition can be guaranteed.
 - Atomicity is not supported in cross-table operations such as MULTI-INSERT.
 - In extreme cases, the following operations may not be atomic:
 - The DYNAMIC INSERT OVERWRITE operation is performed on more than 10,000 partitions.
 - INTO operations fail because the data cleansing fails during transaction rollback. This does not cause raw data loss.
- Consistency
 - OVERWRITE jobs ensure consistency.
 - If an OVERWRITE job fails due to a conflict, data from the failed job may remain.

- Isolation
 - Non-INTO operations ensure that read operations are committed.
 - INTO operations can be performed in scenarios where read operations are not committed.

- Durability

MaxCompute ensures data durability.

16 DataWorks

16.1 What is DataWorks?

DataWorks is an end-to-end big data platform that uses MaxCompute as its compute engine. It integrates all processes from data collection to data display and from data analysis to application running. DataWorks provides various features to help you quickly and effectively complete the entire research and development (R&D) process.

DataWorks supports batch processing, analysis, and mining for large volumes of data. It integrates services such as Data Integration, DataStudio, Administration, Real-Time Analysis, Data Asset Management, Data Quality, Data Protection, and Data Service to enable core data-related processes. It also provides an online API development and management platform and interworks seamlessly with Machine Learning Platform for AI.

In 2018, Forrester, a globally recognized market research company, named Alibaba Cloud DataWorks and MaxCompute as a world-leading cloud-based data warehouse solution. This solution is currently the only product from a Chinese company to receive such an acknowledgment. Building on the success of the previous version, DataWorks V2.0 includes several new additions, such as business processes and components. DataWorks V2.0 also isolates development and production environments, adopts standard development processes, and uses a specific mechanism to reduce errors in code.

16.2 Benefits

16.2.1 Powerful computing capabilities

DataWorks uses MaxCompute as its compute engine, which can process large amounts of data.

- Supports join operations for trillions of records, millions of concurrent jobs, and I/O throughput of up to multiple petabytes (PB) each day.
- Supports batch scheduling of millions of tasks, real-time monitoring, and alerts.
- Provides powerful and easy-to-use SQL and MapReduce engines, and supports the majority of standard SQL syntax.

- Uses triplicate data storage and multiple access control mechanisms, including read/write request authentication, application sandboxing, and system sandboxing. All these mechanisms prevent data loss, leakage, and breaches.

16.2.2 End-to-end platform

- Integrates all processes from data integration, processing, management, and monitoring to output.
- Provides a visualized workflow design tool.
- Adopts a multi-user architecture and enables online job processing. You can create and assign roles with varying permissions to different users.

16.2.3 Data integration from heterogeneous data sources

DataWorks supports batch synchronization between data sources at custom intervals in minutes, days, hours, weeks, or months. More than 400 heterogeneous data source pairs are supported.

16.2.4 Web-based software

DataWorks is an out-of-the-box service. You can use it over the Internet or any internal network without the need of installation and deployment.

16.2.5 Multi-tenant architecture

Data is isolated among different tenants. Tenants independently control permissions, process data, allocate resources, and manage members.

16.2.6 Intelligent monitoring and alerts

You can specify thresholds for the task duration in DataWorks to monitor the task progress. DataWorks provides you with the progress of each task and an overview of the progress.

16.2.7 Easy-to-use SQL editor

The editor supports automatic code and metadata completion, code formatting and folding, and pre-compilation. It also offers two editor themes. These features help to ensure a good user experience.

16.2.8 Comprehensive data quality monitoring

DataWorks supports validation, notification, and management for data sources, batches of data, and real-time data.

16.2.9 Convenient API development and management

The API Gateway service and the Data Service service allow you to easily develop and publish APIs for data sharing.

16.2.10 Secure data sharing

DataWorks enables you to mask sensitive data before sharing them to other tenants, which ensures the security of your big data assets while maximizing their value.

16.3 Architecture

DataWorks is an end-to-end big data platform launched by Alibaba Group, which supports big data processing, management, analysis, mining, sharing, and transmission. It releases you from cluster deployment and management. DataWorks adopts MaxCompute (formerly known as ODPS) as the compute engine to process large volumes of data.

DataWorks is developed based on MaxCompute. DataWorks provides a management console and supports functions such as data processing, management, analysis, and mining.

16.3.1 Services

DataWorks provides the following services:

- Data Integration: big data integration between heterogeneous data sources
- DataStudio: data warehouse design and whole extract, transform, load (ETL) procedure design
- Administration: online ETL job management and monitoring
- Real-Time Analysis: ad hoc query and analysis
- Data Asset Management: metadata management, overall table management, data lineage, and data asset dashboard
- Data Quality: data quality check, monitoring, verification, and grading
- Data Protection: permission management, data management based on security levels, data masking, and data auditing
- Data Service: data sharing and transmission through APIs

16.3.2 System architecture

DataWorks is an end-to-end big data platform that enables you to process data by using Data Integration, DataStudio, Data Management, Data Service, and other services. It serves as a basis for upper-layer applications, which satisfies all user requirements.

16.3.3 Security architecture

The security architecture of DataWorks is composed of error proofing, data security, and optional security tools.

- Error proofing ensures proper running of DataWorks during coding, deployment, and configuration.
- Data security ensures basic security of DataWorks by using features such as resource isolation among tenants, user identity verification, authentication, and log auditing.
- Optional security tools, integrated in DataWorks, enable you to customize security policies for protection and management on your system and data.

16.3.4 Multi-tenant architecture

DataWorks adopts a multi-tenant architecture. The following items are isolated among tenants to ensure resource and data security.

- Storage and compute resources, which are scalable
- Data, permissions, users, and roles

16.4 Services

16.4.1 Data Integration

16.4.1.1 Overview

Data Integration provides stable, efficient, and scalable services for data synchronization. Data Integration supports fast and stable data transmission and synchronization between various heterogeneous data stores in complex networks.

Data Integration provides you with an overview of all data stores, you can monitor and read data in your data stores. Data Integration supports the following types of data stores : relational and NoSQL databases, big data platforms, and File Transfer Protocol (FTP) servers. Data Integration also supports synchronization between heterogeneous data stores in diverse complex networks. Data Integration supports batch, full, and incremental

synchronization. You can also synchronize data at intervals of minutes, hours, days, weeks, or months.

16.4.1.2 Various data sources

16.4.1.2.1 Metadata synchronization

Data Integration collects metadata from more than 20 types of common data stores, such as MySQL, Oracle, and MaxCompute. It generates a clear view of all data assets from the collected metadata. It also helps users to track assets and synchronize data.

16.4.1.2.2 Relational database

Data Integration supports read/write operations on relational databases such as MySQL, Oracle, PostgreSQL, IBM Db2, and ApsaraDB for PPAS.

16.4.1.2.3 NoSQL database

Data Integration supports read/write operations on NoSQL databases such as MongoDB and Table Store.

16.4.1.2.4 MPP database

Data Integration supports read/write operations on Massively Parallel Processor (MPP) databases such as AnalyticDB for PostgreSQL.

16.4.1.2.5 Big data product

Data Integration supports read/write operations on MaxCompute and Hadoop Distributed File System (HDFS).

16.4.1.2.6 Unstructured data storage

Data Integration supports read/write operations on Object Storage Service (OSS) and FTP servers.



Note:

Data Integration supports data synchronization for more than 400 pairs of source and target data stores.

16.4.1.3 Inbound data control

Data Integration supports conversion between various data types. It accurately identifies, filters, collects, and displays dirty data to facilitate inbound data control. Data Integration

enables you to view important statistics such as data volume, data throughput, and job duration. It also can detect dirty data for each job.

16.4.1.4 High transmission rate

Data Integration enhances the performance of one-way synchronization agents and makes full use of the network interface card (NIC) on each server. DataWorks adopts separate reader and writer agents, which ensures transmission of multiple GB or TB data within a short time.

16.4.1.5 Accurate flow control

Data Integration implements accurate flow control on channels, record streams, and byte streams. It also supports fault tolerance. With Data Integration, you can rerun specific threads, processes, and jobs.

16.4.1.6 Synchronization agents

Data Integration provides you with synchronization agents. You can install the agents on data source servers to facilitate data synchronization.

16.4.1.7 Transmission in complicated networks

DataWorks enables data to be transmitted in complicated networks, for example, between local private networks, between VPCs, and across air gaps.



Note:

Transmission of large amounts of data over a long distance is accelerated by adopting certain protocols, which ensures high stability and efficiency.

16.4.2 DataStudio

16.4.2.1 Overview

DataStudio is an integrated development environment (IDE) that allows you to develop ETL and data mining algorithms, and build data warehouses in DataWorks.

Before using DataStudio, you need to add data stores by using Data Integration. Then, you can use DataStudio to process the data retrieved from the data stores.

16.4.2.2 Business flows

16.4.2.2.1 Description

In DataStudio, you can organize nodes in a business flow. The following node types are supported: ODPS SQL, ODPS MR, shell, machine learning, data synchronization, PyODPS, SQL component, and virtual node.

DataStudio provides you with a directed acyclic graph (DAG) for nodes in each business flow. DataStudio also provides professional tools and supports administrative operations for business flows, which promotes intelligent development and management.

You can configure dependencies between nodes within the same business flow or across different business flows. You can schedule a whole business flow or specific nodes.

16.4.2.2.2 Nodes

The following node types are supported: ODPS SQL, ODPS MR, shell, machine learning, data synchronization, PyODPS, SQL component, and virtual node. Tasks are initiated based on either node dependencies or task schedules.

16.4.2.2.3 Configure nodes

After you double-click a node either in the left-side navigation pane or in a DAG, the editor of the node appears. You can configure the node in the editor, for example, entering SQL statements for ODPS SQL nodes and configuring synchronization rules for data integration nodes. If you need to view earlier versions or modify the schedule, dependencies, lineage, and other settings of the node, click the corresponding right-side buttons in the editor.

16.4.2.2.4 Versions

You can view earlier versions of ODPS SQL, ODPS MR, shell, and other nodes. If required, you can roll back a node to an earlier version.

16.4.2.2.5 Publish business flows

In workspaces that adopt standard mode, you can easily submit business flows that have passed checks to the production environment.

16.4.2.3 Solutions

In a DataWorks workspace, you can group multiple business flows in a solution.

You can also add a business flow to multiple solutions. In addition, business flows can be used repeatedly in different solutions, allowing you to assess your solutions from a business perspective.

16.4.2.4 Code editor

16.4.2.4.1 ODPS SQL nodes

DataWorks provides a Web-based SQL editor. The editor supports a variety of features such as automatic SQL statement completion, code formatting and highlighting, and debugging

16.4.2.4.2 ODPS MR nodes

You can upload Java Archive (JAR) files that contain MapReduce code to DataWorks as resources and insert them into ODPS MR nodes.

16.4.2.4.3 Resources

DataWorks supports the following resource types:

- JAR: You can upload JAR files to DataWorks as JAR resources. These JAR resources can then be called by user-defined functions and ODPS MR nodes.
- Python: You can upload Python files to DataWorks as Python resources. These Python resources can then be called by user-defined functions.
- File: You can upload shell scripts, XML and TXT configuration files, and other files to DataWorks as file resources.
- Archive: You can upload compressed files to DataWorks as archive resources. The following formats are supported: .zip, .tgz, .tar.gz, .tar, and .jar. DataWorks automatically identifies the file format based on the extension of uploaded files.

16.4.2.4.4 Register user-defined functions

You can define functions and use them for data processing. Before using user-defined functions, upload JAR files and Python files to DataWorks as resources and register related functions in DataWorks.

16.4.2.4.5 Shell nodes

DataWorks supports online editing and debugging of shell scripts.

16.4.2.5 Code repository and team collaboration

DataWorks enables multiple users to simultaneously work on the same workspace.

DataWorks allows you to create workflows and adopts a lock mechanism, which ensures that each node is edited by only one user at the same time. To edit a node locked by another user, you can force unlock the node and then lock the node yourself. This operation is called steal lock. After you steal the lock of the node, the system sends a notification to the other user.

In addition, DataWorks records each submitted version of your node and workflow. You can compare two versions of a node, and roll back a node to an earlier version.

16.4.3 Administration

16.4.3.1 Overview

Due to the volume, diversity, and complexity of data used in DataWorks, it is necessary to use a scheduling system that supports high concurrency, multiple cycles, and various data processing procedures.

To meet this requirement, Operation Center is developed, which is a centralized data operation and management platform for data developers and administration experts. With Operation Center, you can control and monitor the running of instances, and set node priorities. In Operation Center, you can trace all the nodes committed to the scheduling system, view alerts when nodes do not run as scheduled or fail, and view daily reports of node statistics.

16.4.3.2 Overview

The Overview page displays statistics of nodes, tasks, and task instances. The following statistics items are involved: the trend of task instances run today and in past days, the sorting of tasks by duration, by number of errors, and by number of overtime task instances within 30 days, and the distribution of tasks by status and by type.

16.4.3.3 Task instances

You can perform the following operations on task instances:

- Rerun, terminate, suspend, and set task instances to successful. Set alerts to monitor task execution.
- View task instances in a list or DAG. The DAG clearly shows the relationships between nodes.

- View the status of periodic task instances, ad hoc task instances, and test instances.
- View runtime logs of tasks instances and the code and attributes of corresponding nodes.

16.4.3.4 Monitor

The Monitor feature sends alerts based on specified rules, time, methods, and recipients.

The Monitor feature:

- Reduces your operating expenditure (OPEX).
- Prevents invalid alerts.
- Automatically enables alerts for tasks that are configured with baselines.

The Monitor feature automatically creates alert rules for tasks that are configured with baselines. You can also customize alert rules by completing basic settings.

16.4.4 Workspace Management

16.4.4.1 Description

Workspace Management enables administrators to manage their organizations and workspaces.

Workspaces are the organizational unit for code, member, role, and permission management in DataWorks. Workspaces are isolated from each other. You can view and change code in a workspace only if you are a member of the workspace and have been assigned the required permissions.



Note:

You can add a user to multiple workspaces. The user's permission varies according to the role assigned to the user in each workspace.

16.4.4.2 Organization

The Organization page displays the account, AccessKey ID, and AccessKey Secret of the organization owner. On this page, you can manage all members in an organization.

16.4.4.3 Workspaces

On the Workspaces page, administrators can create, modify, activate, or disable workspaces.

16.4.4.4 Members

The member list shows the name, account, role, and other information of each member.

- Fuzzy match is supported for member search.
- Only administrators can add and remove members from workspaces.



Note:

When you add a user to a workspace, you must assign at least one role to the user.

Only administrators can remove members from workspaces.



Note:

After a user is removed from a workspace, all permissions that have been granted to the user within the workspace are revoked.

16.4.4.5 Authorizations

On the Authorizations page, you can manage roles and specific permissions for all users.

The [Roles and permissions](#) table describes the permissions of each role in DataWorks.

Table 16-1: Roles and permissions

Role	Permissions
Administrator	An administrator can manage the basic information, data sources , compute engine configurations, and members for the workspace . The administrator can also assign users with the administrator, developer, administration expert, deployment expert, and visitor roles.
Developer	A developer can create workflows, script files, resources, and functions. The developer can also create and delete tables, and create publish tasks. However, the developer cannot perform publish operations.
Administration expert	The administration permissions of an administration expert is assigned by an administrator. An administration expert can perform publish and administrative operations, but does not have the permissions of a developer.
Deployment expert	The permissions of a deployment expert is similar to those of an administration expert. However, the deployment expert is not authorized to perform administrative operations.

Role	Permissions
Visitor	A visitor can only view information in the workspace, and cannot change any workflow, code, or other configurations.

16.4.5 Realtime Analysis

Realtime Analysis provides two core features: ad hoc query and private table management . It expedites the analysis process by using the data collection tools of MaxCompute in near real-time mode.

Benefits

The near real-time mode is used by default.

You can run the `set ODPS.service.mode=[all|off|limited]` command to change the configuration.

The advantages of near real-time mode over standard mode are described as follows:

- In near real-time mode, Data Analysis preallocates thread pools based on the job size. The near real-time mode eliminates the need for Job Scheduler to plan jobs and reduces the preparation time to run jobs.
- In near real-time mode, Data Analysis shuffles data from Mappers to Reducers, without transferring the data to Apsara Distributed File System.

Principle

- You can switch to the near real-time mode by setting the `ODPS.service.mode` parameter to `all`. However, if resources in MaxCompute are insufficient to run SQL nodes, Data Analysis switches to the standard mode, and then Job Scheduler is responsible for resource allocation. For example, this occurs if available workers are insufficient for creating instances.
- The near real-time mode includes a complex scheduling mechanism, which is improved based on the mechanism adopted by Job Scheduler to reduce the preparation time for running jobs.
- If the near real-time mode is used, Realtime Analysis first tries to run jobs in near real-time mode. Realtime Analysis uses the standard mode if system resources are insufficient, or if known issues or unknown exceptions occur in near real-time mode.

16.4.6 Data Service

With Data Service, you can manage all your table APIs after you create new APIs or register existing APIs. You can also easily publish your APIs to API Gateway. Together with API Gateway, Data Service provides a secure, stable, cost-effective, and easy-to-use API development and management service. Data Service adopts a serverless architecture and allows you to develop table APIs without thinking about infrastructure such as compute resources. Data Service supports automatic scaling for compute resources, which significantly reduces your OPEX.

Data Service serves the government as a secure, flexible, and reliable platform for data sharing across departments and networks within the government. It also enables the government to share data with the public.

Create an API

You can create APIs for tables in relational databases, NoSQL databases such as Table Store, and analytical database. You can quickly create an API within a few minutes by simply following the wizard provided, and immediately call the API operation after the creation is complete. You can also create APIs by specifying SQL scripts. The script mode supports advanced functions such as associative tables, complex criteria, and aggregate functions.

Register an API

You can register existing Restful APIs to Data Service for unified API management. Four request methods and three data formats are supported. The four request methods are GET, POST, PUT, and DELETE. The three data formats are tables, JSON, and XML.

API Gateway

API Gateway provides API lifecycle management services, including API publishing, management, maintenance, and monetization. It enables low-risk, simple, cost-effective, and fast microservice integration, front and back end separation, and system integration. You can use API Gateway to share functions and data with your partners and third-party developers. API Gateway supports authorization, authentication, flow control, and billing for Data Service.

16.4.7 Data Asset Management

Data Asset Management provides you with an overview of your data assets. Data Asset Management feature requires that data be synchronized by using Data Integration and

processed by using DataStudio before you manage your tables and APIs stored in your business system and DataWorks.

16.4.8 Data Protection

16.4.8.1 Overview

Data Protection is a data security management platform for data asset and sensitive data recognition, classification, masking, risk behavior monitoring, and auditing.

It provides security management services for MaxCompute.

Data Protection provides the following features.

- Sensitive data recognition

Data Protection can recognize sensitive data in enterprises based on self-training algorithms to show the types, distribution, and volume of sensitive data. It also recognizes custom data types.

- Custom data classification

Data Protection allows you to classify data by creating custom levels for better data management.

- Flexible data masking

Data Protection provides diverse and configurable methods for dynamic data masking.

- Risk behavior monitoring and auditing

Data Protection uses various correlation analysis algorithms to identify risk behavior.

Data Protection provides alerts and supports visualized auditing for identified risks.

16.4.8.2 Terms

16.4.8.2.1 Organization

An organization refers to all system settings and resources owned by a single tenant in DataWorks. The system settings and resources include RAM configurations, permission settings, and custom applications.

16.4.8.2.2 Workspace

Workspaces are the organizational unit in DataWorks. Similar to databases in a relational database management system (RDBMS), workspaces isolate resources among different users and offer boundaries for access control. User-defined tables, resources, functions, and nodes are isolated among different workspaces.

16.4.8.2.3 Regular expression

A regular expression is a sequence of characters that define filter criteria. You can use regular expressions to identify sensitive data.

**Note:**

A regular expression consists of metacharacters and literal characters such as letters from a to z.

16.4.8.2.4 Data classification

Data is classified based on value, sensitivity, related risks, legal and regulatory requirements, and the potential impact of data breaches.

16.4.8.2.5 Data recognition

The Data Protection service recognizes sensitive data on end user side based on user-defined rules.

16.4.8.2.6 Data masking

The Data Protection service masks sensitive data based on user-defined rules.

16.4.8.2.7 MaxCompute

MaxCompute is a data processing platform developed by Alibaba Cloud for large-scale data warehousing. It supports storage and compute for batches of structured data, and meets the requirements for big data modeling and analysis in most scenarios.

16.4.8.3 Data recognition rules

You can configure data recognition rules on the **Data Definition** page.

After data recognition rules are configured, you can go to the **Data Recognition Rules** page, or the **Manipulations and Queries** or **Export** tab on the Data Activities page to perform required operations.

16.4.8.4 Data recognition

On the next day after you configure data recognition rules, you can view the recognized data in the Overview, Level, and Fields Recognized sections on the Data Recognition page.

You can filter recognized data by project, rule name, rule type, and risk level.

16.4.8.5 Data activities

Data activities include data manipulations and queries, and data export.

- Data manipulations and queries involve operations such as creating, inserting, and selecting data. Failed operations are not included.
- Data export refers to exporting data from MaxCompute.

Manipulations and queries

On the next day after you configure data recognition rules, you can view data manipulations and queries on the Manipulations and Queries tab on the Data Activities page. The Manipulations and Queries tab displays information about data access activities, including the overview information, trend, and records. You can filter data access activities by project, rule name, rule type, risk level, and user.

Export

On the next day after you configure data recognition rules, you can view data export activities on the Export tab on the Data Activities page. The Export tab displays information about data exported from MaxCompute, such as the overview information, top 5 accounts that have exported most data, and data export details. You can filter the data export activities by rule name, rule type, and minimum exported entries.

16.4.8.6 Data masking

On the Data Masking page, you can create, modify, delete, and test data masking rules.

You can configure data masking rules for each data recognition rule, and configure a whitelist to include recognized sensitive data that does not require data masking.

16.4.8.7 Levels

On the Levels page, you can configure the security levels of rules if the existing configuration cannot meet your needs.

You can create levels, delete levels, and adjust the priority of levels and rules on the **Levels** page.

16.4.8.8 Manual check

On the Manual Check page, you can manually modify recognition results if any sensitive data is recognized incorrectly. You can delete data that is incorrectly recognized, change the type of recognized data, and process multiple data records at a time.

16.4.8.9 Data risks

In Data Protection, data activities are audited manually or based on the risk identification rules and AI-based identification rules configured on the Risk Rules page. The Data Risks page lists data activities that are audited as risky. You can also comment audit results as required.

16.4.8.10 Risk rules

On the Risk Rules page, you can configure rules to identify risks in users' daily accesses to your data and enable AI-based identification rules. The Data Risks page lists data activities with identified risks for further auditing. You can click an activity on the Data Activities page to view the risk rule that is triggered.

16.4.8.11 Data auditing

The Data Auditing page displays statistics of data risks in three sections, including **Overview**, **Trend**, and **Risk Analysis by Dimension**.

16.5 Scenarios

16.5.1 Cloud-based data warehouse

Enterprise customers can create large data warehouses by using DataWorks in Apsara Stack .

DataWorks can integrate petabytes (PB) of data for enterprise customers.

- **Storage:** DataWorks provides a scalable data warehouse for petabytes and exabytes of data.
- **Data integration:** DataWorks supports data synchronization and integration across heterogeneous stores to eliminate data silos.
- **Data processing:** DataWorks uses MaxCompute as the compute engine. It provides a visualized workflow designer and programming framework for SQL and MapReduce.
- **Data management:** DataWorks supports metadata-based data resource management and permission-based resource access control.

- Batch scheduling: DataWorks provides the scheduling of recurring nodes at different intervals, and supports scheduling of millions of concurrent nodes, error alerts, and real-time monitoring of running node instances.

16.5.2 Business intelligence

This topic describes how to create reports by using DataWorks.

You can analyze the following items based on the network logs of your website:

- Page views, unique visitors, and device types such as Android devices, iPads, iPhones, and personal computers. You can also create a daily report based on these statistics.
- Locations of visitors.

A sample log entry is provided as follows:

```
xx.xxx.xx.xxx - - [12/Feb/2014:03:15:52 +0800] "GET /articles/4914.html HTTP/1.1" 200
37666
"http://xxx.cn/articles/6043.html" "Mozilla/5.0 (Windows NT 6.2; WOW64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/xx.x.xxxx.xxx Safari/537.36" -
```

1. Before importing data to MaxCompute, create a target table named ods_log_tracker in MaxCompute.
2. Configure dependencies among the tables to be analyzed.
3. Create a shell node, a sync node, and an ODPS SQL node, respectively.
4. Develop the preceding three nodes.

16.5.3 Data-driven management

- Innovative business: Data mining, data modeling, and real-time decision making can be implemented based on big data analytics results provided by DataWorks.
- Small and medium enterprises: With DataWorks, data can be quickly analyzed and put into commercial use, which helps enterprises to generate operational strategies.

16.6 Limits

None.

16.7 Terms

node

Nodes define operations on data. You can create nodes of following types: ODPS SQL, ODPS MR, shell, machine learning, data synchronization, PyODPS, SQL, and virtual nodes.

instance

When the scheduling system or user triggers a node, an instance is created. The instance is a snapshot of the running node. In Operation Center of DataWorks, you can find the schedule, status, and runtime log of the instance.

commit



Note:

You can commit nodes in the development environment to the scheduling system. The scheduling system runs the code specified in the committed nodes according to the node configurations.

The scheduling system only initiates nodes after you commit them.

manually triggered workflow

You need to manually trigger nodes in this type of workflows, and cannot schedule the nodes. Therefore, you cannot configure parent nodes and outputs for nodes in manually triggered workflows.

17 Realtime Compute

17.1 What is Realtime Compute?

Realtime Compute is an advanced stream processing platform that provides real-time computations over data streams.

Background information

Industries have an increasing demand to obtain real-time insights into their data and simplify procedures. This requires software applications to improve data processing efficiency. In traditional models for big data processing, online transaction processing (OLTP) and offline data analysis are separately performed at different times. These models cannot satisfy the increasing demand for real-time big data processing.

Realtime Compute is designed to respond to the increasing demand for the timeliness of data processing. The value of data rapidly decreases over time. Therefore, data must be processed at the earliest opportunity after it is generated. In traditional models for big data processing, data is accumulated and processed on an hourly or even daily basis. This cannot meet the demand for real-time computations over data streams. Batch processing cannot meet the business needs in the scenarios where an extremely low processing delay is required. These scenarios include real-time big data analysis, early warning and risk control, real-time forecasting, and financial transactions. Realtime Compute enables real-time processing over data streams. With Realtime Compute, you can shorten the data processing delay, easily implement real-time computational logic, and greatly reduce computing costs. All the features of Realtime Compute work together to help you meet your business needs for real-time processing of big data.

Streaming data

Big data can be viewed as a series of discrete events. These discrete events form event streams or data streams along a timeline. Unlike offline data, streaming data is continuously generated by thousands of data sources. Streaming data is typically sent in data records simultaneously and in small sizes. Streaming data is generated from continuous event streams, including:

- Log files
- Online shopping data

- In-game player activities
- Social network information
- Financial transactions
- Geospatial services
- Telemetry data from devices or instruments

Features

Realtime Compute has the following features:

- Real-time and unbounded data streams

Realtime Compute processes data streams in real time, which are continuously generated from data sources. Streaming data is subscribed to and consumed in the chronological order. Data streams continuously flow into the Realtime Compute system. For example, in scenarios where Realtime Compute processes data streams from website visit logs, the log data streams continuously enter the Realtime Compute system while the website is online. In Realtime Compute, unbounded data streams are processed in real time.

- Continuous and efficient computations

Realtime Compute is an event-driven system where unbounded event streams or data streams continuously trigger real-time computations. Each streaming data record triggers a computation task. Realtime Compute performs continuous and real-time computations over data streams.

- Real-time integration of streaming data

Realtime Compute allows you to write the processing result of each streaming data record into the target data store in real time. For example, you can write the result data into the target RDS data store for report display purposes. Realtime Compute enables the result data to be continuously written into the target data store in real time. Therefore, Realtime Compute can be viewed as the data source that generates data streams for the target data store.

17.2 End-to-end real-time computing

Unlike offline or batch computing, end-to-end real-time computing of Alibaba Cloud runs real-time computations over data streams, including real-time data collection, computing

, and integration. The real-time computational logic of Realtime Compute ensures a short processing delay.

1. Data collection

You can use data collection tools to collect and send streaming data in real time to a publish-subscribe system for big data analysis. This publish-subscribe system continuously produces events for Realtime Compute in the downstream to trigger stream processing jobs.

2. Stream processing

Data streams continuously enter Realtime Compute for real-time computing. At least one data stream must enter the Realtime Compute system to trigger a real-time computing job. Each batch of incoming data records initiates a stream processing procedure in Realtime Compute. The computing results for each batch of data records are then instantly provided.

3. Data integration

Realtime Compute allows you to write the result data of stream processing to sinks, such as tables of data stores and message delivery systems. You can also integrate Realtime Compute with the alerting system that is connected to your business applications. This enables you to easily receive alerts if the specified business rules for alerting are satisfied. Unlike batch computing products such as MaxCompute and open source Apache Hadoop, Realtime Compute inherently comes with data integration modules that allow you to write result data to sinks.

4. Data consumption

After the result data of stream processing is written to sinks, the data consumption phase is decoupled from real-time computing. You can use data stores, data transmission systems, or alerting systems to access the result data, send and receive the result data, or send alerts, respectively.

17.3 Differences between real-time computing and batch computing

17.3.1 Overview

Compared with batch processing, stream processing is an emerging technology in the field of big data computing. This section describes the differences between batch processing and stream processing from two aspects: users and products.

**Note:**

For more information, see [Stream processing](#) from Wikipedia.

17.3.2 Batch computing

Batch computing models have been used for most traditional data computing and analysis services. In batch computing models, extract-transform-load (ETL) or online transaction processing (OLTP) systems are used to load data into data stores. The loaded data is then used for online data services, such as ad-hoc queries and dashboard services, based on SQL statements. You can also use SQL statements to obtain results from the analysis.

Batch computing models are widely accepted along with the evolution of relational databases in diversified industries. However, in the era of big data, with the increasing number of human activities being converted to information and then data, more and more data requires real-time and stream processing. The current processing models are facing great challenges in real-time processing.

A typical batch computing model is described as follows:

1. An ETL or OLTP system is used to build data stores and provides raw data for computing and analysis. The batch computing model where users load the data and the batch computing system optimizes queries on the loaded data using multiple methods, such as creating indexes, based on its storage and computing capabilities. In batch computing models, data must be loaded into the batch computing system. Newly arriving data records are collected into a batch and the entire batch is then processed after all data in the batch is loaded.
2. A user or system initiates a computing job, such as a MaxCompute SQL job or Hive SQL job, and submits requests to the ETL or OLTP system. The batch computing system then schedules computing nodes to perform computations on large amounts of data. This may take several minutes or even hours. The mechanism of batch computing determines

that the data to be processed is the accumulated historical data. As a result, the data processing may not be in real time. In batch computing, you can change computational logic using SQL at any time to meet your needs. You can also perform ad-hoc queries instantly after changing the logic.

3. The computing results are returned in the form of data sets when a computing job is completed. If the size of result data is excessively large, the result data is stored in the batch computing system. In this scenario, you can integrate the batch computing system with another system to view the result data. Large amounts of result data lead to a lengthy process of data integration. The process may take several minutes or even hours.

Batch computing jobs are initiated by users or systems and are processed with a long delay. The batch computing procedure is described as follows:

1. You load data into the data processing system.
2. You submit computing jobs. In this phase, you can change computing jobs to meet your business needs, and publish the changed jobs.
3. The batch computing system returns the computing results.

17.3.3 Real-time computing

Unlike batch computing, real-time computing runs real-time computations over data streams and allows for a low processing delay. The differences between real-time computing and batch computing are described as follows:

1. Data integration. For real-time computing, data integration tools are used to send streaming data in real time to streaming data stores such as DataHub. For batch computing, large amounts of data are accumulated and then processed. In contrast, streaming data is sent in micro batches in real time, which ensures a short delay for data integration.

The streaming data is continuously written to data stores in real time. You do not need to preload data for processing. Realtime Compute does not store real-time data that is continuously processed. The real-time data is discarded instantly after it has been processed.

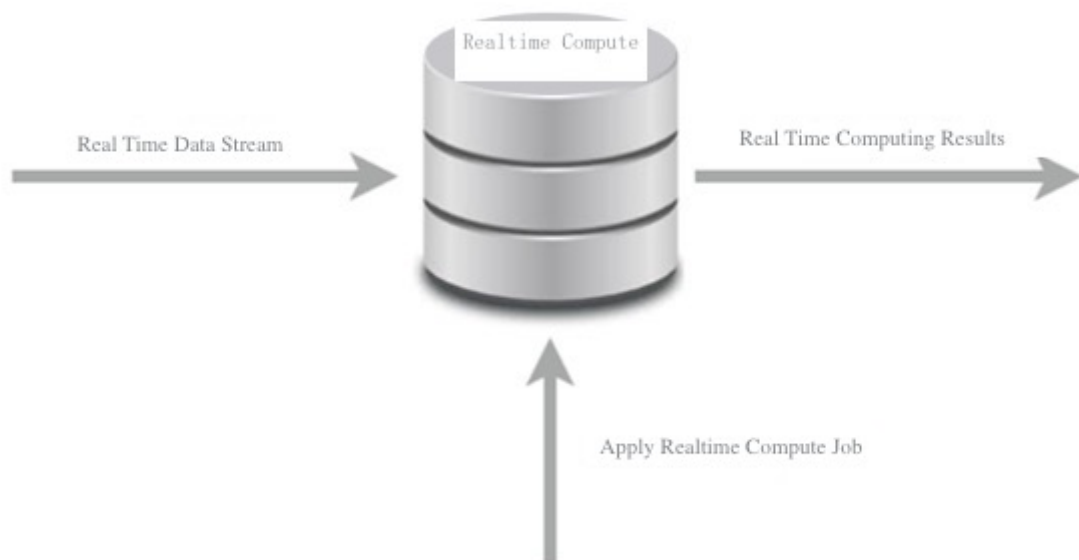
2. Data computing. For batch computing, data is processed only after large amounts of data have been accumulated. In contrast, a real-time computing job is resident in the system and waits to be triggered by events once it is started. Each incoming micro batch of streaming data records initiates a real-time computing job. The computing results are

instantly provided by Realtime Compute. Realtime Compute also divides large batches of data records into smaller batches for incremental computing. This effectively shortens the processing delay.

For real-time computing, you must predefine the computational logic in Realtime Compute. You cannot change the computational logic when real-time computing jobs are running. If you terminate a running job and publish the job after changing the computational logic, the streaming data that has been processed before the change cannot be processed again.

3. Writing result data to target systems. For batch computing, result data can be written to online systems by batch only after all accumulated data has been processed. In contrast, real-time computing allows for writing result data to online and offline systems instantly after each micro batch of data records has been processed. This allows you to view the computing results in real time.

Figure 17-1: Realtime computing



Realtime Compute runs real-time computations over data streams, which are continuously generated from data sources, based on an event-driven mechanism. Realtime Compute allows you to process data streams with a short delay. The real-time computing procedure is described as follows:

1. You publish real-time computing jobs.
2. Streaming data triggers real-time computing jobs.
3. Realtime Compute constantly returns the computing results.

17.3.4 Comparison between real-time computing and batch computing

Table 17-1: Comparison between real-time computing and batch computing shows the differences between real-time computing and batch computing.

Table 17-1: Comparison between real-time computing and batch computing

Item	Batch computing	Real-time computing
Data integration	You load data into the data processing system.	Data is loaded and processed in real time.
Computational logic	The computational logic can be changed, and data can be reprocessed.	After the computational logic is changed, data cannot be reprocessed. This is because streaming data is processed in real time.
Data scope	You can query and process all or most of the data in the data set.	You can query and process the latest data record or the data within the tumbling window.
Data size	It processes large batches of data.	It processes individual records or micro batches consisting of a few records.
Performance	It achieves a processing delay of several minutes or hours.	It achieves a processing delay of several seconds and even milliseconds.
Analysis	You can perform complex data analysis.	You can perform simple analysis, such as simple response functions, aggregates, and rolling metrics.

Realtime Compute uses a simple computing model. Real-time computing of Realtime Compute makes significant improvements to batch computing in most scenarios of big data computing. In particular, in scenarios where event streams need to be processed with an extremely low processing delay, real-time computing is a valuable service for big data computing.

17.4 Benefits

Realtime Compute provides competitive advantages in stream processing, which allows you to easily handle the demand for real-time big data analysis. Realtime Compute offers the following benefits:

Powerful real-time computing functions

Realtime Compute simplifies the development process by integrating a wide range of functions. These functions are described as follows:

- A powerful engine is used. This engine offers the following advantages:
 - Provides the standard Flink SQL that enables automatic data recovery from failures. This ensures accurate data processing when failures occur.
 - Supports multiple types of built-in functions, such as text functions, date and time functions, and statistics functions.
 - Enables an accurate control over computing resources. This ensures complete isolation of each tenant's jobs.
- The key performance metrics of Realtime Compute are three to four times higher than those of Apache Flink. For example, in Realtime Compute, the data processing delay is reduced to seconds or even to sub-second level. The throughput of a job reaches millions of data records per second. A cluster can contain thousands of nodes.
- Realtime Compute integrates cloud-based data stores such as MaxCompute, DataHub, Log Service, ApsaraDB for RDS, Table Store, and AnalyticDB for MySQL. With Realtime Compute, you can read data from and write data to these systems with the least efforts in data integration.

Managed real-time computing services

Unlike open source or user-developed stream processing services, Realtime Compute is a fully managed stream processing engine. You can query streaming data without deploying or managing any infrastructure. With Realtime Compute, you can use streaming data processing services with a few clicks. Realtime Compute integrates services such as development, administration, monitoring, and alerting. This allows you to use cost-effective streaming data services for trial and migrate your data for deployment.

Realtime Compute also enables complete isolation between tenants. This isolation and protection extends from the top application layer to the underlying infrastructure layer. This helps to ensure the security and privacy of your data.

Excellent user experience during development

Realtime Compute provides a standard SQL engine: Flink SQL. It also provides many built-in functions, such as the text functions, date and time functions, and statistics functions. The application of these functions greatly simplifies and accelerates the Flink-based

development. With Flink SQL, even users with limited development knowledge, such as business intelligence (BI) analysts and marketers, can easily perform real-time analysis and processing of big data.

Realtime Compute provides an end-to-end solution for stream processing, including development, administration, monitoring, and alerting. On the Realtime Compute development platform, only three steps are required to publish a job.

Low costs in labors and compute clusters

We have made many improvements to the SQL execution engine, allowing you to create jobs more cost-effectively than to create Flink jobs. Realtime Compute is more cost-effective than open source stream frameworks in both development and production costs. To create an Apache Storm job with complex computational logic, you have to incur high costs and devote a lot of effort, such as writing enormous lines of Java code, debugging, testing, performance tuning, publishing, and long-term administration of open source software applications like Apache Storm and Zookeeper. Realtime Compute allows you to offload the heavy lifting of handling these issues, which helps you focus on your business strategies and rapidly achieve market goals.

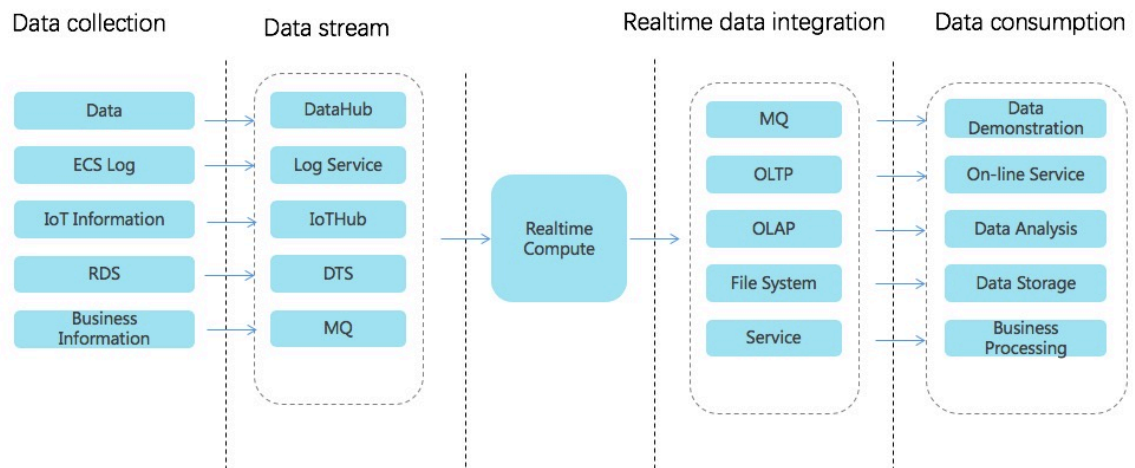
17.5 Product architecture

17.5.1 Business process

We recommend that you have a general knowledge about the stream processing architecture of StreamCompute before using this service. This helps you create effective

plans for the design of stream processing systems. [Figure 17-2: Architecture](#) shows the stream processing architecture of StreamCompute.

Figure 17-2: Architecture



- **Data collection**

You can use data collection tools to collect and send streaming data in real time to a publish-subscribe system for big data analysis. This publish-subscribe system continuously produces events for Realtime Compute in the downstream to trigger real-time computing jobs. The big data ecosystem of Alibaba Cloud offers a wide range of publish-subscribe systems to process streaming data in diversified scenarios. Realtime Compute integrates many of these systems, as shown in the preceding figure. This allows you to easily integrate multiple streaming data stores. To enable compatibility between the computing model of Realtime Compute and that of certain data stores, another data store may be required for data processing. Realtime Compute is seamlessly connected to the following data stores:

- **DataHub**

DataHub allows you to upload data into its system using a wide range of tools and interfaces. For example, you can easily upload logs, binary log files, and IoT streaming data into the DataHub system. DataHub also integrates open source

business software applications. For more information about the data collection tools of DataHub, see DataHub documentation.

- Log Service

Log Service is a one-stop logging service that has been developed by Alibaba Group based on years of experience in addressing challenges involving large amounts of big data experienced by Alibaba Group. Log Service allows you to quickly collect, transfer, query, consume, and analyze log data.

- IoT Hub

IoT Hub is a service that enables developers of IoT applications to implement two-way communications between devices (such as sensors, final control elements, embedded devices, and smart home appliances) and the cloud by creating secure data channels.

You can use the IoT Hub rule engine to easily send IoT data to DataHub, and use Realtime Compute and MaxCompute to process and perform computations on data.

- Data Transmission Service (DTS)

DTS supports data transmission between structured data stores represented by databases. DTS is a data exchange service that streamlines data migration, data synchronization, and data subscription. You can use the data transmission function of DTS to easily parse binary log files such as RDS logs and send data to DataHub. Realtime Compute and MaxCompute allow you to run computations over the data.

- MQ

Message Queue (MQ) is a key service that provides messaging capabilities, such as message publishing and subscription, message tracing, scheduled, and delayed messages, resource statistics, monitoring, and alerting. MQ offers a complete set of enterprise-level messaging functions powered by high-availability (HA) distributed systems and clusters.

- Realtime computing

Data streams continuously enter Realtime Compute for real-time computing. At least one data stream must enter the Realtime Compute system to trigger a real-time computing job. In complex business scenarios, Realtime Compute allows you to perform association queries for static data from data stores and streaming data. For example, you can perform JOIN operations on DataHub and RDS tables based on the primary key of streaming data. You can then perform association queries on DataHub streaming data

and RDS static data. Realtime Compute also enables you to associate multiple data streams. With Flink SQL, you can easily handle large amounts of data and complex business scenarios, such as those experienced by Alibaba Group.

- Realtime data integration

To minimize the data processing delay and simplify data transmission links, Realtime Compute directly writes the result data of real-time computing to data sinks. Realtime Compute allows for a larger Alibaba Cloud ecosystem by integrating the following systems: online transaction processing (OLTP) systems such as ApsaraDB for RDS, NoSQL database services such as Table Store, online analytical processing (OLAP) systems such as AnalyticDB, message queue systems such as DataHub and RocketMQ, and mass storage systems such as Object Storage Service (OSS) and MaxCompute.

- Data consumption

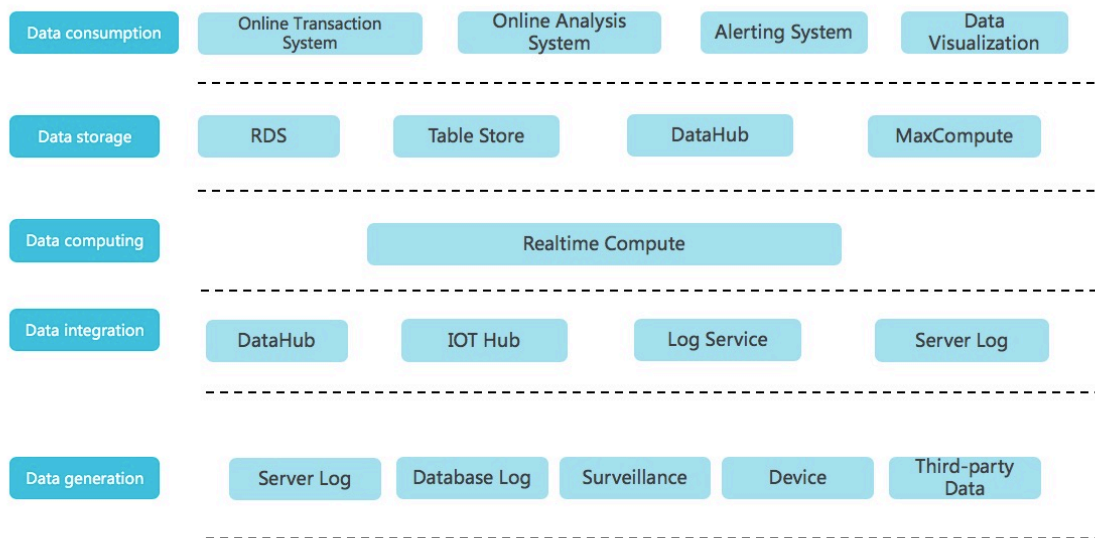
After the result data of real-time computing is written to the sinks, you can consume the data using custom applications.

- You can use data stores to access the result data.
- You can use data transfer systems to send and receive the result data.
- You can use alerting systems to send alerts.

17.5.2 Business architecture

Realtime Compute is a lightweight SQL-enabled streaming engine for real-time processing and analysis of data streams.

Figure 17-3: Business architecture



- **Data generation**

In this phase, streaming data is generated from sources such as server logs, database logs, sensors, and third-party systems. The generated streaming data moves on to the next phase for data integration to drive real-time computing.

- **Data integration**

In this phase, the streaming data is integrated. You can subscribe to and publish the integrated streaming data. The following Alibaba Cloud products can be used in this phase: DataHub for big data computing, IoT Hub for connecting IoT devices, and Log Service for integrating ECS logs.

- **Data computing**

In this phase, the streaming data, which has been subscribed to in the data integration phase, acts as inputs to drive real-time computing in Realtime Compute.

- **Data store**

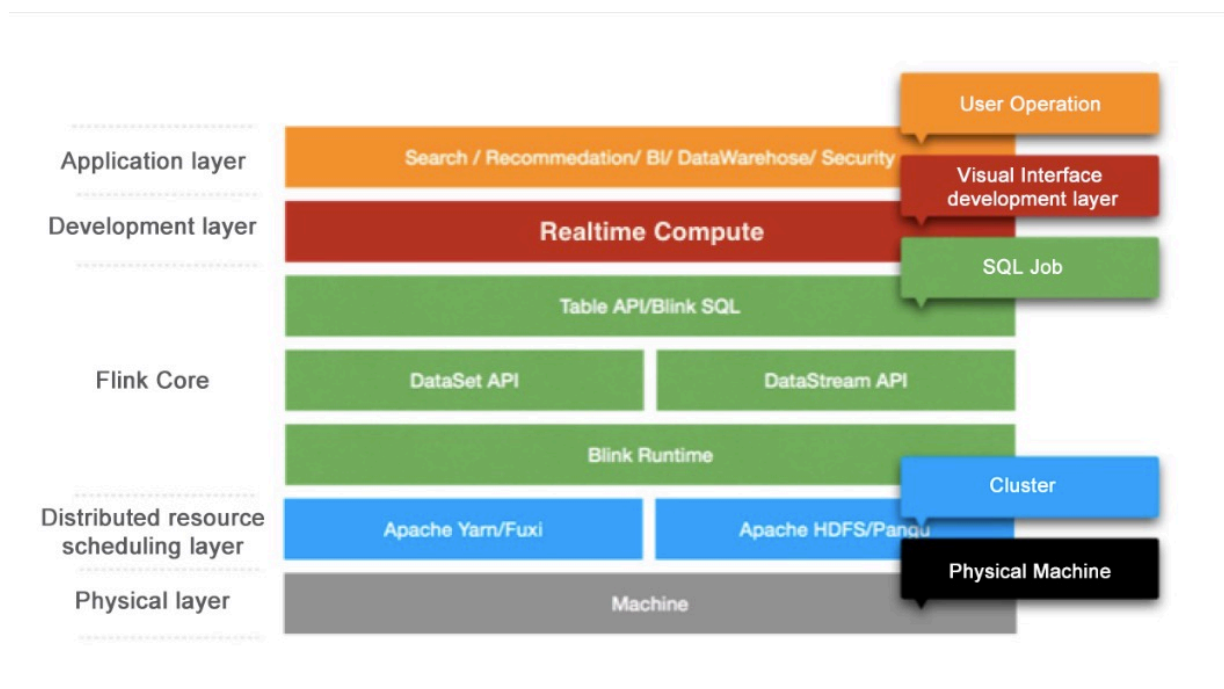
Realtime Compute does not provide built-in data stores. Instead, it writes computing results to external data stores, such as relational databases, NoSQL databases, and online analytical processing (OLAP) systems.

- Data consumption

Realtime Compute supports multiple data store types, which allows you to consume data in various ways. For example, data stores for message queues can be used to report alerts, and relational databases can be used to provide online support.

17.5.3 Technical architecture

Realtime Compute is a real-time data analysis platform for incremental computing. This platform provides statements that are similar to SQL statements and uses the MapReduceMerge (MRM) computing model for incremental computing. Realtime Compute offers a failover mechanism to ensure data accuracy when errors occur.



The Realtime Compute architecture consists of the following five layers.

- Application layer

This layer allows you to create SQL files and publish jobs for real-time data processing based on a development platform. With a well-designed monitoring and alerting system, you would be notified of a processing delay for each job in a timely manner. You can also use systems like Flink UI to view the running information of published jobs and analyze performance bottlenecks. This allows you to quickly and effectively improve job performance.

- Development layer

This layer parses Flink SQL and generates logical and physical execution plans. The execution plans are then conceptualized as executable directed acyclic graphs (DAGs). Based on these DAGs, directed graphs that consist of various models are obtained. Directed graphs are used to implement specific business logic. A model usually contains the following three modules:

- Map: Operations such as data filtering, distribution (GROUP), and join (MAPJOIN) are performed.
- Reduce: Realtime Compute processes streaming data by batch, and each batch contains multiple data records.
- Merge: You can update the state by merging the computing results of the batch, which are produced from the Reduce module, with the previous state. Checkpoints are created after N (configurable) batches have been processed. In this way, the state is stored persistently in a data store, such as Tair and Apache HBase.

- Flink Core

This layer provides a wide range of computing models, Table API, and Flink SQL. You can use DataStream API and DataSet API at the lower sublayer. At the bottom sublayer is Flink Runtime, which schedules resources to ensure that jobs can run properly.

- Distributed resource scheduling layer

Realtime Compute clusters run based on the Gallardo scheduling system. This system ensures that Realtime Compute runs effectively and fault tolerance is provided for recovery.

- Physical layer

This layer provides powerful hardware devices for clusters.

17.6 Features

Realtime Compute has the following features:

- **Data collection and storage**

The premise of running a big data analysis system is that data has been collected into the system. To make full use of your existing streaming data store, Realtime Compute supports integration with multiple upstream streaming data stores, such as DataHub,

Log Service, IoT Hub, Table Store, and MQ. You can use streaming data in existing data stores without operations of data collection and data integration.

You can register data stores on the Realtime Compute development platform. This enables you to leverage the advantages of the one-stop Realtime Compute development platform. Realtime Compute provides the UI for managing different data stores, such as ApsaraDB for RDS, AnalyticDB, and Table Store. Realtime Compute allows you to manage cloud-based data stores in one stop.

- **Data development**

- Realtime Compute provides a fully managed online development platform that integrates a wide range of SQL coding assistance features, such as Flink SQL syntax checking, intelligent code completion, and syntax highlighting.

- **Syntax checking**

On the Development page of Realtime Compute, the revised script is automatically saved. When the script is saved, an SQL syntax check is automatically performed. If a syntax error is detected, the Development page shows the row and column where the error is located, and the cause of the error.

- **Intelligent code completion**

When you enter SQL statements on the Development page of Realtime Compute, auto-completion prompts about keywords, built-in functions, and SQL statements are automatically displayed.

- **Syntax highlighting**

Flink SQL keywords are highlighted in different colors to differentiate data structures.

- The Realtime Compute development platform allows you to manage different versions of SQL code.

Realtime Compute provides key features that help you complete development tasks, such as coding assistance and code version management. On the data development platform, you can manage SQL code versions. Each time you commit code, the system

generates a code version, which can be used for version tracking, modification, and rollback.

- The Realtime Compute development platform allows you to register data stores on its **Development** page for effective data store management, such as data preview and auto DDL generation.

- **Data preview**

The Development page of Realtime Compute allows you to preview the data of multiple data store types. Data preview helps you efficiently analyze upstream and downstream data, identify key business logic, and complete development tasks.

- **Auto DDL generation**

In most cases, the DDL statements for data stores are manually translated into the DDL statements for real-time computing. Therefore, the DDL generation process includes a large number of repetitive tasks. Realtime Compute provides an auto DDL generation feature. This feature simplifies the way that you edit SQL statements for stream processing jobs, reduces the possibility of encountering errors when you manually enter SQL statements, and also improves efficiency.

- Realtime Compute allows you to implement real-time data cleansing, statistics, and analysis using standard SQL. Realtime Compute also supports common aggregation functions, and association queries for streaming data and static data.
- The Realtime Compute development platform provides a simulated running environment where you can customize uploaded data, simulate operations, and check output results.

- **Data operation**

Realtime Compute allows you to manage stream processing jobs on the following tabs under the Administration page: Overview, Curve Charts, FailOver, CheckPoints, JobManager, TaskExecutor, Data Lineage, and Properties and Parameters.

- **Performance tuning**

- Improve performance by automatic configuration

The automatic configuration function of Realtime Compute helps you address performance issues, such as a low throughput of jobs and data piling up in the upstream.

- Improve performance by manual configuration

You can manually configure resources to improve job performance using one of the following methods:

- Optimize resource configuration. You can modify the resources to improve performance by reconfiguring parameters, such as parallelism, core, and heap_memory.
- Improve performance based on job parameter settings. You can specify the job parameters such as miniBatch to improve performance.
- Improve upstream and downstream data stores based on parameter settings. You can specify related parameters to optimize the upstream and downstream data stores for a job.

- **Monitoring and alerting**

This allows you to collect the performance metrics of cloud resources or other custom performance metrics, view service availability, and specify alerts based on the performance metrics. In this way, you can easily view the cloud resource usage and running information of jobs. You can also receive and respond to alerts in a timely manner to ensure that applications can run properly. With Realtime Compute, you can specify alerts for the following performance metrics:

- Processing delay
- Input RPS
- Output RPS
- Failover rate

17.7 Product positioning

Realtime Compute offers Flink SQL to support standard SQL semantics and help you easily implement the computational logic of stream processing. Realtime Compute also provides full-featured UDFs for some authorized users, helping you customize business-specific data

processing logic in scenarios where SQL code functions cannot meet your business needs. In the field of streaming data analysis, you can directly use Flink SQL and UDFs to enable most of the streaming data analysis and processing logic. Realtime Compute focuses on the analysis, statistics, and processing of streaming data. It is less applicable to non-SQL businesses, such as complex iterative data processing and complex rule engine alerts.

Realtime Compute is applicable to the following scenarios:

- Collects the data about page views (PVs) and unique visitors (UVs) in real time.
- Collects the data about the average traffic flow at a traffic checkpoint every 5 minutes.
- Collects and displays the pressure data of hydroelectric dams.
- Reports alerts for financial thefts in online payment services based on fixed rules.

Realtime Compute is inapplicable to the following scenarios for now:

- Replacing Oracle stored procedures with Realtime Compute: Realtime Compute cannot implement all the functions of Oracle stored procedures, because they are designed to handle issues in different fields.
- Seamlessly migrating Spark jobs to Realtime Compute: Currently, you cannot seamlessly migrate Spark jobs to Realtime Compute. However, you can change the stream processing of Apache Spark and migrate this part to Realtime Compute. This eliminates various Apache Spark administration tasks and Spark-based development costs.
- Complex rule engines for alerting: Realtime Compute cannot handle scenarios where multiple complex alerting rules are specified for each data record, and the rules continue to change when the system is running. Specific rule engines need to be used to resolve these issues.

Realtime Compute provides a full set of development tools for streaming data analysis, statistics, and processing based on UDFs and Flink SQL. It allows you to devote the least efforts in developing the underlying code and simply write SQL statements to analyze streaming data. This makes Realtime Compute a good choice for users such as data warehouse developers and data analysts.

17.8 Scenarios

17.8.1 Overview

Realtime Compute uses Flink SQL to provide solutions for streaming data analysis.

- Real-time extract-transform-load (ETL)

Realtime Compute allows you to cleanse, aggregate, and sort streaming data in real time by leveraging the advantages of multiple data channels and flexible data processing capabilities of SQL. Realtime Compute serves as an effective supplement and optimization of offline data warehouses and provides a computing channel for real-time data transmission.

- Real-time reports

Realtime Compute allows you to collect and process streaming data, monitor performance metrics of the business, and view corresponding reports in real time. This enables real-time data administration.

- Monitoring and alerting

Realtime Compute allows you to monitor systems and analyze user behavior in real time, which helps to identify faults and risks in real time.

- Online systems

Realtime Compute allows you to run real-time computations over data streams and view performance metrics in real time. You can shift strategies for online systems in a timely fashion. Realtime Compute can be widely used in various content delivery and intelligent mobile push scenarios.

17.8.2 Management of e-commerce activities

Realtime Compute has evolved into a reliable stream processing platform from Alibaba Group's big data architecture in the e-commerce industry. Realtime Compute is suitable for analyzing various streaming data and providing report support in the e-commerce industry. The e-commerce industry needs to process streaming data in real time in the following scenarios:

- Real-time analysis of user behavior, for example, display of transaction data and user data on big screens. In traditional batch processing models, large amounts of data are processed inefficiently with a long delay. The size of the result data may be excessively large, which poses considerable challenges for online systems that are used for displaying the result data. This may compromise the stability of the online systems.
- Real-time monitoring of users, services, and systems. For example, marketers and engineers can have knowledge of the transactions on the platform over a specified period by viewing the corresponding curve chart. If abnormal fluctuations occur, such as a sharp decrease in transactions, alerts must be instantly triggered and sent to users

- . This helps users effectively respond to abnormal fluctuations and reduces negative impacts on the business.
- Real-time monitoring of major promotional events. For example, marketers need to monitor the metrics of promotional events in real time, such as the Double 11 Shopping Festival created by Alibaba Group and 618 mid-year shopping festival started by JD.com, Inc. This helps marketers effectively decide whether to change strategies.

Integrating with Alibaba Cloud computing and storage systems, Realtime Compute allows you to meet your custom needs for streaming data analysis. Realtime Compute not only satisfies diverse business needs but also simplifies the business development process by using Flink SQL.

17.8.3 Multidimensional analysis of data from IoT sensors

Background

With the economic tidal wave of globalization sweeping over the world, industrial manufacturers are facing increasingly fierce competition. To increase competitiveness, manufacturers in the automotive, aviation, high-tech, food and beverage, textile, and pharmaceutical industries must innovate and replace the existing infrastructure. These industries have to address many challenges during the innovation process. For example, the existing traditional devices and systems have been used for decades, which results in high maintenance costs. However, replacing these systems and devices may slow down the production process and compromise the product quality.

These industries face two additional challenges, which are high security risks and the urgent need for complex process automation. The manufacturing industry has prepared to replace the existing traditional devices and systems. In this industry, high reliability and availability systems are needed to ensure the safety and stability of real-time operations. A manufacturing process involves a wide range of components, such as robotic arms, assembly lines, and packaging machines. This requires remote applications that can seamlessly integrate each stage of the manufacturing process, including the deployment, update, and end-of-life management of devices. The remote applications also need to handle failover issues.

Another requirement for these next-generation systems and applications is that they be able to capture and analyze the large amounts of data generated by devices, and respond appropriately in a timely manner. To increase competitiveness and accelerate development, manufacturers need to optimize and upgrade their existing systems and devices. The

application of Realtime Compute and Alibaba Cloud IoT solutions allows you to analyze device running information, detect faults, and predict yield rates in real time. This topic describes a use case as an example. In this use case, a manufacturer uses Realtime Compute to analyze the large amounts of data collected from sensors in real time. Realtime Compute is also used to cleanse and aggregate data in real time, write data to an online analytical processing (OLAP) system in real time, and monitor the key metrics of devices in real time.

Scenario description

In this use case, the manufacturer has more than 1,000 devices from multiple factories in many cities. Each device is equipped with 10 types of sensors. These sensors send the collected data every 5 seconds to Log Service. The data collected from each sensor follows the format described in the following table.

s_id	s_value	s_ts
The ID of the sensor.	The current value from the sensor.	The time when the data was sent.

The sensors are distributed across devices from multiple factories. The manufacturer creates an RDS dimension table to display the distribution of sensors across devices and factories.

s_id	s_type	device_id	factory_id
The ID of the sensor.	The type of the sensor.	The ID of the device.	The ID of the factory.

The information included in this dimension table is stored in the RDS system. The manufacturer needs to organize the data from sensors based on this dimension table, and sort the data by device. To meet this need, Realtime Compute provides a summary table where the data sent from sensors is logically aggregated by device every minute.

ts	device_id	factory_id	device_temp	device_pres
The time when the data was sent.	The ID of the device.	The ID of the factory.	The temperature of the device.	The pressure of the device.

Assume that there are only two types of sensors in this use case: temperature and pressure. The computational logic is described as follows:

1. Realtime Compute identifies the devices whose temperatures are higher than 80°C and triggers alerts at the downstream nodes. In this use case, Realtime Compute sends the data of the identified devices to MQ. MQ then triggers alerts that the manufacturer has specified in the downstream alerting system.
2. Realtime Compute writes the data to an OLAP system. In this use case, the manufacturer uses HybridDB for MySQL. To integrate with HybridDB for MySQL, the manufacturer has developed a set of business intelligence (BI) applications for multidimensional data display.

FAQ

- How can I aggregate data into a summary table?

In most cases, each sensor only collects the IoT data of one dimension. This poses challenges for subsequent data processing and analysis. To create a summary table, Realtime Compute aggregates data based on windows and organizes data by dimension.

- Why is MQ used to trigger alerts?

Realtime Compute allows you to write data to any type of storage system. We recommend that you use message storage systems like MQ for sending alerts and notifications. This is because the application of these systems helps to prevent the errors encountered by user-defined alerting systems. These errors may cause failures to report certain alerts and notifications.

Code description

Send the data uploaded from sensors to Log Service. The data format of a row is shown as follows:

```
{
  "sid": "t_xxsfdsad",
  "s_value": "85.5",
  "s_ts": "1515228763"
}
```

Define a Log Service source table `s_sensor_data`.

```
CREATE TABLE s_sensor_data (
  s_id VARCHAR,
  s_value VARCHAR,
  s_ts VARCHAR,
  ts AS CAST(FROM_UNIXTIME(CAST(s_ts AS BIGINT)) AS TIMESTAMP),
  WATERMARK FOR ts AS withOffset(ts, 10000)
) WITH (
  TYPE='sls',
  endPoint='http://cn-hangzhou-corp.sls.aliyuncs.com',
```

```

    accessId = '*****',
    accessKey = '*****',
    project = 'ali-cloud-streamtest',
    logStore = 'stream-test',
);

```

Create an RDS dimension table `d_sensor_device_data`. This dimension table stores the mappings between sensors and devices.

```

CREATE TABLE d_sensor_device_data (
  s_id VARCHAR,
  s_type VARCHAR,
  device_id BIGINT,
  factory_id BIGINT,
  PRIMARY KEY(s_id)
) WITH (
  TYPE='RDS',
  url='',
  tableName='test4',
  userName='test',
  password='*****'
);

```

Create an MQ result table `r_monitor_data`. This table specifies the logic for triggering alerts.

```

CREATE TABLE r_monitor_data (
  ts VARCHAR,
  device_id BIGINT,
  factory_id BIGINT,
  device_TEMP DOUBLE,
  device_PRES DOUBLE
) WITH (
  TYPE='MQ'
);

```

Create a HybridDB for MySQL result table `r_device_data`.

```

CREATE TABLE r_device_data (
  ts VARCHAR,
  device_id BIGINT,
  factory_id BIGINT,
  device_temp DOUBLE,
  device_pres DOUBLE,
  PRIMARY KEY(ts, device_id)
) WITH (
  TYPE='HybridDB'
);

```

Aggregate the data collected from sensors by minute and create a summary table based on the aggregated data. To clearly view the code structure and facilitate subsequent administration, we create views in this use case.

```

// Create a view to obtain the device and factory mapping each sensor.
CREATE VIEW v_sensor_device_data
AS
SELECT
  s.ts,

```

```

    s.s_id,
    s.s_value,
    s.s_type,
    s.device_id,
    s.factory_id
FROM
    s_sensor_data s
JOIN
    d_sensor_device_data d
ON
    s.s_id = d.s_id;

// Aggregate the data collected from sensors.
CREATE VIEW v_device_data
AS
SELECT
    // Specify the start time of a tumbling window as the time for the record.
    CAST(TUMBLE_START(v.ts, INTERVAL '1' MINUTE) AS VARCHAR) as ts,
    v.device_id,
    v.factory_id,
    CAST(SUM(IF(v.s_type = 'TEMP', v.s_value, 0)) AS DOUBLE)/CAST(SUM(IF(v.s_type = '
TEMP', 1, 0)) AS DOUBLE) device_temp, // Compute the average temperature by minute.
    CAST(SUM(IF(v.s_type = 'PRES', v.s_value, 0)) AS DOUBLE)/CAST(SUM(IF(v.s_type = 'PRES
', 1, 0)) AS DOUBLE) device_pres // Compute the average pressure by minute.
FROM
    v_sensor_device_data v
GROUP BY
    TUMBLE(v.ts, INTERVAL '1' MINUTE), v.device_id, v.factory_id;

```

In the preceding core computational logic, the average temperature and pressure by minute are computed as the output. Tumbling windows are used in this use case. A new window is started every minute, and a new set of data is generated every minute. The generated data is then filtered and written to the MQ result table and HybridDB result table.

```

// Identify the sensors whose temperatures are higher than 80°C and write the data to
the MQ result table to trigger alerts.
INSERT INTO r_monitor_data
SELECT
    ts,
    device_id,
    factory_id,
    device_temp,
    device_pres
FROM
    v_device_data
WHERE
    device_temp > 80.0;
// Write the result data to the HybridDB for MySQL result table for analysis.
INSERT INTO r_device_data
SELECT
    ts,
    device_id,
    factory_id,
    device_temp,
    device_pres
FROM

```

```
v_device_data;
```

17.8.4 Big screen service for the Tmall Double 11 Shopping Festival

The annual Tmall Double 11 Shopping Festival has become the largest sales event for online shopping in the world. A large number of netizens demonstrate a strong desire to purchase products during the sales event each year. One of the key highlights of this event has been the increase in the overall turnover that is displayed on the Tmall big screen in real time. The real-time display of turnover on the big screen is a result of our senior engineers' hard work over several months. The big screen service excels in key performance metrics. For example, the end-to-end delay has been reduced within 5 seconds, from placing orders on the Tmall platform, to data collection, processing, verification, and to displaying the sales data on the big screen. As for the processing capability, hundreds of thousands of orders can be processed during the peak hours around 00:00 on November 11. Additionally, to ensure fault tolerance, multiple channels have been used to back up data.

Realtime Compute provides key support for the big screen service. The stream processing of the big screen service was previously based on the open source Apache Storm. The Storm-based development process took around one month. The application of Flink SQL shortened the development process of the big screen service to one week. The underlying layer of Realtime Compute removes the Apache Storm modules that are designed for execution optimization and troubleshooting. This enables higher efficiency and faster processing for Realtime Compute jobs.

- Online shopping rush

During the Double 11 Shopping Festival, an enormous number of netizens join the online shopping rush on the Tmall platform. During the peak hours when "seckill" activities occur, such as 00:00 on November 11, hundreds of thousands of sales orders need to be processed in real time. The word "seckill" vividly describes fighting among buyers, which means that a buyer wins or loses all in a matter of seconds.

- Real-time data collection

The data collection system collects and sends the logs of database changes to the DataHub system. With the application of Data Transmission Service (DTS), the data from online transaction processing databases can be written to DataHub tables within seconds at the peak hours around 00:00 on November 11.

- Real-time data computing

Realtime Compute subscribes to the DataHub streaming data, continuously analyzes the streaming data, and calculates the total turnover up to the current time. In Realtime Compute, a cluster can contain up to thousands of nodes. The throughput of a job reaches millions of data records per second, fully meeting the system requirements of processing hundreds of thousands of transactions per second in Tmall. Realtime Compute subscribes to data and writes the result data to an online RDS system in real time.

- Frontend data visualization

We also provide advanced data visualization components for the Tmall Double 11 Shopping Festival. These components allow you to view the total turnover on a dashboard, and the distribution of global transaction activities across the world in real time. To achieve astounding visual effects for the big screen, the frontend server performs periodic polling operations on the RDS system, and advanced web frontend applications are used.

17.8.5 Mobile data analysis

Realtime Compute allows you to analyze the data of mobile apps in real time. With Realtime Compute, you can analyze performance metrics of mobile apps, such as crash detection and distribution, and distribution of app versions. Mobile Analytics is a product provided by Alibaba Group to analyze the data of mobile apps. This product allows you to analyze user behavior and logs from multiple dimensions. It also helps mobile developers implement fine-grained operations based on big data analysis, improve product quality and customer experience, and enhance customer stickiness. The underlying big data computing of Mobile Analytics is implemented based on big data products of Alibaba Cloud, such as Realtime Compute and MaxCompute. Mobile Analytics uses Realtime Compute as the underlying engine for streaming data analysis. This allows Mobile Analytics to offer a wide range of real-time data analysis and reporting services for mobile apps.

- Data collection

To collect data, developers can include the software development kit (SDK) provided by Mobile Analytics into an app installation package. This SDK offers data collection components based on mobile operating systems. These components collect and send the data about mobile phones and user behavior to the backend of Mobile Analytics for analysis.

- Data reporting

The backend of Mobile Analytics offers a data reporting system, which allows you to collect the data reported by mobile phones using the specified SDK. The data reporting system preliminarily removes dirty data, and sends the processed data to DataHub.

**Note:**

In the future, DataHub provides an SDK for mobile phones to directly report data. The removal of dirty data is performed in Realtime Compute instead of Mobile Analytics, reducing the host costs of Mobile Analytics.

- Stream processing

Realtime Compute continuously subscribes to the DataHub streaming data. It also continuously reads and runs computations over the data about the performance metrics of mobile apps. Realtime Compute then writes the result data of stream processing during each period to an online RDS or Table Store system.

- Data display

Mobile Analytics provides a complete set of performance metrics that allow you to quickly view the running information and usage of mobile apps. For example, you can quickly know user locations, visited pages, browsing duration, end devices and network environments, and slow responses or crashes. With Mobile Analytics, you can also analyze crashes by device, and view the details of crashes. The data display is based on the result data that is obtained in the stream processing phase.

17.9 Restrictions

None

17.10 Terms

Project

In Realtime Compute, a project is a basic unit for managing clusters, jobs, resources, and users. Project administrators can create projects, or add users to other existing projects . Realtime Compute projects can be collaboratively managed by Apsara Stack tenant accounts and RAM users.

Job

Similar to a MaxCompute or Hadoop job, a Realtime Compute job implements the computational logic of stream processing. A job is a basic unit for stream processing.

CU

In Realtime Compute, a compute unit (CU) defines the minimum capabilities of stream processing for a job with the specified CPU cores, memory, and input/output capacities. A Realtime Compute job can use one or more CUs.

Currently, a CU is assigned with one CPU core and 4 GB memory.

Flink SQL

Unlike most open source stream processing systems that provide basic APIs, Realtime Compute offers Flink SQL that includes standard SQL semantics and advanced semantics for stream processing. Flink SQL is designed to satisfy diversified business needs, and it allows developers to perform stream processing by using standard SQL. With Realtime Compute, even users with limited technological skills, such as data analysts, can quickly and easily process and analyze streaming data.

UDF

Realtime Compute allows you to use user-defined functions (UDFs) that are similar to Apache Hive UDFs. We recommend that you use UDFs to implement your custom computational logic. UDFs are a supplement to Flink SQL that can be used for standard stream processing. Currently, Realtime Compute only supports Java UDFs.

Resource

Currently, Realtime Compute only supports Java UDFs. A JAR file uploaded by a user is defined as a resource.

Data collection

During a typical data collection process, data is collected from sources and ingested into a big data processing engine. The data collection process of Realtime Compute focuses on the phases where data is collected from the source and then transferred into a data bus.

Data store

Realtime Compute is a lightweight computing engine without built-in data stores. Data sources and sinks of Realtime Compute are based on external data stores. For example, you can use RDS to store result tables for Realtime Compute.

Data development

During the data development process, you edit Flink SQL statements to create a Realtime Compute job. Realtime Compute offers an online integrated development environment (IDE) where you can edit SQL statements and debug data before publishing a Realtime Compute job.

Data administration

The data administration page of the Realtime Compute platform allows for online management of jobs. Realtime Compute helps you easily and effectively manage stream processing jobs.

18 Machine Learning Platform for AI

18.1 What is machine learning?

Machine learning is a process of using statistical algorithms to learn large amounts of historical data and generate an empirical model to provide business strategies.

Apsara Stack Machine Learning Platform for AI is a set of data mining, modeling, and prediction tools. It is developed based on MaxCompute (also known as ODPS). Machine Learning Platform for AI supports the following functions:

- Provides an all-in-one algorithm service covering algorithm development, sharing, model training, deployment, and monitoring.
- Allows you to complete the entire procedure of an experiment either through the GUI or by running PAI commands. This function is typically intended for data mining personnel, analysts, algorithm developers, and data explorers.
- In Apsara Stack, Machine Learning Platform for AI runs on MaxCompute. Machine Learning Platform for AI allows you to call algorithms to decouple the applications and compute engines after you have deployed algorithm packages in MaxCompute clusters.
- Provides various algorithms and reliable technical support, providing more options to resolve service issues. In the Data Technology (DT) era, you can use Machine Learning Platform for AI to implement data-driven services.

Machine Learning Platform for AI can be applied in the following scenarios:

- Marketing: commodity recommendations, user profiling, and precise advertising.
- Finance: loan delivery prediction, financial risk control, stock trend prediction, and gold price prediction.
- Social network sites (SNSs): microblog leader analysis and social relationship chain analysis.
- Text: news classification, keyword extraction, text summarization, and text analysis.
- Unstructured data processing: image classification and image text extraction through OCR.
- Other prediction cases: rainfall forecast and football match result prediction.

Machine learning can be divided into three types:

- **Supervised learning:** Each sample has an expected value. You can create a model and map input feature vectors to target values. Typical examples of this learning mode include regression and classification.
- **Unsupervised learning:** No samples have a target value. This learning mode is used to discover potential regular patterns from data. Typical examples of this learning mode include simple clustering.
- **Reinforcement learning:** This learning mode is complex. A system constantly interacts with the external environment to obtain external feedback and determines its own behavior to achieve a long-term optimization of targets. Typical examples of this learning mode include AlphaGo and driverless vehicles.

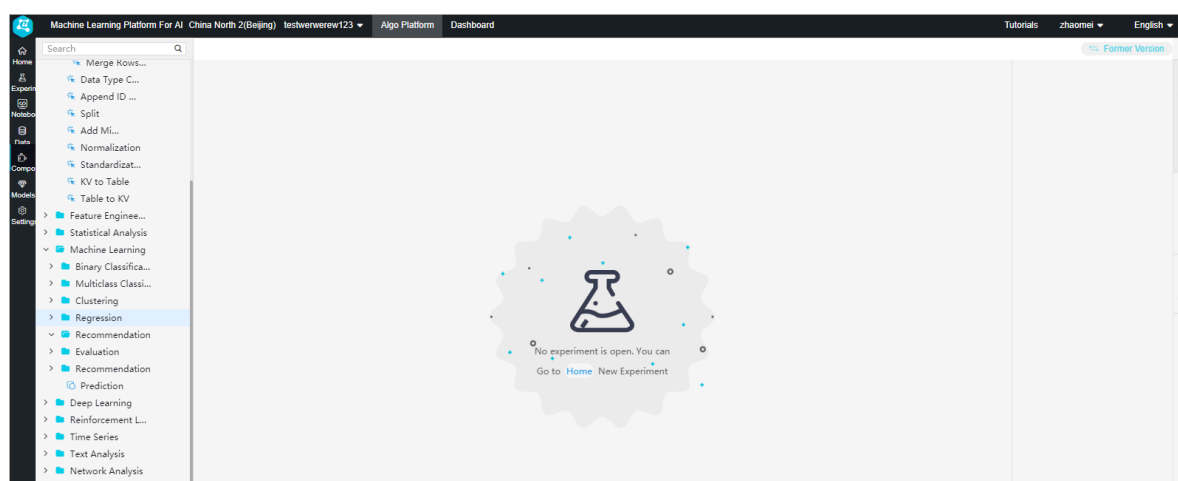
18.2 Benefits

Alibaba Cloud Machine Learning Platform for AI has the following benefits:

All-in-one visual user interface

- Machine Learning Platform for AI provides a Web interface for you to mine data by dragging and dropping components without programming, like piling up blocks, as shown in [Figure 18-1: User interface](#).

Figure 18-1: User interface



- Machine Learning Platform for AI provides the data model visualization function. It allows you to use charts to view data analysis results and algorithm evaluation.
- Machine Learning Platform for AI provides an all-in-one solution for data processing, model training, prediction, evaluation, model deployment, service building, and task scheduling.

- In addition to the Web interface, Machine Learning Platform for AI also provides command line tools to easily integrate algorithms into your projects.

Multiple high-performance machine learning algorithms

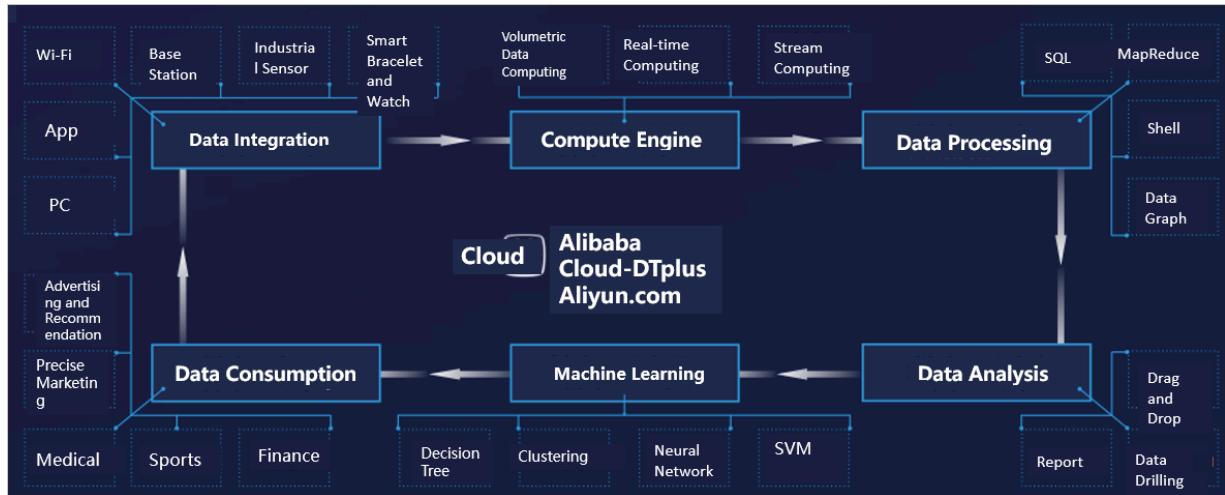
- Machine Learning Platform for AI provides nearly 100 machine learning algorithms that can be applied to multiple business scenarios, such as data preprocessing, clustering, regression, text analysis, and feature processing algorithms.
- Compared with traditional software, Machine Learning Platform for AI adopts the latest and optimal algorithms in the machine learning industry to improve the computing capability and accuracy.
- Machine Learning Platform for AI supports deep learning and GPU job scheduling. Machine Learning Platform for AI integrates and completely optimizes the TensorFlow framework. You can get started with TensorFlow for model training.
- Machine Learning Platform for AI provides open-source algorithms that are developed based on years of experience of Alibaba Cloud in big data mining and utilization. This greatly shortens the data modeling, model deployment, and model utilization period.

Full compatibility with Alibaba Cloud services

- Apsara Stack has established a big data ecosystem, such as Machine Learning Platform for AI. All services are ready for use after you activate them.
- Machine Learning Platform for AI runs on MaxCompute and is integrated with DTplus DataWorks to help data mining, parent and child node data collection, experiment scheduling, and data utilization, as shown in [Figure 18-2: Alibaba Cloud DTplus services](#).

- Based on the MPI, PS, graph algorithms, and MapReduce computing frameworks and distributed algorithms, Machine Learning Platform for AI easily handles a large amount of data.

Figure 18-2: Alibaba Cloud DTplus services



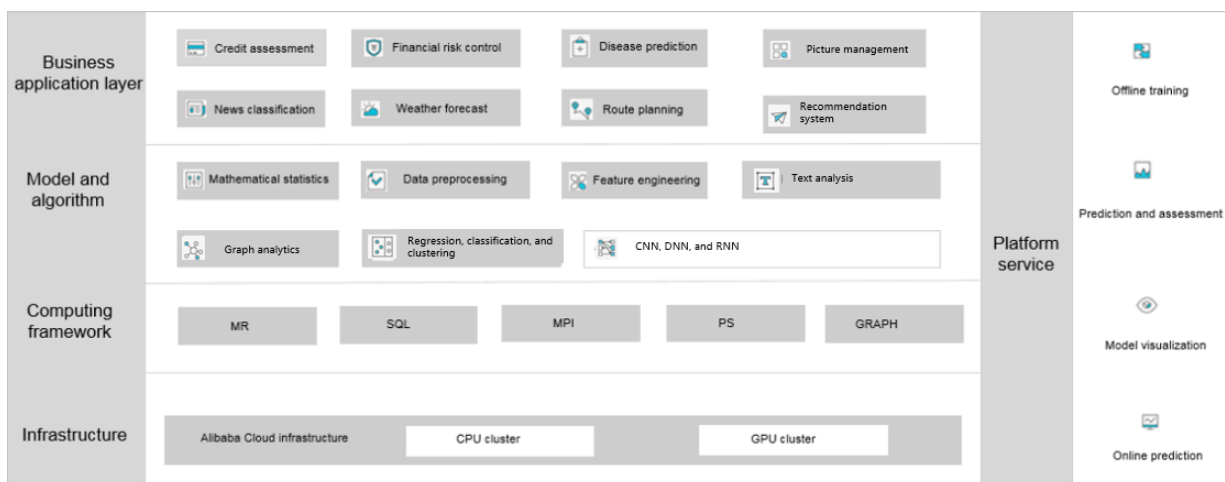
High-quality technical support

Machine Learning Platform for AI is supported by Alibaba algorithm scientists and Apsara Stack technical support. If you have any issues, submit a ticket through the ticket system or contact Apsara Stack technical personnel.

18.3 Architecture

Figure 18-3: Basic architecture of Machine Learning Platform for AI shows the basic architecture of Machine Learning Platform for AI.

Figure 18-3: Basic architecture of Machine Learning Platform for AI



The architecture is composed of the following layers:

- Infrastructure layer: provides the cluster resources for computing. You can choose CPU or GPU computing clusters based on the algorithm type.
 - The CPU cluster runs machine learning algorithms and provide computing resources such as CPU and memory resources. Computing resources are centrally managed by an algorithm framework. After jobs are submitted, the algorithm framework schedules compute nodes in the CPU cluster and dispatches jobs to the compute nodes.
 - The GPU cluster runs deep learning framework jobs and provides computing resources such as GPU and graphics memory. Computing resources are centrally managed by an algorithm framework. After jobs are submitted, the algorithm framework schedules compute nodes in the GPU cluster. For a task that requires multiple workers and GPUs, a virtual network is automatically created to dispatch the jobs to the compute nodes in the virtual network.
- Computing framework layer: manages the CPU resources, GPU resources, and a basic runtime environment for algorithms, such as the MapReduce runtime library, MPI runtime library, PS runtime library, and TensorFlow framework.

The deep learning framework TensorFlow supports the open-source version 1.4. The computing framework layer also optimizes the performance and I/O interfaces. You can use TensorFlow to read files from and write models to OSS buckets. When TensorFlow is running, you can start TensorBoard to display the status of parameter convergence during convolution.

- Model and algorithm layer: provides basic components such as data preprocessing , feature engineering, and machine learning algorithm components. All algorithm components come from the Alibaba Group algorithm system and have been tested on petabytes of service data.
- Business application layer: The Alibaba search system, recommendation system, Ant Financial, and other projects use Machine Learning Platform for AI for data mining. Machine Learning Platform for AI can be applied to industries such as finance, medical care, education, transportation, and security.

Based on this architecture, Machine Learning Platform for AI provides a Web-based visual algorithm experiment console. The Web GUI allows you to perform offline training, prediction, and evaluation, visualize models, deploy online prediction services, or release experiments to the scheduling system of DataWorks.

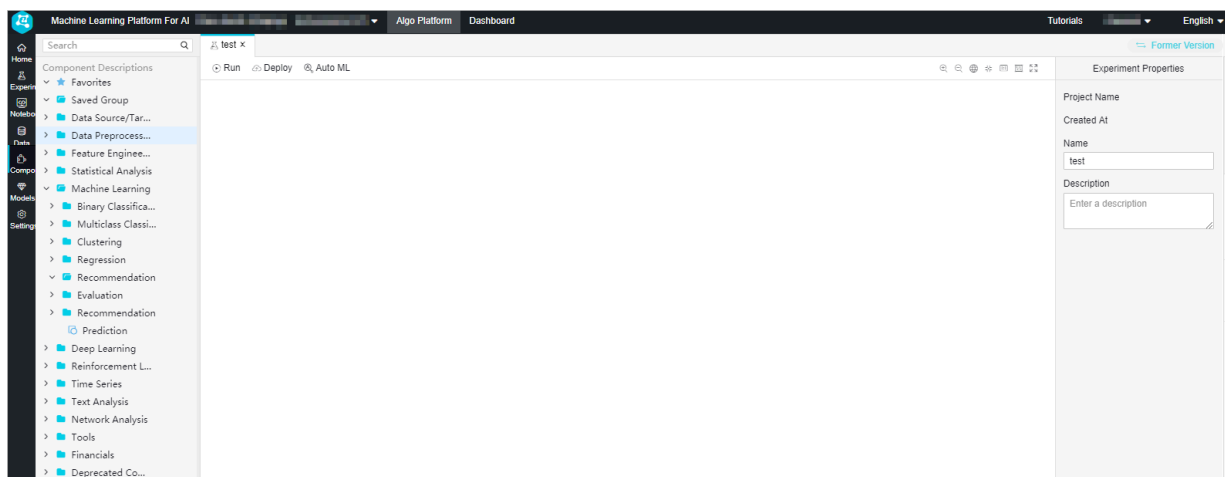
18.4 Features

18.4.1 Visualized modeling

Machine Learning Platform for AI provides the easy-to-use visual modeling feature, which allows you to view the logic of the procedure.

The visual modeling pages include the algorithm platform page and online model service page. [Figure 18-4: Algorithm platform page](#) shows the function section on the algorithm platform page. [Figure 18-5: Online model service page](#) shows the online model service page.

Figure 18-4: Algorithm platform page



Sections on the algorithm platform page

The algorithm platform page includes the following sections:

- **Features section:** displays machine learning features and information, such as experiments, components, data sources, and models, in a tree structure.
- **Canvas section:** You can drag and drop components to the canvas to build a directional workflow in order to complete data mining tasks, such as the metadata collection, data processing, modeling, and model deployment.
- **Properties section:** You can configure component parameters in this section.

Features section

The features section on the algorithm platform page provides the following menus:

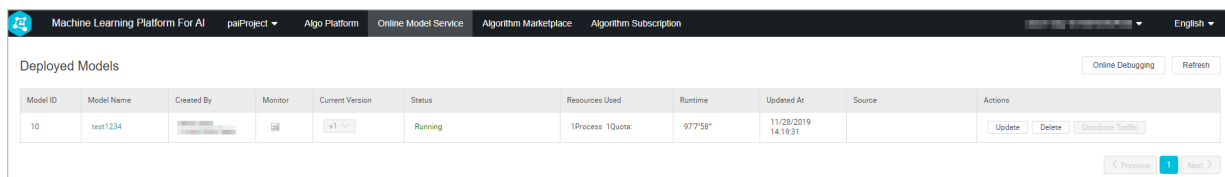
- **Search:** You can search data, tables, and experiments.

- **Experiments:** After you double-click the name of an experiment, the canvas displays the directional flowchart of the experiment. You can continue modifying the experiment.
- **Data sources:** allows you to view and manage all data tables.
- **Components:** provides multiple key features of machine learning, such as machine learning components.
- **Models:** allows you to manage all models.
- **Developer tool:** allows you to view the experiment runtime log and troubleshoot experiment issues based on returned error messages and alerts.

Online model service

In the upper section of the page, select Online Model Service to go to the online model service page. This page displays user-created online services. You can monitor or select an action for these algorithm services.

Figure 18-5: Online model service page

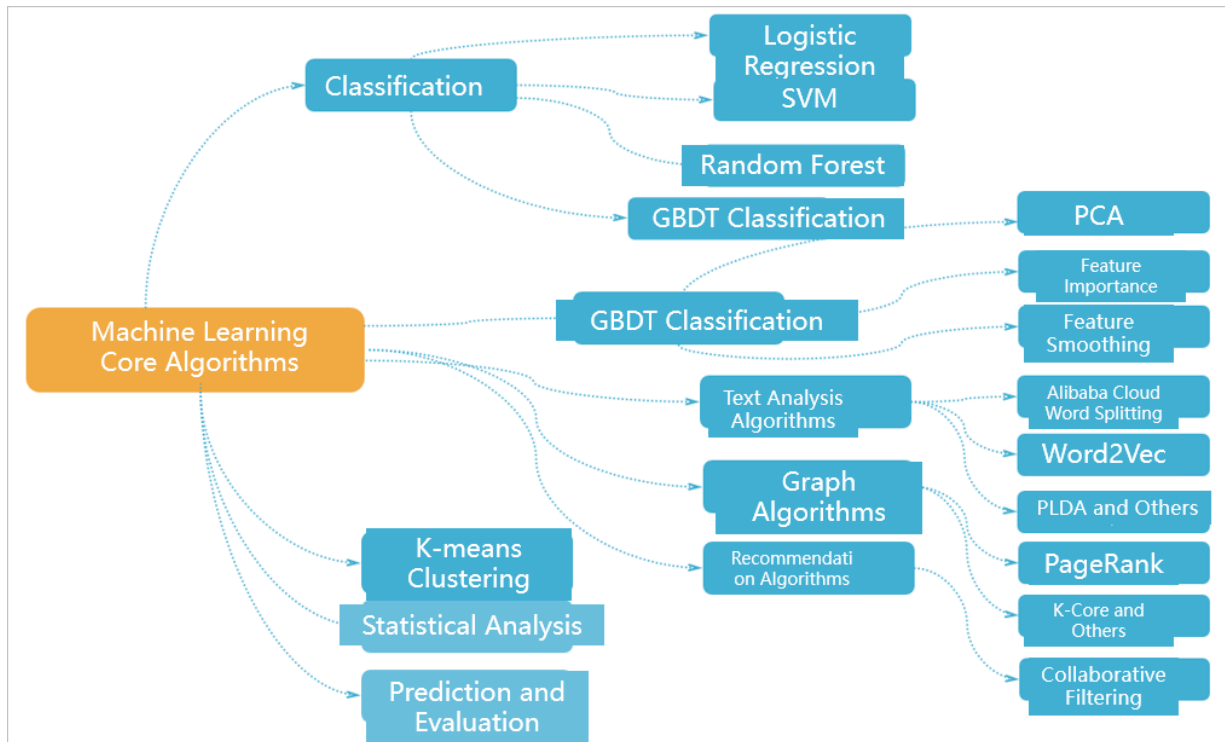


Model ID	Model Name	Created By	Monitor	Current Version	Status	Resources Used	Runtime	Updated At	Source	Actions
10	test1234			v1.0	Running	1Process 1Quota	977'58"	11/28/2019 14:19:31		Update Delete Distribute Traffic

18.4.2 All-in-one experience

Machine Learning Platform for AI has a complete algorithm library, as shown in [Figure 18-6: Algorithm library of Machine Learning Platform for AI](#).

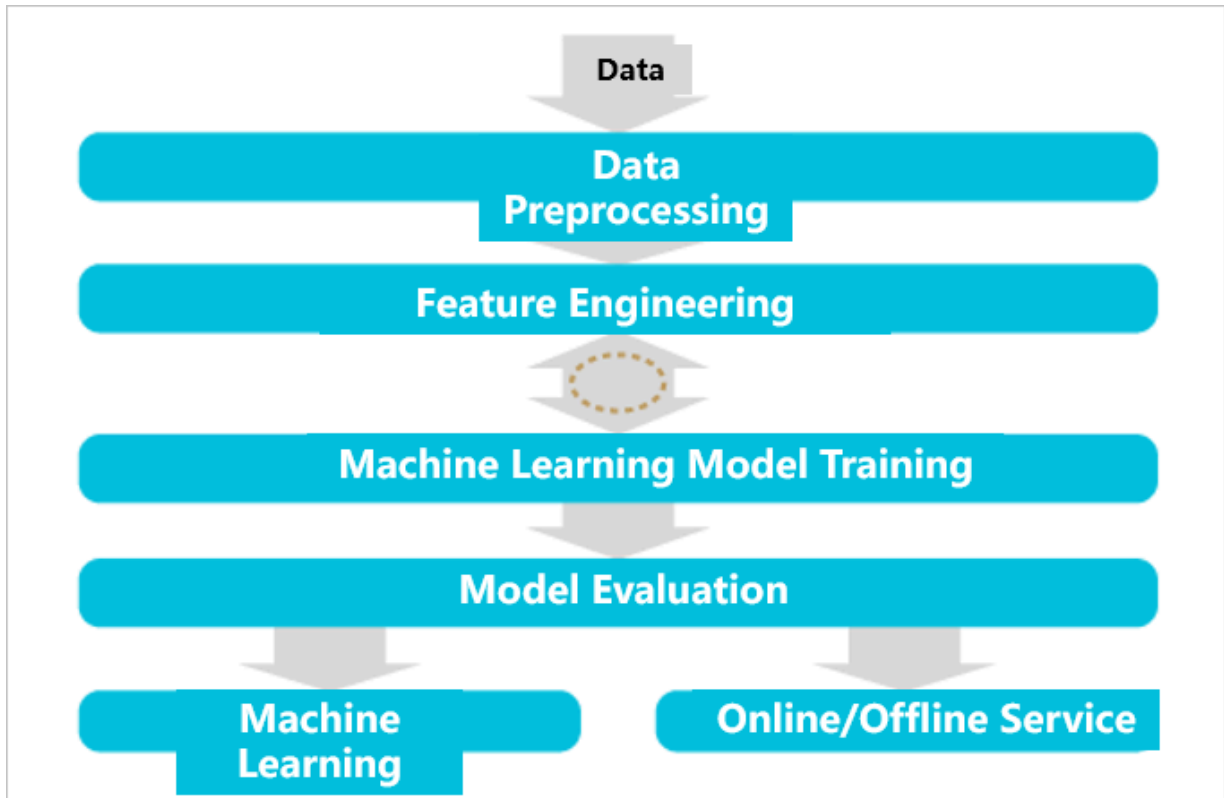
Figure 18-6: Algorithm library of Machine Learning Platform for AI



Typically, you need to perform many operations to complete data mining or model training such as data extract, transform, and load (ETL), data preprocessing, feature engineering, model training, evaluation, and deployment, as shown in [Algorithm development process of atypical model](#). Machine Learning Platform for AI provides an all-in-one development environment with a complete set of components and tools for you to complete the entire data mining or model training task, such as metadata processing and model deployment. With these basic components, you can import data to the platform, create an experiment,

and create solutions to resolve issues in different scenarios, and save costs on environment switching.

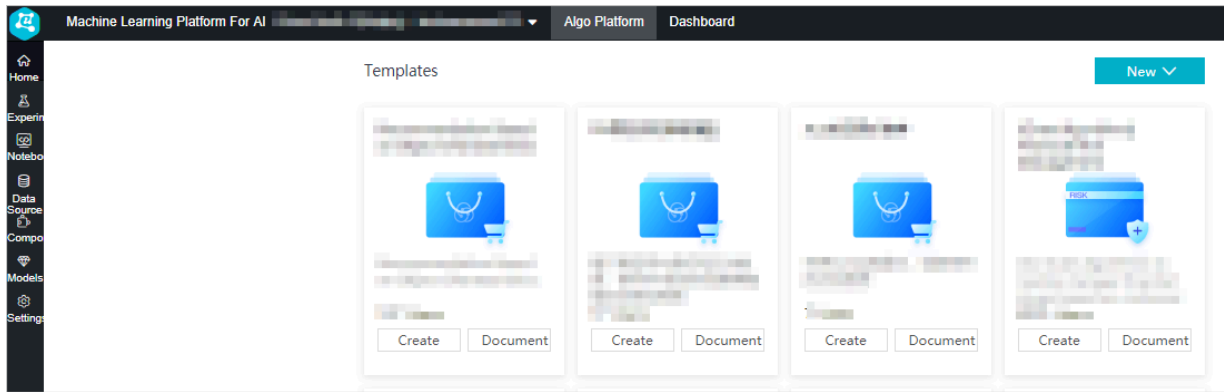
Figure 18-7: Procedure of developing algorithms for typical machine learning models



18.4.3 Multiple templates on the homepage

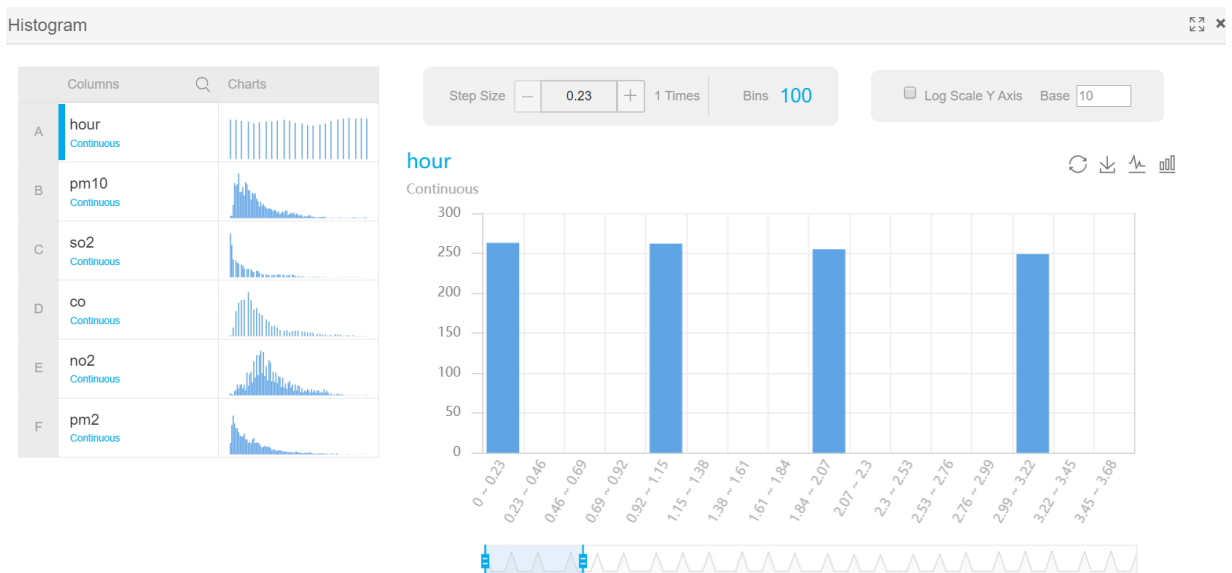
To help data analysts get started with the service, Machine Learning Platform for AI provides a set of experiment templates for scenarios such as product recommendations, text analysis, financial risk management, and weather prediction. These templates contain configurations and data that you can run the experiment with.

You can create experiments from the templates provided on the homepage. You can learn information about how an experiment is configured, how machine learning works, and how data is processed.



18.4.4 Data visualization

You can right-click an output component to view the visual output model. For example, you can view the model evaluation report and data analysis results. Visualized output can be displayed in multiple forms such as line charts, dot charts, and bar charts.



18.4.5 Model management

Visualized model management:

1. In the left-side navigation pane, click **Models**.
2. Expand the **Models** folder to locate the model built from a specified experiment.

3. Right-click the model and then select **Show Model**.

Binary Logistic Regression

If the input is sparse data, feature columns with zero weight are not displayed.

Column Name ▲	Weight	
	1 ▲	0 ▲
pm10	15.7360985405929	-
so2	1.474706135267199	-
co	0.8739231801649705	-
no2	8.652489608506198	-
Constants	-14.47551218333226	0

Save to MaxCompute: pai_lr_coefficient_xlab_m_logisticregres_1165164_v0

Save

Close



Note:

You can also right-click the model to perform other actions. For example, you can export the PMML file or deploy the model.

18.4.6 Algorithm components

The current version of Machine Learning Platform for AI on Apsara Stack provides up to 89 algorithm components. These components are classified into 11 categories. The following table lists the algorithm components.

Level 1	Level 2	Level 3	Description
1. Data source /target	1.1 Read MaxCompute table	N/A	Reads a MaxCompute table.
	1.2 Write MaxCompute table	N/A	Writes a MaxCompute table.
2. Data preprocessing	2.1 Sampling and filtering	2.1.1 Weighted sampling	Collects samples according to the specified sampling fraction .
		2.1.2 Random sampling	Collects samples randomly.
		2.1.3 Filtering and mapping	Uses the SQL WHERE clause to filter data.
		2.1.4 Stratified sampling	Collects samples by stratum.

Level 1	Level 2	Level 3	Description
	2.2 Data merge	2.2.1 JOIN	Uses the SQL JOIN clause to merge data.
		2.2.2 Merge columns	Merges two columns from two tables.
		2.2.3 Merge rows (UNION)	Uses the SQL UNION operator to merge rows.
	2.3 Others	2.3.1 Standardization	Standardizes a column in a table.
		2.3.2 Splitting	Splits data according to the specified ratio.
		2.3.3 Normalization	Normalization is a method to simplify computation. Normalization converts a dimensional expression into a dimensionless expression (scalar).
		2.3.4 Missing data imputation	Replaces a null or specified value with the maximum, minimum, average, or custom value.
		2.3.5 KV to Table	Converts key:value (KV) pairs to a regular table.
		2.3.6 Table to KV	Converts a regular table to KV pairs.
		2.3.7 Append ID column	Adds an auto-increment ID column to a table.
3. Feature engineering	3.1 Feature transformation	3.1.1 PCA	Dimension reduction algorithm .
	3.2 Feature importance evaluation	3.2.1 Linear model feature importance evaluation	Evaluates the importance of the features in a linear model.
		3.2.2 Random forest feature importance evaluation	

Level 1	Level 2	Level 3	Description
4. Statistical analysis	4.1 Percentile	N/A	Calculates the percentile of a column.
	4.2 Data pivoting	N/A	Supports the data pivoting function.
	4.3 Covariance	N/A	Measures the covariance of two given values.
	4.4 Empirical probability density chart	N/A	Returns the nonparametric density based on the estimated probability density.
	4.5 Chi-square goodness of fit test	N/A	Determines the differences between the observed and expected frequencies of each class for a multiclass nominal variable.
	4.6 Chi-square test of independence	N/A	Checks whether two factors (each has two or more classes) are mutually independent. The null hypothesis is that the two factors are independent of each other.
	4.7 Two sample T test	N/A	The two sample T test includes the independent sample T test and the paired sample T test. The two samples independent of each other are called independent samples. An independent sample T test checks whether two samples are significantly different from each other. A paired sample T test checks whether the mean values from two paired populations are significantly different from each other.
	4.8 One sample T test	N/A	One sample T test checks whether the mean of a normally distributed population differs significantly from the target value.

Level 1	Level 2	Level 3	Description
	4.9 Normality test	N/A	Determines whether observed values are normally distributed.
	4.10 Lorenz curve	N/A	Illustrates the distribution of wealth across a population.
	4.11 Whole table statistics	N/A	Calculates the statistical information of each column in a table, including the default value, maximum value, minimum value, variance, and deviation.
	4.12 Pearson coefficient	N/A	Calculates the Pearson coefficient of two numerical columns.
	4.13 Histogram	N/A	Shows metrics in a histogram.
	4.14 Scatter plot	N/A	A chart where data points are distributed on the Cartesian coordinate plane.
	4.15 Correlation coefficient matrix	N/A	Calculates a matrix of correlated coefficients for multiple columns.
5. Machine learning	5.1 Binary classification	5.1.1 Linear SVM	A supervised machine learning algorithm used to identify and classify models, and then perform regression analysis.
		5.1.2 Logistic regression for binary classification	A supervised machine learning algorithm that uses logistic regression to train a binary classification model.
		5.1.3 GBDT binary classification	GBDT is an iterative decision tree algorithm that calculates results based on the final conclusions of multiple decision trees.
	5.2 Multiclass classification	5.2.1 Logistic regression for multiclass classification	A linear regression algorithm used for multiclass classification.

Level 1	Level 2	Level 3	Description
		5.2.2 Random forest	A type of classifier that uses multiple trees and sample data to generate models and make predictions.
		5.2.3 KNN	For a row in a prediction table , this component selects K nearest records from the training table. It then adds the row to the class that is most common among the K records.
		5.2.4 Naive Bayes	A family of classification algorithms based on the theorem of Bayes and independent hypothesis of feature conditions.
	5.3 Regression	5.3.1 GBDT regression	An algorithm that uses the GBDT structure for regression.
		5.3.2 PS linear regression	An algorithm that supports a large amount of training data by using parameter servers.
		5.3.3 Linear regression	Analyzes the linear relationship between a dependent variable and multiple independent variables.
	5.4 Clustering	5.4.1 K-means clustering	Clustering similarity is calculated based on a central object obtained by using mean values of objects in different clusters.
	5.5 Evaluation	5.5.1 Binary classification evaluation	Evaluates a binary classification model and uses it to make predictions.
		5.5.2 Multiclass classification evaluation	
		5.5.3 Regression model evaluation	

Level 1	Level 2	Level 3	Description
		5.5.4 Clustering model evaluation	
		5.5.5 Confusion matrix	
	5.6 Prediction	5.6.1 Prediction	
	5.7 Collaborative recommendation	5.7.1 Collaborative filtering (etrec)	Etrec is an item-based collaborative filtering algorithm that uses two input columns and outputs the top K items with the highest similarity.
6. Time series	6.1 x13_arima	N/A	ARIMA is short for Autoregressive Integrated Moving Average Model. x13-arima is an ARIMA algorithm based on the open-source X-13ARIMA-SEATS seasonal adjustment.
	6.2 x13_auto_arima	N/A	Automatically selects an ARIMA model based on the Gomez and Maravall processes.
7. Text analysis	7.1 Word splitting	N/A	An algorithm that is used to split words in the specified text. Currently, only Chinese is supported for the Taobao and Internet word splitting models.
	7.2 Word frequency statistics	N/A	After word splitting, words are listed in the same order of document IDs. The frequency of a word appears in each document is calculated.
	7.3 TF-IDF	N/A	A statistical method for evaluating the importance of a word to a document in a collection or corpus.
	7.4 PLDA	N/A	Outputs the probability density of the topics in each document.
	7.5 Word2Vec	N/A	Converts words to vectors.

Level 1	Level 2	Level 3	Description
	7.6 Convert rows, columns, and values to KV pairs	N/A	Converts a set of row, column , and value to a KV pair (row,[col_id,value]).
	7.7 Text summarization	N/A	Uses machine learning algorithms to create a summary on a document.
	7.8 Keyword extraction	N/A	Extracts words from a document. The extracted words are most correlated to the meaning of the document.
	7.9 Sentence splitting	N/A	Splits sentences by punctuation.
	7.10 Deprecated word filtering	N/A	A preprocessing method in text analysis. This method is used to filter out the noise in word splitting results, such as of, yes , and ah. Custom dictionaries are supported.
	7.11 String similarity		String similarity calculation is a basic operation in machine learning. It is typically used in industries such as information retrieval, natural language processing, and bioinformatics. This algorithm supports these similarity calculation methods: Levenshtein distance , longest common substring , string subsequence kernel, cosine, and simhash_hamming . It also supports these input methods: string-to-string and top N.
	7.12 String similarity-top N	N/A	Checks whether a given string is one of the top N predictions.

Level 1	Level 2	Level 3	Description
	7.13 Semantic vector distance	N/A	Outputs words (sentences) that are nearest to a given word (sentence) based on the semantic vectors calculated by an algorithm component, such as Word2vec. For example , you can generate a list of words that are most similar to a given word based on the semantic vectors returned by the Word2vec component.
	7.14 N-gram counting	N/A	N-grams are generated based on words.The number of the corresponding N-grams in all corpora is counted.
	7.15 PMI	N/A	Counts the co-occurrence of all words in several documents and calculates the Pointwise Mutual Information (PMI) between every two words.
	7.16 Document similarity	N/A	Calculates the similarity between documents or sentences that are separated with spaces based on the similarity of strings. The document similarity is calculated in same way as string similarity calculation.
8. Deep learning	8.1 TensorFlow 1.4	N/A	Uses the PAI-Tensorflow framework to implement deep learning.
	8.2 Read OSS buckets	N/A	Specifies the authorization information and endpoint that are required when Machine Learning Platform for AI reads or writes the corresponding Object Storage Service (OSS) bucket.

Level 1	Level 2	Level 3	Description
9. Network analysis	9.1 K-Core	N/A	The k-core of a graph is the subgraph that remains after all vertices with a degree less than or equal to K are removed .
	9.2 Single-source shortest path	N/A	Calculates the shortest path between two points.
	9.3 Page ranking	N/A	Calculates the rank of a Web page.
	9.4 Label propagation clustering	N/A	A graph-based semi-supervised machine learning algorithm. A node is labeled based on the labels of its neighboring nodes. The scale at which labels are propagated is determined by the similarity between the node and its neighbors. Labels are propagated iteratively and updated over time.
	9.5 Label propagation classification	N/A	A semi-supervised classification algorithm that uses the label information of labeled nodes to predict that of unlabeled nodes.
	9.6 Modularity	N/A	A measure of the structure of networks. It measures the closeness of communities divided from a network structure. A value larger than 0.3 represents an obvious community structure.
	9.7 Maximal connected subgraph	N/A	A maximal connected subgraph of an undirected graph G is a connected subgraph of G, where all vertices are connected with the least edges.

Level 1	Level 2	Level 3	Description
	9.8 Vertex clustering coefficient	N/A	Calculates the peripheral density of nodes near a node in an undirected graph. The density of a star network is 0. The density of a fully meshed network is 1.
	9.9 Edge clustering coefficient	N/A	Calculates the peripheral density of each edge in an undirected graph.
	9.10 Counting triangle	N/A	Generates all triangles in an undirected graph.
	9.11 Tree depth	N/A	Generates the depth and tree ID of each node in a network composed of many trees.
10. Tools	10.1 MaxCompute SQL	N/A	Runs MaxCompute SQL statements.
11. Financials	11.1 Binning	N/A	Performs data binning by using the equal frequency, equal width, or auto binning mode. The input is continuous or discrete features. The output is binning rules for all features .
	11.2 Data conversion module	N/A	Uses binning results to convert features that you input. The data conversion module supports normalization, Weight of Evidence (WoE), and discretization. Normalization outputs values between 0 to 1 . WoE replaces feature values with WOE of bins. Discretization converts variables to dummy variables based on binning results. All output data is in KV format.

Level 1	Level 2	Level 3	Description
	11.3 Scorecard training	N/A	The scorecard is a modeling tool commonly used in credit risk evaluation. Scorecard modeling performs original variable discretization through binning and uses linear models (such as logistic regression and linear regression) to conduct model training. The scorecard supports various features, including feature selection and score conversion. In addition, it allows you to add constraints to variables during model training.
	11.4 Scorecard prediction	N/A	The scorecard prediction component uses the model generated by the scorecard training component to predict scores.
	11.5 Population stability index	N/A	Population stability index (PSI) is an important metric to identify a shift in the population for credit scorecards, for example, the changes in the population within two months. A PSI value smaller than 0.1 indicates insignificant changes. A PSI value between 0.1 and 0.25 indicates minor changes. A PSI value larger than 0.25 indicates major changes in the population.

18.5 Scenarios

Machine Learning Platform for AI can be applied to the following scenarios:

Marketing

- Use cases: commodity recommendations, user profiling, and precise advertising.
- Example: Machine Learning Platform for AI associates user shopping behavior data with commodities to offer commodity recommendations and evaluate the recommendation results.

Finance

- Use cases: loan delivery prediction, financial risk management, stock trend prediction, and gold price prediction.
- Example 1: Agricultural loan delivery is used in a typical example of data mining. A lender uses machine learning to build an empirical model based on historical data such as the lendee's annual income, cultivated crop type, and debit and credit history. This model is then used to predict the lendee's capacity.
- Example 2: Users' credit card expense records are processed by a machine learning algorithm. After raw data binning and feature engineering transformation, data is used to build a linear model. The final credit score of each user is determined by the model predictions, and can be used in a variety of loan and finance related credit checks.

Text

- Use cases: news classification, keyword extraction, text summarization, and text analysis .
- Example: A simple system for automatic commodity label classification is built using the text analysis function of Machine Learning Platform for AI.

Take online shopping as an example. A commodity typically has labels for multiple dimensions. For example, the commodity description of a pair of shoes may be "Korean Girl Dr. Martens Women's Preppy/British-style Lace-up Dull-polish Ankle High Platform Leather Boots." A bag may be described as "Discount Every Day 2016 Autumn and Winter New Arrival Women's Korean-style Seashell-shaped Tassel Three-way Bag as a Messenger Bag, Hand Carry Bag, and Shoulder Bag."

Each product description contains multiple dimensions such as the time, place of origin , and style. E-commerce platforms face the daunting challenge of how to classify hundreds of thousands of products based on these specified dimensions. The biggest

challenge is determining which labels constitute the dimensions of each product. A label classification system can be built using a machine learning algorithm to automatically learn label terms. For example, the system can learn location-related labels such as Japan, Fujian, and Korea.

Unstructured data processing

- Use cases: image classification and image text extraction by using optical character recognition (OCR).
- Example: A prediction model can be quickly built for image recognition by using the TensorFlow deep learning framework. The TensorFlow deep learning framework can recognize images and return image classification results within half an hour. Image recognition by using deep learning can also be used in illicit image filtering, facial recognition, and object detection.

Other prediction cases

- Use cases: rainfall forecast, football match result prediction, microblog leader analysis, and social relationship chain analysis.
- Example: Air quality can be predicted by Machine Learning Platform for AI based on historical air quality index data such as PM 2.5, carbon monoxide concentration, and nitrogen dioxide concentration. The prediction results can then be used to determine which air quality index has the greatest impact on PM 2.5 levels.

18.6 Limits

This topic describes the limits of Machine Learning Platform for AI.

Item	Description
MaxCompute service deployment	The computing service of Machine Learning Platform for AI relies on MaxCompute to store tables and perform some SQL-related computations. Therefore, MaxCompute must be deployed before you can use the machine learning service.
OSS service deployment (deep learning)	The deep learning service of Machine Learning Platform for AI relies on OSS to store, read, and write data. Therefore, OSS must be deployed.

Item	Description
Limit to component use	For more information, see parameter configurations for each component.

18.7 Terms

This topic describes terms used in Machine Learning Platform for AI.

data mining

A broad definition that describes the use of algorithms to explore useful information from large amounts of data. Typically, data mining uses machine learning algorithms.

Alibaba Cloud DTplus

The big data platform of Alibaba Cloud. DTplus provides enterprises with a complete set of end-to-end big data solutions for fields such as enterprise data warehouses, BI, machine learning, and data visualization. These solutions help enterprises become more agile, smarter, and more perceptive in the data technology (DT) era.

table

Data storage units of MaxCompute. Tables used by machine learning are stored in MaxCompute. Logically, a table is a two-dimensional structure that consists of rows and columns. Each row represents a record. Each column represents a field of the same data type. One record can contain one or more columns. The schema of a table consists of column names and column types.

On Machine Learning Platform for AI, you can create a table, add the table to favorites, and import data to the table. The table is automatically stored in MaxCompute. To delete a table, you must log on to MaxCompute.

partition

Certain columns specified in a table when the table is created. In most cases, you can consider a partition as a directory in a file system.

Tables are stored in MaxCompute. MaxCompute uses each value in a partition column as a partition (directory). You can specify multiple hierarchies of partitions by using multiple table columns as table partitions. The relationships between partitions are similar to those between multiple hierarchies of directories.

When using data, if you specify the name of a partition, only the data in the specified partition is read. This removes the need to scan the entire table for data, improves processing efficiency, and minimizes costs.

lifecycle

The period of time that determines how long a table partition is retained since it was last updated. If a table (partition) is not updated within the specified time period, MaxCompute automatically deletes it.

sparse data format

Datasets in which most data entries are null or have a value of 0. Sparse data can be utilized effectively if efficient methods are used to explore the useful information that exists in the relatively incomplete sparse data set.

On Machine Learning Platform for AI, if the data of a feature in a sample is in the sparse format, you must convert the format to the LibSVM format, select **key:value**, **key:valuesparse data format** on the parameter setting page, and then upload the data.

feature

An attribute that is used to describe an object. For example, a person can be described by age, gender, occupation, and other attributes. Each of these attributes is a feature of the person.

On Machine Learning Platform for AI, a dataset is stored as a table. A column in the table is a feature of this dataset. The features of data are important to machine learning. Data and its features determine the upper limit of machine learning capabilities. Models and algorithms are used to help machine learning reach the upper limit. Therefore, features must be processed before a machine learning experiment can be executed. Typical feature processing methods include data preprocessing, feature selection, and dimension reduction.

dimension reduction

A method that is used to remove the dimensions that have minor impacts and extract key features from a large number of features. A dimension is the way something is observed. In machine learning, dimensions describe the features of a dataset. If a dataset has millions of features, the training model for machine learning will be complex and the training will take a long period of time. In this case, dimension reduction is required. The dimension reduction algorithms on Machine Learning Platform for AI include PCA and LDA.

19 Quick BI

19.1 What is Quick BI?

Quick BI is a flexible, lightweight self-service BI platform based on cloud computing.

Quick BI supports various data sources:

- MaxCompute (formerly known as ODPS) and AnalyticDB for PostgreSQL
- User-created MySQL databases that are hosted on ECS
- Data sources in VPC networks

Quick BI provides real-time online analysis of a large volume of data. Quick BI helps significantly reduce data retrieval costs and is easy to use with the support of intelligent data modeling tools. Drag-and-drop operations and various visual charts allow you to create reports and use pivot charts and tables, downloads, data exploration, and BI portals to easily analyze data.

Quick BI enables everyone to be both a data viewer and a data analyst to achieve data-driven operations for enterprises.

19.2 Benefits

Benefits of Quick BI can be summarized as follows.

High compatibility

Connects to various Alibaba Cloud data sources, such as MaxCompute and AnalyticDB for PostgreSQL.

Quick response

Responds in seconds to hundreds of millions of data queries.

Powerful capabilities

Allows users to easily create complex reports by using workbooks.

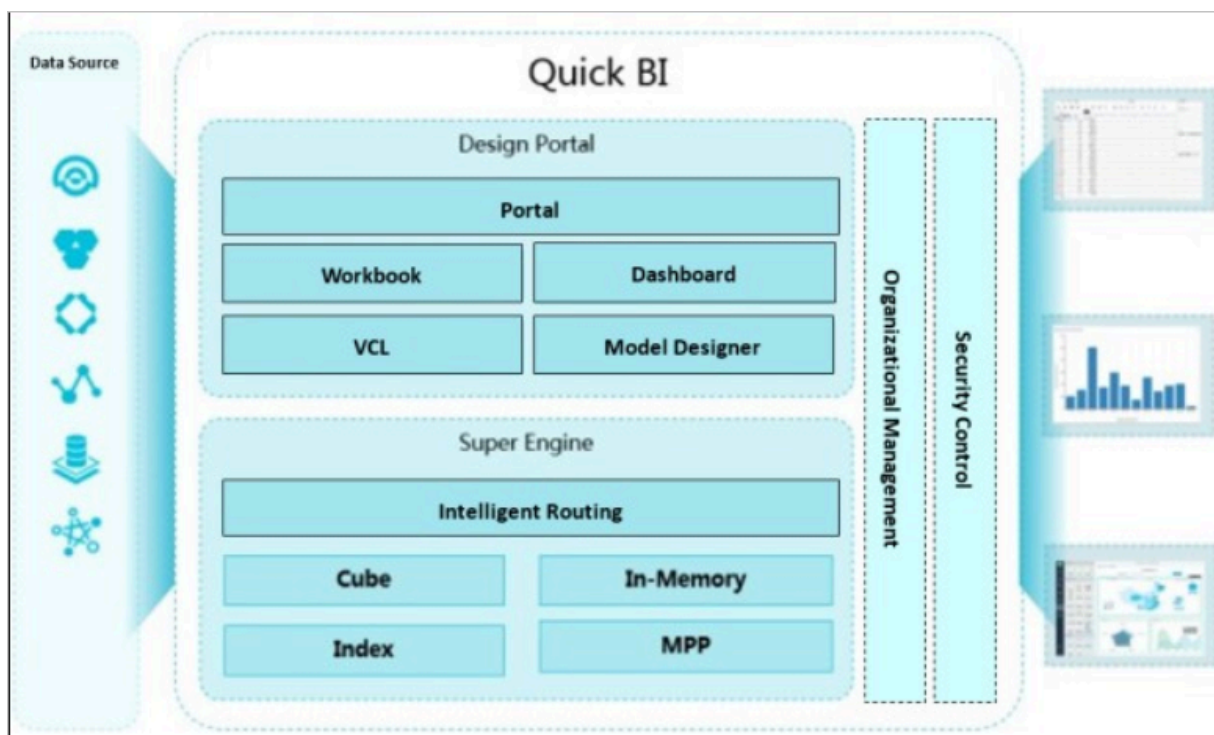
User-friendliness

Provides various data visualization functions and automatically identifies data properties to generate the most appropriate charts for users.

19.3 Architecture

This topic describes the overall architecture of Quick BI, and its main modules and functions.

The following figure shows the architecture of Quick BI.



Modules and features of Quick BI:

- **Data connection module**

Connects to various Alibaba Cloud data sources, such as MaxCompute and AnalyticDB for PostgreSQL. This module provides APIs to query metadata or data from data sources.

- **Data preprocessing module**

Provides lightweight ETL processing for data sources. Currently, Quick BI supports custom SQL of MaxCompute. Quick BI will support data preprocessing for more data sources in the future.

- **Data modeling module**

Takes charge of OLAP modeling of data sources and transforms data sources into multi-dimensional analysis models. It supports standard semantics such as dimensions (such as Date type dimensions and Geo type dimensions), measures, and galaxy schemas. It also supports calculated fields, and allows you to process dimensions and measures by using SQL syntax for existing data sources.

- **Workbook**

Provides workbook functions. This module enables data analysis, such as row and column filtering, standard and advanced filtering, subtotal and total calculation, and conditional formatting. It also supports operations such as data export, text processing, and sheet processing.

- **Dashboard**

Assembles visual charts into dashboards. Supports various types of charts, including line chart, pie chart, vertical bar chart, funnel chart, hierarchy chart, bubble map, colored map, and kanban. It also supports inter-chart field dependency and five basic widgets including filter bar, tab, iFrame, image, and text area.

- **BI portal**

Assembles dashboards into a BI portal, and supports built-in links (dashboards), external links (third-party links), and basic settings of templates and the menu bar.

- **Query engine**

Queries data that is stored in data sources.

- **Organization permission management**

Configures permissions based on organizations, workspaces, and workspace-specific user roles. Quick BI allows you to grant your members different permissions on a report as needed.

- **Row-level permission control**

Controls row-level permissions of data. Different members can view different data in a report based on the permissions they are granted.

- **Share and publish**

Shares workbooks, dashboards, and BI portals with other members. You can also publish dashboards to the Internet to share with the public.

19.4 Features

This topic describes the features of Quick BI.

Quick BI provides the following features:

Seamless integration with cloud databases

Supports various Alibaba Cloud data sources, such as MaxCompute and AnalyticDB for PostgreSQL.

Charts

Provides diverse options for data visualization. To meet data presentation needs in different scenarios, Quick BI supports various built-in visual charts, such as vertical bar charts, line charts, pie charts, radar charts, and scatter charts. Quick BI automatically identifies data properties and intelligently recommends visualization solutions.

Analysis

Enables multidimensional data analysis. Quick BI is a web-based data analysis system. It supports drag-and-drop operations in workbooks, clear data presentation similar to EXCEL files, one-click data import, and real-time analysis. You can use Quick BI to analyze data from different perspectives without repetitive modeling.

Quick building of BI portals

Provides drag-and-drop operations, powerful data modeling capabilities, and multiple visual charts to help you build BI portals in a short time.

Real-time analysis

Supports online analysis for a large volume of data without the need of preprocessing, significantly improving the efficiency of data analysis.

Data permissions

Supports member management and row-level data permission control. This enables users of different roles to view different reports and view different data of the same report.

19.5 Scenarios

19.5.1 Instant data analysis and effective decision-making

Quick BI can instantly analyze a large volume of data and make decisions.

Business goals:

- Convenient data retrieval

Quick BI eliminates the reliance on IT professionals to write SQL statements for multidimensional data analysis.

- Convenient report generation and maintenance

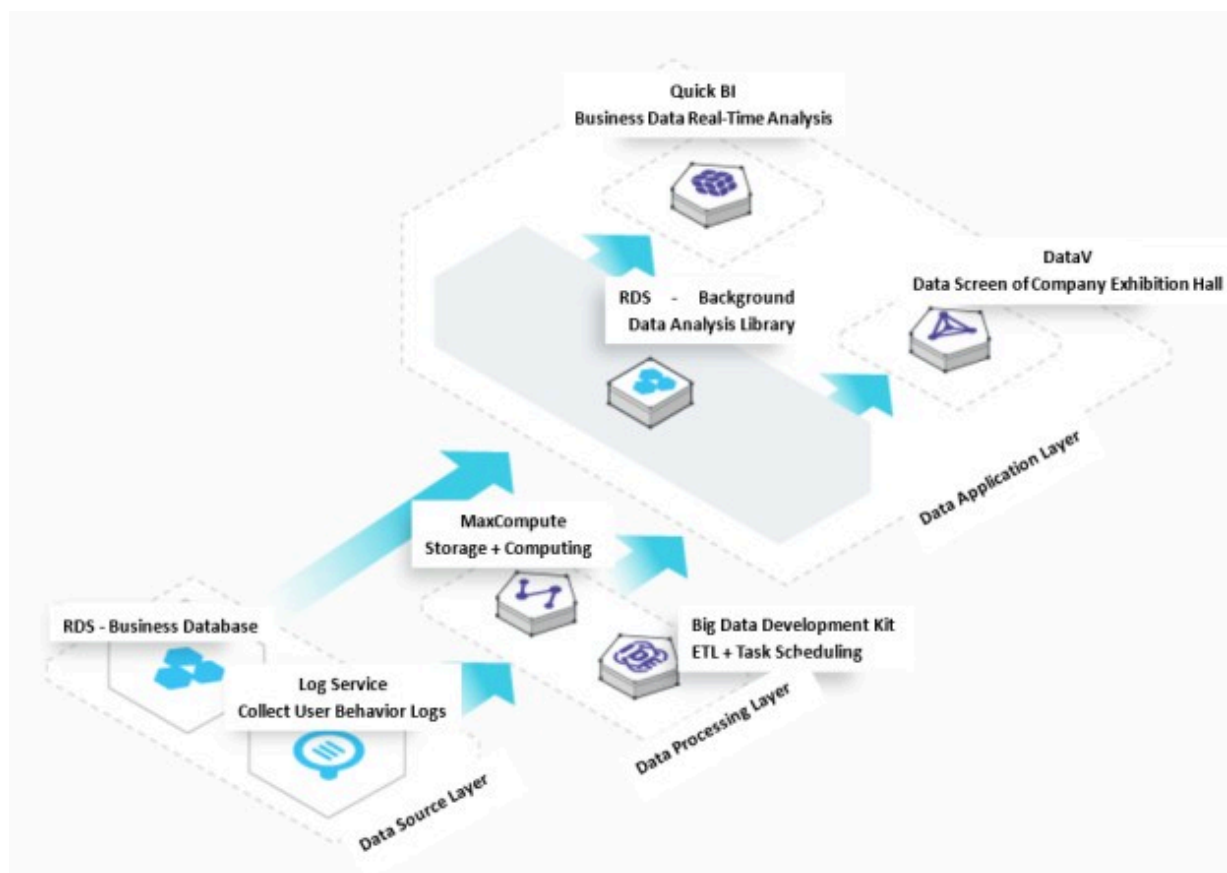
Quick BI simplifies and shortens the process of delivering updates and new code to an analytics system.

- Low human resource costs

Quick BI provides easy-to-use user interfaces, reducing your maintenance costs.

Recommended combination: relational database and Quick BI

Figure 19-1: Instant data analysis and effective decision-making



19.5.2 Integration with existing systems

The Quick BI report system can be integrated with your own systems, such as the OA system and internal management system, to efficiently present data.

Business goals:

- Easy adoption

Quick BI is a user-friendly and easy-to-use service for users from different backgrounds, which satisfies data analysis needs of personnel in various departments.

- High efficiency for data visualization

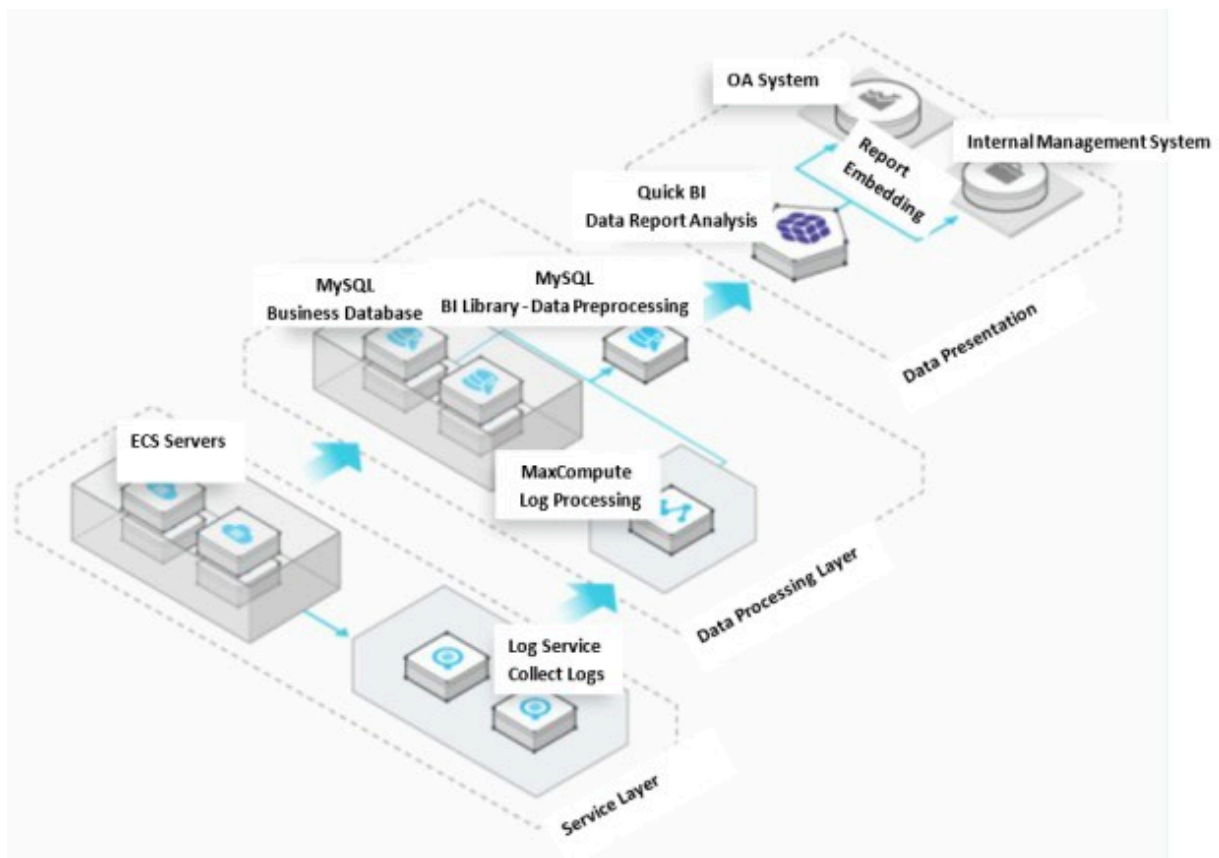
Integration with existing systems allows for quick data analysis and improves the efficiency of viewing data.

- Unified management platform

You can access and manage data by using a unified platform.

Recommended combination: relational database and Quick BI

Figure 19-2: Integration with existing systems



19.5.3 Permission control of transaction data

Quick BI controls over transaction data permissions and row-level permissions.

Business goals:

- Row-level permission control

You can easily create a report for all members, but one member can only view data related to their marketplace.

- Dynamic business requirements

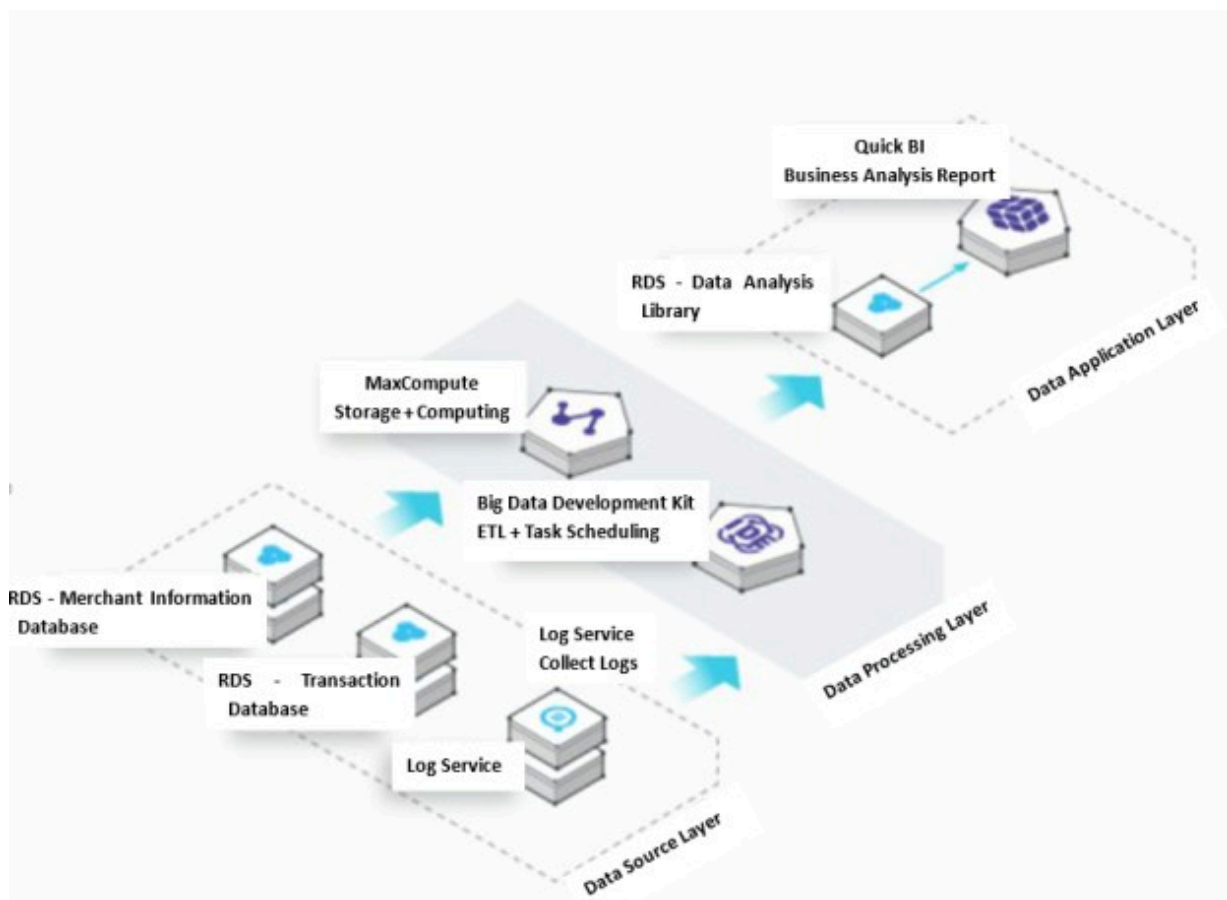
Quick BI responds quickly to numerous changes in statistical indicators as business grows.

- Consistent computing performance across multiple data sources

You can leverage the BI capabilities of Alibaba Cloud to resolve the issues arising from cross-source data analysis and the computing performance bottleneck.

Recommended combination: Log Service, relational database, Quick BI, and MaxCompute

Figure 19-3: Permission control of transaction data



19.6 Limits

None.

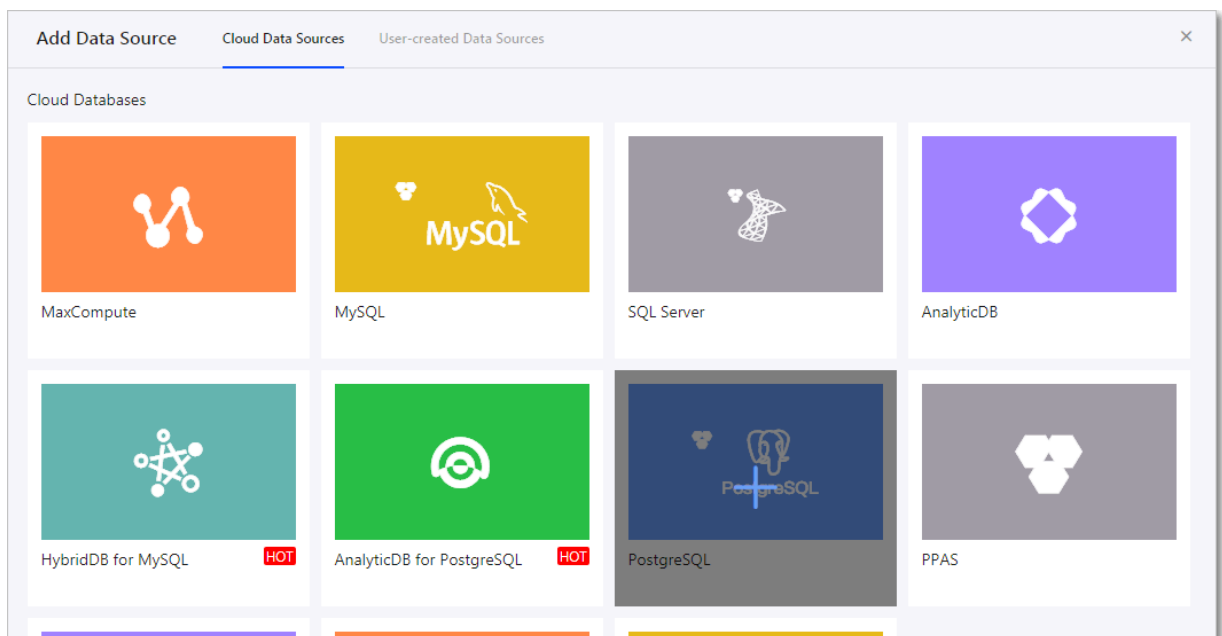
19.7 Terms

This topic describes commonly used terms and concepts in Quick BI.

Data source

When you use Quick BI for data analysis, you must first specify the source of the raw data. A data source is where data is stored. You can add data sources by using either of the following methods:

- Add data sources from cloud databases.
- Add data sources from user-created databases.



Dataset

You can use tables from data sources to create datasets. You can edit, move, or delete a dataset from the dataset list.

Dashboard

Dashboards employ a flexible tile layout to allow you to create interactive reports. Moreover, dashboards provide data filtering and data query functions, and use multiple charts to display data.

You can drag and drop fields or double-click them to add the data of the fields to charts in dashboards and view the data more clearly. Dashboards provide user-friendly interfaces, improving user experience.

Workbook

Workbooks display analyzed and processed data in a dataset. You can use workbooks in both personal and group workspaces. To analyze data in a workbook, you can select the dataset where the data is located and perform required operations.

BI portal

A BI portal is a set of dashboards organized in the form of menus. You can build a business analysis system by using BI portals. A BI portal references analyzed data from Quick BI and supports external links.

20 Apsara Big Data Manager (ABM)

20.1 What is Apsara Big Data Manager?

Apsara Big Data Manager (ABM) is an operations and maintenance platform tailored for big data products.

Currently, ABM supports the following products:

- MaxCompute
- DataWorks
- StreamCompute
- Quick BI
- DataHub
- Machine Learning Platform for AI

ABM supports operations and maintenance of big data products from perspectives such as business, services, clusters, and hosts. You can also install patches for big data products, customize alert configurations, and view O&M history through the ABM console.

ABM allows on-site Apsara Stack engineers to manage big data products with ease. For example, they can view performance metrics in real time, modify runtime configurations, and check and handle alerts in a timely manner.

20.2 Benefits

This topic describes the benefits of Apsara Big Data Manager in the following aspects: cluster health monitoring, resource usage analysis, and graphical O&M management.

Cluster health monitoring

Allows you to monitor and configure the devices, resources, and services that are used in the clusters of big data products, and collects performance metrics in real time for dynamic display.

Resource usage analysis

Collects the runtime statuses of cluster devices, resources, and services in real time, and supports data aggregation and analysis to help you evaluate the health status of the

cluster. If the evaluation result indicates potential risks in a cluster, responsible engineers would be notified immediately.

Graphical management interface

Provides a graphical user interface for performance metrics visualization and common O&M operations.

20.3 Architecture

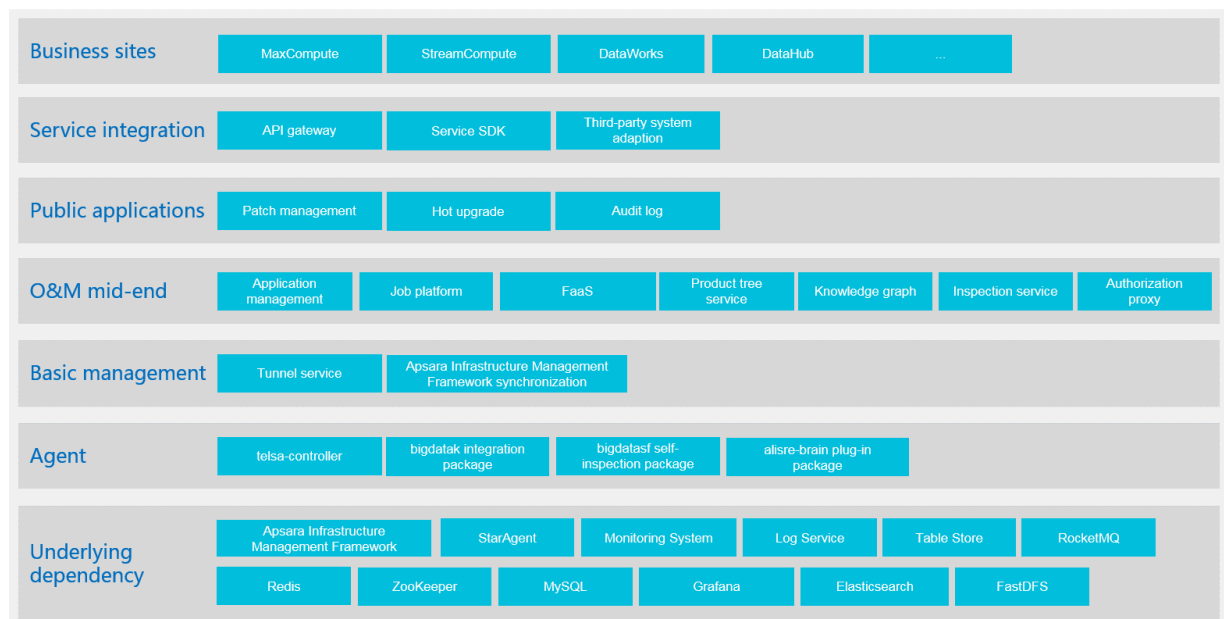
20.3.1 System architecture

This topic describes the system architecture of Apsara Big Data Manager (ABM) and the functions of each component.

ABM uses a microservice architecture that supports data integration, interface integration, and feature integration through a unified platform, and provides standard service interfaces. This architecture enables a consistent user interface, which means that O&M operations are the same for all products. This reduces training costs and lowers O&M risks.

The ABM system consists of the following components: underlying dependency, agent, basic management, O&M mid-end, public applications, service integration, and business sites.

Figure 20-1: Architecture



Underlying dependency

ABM depends on open source systems from Alibaba and third parties.

- Uses StarAgent and Monitoring System of Alibaba to run remote commands and remote data collection instructions.
- Uses ZooKeeper to coordinate primary and secondary services. This guarantees high availability of services.
- Uses RDS to store metadata, Redis to store cache data, and Table Store to store large amounts of self-test data. This improves service throughput.

Agent

The agent provides client SDKs, scripts, and monitoring packages to be deployed on managed servers.

O&M mid-end and basic management

The O&M mid-end and basic management components form the base of the ABM system. Each service in these two components provides different capabilities for business sites. This enables quick construction of business sites and makes the capabilities of each business site complete.

Public applications

Public applications are developed based on the O&M mid-end and designed with special purposes. These applications are adaptive to all big data products supported by ABM.

Service integration

Service integration links business sites with underlying components. It integrates interfaces of all internal services, adapts to various third-party systems, and provides a unified SDK for users.

Business sites


Business sites are built based on the O&M mid-end and cover all big data products, including MaxCompute, StreamCompute, DataWorks, and DataHub. Each business site provides comprehensive O&M capabilities for one product.

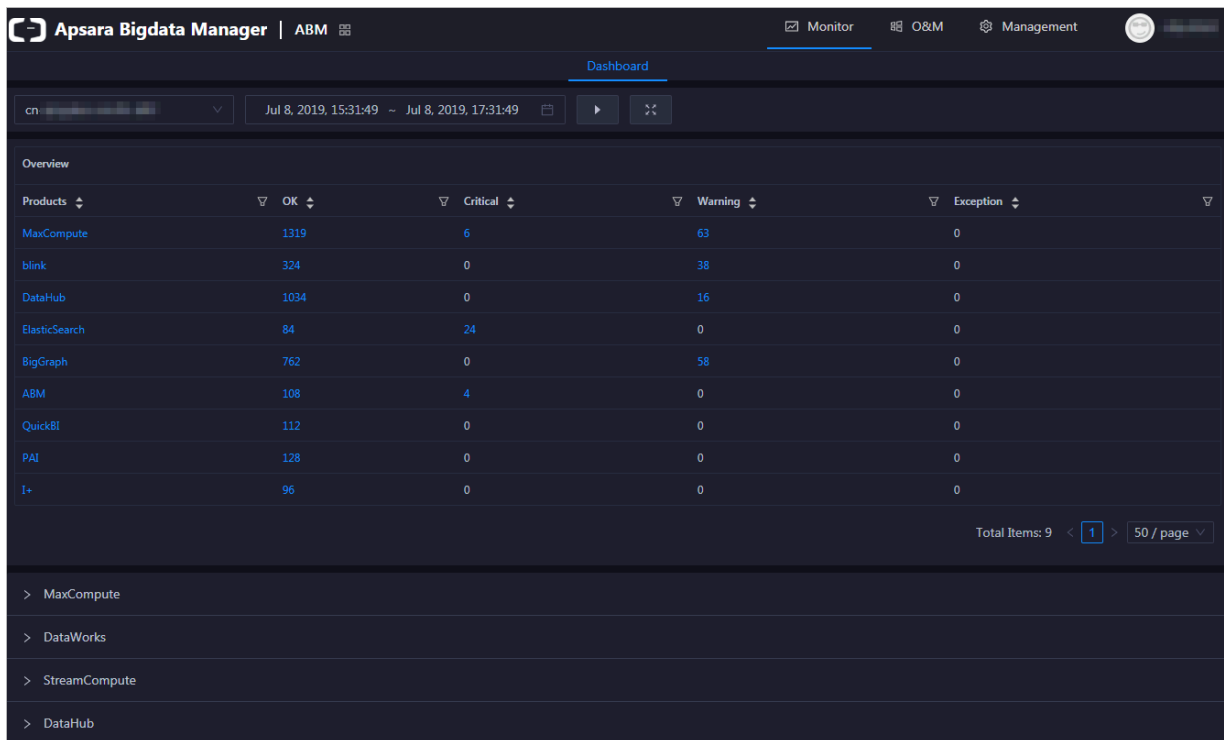
20.4 Features

20.4.1 Dashboard

The Dashboard module is the homepage of the ABM console. It displays key performance metrics for MaxCompute, DataWorks, StreamCompute, and DataHub, and provides alerts for all big data products. This allows you to understand the overall runtime performance of all big data products.

Dashboard page

After you log on to the ABM console, the **Dashboard** page appears by default. To return to the **Dashboard** page from any other page, click  in the upper-left corner and select **ABM**.



Products	OK	Critical	Warning	Exception
MaxCompute	1319	6	63	0
blink	324	0	38	0
DataHub	1034	0	16	0
ElasticSearch	84	24	0	0
BigGraph	762	0	58	0
ABM	108	4	0	0
QuickBI	112	0	0	0
PAI	128	0	0	0
I+	96	0	0	0

Total Items: 9 < 1 > 50 / page

On the **Dashboard** page, you can select a region from the **Dashboard** drop-down list in the upper-left corner. You can then view the runtime performance of big data products in the selected region.

Alerts overview

In the **Overview** section, you can view the number of alerts about each big data product. You need to pay close attention to **Critical** and **Warning** alerts, which must be cleared in a timely manner.

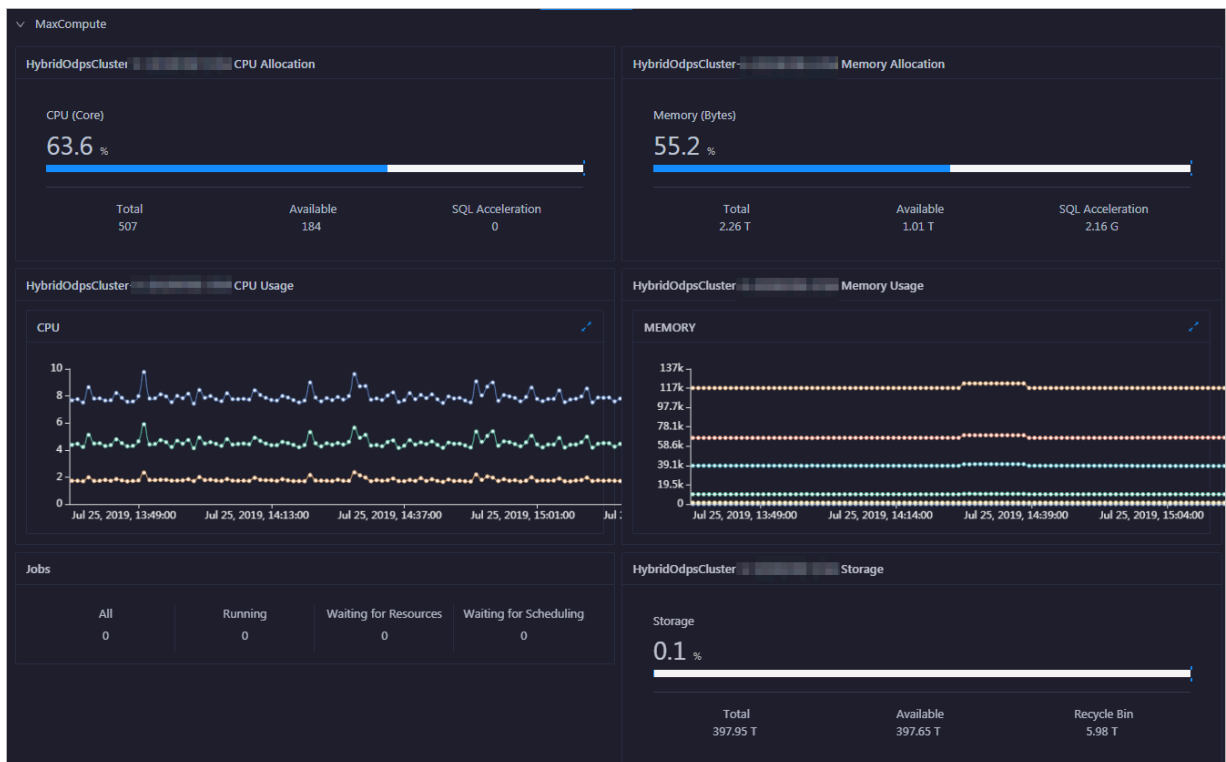
Overview				
Products	OK	Critical	Warning	Exception
MaxCompute	1319	6	63	0
blink	324	0	38	0
DataHub	1034	0	16	0
ElasticSearch	84	24	0	0
BigGraph	762	0	58	0
ABM	108	4	0	0
QuickBI	112	0	0	0
PAI	128	0	0	0
I+	96	0	0	0

Total Items: 9 < 1 > 50 / page

In the **Overview** section, you can click a product name or alert count to go to the O&M page of the product.

MaxCompute metrics

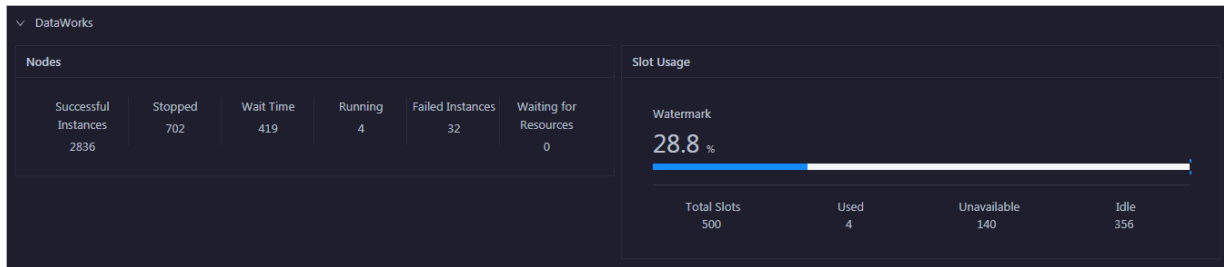
The Dashboard page displays key performance metrics for MaxCompute. To view these metrics, click **MaxCompute** in the Overview section of the **Dashboard** page.



The **MaxCompute** section displays jobs overview, real-time capacity for control system, data traffic, compute resource usage, storage resource usage, and the trends of logical and physical CPU usage.

DataWorks metrics

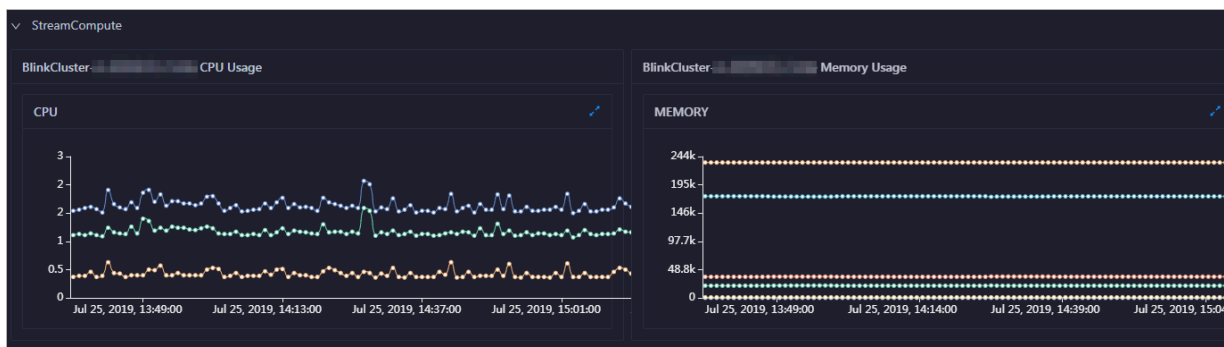
The Dashboard page displays key performance metrics for DataWorks. To view these metrics, click **DataWorks** in the **Overview** section of the **Dashboard** page.



The **DataWorks** section displays nodes overview, slot usage overview, and the trend of finished tasks.

StreamCompute metrics

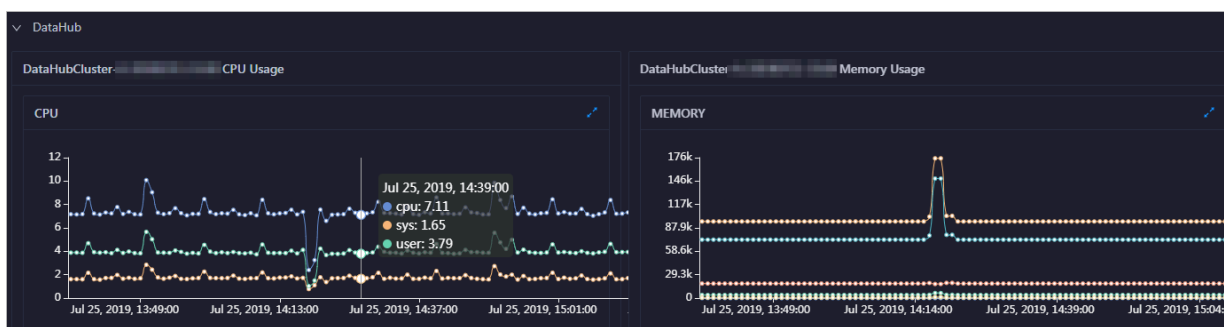
The Dashboard page displays key performance metrics for StreamCompute. To view these metrics, click **StreamCompute** in the **Overview** section of the **Dashboard** page.



The **StreamCompute** section displays the following trends of cluster jobs: TPS, failover rate, CPU usage, and memory usage.

DataHub metrics

The Dashboard page displays key performance metrics for DataHub. To view these metrics, click **DataHub** in the **Overview** section of the **Dashboard** page.



The **DataHub** section displays read/write latency, the numbers of read/write records, read/write request rates, read/write throughputs, and the trends of CPU and memory usage.

20.4.2 Repository


The Repository module displays the resource usage in MaxCompute, DataWorks, and DataHub clusters.

Repository page

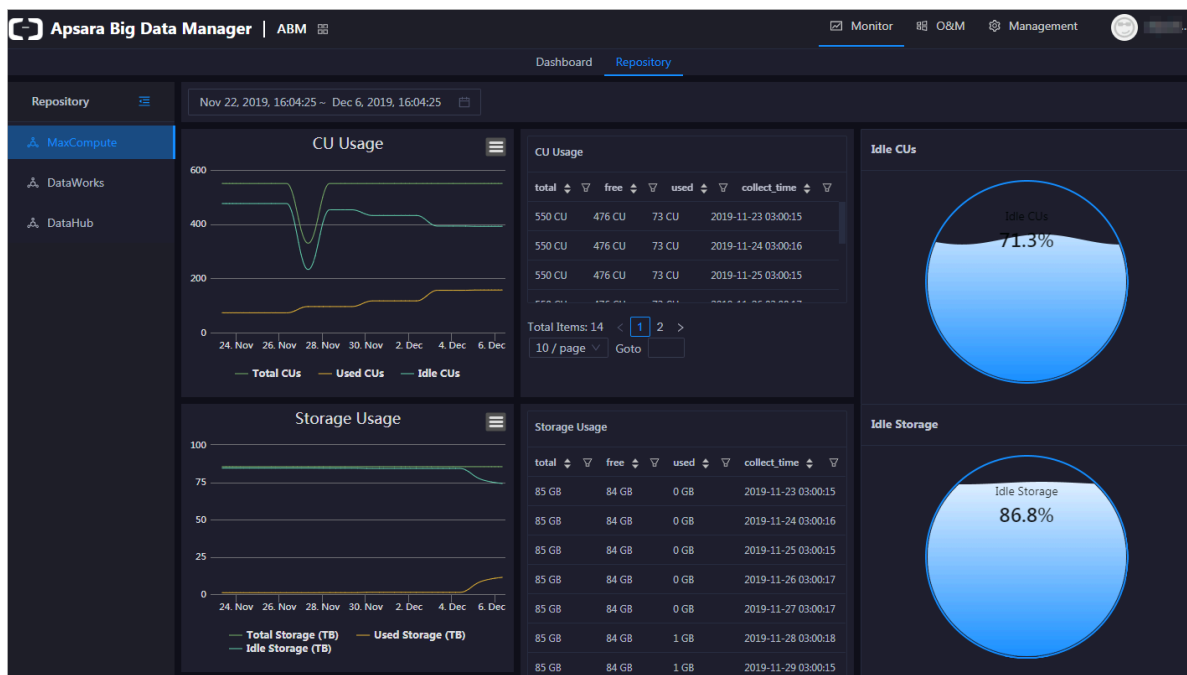
1. Log on to the ABM console. The **Dashboard** page appears by default.



Note:

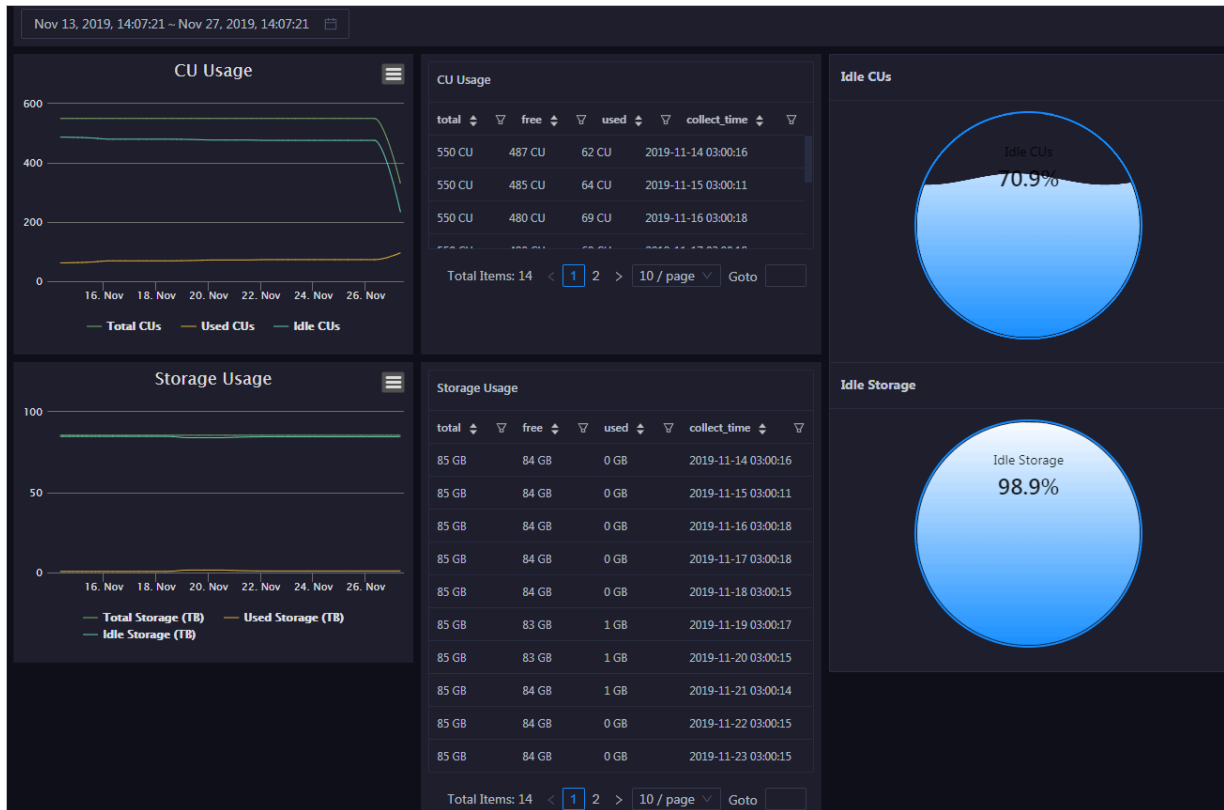
After you log on to the ABM console, the **Dashboard** page appears by default. To return to the **Dashboard** page from any other page, click  in the upper-left corner and select **ABM**.

2. On the **Dashboard** page, click the **Repository** tab to go to the **Repository** page.



MaxCompute repository

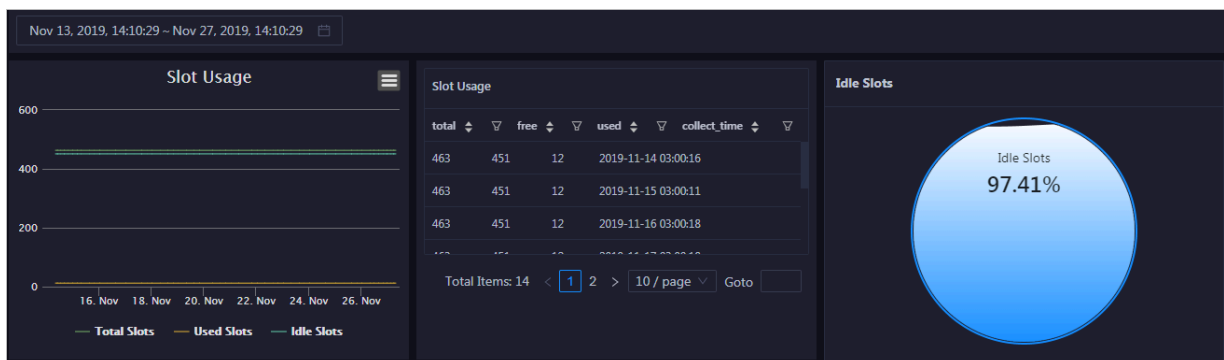
In the left-side navigation pane of the **Repository** page, click **MaxCompute** to view the resource usage in MaxCompute.



This page displays the trends and details of CU and storage usage, and the percentages of idle CUs and storage.

DataWorks repository

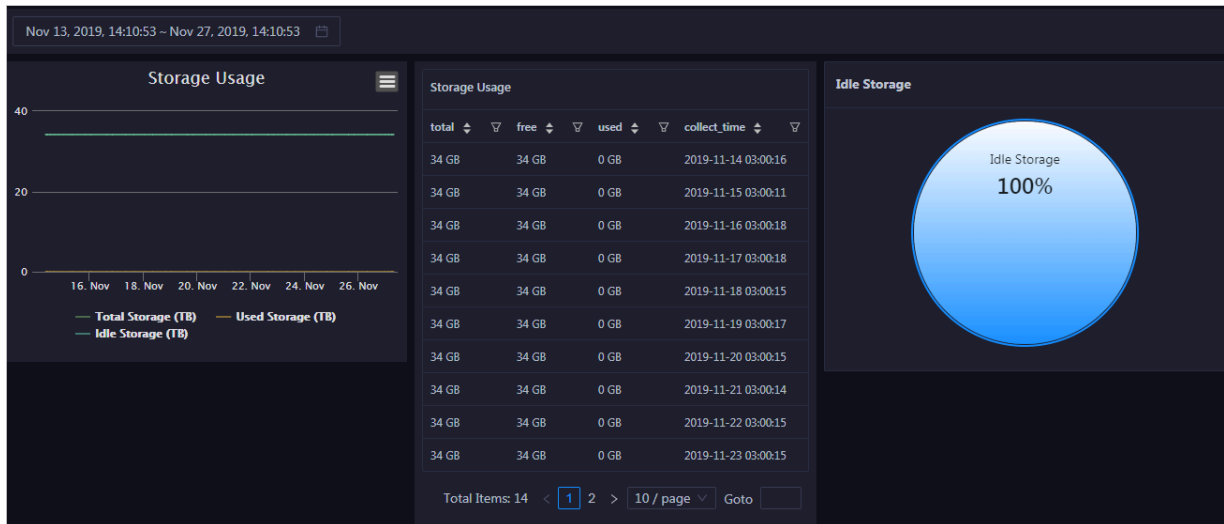
In the left-side navigation pane of the **Repository** page, click **DataWorks** to view the resource usage in DataWorks.



This page displays the trend and details of slot usage, and the percent of idle slots.

DataHub repository

In the left-side navigation pane of the **Repository** page, click **DataHub** to view the resource usage in DataHub.



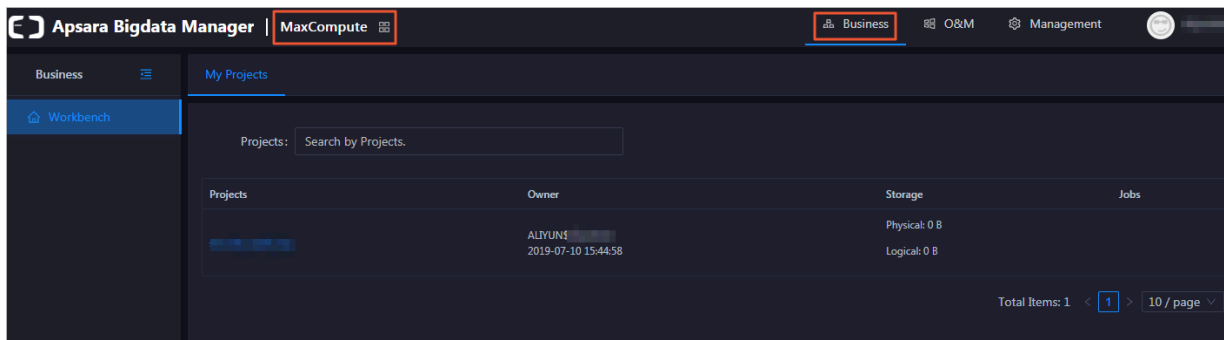
This page displays the trend and details of storage usage, and the percent of idle storage.

20.4.3 Business

The Business module is exclusive for MaxCompute. It provides a workbench where you can view all MaxCompute projects and their details, and configure project parameters.

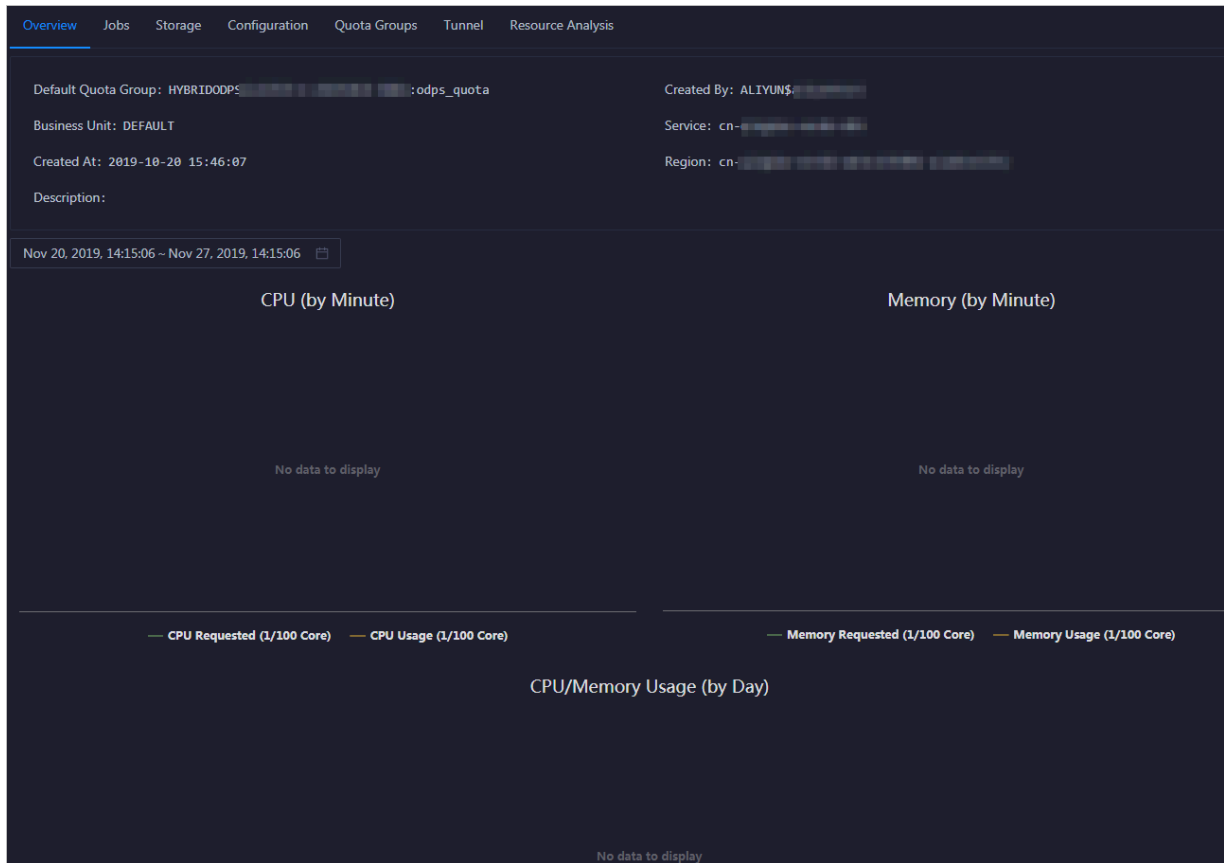
Workbench page

By default, the **My Projects** tab appears when you open the workbench. You can click a project name to view project details.



Project details

Project details include the project overview and information about jobs, storage, configuration, and quota groups.



- Overview page: displays basic information, and CPU and memory usage of the project.
- Jobs page: allows you to view job snapshots in the last week by day and provides log data to help you locate job failures.
- Storage page: displays the storage usage by percentage, including used storage, storage quota, and available storage.
- Configuration page: allows you to configure various properties and the encryption algorithm of the project. These properties include general properties, sandbox, SQL, MapReduce, access control, and resource recycling settings.
- Quota Groups page: displays information about quota groups of the project.
- Tunnel page: displays the tunnel throughputs in bytes/minute, including inbound throughput and outbound throughput.
- Resource Analysis page: allows you to analyze the resources in MaxCompute clusters from different perspectives and monitor the situation of data storage in MaxCompute. You can now analyze resources by table, task, execution time, start time, and engine.

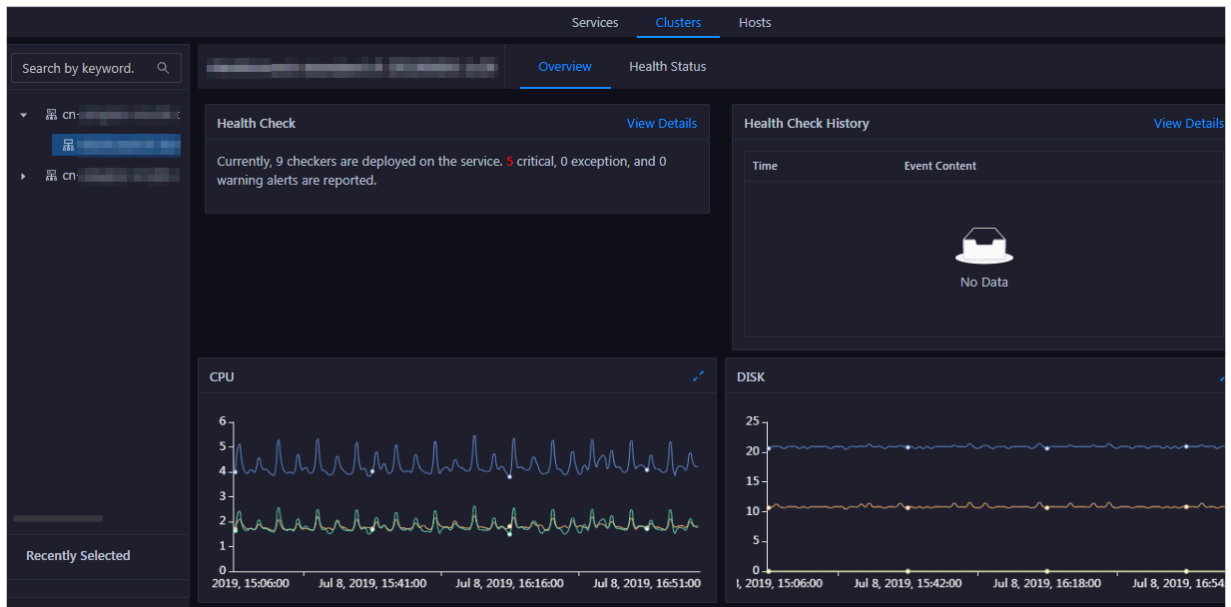
20.4.4 O&M

ABM allows you to perform operations and maintenance on big data products such as MaxCompute, DataWorks, StreamCompute, and DataHub from perspectives such as

business, clusters, services, and hosts. It also provides exclusive custom features for certain products.

Clusters

To help you implement operations and maintenance on clusters, ABM provides two major features: cluster overview and cluster health.



- Overview page: displays the overall performance information about a cluster, including host statuses, service statuses, health check results, and health check history. This page also provides trends of CPU usage, disk usage, memory usage, load, and package counts.
- Health Status page: displays the check results of all checkers in a cluster. Check results are divided into three types: Critical, Warning, and Exception.

In addition to the preceding features, ABM also provides the following custom features for MaxCompute, StreamCompute, and DataHub:

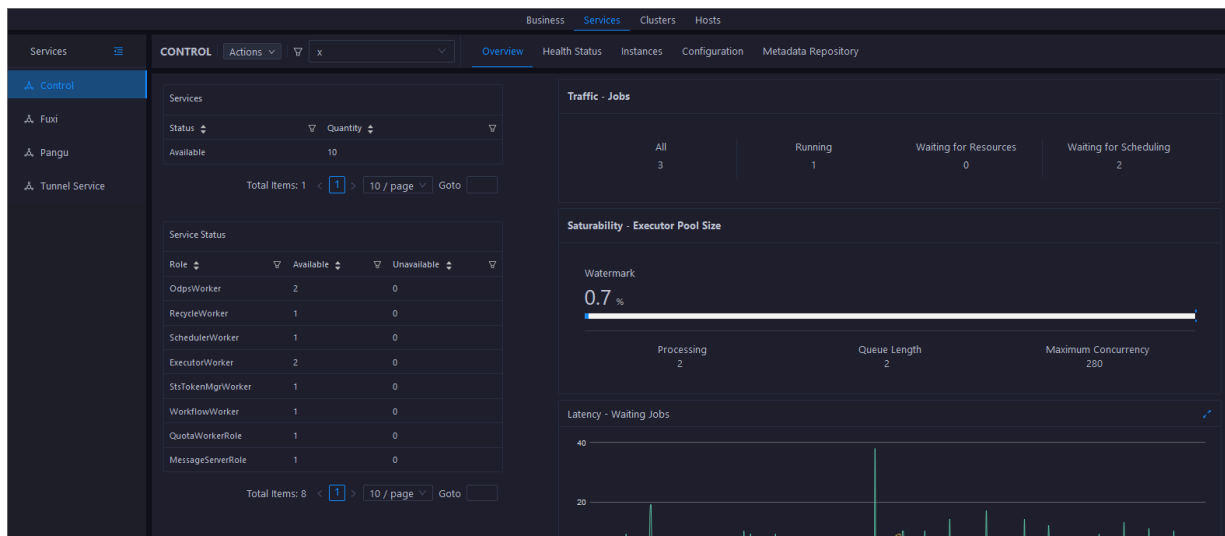
- Servers page: displays information about all servers in a cluster, including CPU usage, memory usage, root disk usage, packet loss rate, and packet error rate.
- Scale in Cluster and Scale out Cluster actions: allow you to scale in or scale out a cluster by removing or adding physical servers.
- Reverse Parse Request ID action (exclusive for DataHub): allows you to reverse parse RequestId in DataHub to obtain the time when a job was run and the IP address of the host. This enables you to query logs for troubleshooting.

- Delete Topic from Smoke Testing action (exclusive for DataHub): allows you to delete topics from a DataHub test project and view execution history.

Services

To help you implement operations and maintenance on services, ABM provides exclusive features for MaxCompute, DataWorks, DataHub, and StreamCompute.

For MaxCompute, ABM supports operations and maintenance on the following services: control service, Job Scheduler, Apsara Distributed File System, and tunnel service.

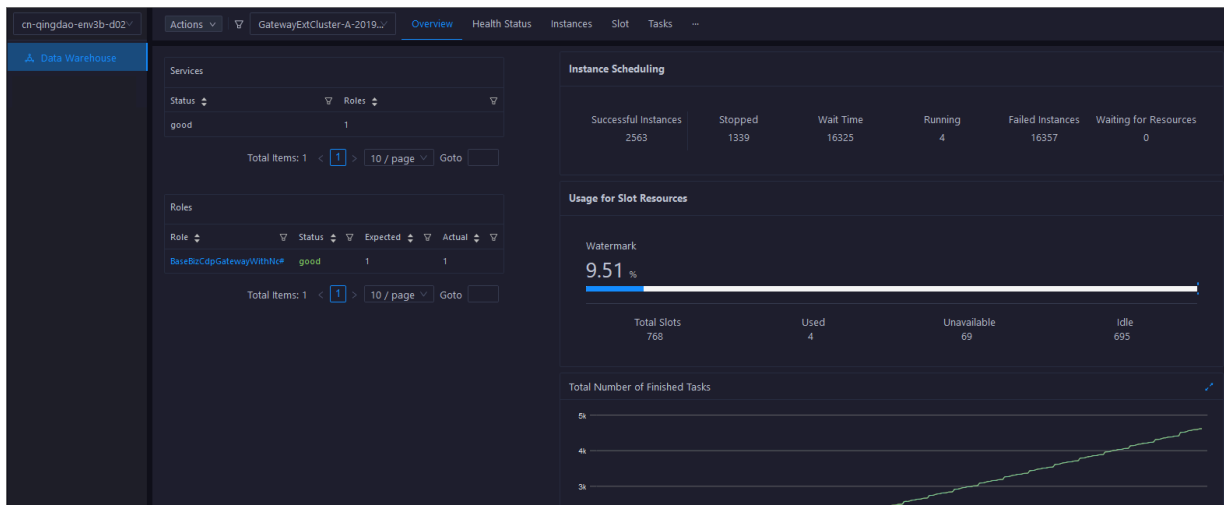


- Control Service page: displays general information about services, health checkers and instances. You can also configure services for clusters and enable or disable service roles.
- Job Scheduler page: displays general information about Job Scheduler, health checkers, and instances. You can also manage quota groups, set compute nodes to read-only or read-write, add or remove compute nodes to or from the blacklist, enable or disable SQL acceleration, and restart master nodes.
- Apsara Distributed File System page: displays general information about Apsara Distributed File System, health checkers, and instances. You can also set storage node status to disabled or normal, set disk status to error or normal, change the primary master node, empty the recycle bin, enable or disable data rebalancing, and run checkpoints on master nodes.
- Tunnel Service page: displays general information about the tunnel and instances. You can also restart the tunnel server.

For DataWorks, ABM supports operations and maintenance on DataWorks services.

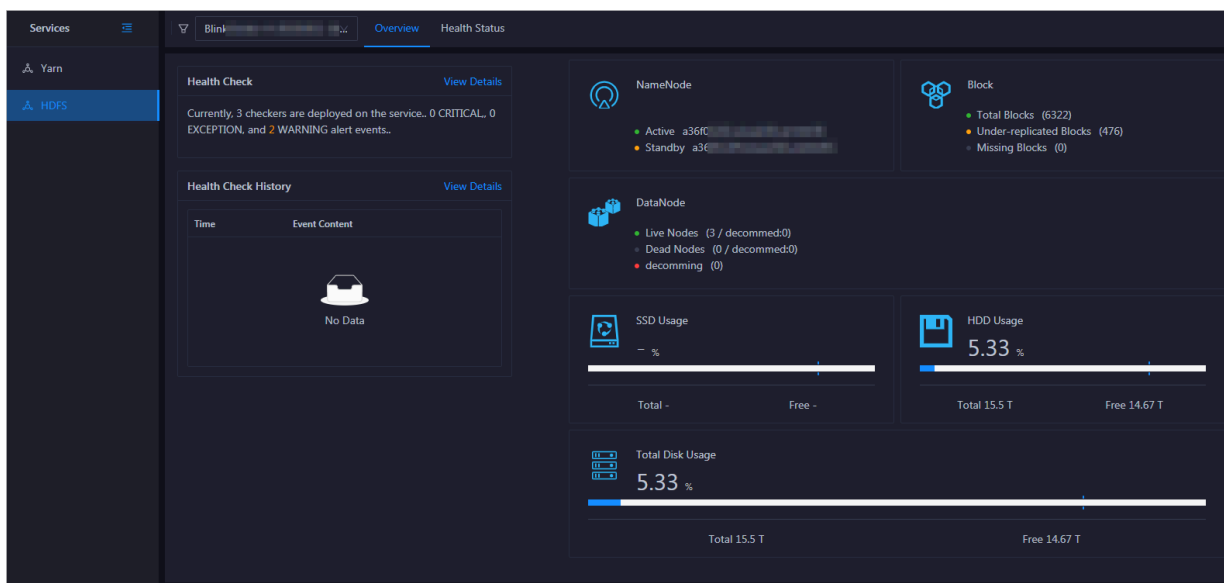
DataWorks page: displays general information about services, health statuses, instances,

slots, and service configurations. You can also add or remove servers to scale out or in the cluster.



For DataHub, ABM supports operations and maintenance on control service, Job Scheduler, and Apsara Distributed File System, which are the same as those for MaxCompute.

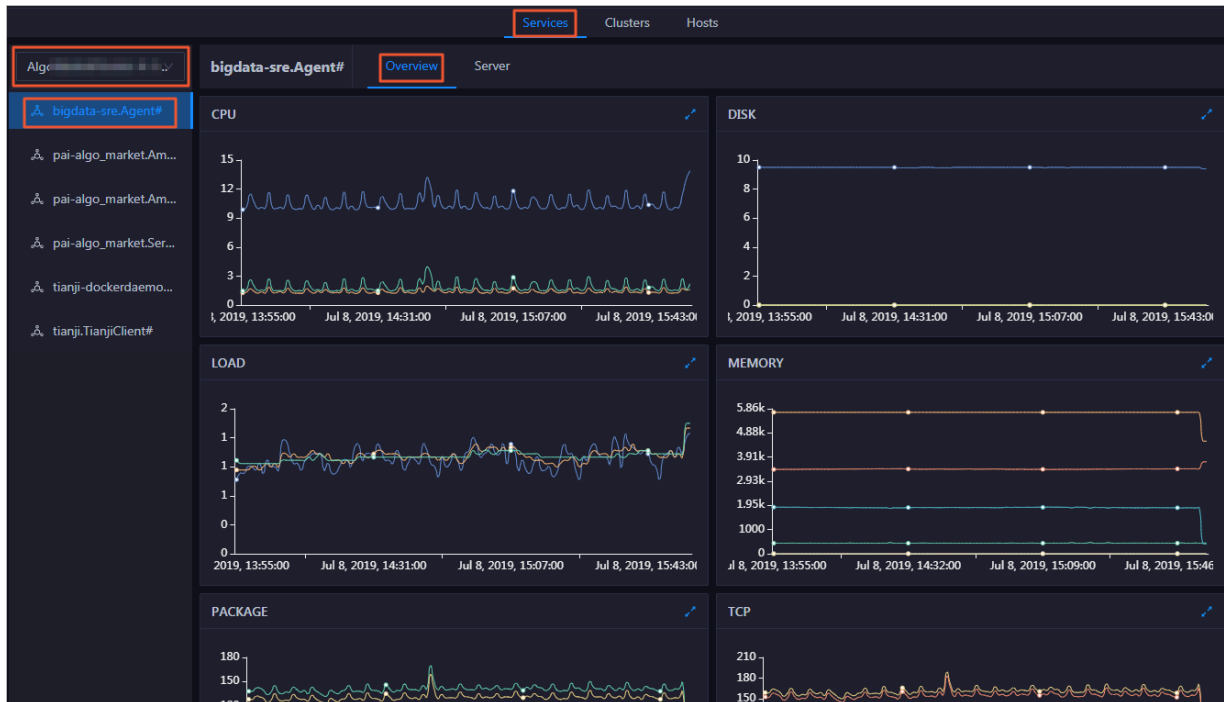
For StreamCompute, ABM supports operations and maintenance on Blink, Yet Another Resource Negotiator (YARN) and Hadoop Distributed File System (HDFS).



- Blink page: displays information about Blink services, service statuses, health check results, health check history, and core metrics such as TPS and Failover Rate.
- YARN page: displays information about YARN services, checkers, health check results, health check history, applications, containers, nodes, logical CPU usage, and logical memory usage.

- HDFS page: displays information about HDFS services, checkers, health check results, health check history, NameNode, blocks, DataNode, SSD usage, HDD usage, and total disk usage.

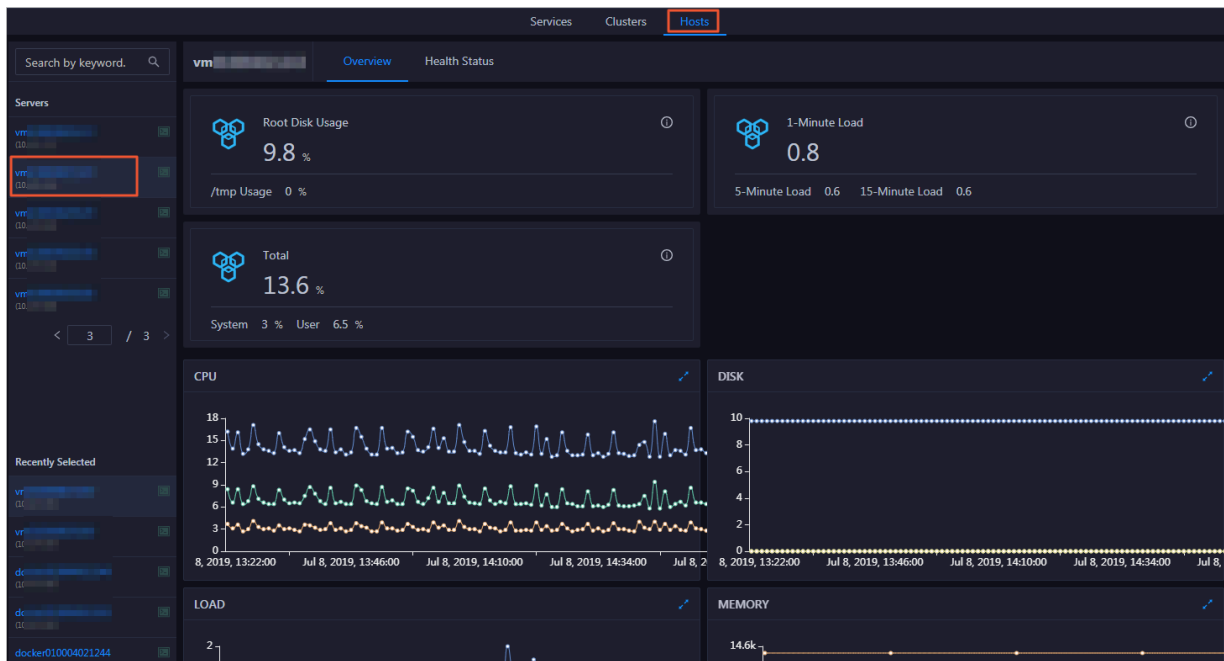
For other big data products, ABM provides information about all service roles in the cluster and the resource usage trend of each service role.



You can select a service from the left-side service list and view the trends of CPU usage, disk usage, memory usage, load, package count, TCP connections, and root disk usage on the right.

Hosts

To help you implement operations and maintenance on hosts, ABM provides two major features: host overview and health status.



- Overview page: displays general information about hosts in a cluster, including host details, service role statuses, health check results, health check history, and trends of CPU usage, memory usage, storage usage, load, and packet loss rate.
- Health Status page: displays the check results of all checkers on a host. Check results are divided into three types: Critical, Warning, and Exception.

In addition to the preceding features, ABM provides the following custom features for MaxCompute, DataHub, and StreamCompute:

- Charts page: displays trends of CPU usage, memory usage, storage usage, load, and package counts in enlarged charts. These trends are the same as the ones displayed on the host overview page.
- Services page: displays the cluster where a host belongs, the services running on the host, and the roles of the services.

Business

To help you implement operations and maintenance on business, ABM provides exclusive custom features for MaxCompute, StreamCompute, Elasticsearch, and DataHub. For MaxCompute, ABM supports the following features: project management, job management, and business optimization.

- Project management
 - Projects page: displays information about all projects in a MaxCompute cluster. You can filter, query, and sort projects based on needs and modify the quota group of a

project. If disaster recovery is enabled, you can set resource replication parameters and enable or disable resource replication for a project.

- Authorize Package for Metadata Repository: grants project members access to metadata repositories in MaxCompute.
- Encryption at Rest: sets whether to encrypt data in MaxCompute projects.
- Disaster Recovery: allows you to view cluster status when disaster recovery is enabled . You can enable switchover between primary and secondary clusters, and enable or disable scheduled tasks to synchronize resources between clusters.
- Job Snapshots: displays information about all jobs in a MaxCompute cluster. You can filter and query jobs based on needs. You can also view job logs, terminate a running job, and collect job logs.
- Business Optimization: provides features such as file merging, file archiving, and resource analysis.

For DataHub, ABM provides information about projects and topics in DataHub clusters.

- Projects page: displays all projects and project details, which include the project overview and related topics.
- Topics page: displays all topics and topic details. Topic details include the topic overview and information about monitoring metrics, shards, subscriptions, DataConnectors, and schemas.

For StreamCompute, ABM provides information about projects, jobs, and queues in StreamCompute clusters.

- Projects page: displays all projects.
- Jobs page: displays all jobs and allows you to diagnose jobs to troubleshoot issues.
- Queues page: displays all queues, queue resources, and tasks under a queue to help you perform analysis on the queue.

20.4.5 Management

The Management module allows you to manage configurations in a comprehensive manner. This module supports features such as job management, package management, hot upgrades, health management, and operation audit.

Jobs

ABM executes jobs to perform O&M operations on big data products. ; Jobs are divided into two types: cron jobs and ordinary jobs. Cron jobs are either automatically executed based on schedule or manually executed. Ordinary jobs are all manually executed.

ABM offers multiple O&M schemes to meet the needs of most scenarios. A scheme is a job template. You can easily generate and execute jobs by using schemes.

Apart from schemes, ABM also provides an atom library that contains most common O&M operations. An atom is a template of an atomic step. When you generate a job from a scheme, you can directly use atoms as job steps.

Packages

The Packages page allows you to apply patches to Docker containers of big data products . Docker is an application container engine. It allows you to quickly upgrade product software by replacing only files that need to be updated.

Hot Upgrades

The Hot Upgrades page allows you to upgrade monitoring configurations and checkers without interrupting services.

Health Management

ABM provides a wide variety of built-in checkers for each big data product. These checkers are used to check product faults and send alerts, making it easy to detect and fix faults in time.

Scheduling: Checkers run scheduling scripts on servers of specific Tianji roles and generate raw alert data, which provides information about the checker, server, alert severity, and alert content. Raw alert data is stored in ABM database.

Monitoring: Checkers can mount raw alert data on the corresponding product page for monitoring. You can configure filtering policies to display alerts about checkers that are of high priority.

ABM allows you to customize the execution interval, parameters, and mount point of a checker, and enable or disable a checker.

Operation Audit

The Operation Audit page records O&M history and the details of all O&M operations. This allows you to track and locate faults in the future.

20.5 Scenarios

If you are using Apsara Stack and have deployed one or more big data products, you need to use Apsara Big Data Manager (ABM) to perform O&M operations on these big data products.

Apsara Stack Enterprise and big data products

If you are using Apsara Stack Enterprise and have deployed one or more big data products , such as MaxCompute and DataWorks, you need to use Apsara Big Data Manager (ABM) to perform O&M operations on these big data products.

20.6 Limits

None.

20.7 Concepts

This topic describes basic concepts of ABM.

Product

A group of clusters. A product provides services for users.

Cluster

A group of physical hosts. A cluster provides services logically and is used to deploy software of a product. A cluster belongs to only one product. You can deploy multiple services on a cluster.

Service

A group of software used to provide an independent feature. A service contains one or more service roles. You can deploy a service on multiple clusters.

Service role

One or multiple indivisible function units of a service. A service role contains one or more applications. If you deploy a service on a cluster, you must deploy all service roles of the service on hosts in the cluster.

Service role instance

A service role on a specific host. A service role can be deployed on multiple hosts. The service role on a specific host is called a service role instance.

Application

A software entity, which is the minimum unit for starting software. Generally, an application is an executable file or a Docker container. If you deploy a service role on a host, you must deploy all applications of the service role on the host.

Service tree

The overall organizational structure of a product. Each product is an independent entity consisting of a certain number of services. The hierarchy of a product's services forms a service tree.

Workflow

A packaged framework that consists of a sequence of processes predetermined based on specific rules. A workflow supports automatic execution. You can use workflows to perform repetitive tasks.

Job

A product O&M task created by users.

Atom

A template of an atomic step. Atoms can be used to create jobs.

Atomic step

An atom that is directly included as a step when you use schemes to create jobs.

Scheme

A job template. You can use schemes to create jobs.