Alibaba Cloud
Apsara Stack Enterprise

Product Introduction

Version: 1808..

Issue: 20180831

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 disseminat ed 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, adjustment s, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies . However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.
5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified,

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

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

Generic conventions

Table -1: Style conventions

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

Contents

1 Apsara Stack overview

1.1 What is Apsara Stack

Private cloud

The private cloud is a cloud computing system built within enterprises by cloud computing service providers. It places cloud infrastructures and software and hardware resources within firewalls to allow departments within an organization or 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 to public clouds, private clouds provide better privacy and exclusivity.

Private clouds are divided into two types by the sizes of enterprises or business requirements:

- Multi-tenant comprehensive private clouds for industries and large groups: A full stack cloud system created in a top-down manner to run hyper-scale digital applications. It satisfies IT requirements, such as continuous integration and development of DevOps applications, and operation support of production environments.
- Single-tenant basic private clouds 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 can also perform local computing tasks.

Apsara Stack

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

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

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 resources to minimize costs and maximize performance and stability.
- **Agile**

  Integrates business with 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 characteristics

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

- 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.
- Centralized access management: Isolates permissions of different roles for easy subsequent O &M management.

Development history

Apsara Stack has evolved to V3.0 since V1.0 in July 2015. Apsara Stack is developing towards a more open, reliable, and controllable full stack cloud platform for enterprises, and continually brings service values to users.

Table 1-1: Development history of Apsara Stack

Version	Release date	Related content
V1.0	Jul-2015	• Scenario: Big data • Key products: IAAS + basic big data products • Platform features: Semi-automated deployment and ERMS
V2.0	May-2016	• Scenarios: Internet-based shared service + big data • Key products: IAAS + big data + Aliware middleware • Platform features: Centralized O&M management, reduced number of management nodes, and effective control over the output server volumes
V3.0	Jul-2017	• Scenario: Internet Finance (big data involved) • Key products: IAAS + big data products + middleware (including Docker) + Alibaba Cloud Security Advanced Edition • Platform features: Level-3 security protection and three centers in two locations

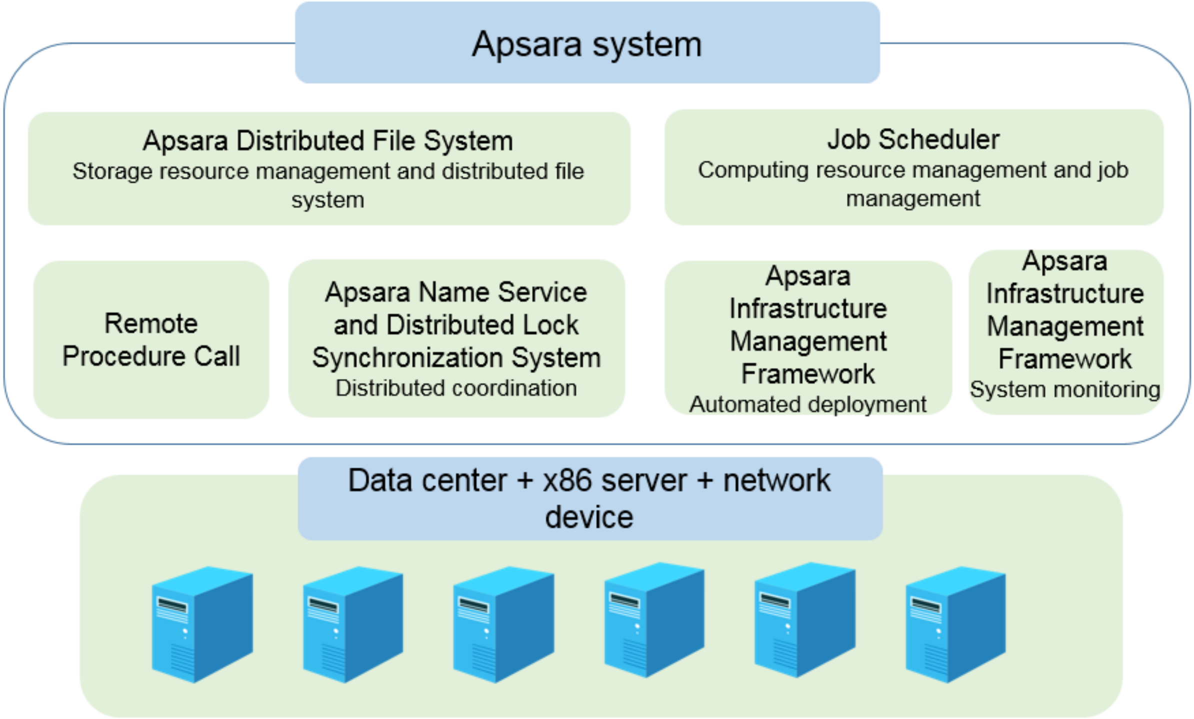
1.2 Why Apsara Stack

As the only large-scale commercial private cloud platform based on the Cloud Native architecture, Apsara Stack provides the following benefits.

1.2.1 Hyper-scale distributed cloud operating system

Apsara Stack is based on the same underlying architecture (Apsara's large-scale distributed computing system kernel) as Alibaba Cloud public clouds. 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, providing 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


Apsara system
Apsara Distributed File System
Storage resource management and distributed file system
Job Scheduler
Computing resource management and job management
Remote Procedure Call
Apsara Name Service and Distributed Lock Synchronization System
Distributed coordination
Apsara Infrastructure Management Framework
Automated deployment
Apsara Infrastructure Management Framework
System monitoring
Data center + x86 server + network device

The modules of the Apsara system 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 services. 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, and automatically protect against software and hardware faults to guarantee uninterrupted data access. With the support for incremental expansion and automatic data balancing, the modules provide APIs, which are similar to portable operating system interfaces (POSIXs), 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 both online services that rely heavily on the response speed and offline jobs that require high data processing throughput. The modules automatically detect faults and hot spots in systems, and guarantee 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 status of clusters, and the running status and performance metrics of upper-layer application services to send alarm notifications and record exception events. The modules enable the O&M 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

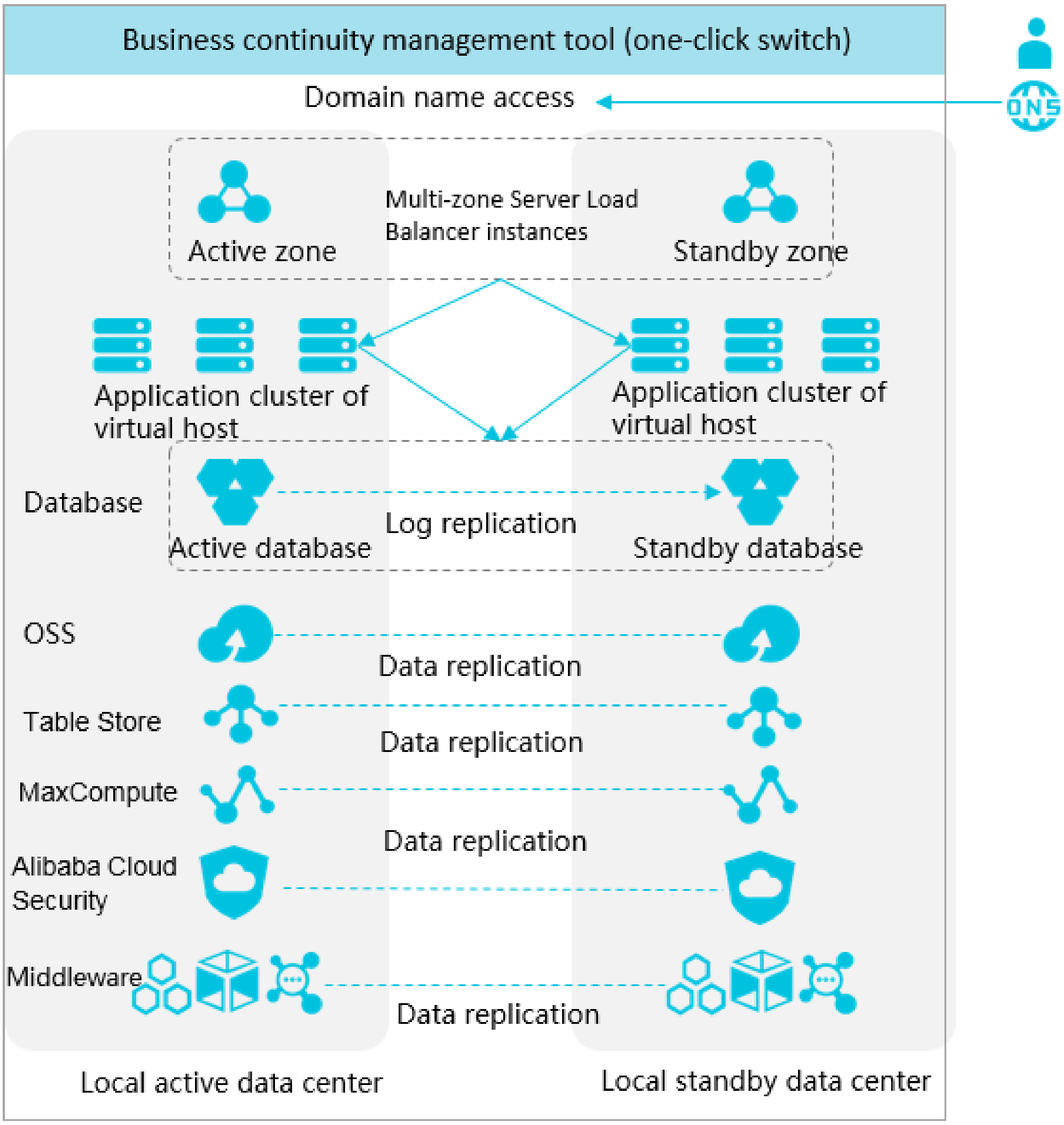
Apsara Infrastructure Management Framework provides the centralized deployment, authentication, authorization, and control for cloud products, and provides the cloud services with basic support. Apsara Infrastructure Management Framework contains deployment framework, resource library, metadatabase, Alibaba Cloud Security, authentication and authorization component, interface gateway, Log Service, and control module.

- The deployment framework provides all cloud services with unified access platform deployment and a management function that handles the dependencies between services.
- The resource library stores the execution files of all cloud services and their dependent components.
- Alibaba Cloud Security protects cloud services from Web attacks.
- The authentication and authorization component provides the access control capabilities for cloud services and isolates 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 module monitors the basic health of cloud services and supports the O&M system of the cloud platform.

1.2.3 High-reliability disaster recovery solutions

Apsara Stack disaster recovery solutions are designed and developed based on Alibaba Cloud's own cloud computing capabilities. They comply with common international disaster recovery standards. The standby data center must be within a 50-kilometer radius from the active data center in the same city, with a network latency of less than 0.6ms. The Apsara Stack platform deploys the network access layer and user application layer in active-active mode and the data persistence layer in active-standby mode.

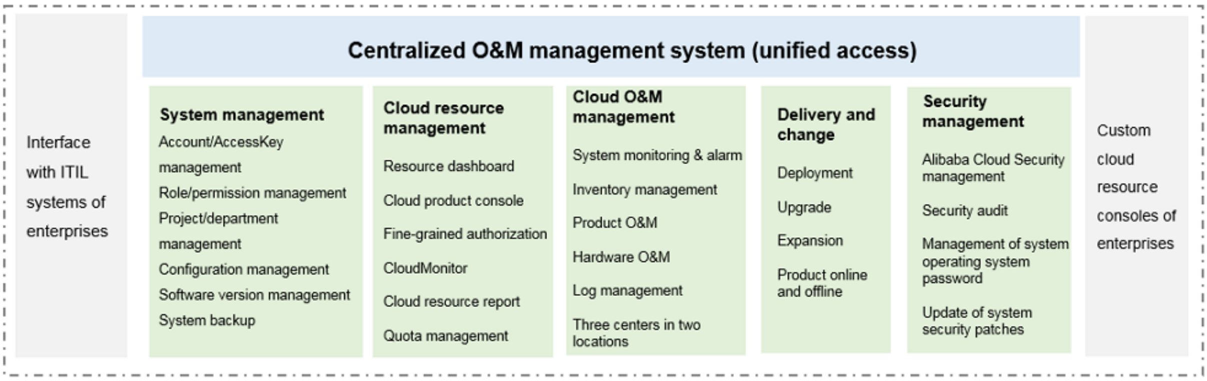
Figure 1-2: Local disaster recovery



1.2.4 Centralized O&M management and automated O&M capabilities

Apsara Stack provides a centralized O&M management system to configure different management permissions for user roles. Users can gain access to O&M management capabilities by using APIs and customize their own cloud resource consoles. To interoperate and integrate with existing IT systems of various enterprises synchronously, Apsara Stack can interface with information technology infrastructure library (ITIL) systems of enterprises.

Figure 1-3: Centralized O&M management


Centralized O&M management system (unified access)
Interface with ITIL systems of enterprises
System management
Account/AccessKey management
Role/permission management
Project/department management
Configuration management
Software version management
System backup
Cloud resource management
Resource dashboard
Cloud product console
Fine-grained authorization
CloudMonitor
Cloud resource report
Quota management
Cloud O&M management
System monitoring & alarm
Inventory management
Product O&M
Hardware O&M
Log management
Three centers in two locations
Delivery and change
Deployment
Upgrade
Expansion
Product online and offline
Security management
Alibaba Cloud Security management
Security audit
Management of system operating system password
Update of system security patches
Custom cloud resource consoles of enterprises

1.2.5 Open cloud service interface

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

1.3 Product architecture

1.3.1 Types of private cloud architectures

Private cloud has two types of architectures: native cloud architecture and integrated cloud architecture.

- **Native cloud architecture**

  The native cloud architecture evolved 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 expanded 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 has been open-sourced by the OpenStack and become the mainstream private cloud architecture.

Apsara Stack uses the native cloud architecture based on Alibaba Cloud's self-developed distributed technologies and products. A 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 range of disaster recovery and backup capabilities, and full autonomous control.

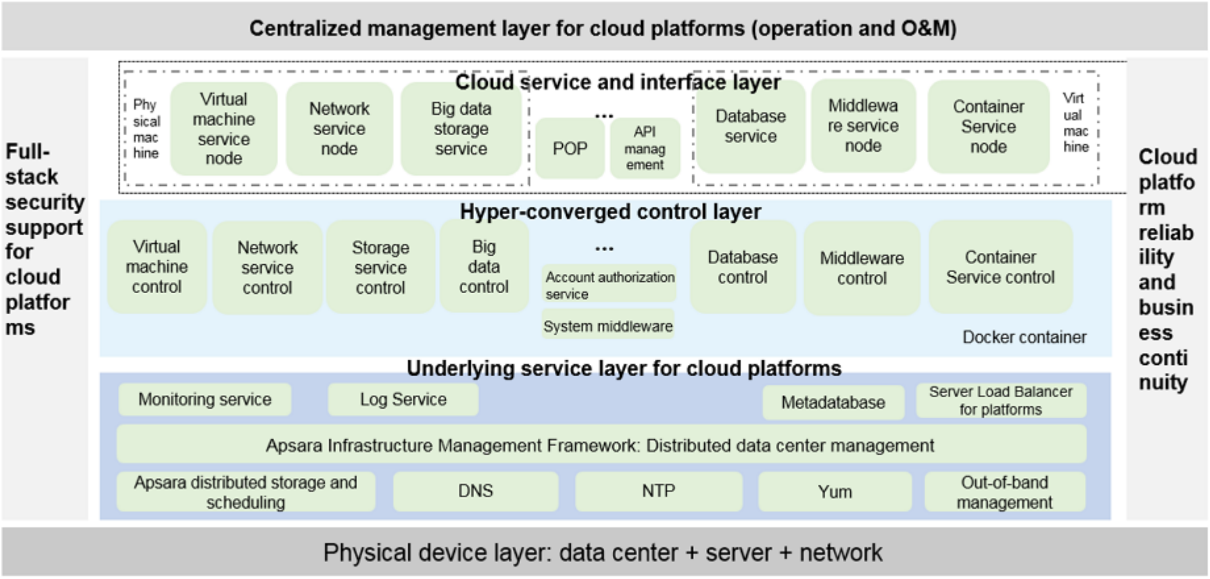
1.3.2 System architecture

The Apsara Stack system architecture consists of the following components, as shown in *Figure 1-4: Apsara Stack system architecture*:

- Physical device layer: Includes physical data centers and hardware devices, such as servers and network devices.
- Underlying service layer for cloud platforms: Provides underlying services for upper-layer applications based on the underlying physical environment.
- Hyper-converged control layer: Schedules upper-layer applications or services uniformly based on the hyper-converged control architecture.
- Cloud service and interface layer: Provides centralized management and O&M for virtual machines and physical machines on converged service nodes, and uses the open API platform to unify the interfaces and support custom development.
- Centralized management layer for cloud platforms: Provides centralized operation and O&M management.

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

Figure 1-4: Apsara Stack system architecture

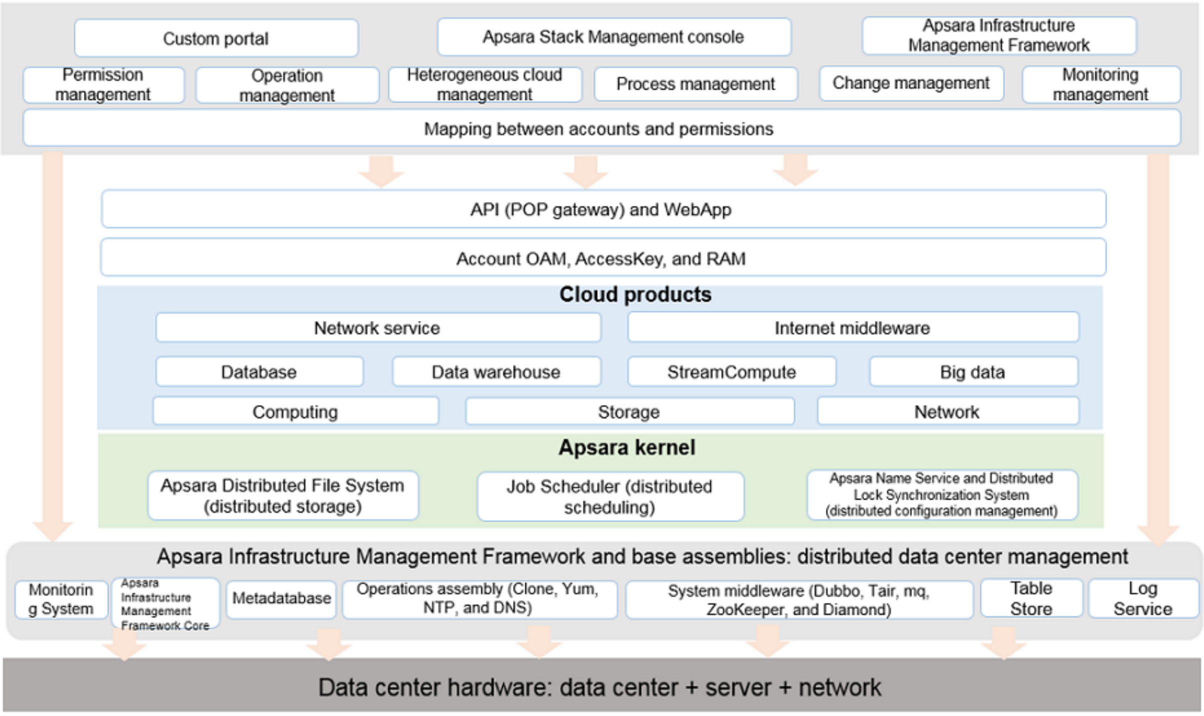

Centralized management layer for cloud platforms (operation and O&M)
Full-stack security support for cloud platforms
Cloud service and interface layer
Physical machine
Virtual machine service node
Network service node
Big data storage service
...
POP
API management
Database service
Middleware service node
Container Service node
Virtual machine
Cloud platform reliability and business continuity
Hyper-converged control layer
Virtual machine control
Network service control
Storage service control
Big data control
...
Account authorization service
System middleware
Database control
Middleware control
Container Service control
Docker container
Underlying service layer for cloud platforms
Monitoring service
Log Service
Metadatabase
Server Load Balancer for platforms
Apsara Infrastructure Management Framework: Distributed data center management
Apsara distributed storage and scheduling
DNS
NTP
Yum
Out-of-band management
Physical device layer: data center + server + network

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 users' applications, and can be interconnected with users' existing account systems and monitoring O&M 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 offer various cloud products.
- All cloud products are required to follow a uniform API framework, management and O&M system (accounts, authorization, monitoring, and logs), and security system.
- Make sure that all cloud products have a consistent user experience.

Figure 1-5: Apsara Stack logical architecture


Custom portal
Apsara Stack Management console
Apsara Infrastructure Management Framework
Permission management
Operation management
Heterogeneous cloud management
Process management
Change management
Monitoring management
Mapping between accounts and permissions
API (POP gateway) and WebApp
Account OAM, AccessKey, and RAM
Cloud products
Network service
Internet middleware
Database
Data warehouse
StreamCompute
Big data
Computing
Storage
Network
Apsara kernel
Apsara Distributed File System (distributed storage)
Job Scheduler (distributed scheduling)
Apsara Name Service and Distributed Lock Synchronization System (distributed configuration management)
Apsara Infrastructure Management Framework and base assemblies: distributed data center management
Monitorin g System
Apsara Infrastructure Management Framework Core
Metadatabase
Operations assembly (Clone, Yum, NTP, and DNS)
System middleware (Dubbo, Tair, mq, ZooKeeper, and Diamond)
Table Store
Log Service
Data center hardware: data center + server + network

1.3.3 Network architecture

The Apsara Stack network architecture defines two logical areas, the business service area and the integrated access area, as shown in *Figure 1-6: Logical areas*.

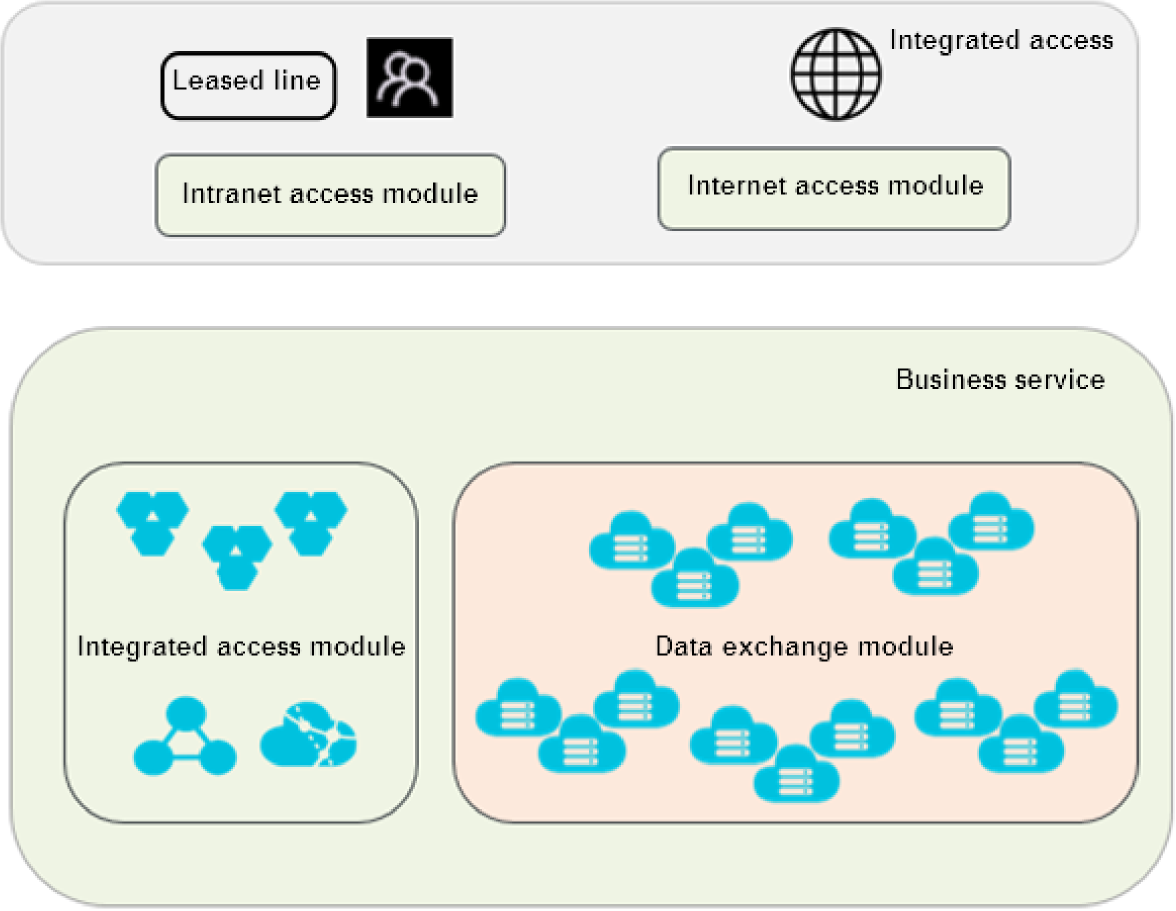
- **Business service area**

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

- **Integrated access area**

  Customize this area based on the actual deployment requirements. As an extension of the business service area, the integrated access area provides a channel for user-managed networks, private networks, and the Internet to access Apsara Stack networks.

Figure 1-6: Logical areas



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

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

Role	Module	Purpose
Customer switch (CSW)	Intranet access module	CSW facilitates the access to users' internal backbone networks. It performs route distribution and interaction between the inside and outside of cloud networks, including access to VPC instances by using leased lines.
Distributed switch (DSW)	Data exchange module	DSW functions as a core switch to connect all access switches.
Access switch (ASW)	Data exchange module	ASW provides access to cloud servers and is uplinked with the core switch DSW.
Integrated access switch (LSW)	Integrated access module	LSW provides access to cloud products, such as VPC and Server Load Balancer.

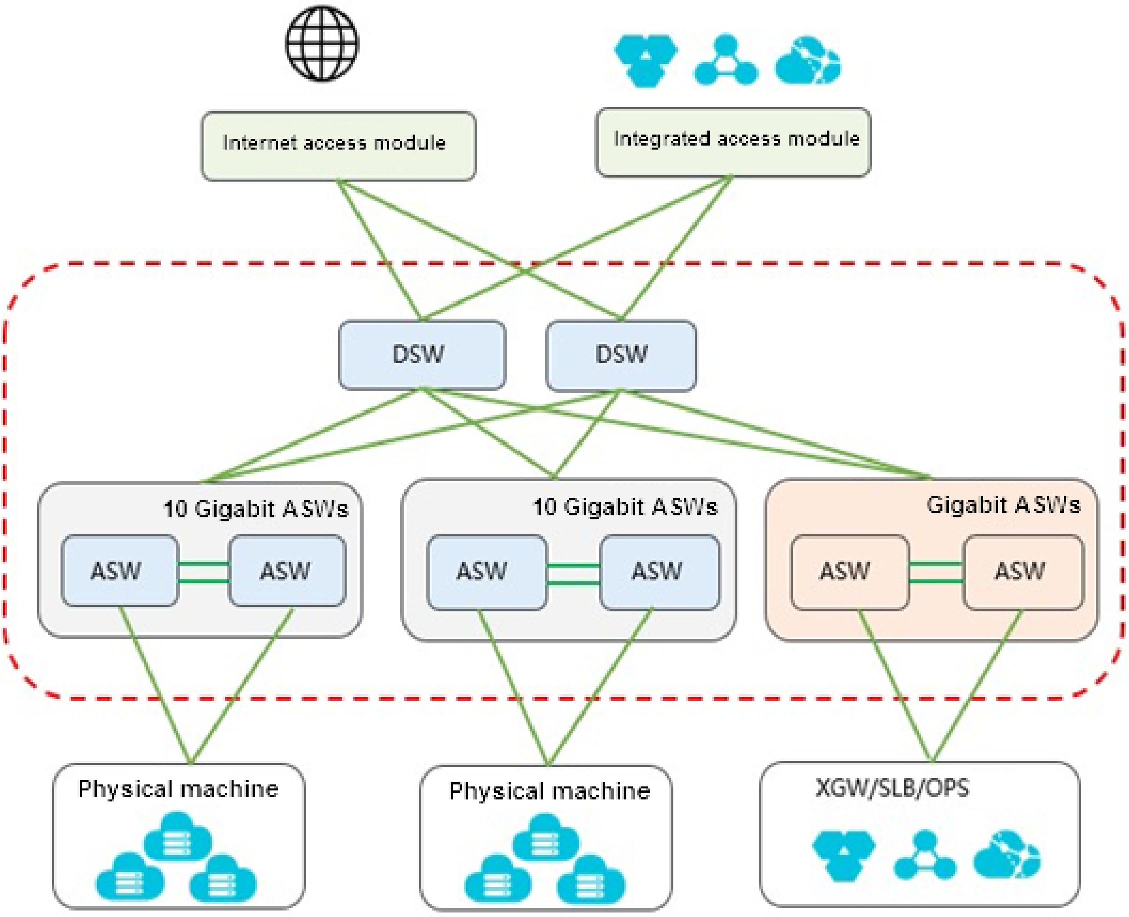
1.3.3.1 Business service area

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

- **Data exchange module**

  The data exchange module has a layer-2 CLOS architecture that is comprised of DSWs and ASWs. Each ASW pair forms a stack as a leaf node. According to network sizes, this node selects 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 uses EBGP to interact with other modules, receives the Internet routes from ISWs, and releases the address segments of cloud products to the ISWs.

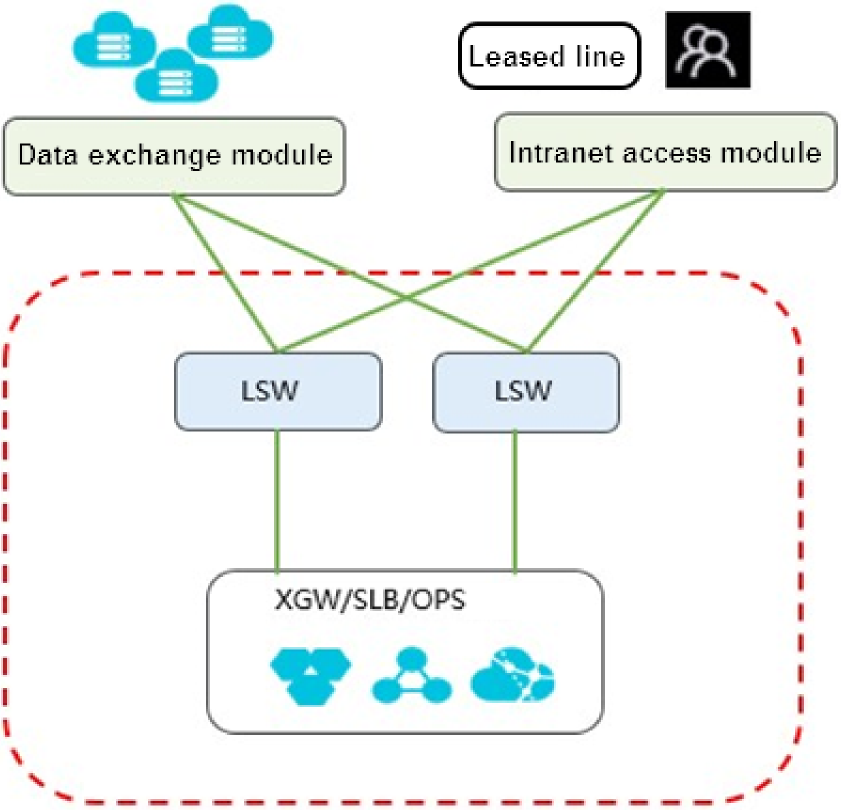
Figure 1-7: Data exchange module



- **— Integrated service module**

  Each cloud service server (XGW/SLB/OPS) is connected to two LSWs. These servers use Open Shortest Path First (OSPF) to exchange routing information. The two LSWs use Internal Border Gateway Protocol (IBGP) to exchange routing information between each other, and use EGBP to exchange routing information with DSWs and CSWs.

Figure 1-8: Integrated service module



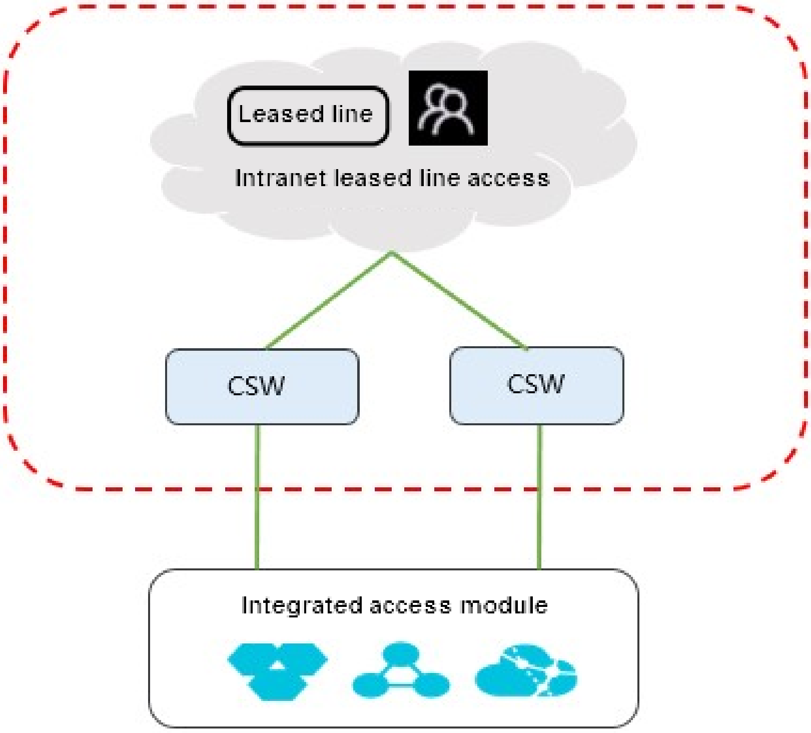
1.3.3.2 Integrated access area

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 instances and general cloud services. For access to VPC instances, the CSWs set up a map from internal users to VPC instances, and import these users into different VPC instances. Different user groups remain isolated from each other on the CSWs. For access to general cloud services, the CSWs are connected to the integrated service module by using EBGP and allow direct access to all the resources in the business service area.

Figure 1-9: Intranet access module

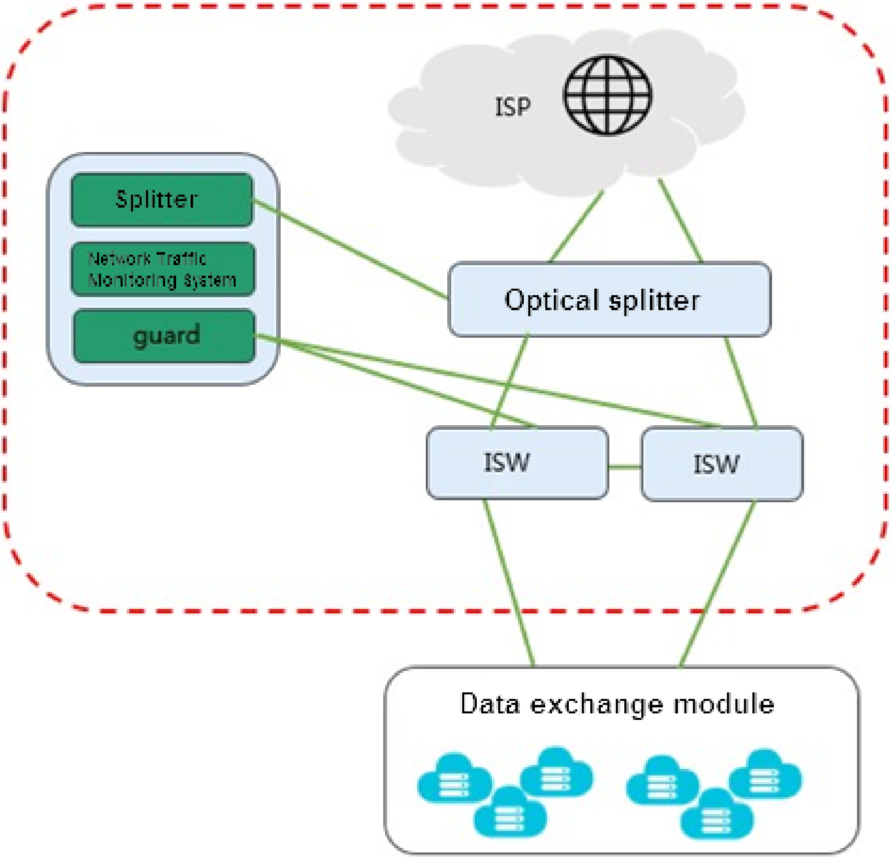


- **Internet access module**

This module consists of two ISWs. It facilitates the access to ISPs or users' public backbone networks. It performs route distribution and interaction between the inside and outside of cloud networks. The two ISWs run IBGP to back up routes between each other. Based on actual conditions, the ISWs use either static routing or EBGP to uplink with 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 the ISWs use BGP to connect to multiple carriers, each of which has 2*10 GE lines, to improve the reliability. The Internet access module uses EBGP to exchange routes with the data exchange module, releases Internet routes to the data exchange module, and receives the internal cloud service routes 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. Use Internet to access the cloud networks and the generated traffic is diverted to Network Traffic Monitoring System by using an optical splitter. When Network Traffic Monitoring System detects malicious traffic, it releases the corresponding route by using Alibaba Cloud Security to divert the malicious traffic to Alibaba Cloud Security for cleaning. The cleaned traffic is injected back into the Internet access module.

Figure 1-10: Internet access module

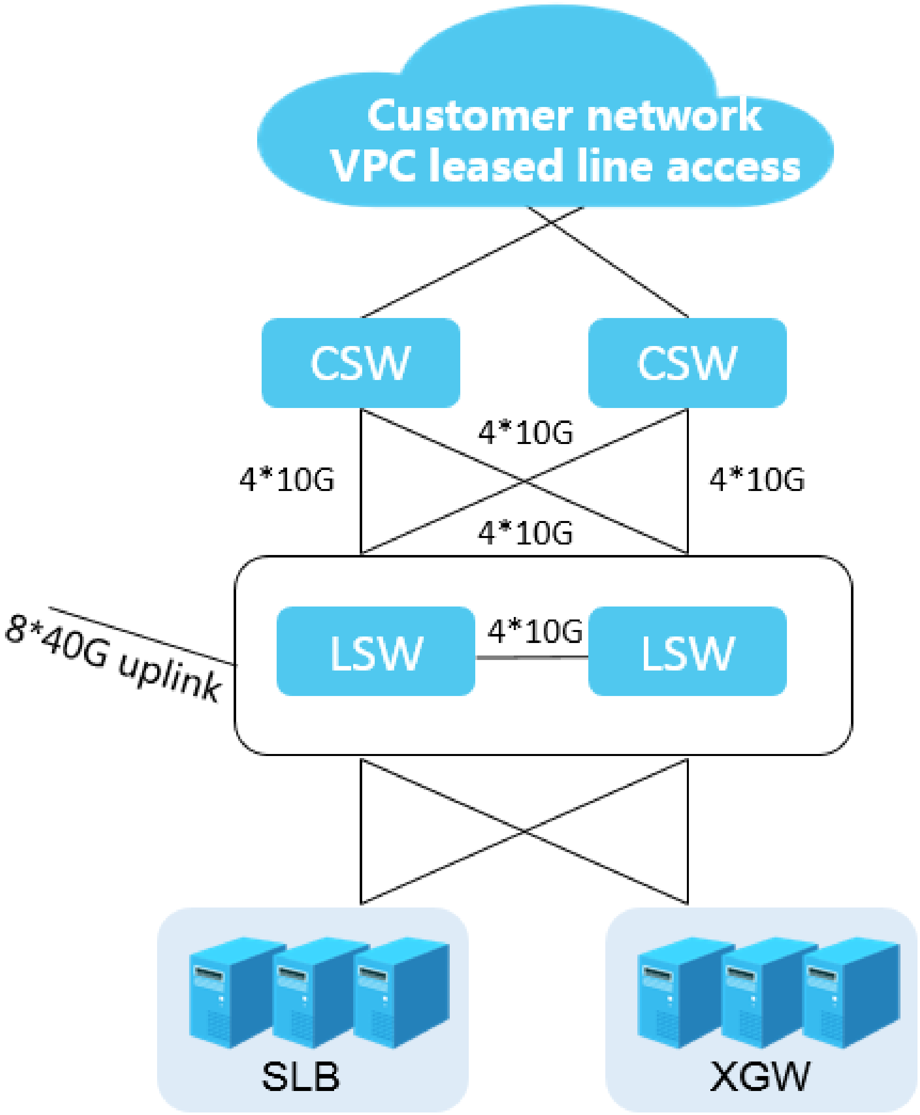


1.3.3.3 VPC leased line access

The VPC leased line access solution gives users full control over their own virtual networks, such as selecting their own IP address ranges and configuring the route tables and gateways. In addition, users can connect their VPC instances to a traditional data center by using leased lines or VPN connections to create a custom network environment. This enables smooth migration of applications to the cloud.

Each cloud service server (XGW/SLB) is connected to two LSWs. These servers use OSPF to exchange routing information. The two LSWs use IBGP to exchange routing information between each other, and use EGBP to exchange routing information with CSWs.

Figure 1-11: VPC leased line access

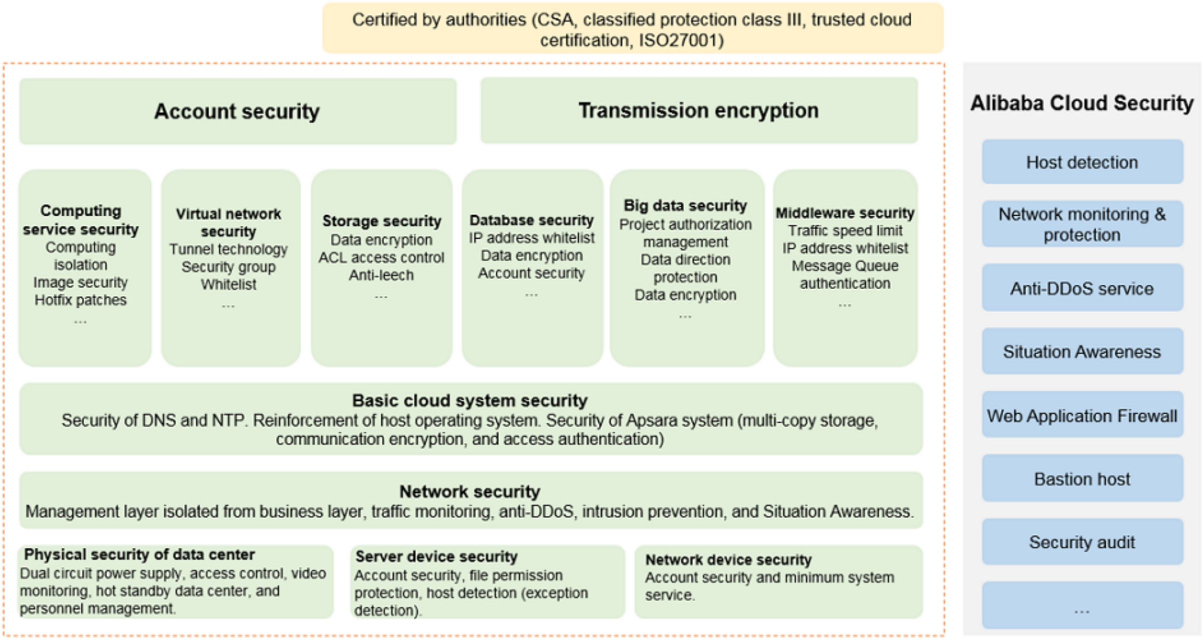


1.3.4 Security architecture

Apsara Stack provides all-around security capabilities, from underlying communication protocols to upper-layer applications, to guarantee the user access and data security. Access to every console in Apsara Stack is allowed only with HTTPS certificates. Apsara Stack provides a comprehensive role authorization mechanism to guarantee the secure and controllable access to resources in multi-tenant mode. It supports different security roles, such as security administrators , system administrators, and security auditors.

Apsara Stack has incorporated Alibaba Cloud Security since version 3, providing users with a multi-level and integrated cloud security protection solution.

Figure 1-12: Hierarchical security architecture of Apsara Stack


Certified by authorities (CSA, classified protection class III, trusted cloud certification, ISO27001)
Account security
Transmission encryption
Computing service security
Computing isolation
Image security
Hotfix patches
...
Virtual network security
Tunnel technology
Security group
Whitelist
...
Storage security
Data encryption
ACL access control
Anti-leech
...
Database security
IP address whitelist
Data encryption
Account security
...
Big data security
Project authorization management
Data direction protection
Data encryption
...
Middleware security
Traffic speed limit
IP address whitelist
Message Queue authentication
...
Basic cloud system security
Security of DNS and NTP. Reinforcement of host operating system. Security of Apsara system (multi-copy storage, communication encryption, and access authentication)
Network security
Management layer isolated from business layer, traffic monitoring, anti-DDoS, intrusion prevention, and Situation Awareness.
Physical security of data center
Dual circuit power supply, access control, video monitoring, hot standby data center, and personnel management.
Server device security
Account security, file permission protection, host detection (exception detection).
Network device security
Account security and minimum system service.
Alibaba Cloud Security
Host detection
Network monitoring & protection
Anti-DDoS service
Situation Awareness
Web Application Firewall
Bastion host
Security audit
...

1.3.5 Base assembly

Apsara Stack base consists of three types of assemblies, providing support for the deployment and O&M of the cloud platform.

Table 1-2: Base assembly

Assembly		Function
Operations assembly	Yum	Installation software package The software source is deployed during the initial installation phase. This package is mainly used to install the operating system and deploy Apsara Stack's application software packages and their dependent components, such as the Apsara platform and ECS, on physical machines.
	Clone	Machine cloning service
	NTP	Clock source service The physical machines deployed on Apsara Stack synchronize time from a standard NTP time source and provide the time to other hosts.
	DNS	Domain name resolution service DNS provides forward and reverse resolution of domain names for the internal Apsara Stack

Assembly		Function
		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.
Base middleware	Dubbo	Distributed RPC service
	Tair	Cache service
	mq	Message Queue service
	ZooKeeper	Distributed collaboration
	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	Cloud platform Log Service
	Metadatabase	Metadatabase
	POP	Open 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), Log Service, 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), and Table Store.

- **Database**

  Apsara Stack is equipped with diversified data engines. These data engines are able to interoperate with each other. The main products include ApsaraDB for MySQL, ApsaraDB for PostgreSQL, ApsaraDB for SQLServer, ApsaraDB for MongoDB, ApsaraDB for Redis, and ApsaraDB for Memcache.

- **Big data processing**

  Apsara Stack provides the various big data analysis, applications, and visualization capabilities to give value to data. The main products include StreamCompute, E-MapReduce, Quick BI, and Dataphin.

- **Security**

  Apsara Stack provides all-around protection in the form of Alibaba Cloud Security, from underlying communication protocols to upper-layer applications, to guarantee the user access and data security.

1.5 Scenarios

Apsara Stack provides flexible and scalable industrial solutions for users of different scales in the same sector. Based on the business traits of different sectors, such as industry, agriculture , transportation, government, finance, and education, Apsara Stack creates custom solutions to provide users with one-stop products and services. This section focuses on 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, forming the urban data brain.

Values and features

- A breakthrough of urban governance mode. With the urban data as a resource, City Brain improves the government management capabilities, resolves outstanding issues of urban governance, and achieves an intelligent, intensive, and humane form of governance.
- A breakthrough of urban service mode. City Brain provides services for enterprises and individuals more accurately and conveniently, 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.

Financial cloud

Financial 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), and provide more professional and comprehensive services for financial users. Enterprises can build financial cloud independently or with Alibaba Cloud. Financial cloud meets the requirements of large- and medium-sized financial organizations for independent data centers that are completely physically isolated, and can output the cloud computing and big data platforms to users' data centers.

Values and features

- Independent resource clusters
- Stricter data center management
- Better disaster recovery
- Stricter requirements for network security isolation
- Stricter access control
- Compliance with the security supervision requirements and compliance requirements for banks
- Dedicated security operation team, security compliance team, and security solution team for the financial cloud sector
- Dedicated financial cloud account managers and cloud architects
- Stricter user access mechanism

1.6 Compliance security solution

On June 1, 2017, the Cybersecurity Law of the People's Republic of China was formally implemented, which has made clear provisions for classified protection compliance. To help enterprise users quickly meet the requirements of the classified protection compliance, Alibaba Cloud integrates the technical advantages of Alibaba Cloud Security to establish the "Classified Protection Compliance Ecology". Alibaba Cloud works with its cooperative assessment agencies and security consulting manufacturers in various places to provide you with one-stop classified protection assessment. The complete attack protection, data audit, encryption, and security management help you quickly and easily pass the classified protection compliance assessment.

1.6.1 Interpretations on keys

Network and communication security

Interpretations on clauses

- The network is divided into different security domains by server role and importance.
- An access control policy is set at the security domain boundary between the intranet and Internet, which must be configured to specific ports.
- Intrusion prevention means must be deployed at the network boundary to prevent against and record intrusion behaviors.
- User behaviors must be logged and security events must be recorded and audited in the network.

Coping strategies

- We recommend that you use VPC and security group of Alibaba Cloud to divide a network into different security domains and control the access reasonably.
- Web Application Firewall is used to prevent against network intrusion.
- Use the logging function to log user behaviors and record, analyze, and audit security events.
- If a system is frequently threatened by DDoS, an advanced anti-DDoS service can be used to filter and clean abnormal traffic.

Device and computing security

Interpretations on clauses

- It is the basic security requirement to record and audit O&M operations, and avoid sharing accounts.

- Necessary security measures are taken to guarantee the security of the system layer and prevent against intrusion to servers.

Coping strategies

- Audit the operations on servers and data. Create an independent account for each O&M personnel to avoid sharing accounts.
- Use Server Guard to conduct complete management of server vulnerabilities, baseline inspection, and intrusion prevention.

Application and data security

Interpretations on clauses

- An application is the direct implementation of specific business. Unlike network and system , applications do not have the relative standard characteristics. Most applications include functions such as identity authentication, access control, and operation audit, which are difficult to be replaced by third-party products.
- Besides security prevention at other levels, encryption is the most effective method for data integrity and confidentiality.
- Remote data backup is one of the most important requirements that distinguish class III protection from class II protection, and is the most foundational technical safeguard for business continuity.

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, functions such as account authentication, user permission classification, and log audit are designed to satisfy classified protection requirements.
- For data security, use HTTPS to make sure that data is encrypted in transmission.
- For data backup, we recommend that you use an RDS remote disaster tolerance instance to automatically back up data. You can also manually synchronize the database backup files to Alibaba Cloud servers in other regions.

Security management policies

Interpretations on clauses

- Security policies, regulations, and management personnel are foundational for sustainable security. The policy guides a security direction, the regulation identifies a security process, and persons fulfill security responsibilities.
- Classified protection requirements provide a methodology and best practice. Security can be continuously constructed and managed according to the classified protection methodology.

Coping strategies

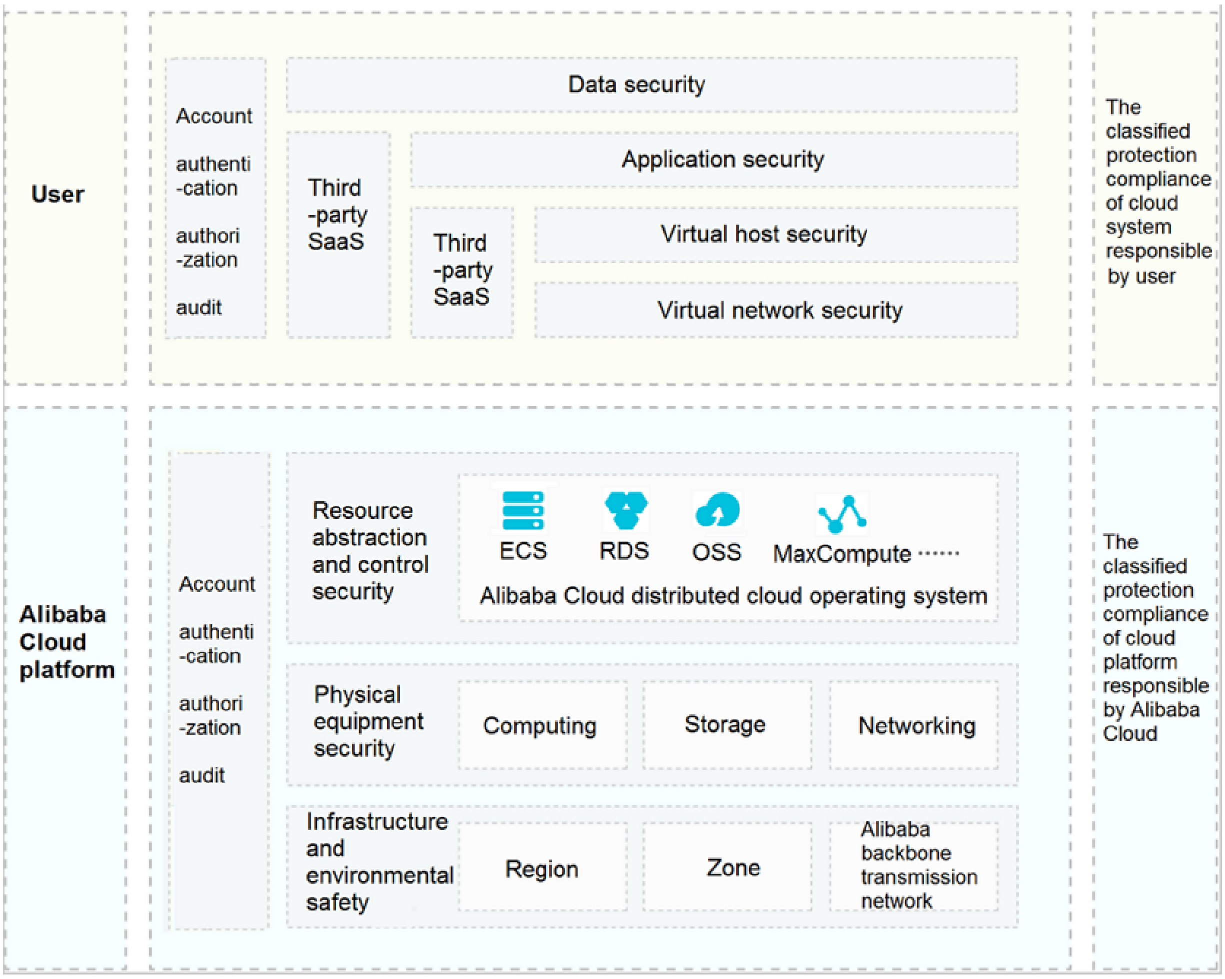
- Security policies, regulations, and management personnel must be arranged, prepared, and fulfilled by customers' management level according to actual conditions of enterprises, and special documents must be prepared.
- For technical measures required in the process of vulnerability management, we recommend that you use Alibaba Cloud Server Guard to quickly detect system vulnerabilities on the cloud and handle them in time.

1.6.2 Cloud-based classified protection compliance

Shared compliance responsibilities

The Alibaba Cloud platform and cloud tenant systems must be rated and assessed respectively. Assessment conclusions of the Alibaba Cloud platform can be reused by tenant systems in assessment.

Figure 1-13: Shared compliance responsibilities


User
Account
authenti-cation
authori-zation
audit
Third-party SaaS
Data security
Application security
Third-party SaaS
Virtual host security
Virtual network security
The classified protection compliance of cloud system responsible by user
Alibaba Cloud platform
Account
authenti-cation
authori-zation
audit
Resource abstraction and control security
ECS
RDS
OSS
MaxCompute ……
Alibaba Cloud distributed cloud operating system
Physical equipment security
Computing
Storage
Networking
Infrastructure and environmental safety
Region
Zone
Alibaba backbone transmission network
The classified protection compliance of cloud platform responsible by Alibaba Cloud

Alibaba Cloud provides the following contents:

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

Detailed interpretations on shared responsibility 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 standard. The public cloud and e-Government cloud pass class III protection archival filing and assessment. The financial cloud passes class IV protection archival filing and assessment.
- When tenant systems of Alibaba Cloud pass classified protection assessment, physical security , partial network security, and security management conclusions can be reused. Alibaba

Cloud can provide explanations according to conclusion reuse rules issued by the supervision authorities.

- Complete security technologies and management architecture of the Alibaba Cloud platform and Alibaba Cloud Security protection system facilitate tenants to pass classified protection assessment better.

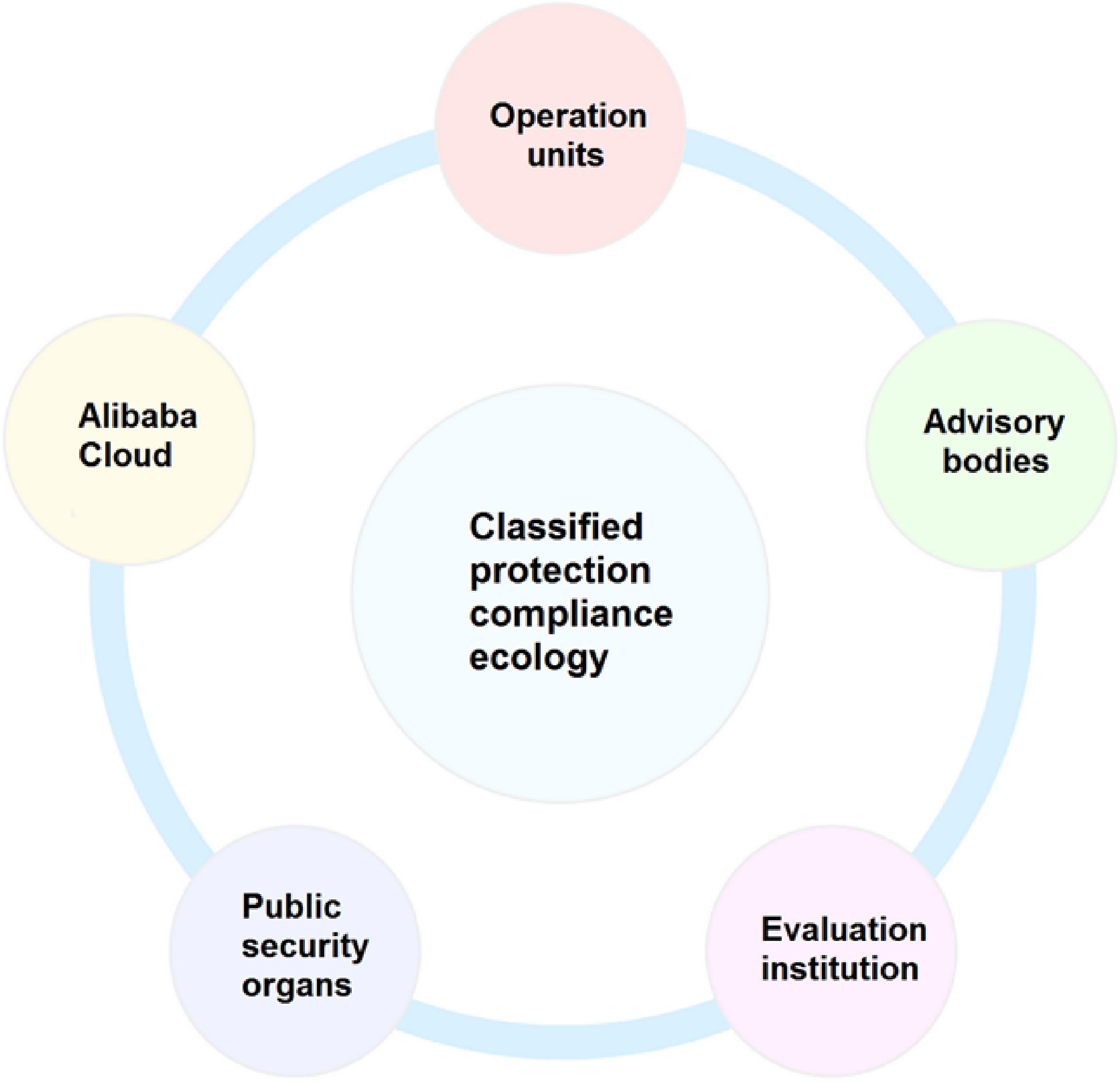
Classified protection compliance ecology

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 communication with supervision authorities.
- Security systems lag behind business development.

To facilitate cloud-based systems to quickly pass classified protection assessment, the Classified Protection Compliance Ecology is established by Alibaba Cloud to provide one-stop classified protection compliance solution.

Figure 1-14: Classified protection compliance ecology



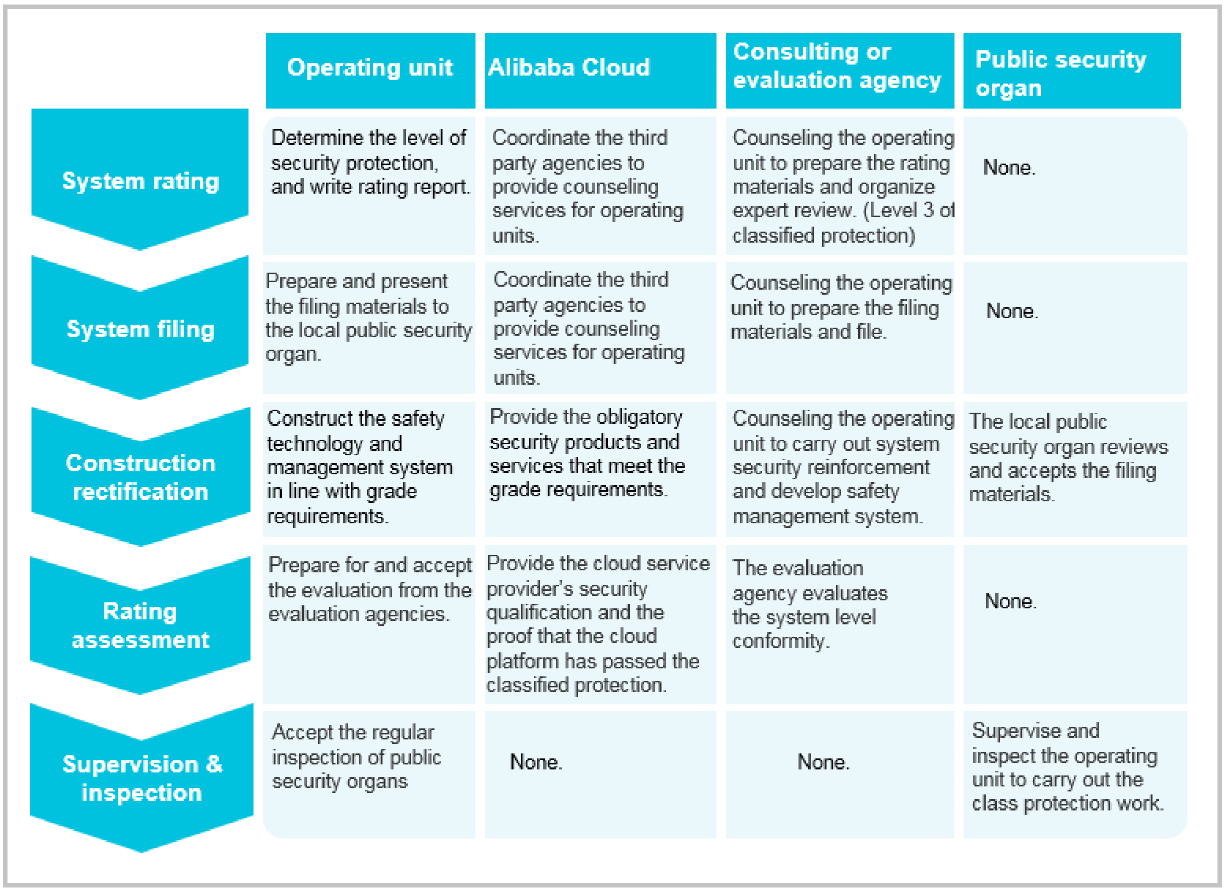
Work division of classified protection:

- 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 authorities: Responsible for archival filing review, supervision, and inspection.

1.6.3 Classified protection implementation process

The classified protection implementation process is as shown in *Figure 1-15: Classified protection implementation process* .

Figure 1-15: Classified protection implementation process



	Operating unit	Alibaba Cloud	Consulting or evaluation agency	Public security organ
System rating	Determine the level of security protection, and write rating report.	Coordinate the third party agencies to provide counseling services for operating units.	Counseling the operating unit to prepare the rating materials and organize expert review. (Level 3 of classified protection)	None.
System filing	Prepare and present the filing materials to the local public security organ.	Coordinate the third party agencies to provide counseling services for operating units.	Counseling the operating unit to prepare the filing materials and file.	None.
Construction rectification	Construct the safety technology and management system in line with grade requirements.	Provide the obligatory security products and services that meet the grade requirements.	Counseling the operating unit to carry out system security reinforcement and develop safety management system.	The local public security organ reviews and accepts the filing materials.
Rating assessment	Prepare for and accept the evaluation from the evaluation agencies.	Provide the cloud service provider's security qualification and the proof that the cloud platform has passed the classified protection.	The evaluation agency evaluates the system level conformity.	None.
Supervision & inspection	Accept the regular inspection of public security organs	None.	None.	Supervise and inspect the operating unit to carry out the class protection work.

1.6.4 Security compliance architecture

Quickly access Alibaba Cloud Security and quickly complete security correction. Satisfy technical requirements for foundational compliance in classified protection with minimal security investment s.

Basic requirements of classified protection

- Physical and environmental security: Including measures such as data center power supply, temperature and humidity control, wind prevention, rain prevention, and thunder prevention. Assessment conclusions of Alibaba Cloud can be directly reused.
- Network and communication security: Including network architecture, boundary protection, access control, intrusion prevention, and communication encryption.
- Device and computing security: Including intrusion prevention, malicious code prevention, identity authentication, access control, centralized management and control, and security audit.
- Application and data security: Including security audit, data integrity, and data confidentiality.

1.6.5 Solution benefits

One-stop classified protection assessment service

Select and cooperate with local consulting and assessment agencies that offer high-quality services, provide one-stop and whole-process compliance, and greatly reduce operator investments.

- Avoid multi-point communications and repeated work to reduce operator investments.
- Greatly improve efficiency and complete assessment in minimal two weeks.
- Alibaba Cloud provides cloud security and compliance best practices.

Complete security prevention system

With a complete Alibaba Cloud Security architecture, an operator can find corresponding products on Alibaba Cloud, correct non-conformances, and completely satisfy classified protection requirements.

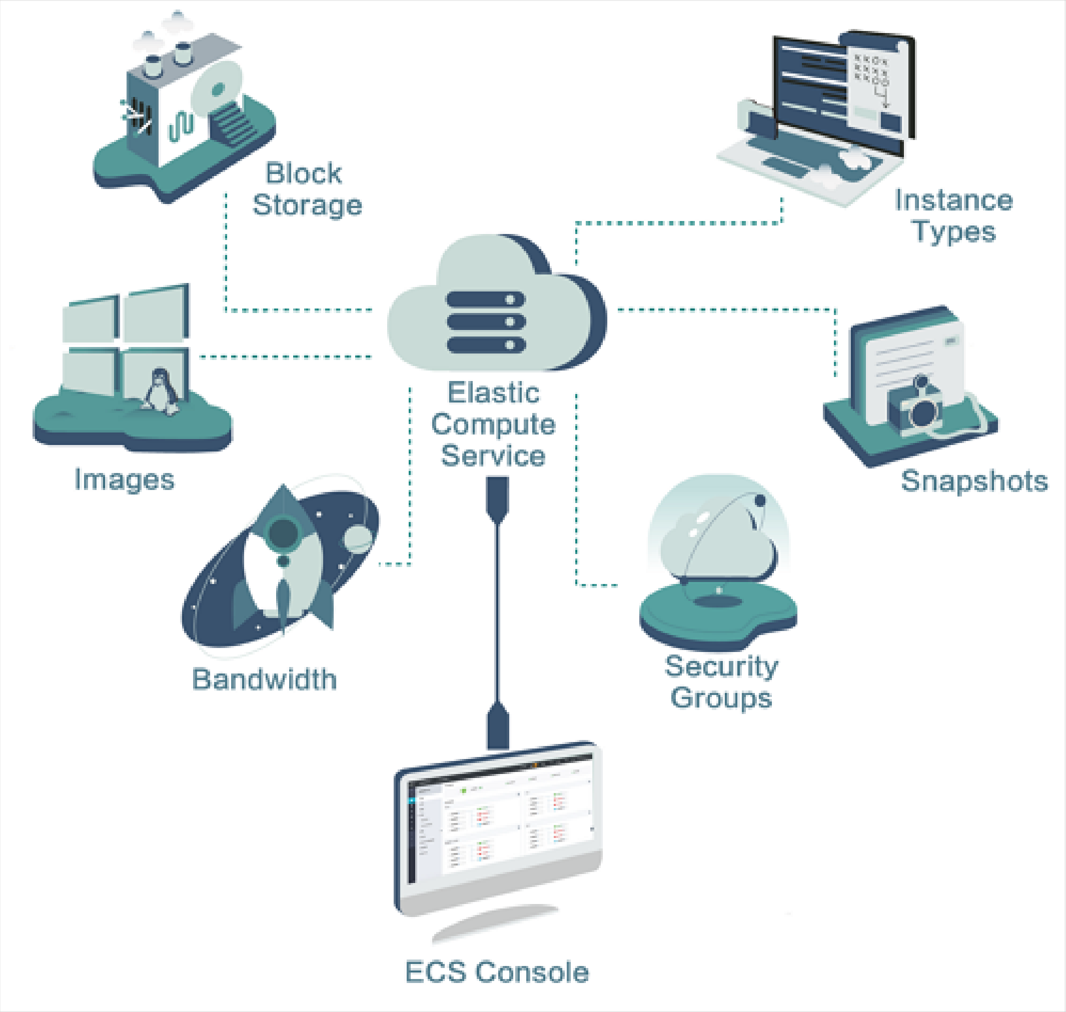
2 Elastic Compute Service (ECS)

2.1 What is ECS

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

As a virtual computing environment made up of the basic components such as CPU, memory, and storage, an ECS instance is provided by ECS for you to carry out relevant operations. It is the core concept of ECS and you can perform actions on ECS instances on the ECS console. As for other resources such as block storage, images, and snapshots, they cannot be used until being integraed with ECS instances. *Figure 2-1: Concept of an ECS instance* illustrates the services supported by an ECS instance.

Figure 2-1: Concept of an ECS instance



2.2 Benefits

Compared to the traditional Internet Data Centers (IDCs) or servers, ECS has the following advantages:

- *High availability*
- *Security*
- *Elasticity*

High availability

Alibaba Cloud adopts more stringent IDC standards, server access standards, and O&M standards to guarantee data reliability and high availability of cloud computing infrastructure and cloud servers.

When even higher availability is needed, you can build active/standby or active/active services in multiple zones. For a finance-oriented solution with three IDCs in two regions, you can deliver services of higher availability with multiple regions and zones. For such services as disaster tolerance and backup, mature solutions are readily available in Alibaba Cloud.

Alibaba Cloud provides you with the following support services:

- Products and services for availability improvement, including cloud servers, server load balancers, multi-backup databases, and Data Transport Services (DTS).
- Industry partners and ecosystem partners that help you build a more advanced and stable architecture and guarantee service continuity.
- Diverse training services that enable you to deliver high availability from the business level to the underlying service level.

Security

Users of cloud computing are most concerned about security and stability. Alibaba Cloud has recently passed a host of international information security certifications, including ISO 27001 and MTCS, which demand strict confidentiality of user data and user information and user privacy protection.

- **Alibaba Cloud VPC offers more business possibilities.** You only need to perform simple configuration to connect your business environment to global IDCs, making your business more flexible, stable, and extensible.
- **Alibaba Cloud VPC can connect to your IDC** through a leased line to build a hybrid cloud architecture. You can build a more flexible business with the powerful network functions from Alibaba Cloud’s various hybrid cloud solutions and network products. A superior business ecosystem is possible with Alibaba Cloud’s ecosystem.
- **Alibaba Cloud VPC is more stable and secure.**
    - **Stable:** After building your business on VPC, you can update your network architecture and functions on a daily basis as the network infrastructure evolves constantly, allowing your business to run steadily. You can divide, configure, and manage your network on VPC according to your needs.
    - **Secure:** VPC is endowed with traffic isolation and attack isolation to protect your services from endless attack traffic on the Internet. After you build your business on VPC, the first line of defense is established immediately.

VPC provides a stable, secure, fast-deliverable, self-managed, and controllable network environment. With the capability and architecture of VPC hybrid cloud, the technical advantages of cloud computing are open to all industries and enterprises.

Elasticity

Elasticity is the biggest advantage of cloud computing.

- **Elastic computing**
  - **Vertical scaling**. Vertical scaling involves modifying the configuration of an individual server, which is hard for traditional IDCs. This, however, is just an easy task for Alibaba Coud as you can scale up or down ECS or storage capacity based on your transaction volume.
  - **Horizontal scaling**. During peak hours for gaming or live video streaming apps, your hands may be tied when a request for additional resources arises in the traditional IDC mode. On the contrary, cloud computing can leverage elasticity to tide you over that period. When the period ends, you can release unnecessary resources to reduce your business cost. With horizontal scaling and auto-scaling, you can determine how and when to scale your resources or implement scaling based on business loads.
- **Elastic storage**

  Alibaba Cloud has a powerful elastic storage. When more storage space is required, you can only add servers in the traditional IDC mode, which has a limit on the number of servers that can be added. In the cloud computing mode, however, the sky is the limit. You can order as needed to guarantee sufficient storage space.
- **Elastic network**

  Alibaba Cloud also features an elastic network. When you purchase the Alibaba Virtual Private Cloud (VPC), you can have the same network configuration as that of IDCs. In addition, you can have the following benefits:

  - Interconnection between data centers
  - Secure domains isolated among data centers
  - Flexible network configuration and planning within the VPC

The elasticity of Alibaba Cloud is reflected in computing, storage, network, and the ability to redesign business architecture. By using Alibaba Cloud, you can work out your business portfolio as you wish.

2.3 Architecture

Built upon the Apsara system developed by Alibaba Cloud on its own, ECS allows you to create a virtual machine instance via the KVM technology and uses the Apsara Distributed File System for data storage.

Product architecture

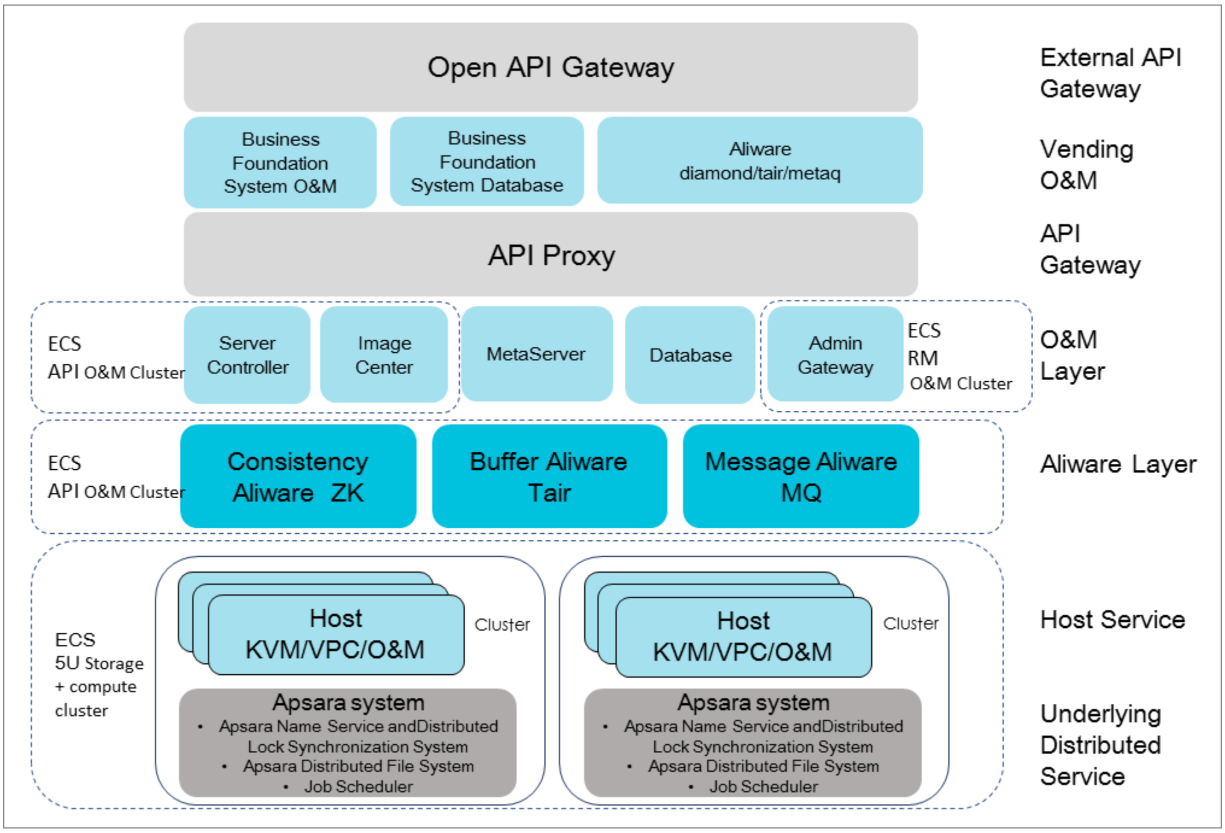


Table 2-1: Description of the architecture

Component	Description
Apsara Name Service and Distributed Lock Synchronization System	This is the fundamental module in the Alibaba Cloud family that provides the distributed consistency service . As the critical distributed coordination system, this module provides three types of basic services: the distributed lock synchronization service, the distribute d publish/subscribe notification service, and the lightweight metadata storage service.
Apsara Distributed File System	This is a distributed storage system developed by Alibaba Cloud on its own. As of 2017, the Apsara Distributed File System has been adopted to deploy hundreds of clusters in production environments,

Component	Description
	managing hundreds of thousands of storage nodes and dozens of EBs of disk space.
Job Scheduler	This is a distributed resource scheduler intended for managing and dispatching resources in a distributed system.
Server Controller	This is the scheduling system for ECS. It schedules such resources as hosts, IPs and storage and delivers virtual machine instances to users.
Major flow	OpenAPI --> Service Layer --> Server Controller --> Host Service
OpenAPI Gateway	Provides the basic services such as authentication and forwarding requests.
Business Foundation System O&M	This module handles the vending requests, creates and releases instances, takes snapshot polices as scheduled and provides the OpenAPI service to the outside.
API Proxy	Forwards a request to the corresponding region based on the region_id.
Server Controller Database	Stores the information about control data, status data, and so on.
Server Controller	As the core of the entire control system, this module is responsible for handling hosts, IPs and storage.
Tair	Provides the cache service for the Server Controller.
Zookeeper	This module provides the distributed lock synchroniz ation service for the Server Controller.
Message Aliware MQ	This module provides the queuing service for virtual machine status messages.
Image Center	This module provides images management service, such as importing/copying images.
MetaSever	This module provides the meta data management service for ECS instance.
Host	KVM (virtualized computation), VPC (a virtualized network) and O&M (a control process that interacts with Libivrt).

Component	Description
AG (Admin Gateway)	This is equivalent to a jump server for logging on to the NC during O&M management.
ECS Decider	This is an O&M node and responsible for deciding which NC is used to deploy ECS.

2.4 Features

2.4.1 Instances

An ECS instance is equivalent to a virtual machine that includes CPU, memory, operating system , bandwidth, disks, and other basic computing components. You can easily customize and change the configuration of an instance and you have full control over such a virtual machine. Unlike a local server, you can use ECS instances and perform such operations as independent management and top-level configuration so long as you log in to Alibaba Cloud.

2.4.1.1 Instance type families

An ECS instance is the minimal unit that can provide computing capabilities and services for your business. ECS instances are available in several type families based on their configuration and business purposes they serve.

Note:

All instance type families listed in this document are for reference purpose only. The specific configurations of your instances are determined by the physical servers that the instances are hosted on.

Table 2-2: Instance type families

Type	Feature	Ideal for
G5, general-purpose type family	• vCPU : Memory = 1:4 • Ultra high packet forwarding rate • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors • Higher computing specificat ions matching higher network performance	• Scenarios of receiving and transmitting a large volume of packets, such as live commenting on videos, retransmission of telecommunication services, and so on • Enterprise-level applications of various types and sizes

Type	Feature	Ideal for
		• Small and medium database systems, caches, and search clusters • Data analysis and computing • Computing clusters and data processing depending on the memory
SN2NE, general-purpose type family with enhanced network performance	• vCPU : Memory = 1:4 • Ultra high packet forwarding rate • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors • Higher computing specificat ions matching higher network performance	• Scenarios that require receiving and transmitting a large volume of packets, such as live commenting on videos, retransmission of telecommunication services • Enterprise-level applications of various types and sizes • Small and medium database systems, caches, and search clusters • Data analysis and computing • Computing clusters and data processing depending on the memory
C5, compute instance type family	• vCPU : Memory = 1:2 • Ultra high packet forwarding rate • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors • Higher computing specificat ions matching higher network performance	• Scenarios that require receiving and transmitting a large volume of packets, such as live commenting on videos, retransmission of telecommunication services • Web front-end servers • Massively Multiplayer Online (MMO) game front-ends • Data analysis, batch compute, and video coding • High performance science and engineering applications
SN1NE, compute optimized type family with enhanced network performance	• vCPU : Memory = 1:2 • Ultra high packet forwarding rate	• Scenarios that require receiving and transmitting a large volume of packets, such as live commenting on

Type	Feature	Ideal for
	• 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors • Higher computing specificat ions matching higher network performance	videos, retransmission of telecommunication services • Web front-end servers • Massively Multiplayer Online (MMO) game front-ends • Data analysis, batch compute, and video coding • High performance science and engineering applications
R5, memory instance type family	• Ultra high packet forwarding rate • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors • Higher computing specificat ions matching higher network performance	• Scenarios that require receiving and transmitting a large volume of packets, such as live commenting on videos, retransmission of telecommunication services • High performance databases and high memory databases • Data analysis and mining , and distributed memory caches • Hadoop, Spark, and other enterprise-level applicatio ns with large memory requirements
SE1NE, memory optimized type family with enhanced network performance	• vCPU : Memory = 1:8 • Ultra high packet receiving and forwarding rate • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors • Higher computing specificat ions matching higher network performance	• Scenarios that require receiving and transmitting a large volume of packets, such as live commenting on videos, retransmission of telecommunication services • High performance databases and high memory databases • Data analysis and mining , and distributed memory caches • Hadoop, Spark, and other enterprise-level applicatio ns with large memory requirements

Type	Feature	Ideal for
SE1, memory optimized type family	- vCPU : Memory = 1:8 - 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell), or E5-2680 v4 (Haswell) processors - The latest DDR4 memory - Higher computing specifications matching higher network performance - I/O optimized by default	As an instance that uses the memory exclusively, SE1 features a greater ratio of memory to vCPU. It is intended for the scenarios that require fixed performance of computing, such as Cache/Redis, searching, memory databases, high I/O databases (e.g., Oracle, MongoDB), Hadoop clusters, and so on
D1NE, big data type family with enhanced network performance	- High-volume local SATA HDD disks with high I/O throughput and up to 35 Gbit/s of bandwidth for a single instance - vCPU : Memory = 1:4, designed for big data scenarios - 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors - Higher computing specifications matching higher network performance **Note:** - Elastic Network Interfaces (ENIs) are supported by the enterprise-level types with 2 or more vCPUs - This type family does not support changing configuration currently	- Hadoop MapReduce, HDFS, Hive, HBase, and so on - Spark in-memory computing, MLlib, and so on - Enterprises that require big data computing and storage analysis to store and compute massive data, for example, companies in the Internet and finance industries - Elasticsearch, logs, and so on
D1, big data type family	- High-volume local SATA HDD disks with high I/O throughput and up to 17 Gbit/s of bandwidth for a single instance - vCPU : Memory = 1:4, designed for big data scenarios	- Hadoop MapReduce, HDFS, Hive, HBase, and so on - Spark in-memory computing, MLlib, and so on - Enterprises that require big data computing and storage analysis to store

Type	Feature	Ideal for
	- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors - Higher computing specificat ions matching higher network performance	and compute massive data , for example, companies in the Internet and finance industries - Elasticsearch, logs, and so on
GN5, compute optimized type family with GPU	- NVIDIA P100 GPU processors - Various ratios of vCPU to memory - High-performance NVMe SSD disks - 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors - Higher computing specificat ions matching higher network performance	- Deep learning - Scientific computing, such as computational fluid dynamics, computational finance, genomics, and environmental analysis - High performance computing , rendering, multi-media coding and decoding, and other server-side GPU compute workloads
GN5i, compute optimized type family with GPU	- NVIDIA P4 GPU processors - vCPU : Memory = 1:4 - 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors - Higher computing specificat ions matching higher network performance	- Deep learning - Multi-media coding and decoding and other server-side GPU compute workloads
GN4, compute optimized type family with GPU	- NVIDIA M40 GPU processors - Various ratios of CPU to memory - 2.5 GHz Intel Xeon E5-2680 v4 (Haswell) processors - Higher computing specificat ions matching higher network performance	- Deep learning - Scientific computing, such as computational fluid dynamics, computational finance, genomics, and environmental analysis - High performance computing , rendering, multi-media coding and decoding, and other server-side GPU compute workloads

Type	Feature	Ideal for
GA1, visualization compute type family with GPU	• AMD S7150 GPU processors • vCPU : Memory = 1:2.5 • 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors • High-performance local NVMe SSD disks • Higher computing specificat ions matching higher network performance	• Rendering, multimedia coding and decoding • Machine learning, high-performance computing, and high performance databases • Other server-end business scenarios that require powerful concurrent floating-point compute capabilities
N4, general entry-level instance type family	• vCPU : Memory = 1:2 • 2.5 GHz Intel Xeon E5-2680 v3 (Broadwell) processors • The latest DDR4 memory • I/O optimized by default	• Small and medium-sized Web servers • Batch processing • Distributed analysis • Advertisement services
XN4, compact entry-level instance type family	• vCPU : Memory = 1:1 • 2.5 GHz Intel Xeon E5-2680 v4 (Haswell) or E5-2682 v4 (Broadwell) processors • The latest DDR4 memory • I/O optimized by default	• Small-sized Web applicatio ns • Small-sized databases • Applications for development or testing environments • Code repositories
MN4, balanced entry-level instance type family	• vCPU : Memory = 1:4 • 2.5 GHz Intel Xeon E5-2680 v3 (Broadwell), E5-2680 v4 (Haswell), E5-2682 v4 (Broadwell), or E5-2650 v2 (Haswell) processors • The latest DDR4 memory • I/O optimized by default	• Medium-sized Web servers • Batch processing • Distributed analysis • Advertisement services • Hadoop clusters
E4, memory instance type family	• vCPU : Memory = 1:8 • 2.5 GHz Intel Xeon E5-2680 v4 (Broadwell), E5-2680 v3 (Broadwell), E5-2650 v2 (Haswell), or E5-2682 v4 (Broadwell) processors • I/O optimized by default	Applications that involve numerous operations in the memory, searching and computing, for example, Cache /Redis, searching, in-memory database, and so on

Type	Feature	Ideal for
F3, compute optimized type family with FPGA	• Self-developed compute cards based on Xilinx Virtex UltraScale+ VU9P • vCPU : Memory = 1:4 • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors • Higher computing specificat ions matching higher network performance	• Deep learning • Genomics research • Video coding and decoding • Chip prototype verification • Database acceleration
EBMG5, general-purpose ECS Bare Metal Instance type family	• vCPU : Memory = 1:4 • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96-core vCPU, up to 2.9 GHz Turbo Boot • High network performance: 4.5 million pps packet forwarding rate • Supports SSD Cloud Disks and Ultra Cloud Disks	• Deployment of OpenStack , ZStack, and other private cloud services • Deployment of Docker containers and other services • Scenarios that require receiving and transmitting a large volume of packets, such as live commenting on videos, retransmission of telecommunication services • Enterprise-level applications of various types and sizes • Medium and large database systems, caches, and search clusters • Data analysis and computing • Computing clusters and data processing depending on memory
I2, type family with local SSD disks	• High-performance local NVMe SSD disks with high IOPS, high I/O throughput, and low latency • vCPU : Memory = 1:8, designed for high performance databases • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors	• OLTP and high performance relational databases • NoSQL databases, such as Cassandra and MongoDB • Search applications, such as Elasticsearch

Type	Feature	Ideal for
	• Higher computing specifications matching higher network performance	
RE5, type family with enhanced memory	• Optimized for memory-intensive enterprise applications that involve high-performance databases and in-memory databases • 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors • vCPU : Memory = 1:16, up to 2970 GiB of memory	• High performance databases and in-memory databases • Memory intensive applications • Big data engines like Apache Spark and Presto

2.4.1.2 Instance types

For an ECS instance, its type specifies two attributes, its CPU (such as model and clock speed) and memory. To definitely determine the application scenario, however, you must select the image, disk, and network service at the same time. *Table 2-3: Instance types* details the instances of various attributes within each ECS instance type family.

Table 2-3: Instance types

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
N4	ecs.n4.small	N/A	1	2.0	1
	ecs.n4.large	N/A	2	4.0	1
	ecs.n4.xlarge	N/A	4	8.0	2
	ecs.n4.2xlarge	N/A	8	16.0	2
	ecs.n4.4xlarge	N/A	16	32.0	2
	ecs.n4.8xlarge	N/A	32	64.0	2
MN4	ecs.mn4.small	N/A	1	4.0	1
	ecs.mn4.large	N/A	2	8.0	1
	ecs.mn4.xlarge	N/A	4	16.0	2

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.mn4.2xlarge	N/A	8	32.0	3
	ecs.mn4.4xlarge	N/A	16	64.0	8
	ecs.mn4.8xlarge	N/A	32	128.0	8
E4	ecs.e4.small	N/A	1	8.0	1
	ecs.e4.large	N/A	2	16.0	1
	ecs.e4.xlarge	N/A	4	32.0	2
	ecs.e4.2xlarge	N/A	8	64.0	3
	ecs.e4.4xlarge	N/A	16	128.0	8
XN4	ecs.xn4.small	N/A	1	1.0	1
gn5	ecs.gn5-c4g1.xlarge	440	4	30.0	2
	ecs.gn5-c8g1.2xlarge	440	8	60.0	3
	ecs.gn5-c4g1.2xlarge	880	8	60.0	3
	ecs.gn5-c8g1.4xlarge	880	16	120.0	8
	ecs.gn5-c28g1.7xlarge	440	28	112.0	8
	ecs.gn5-c8g1.8xlarge	1,760	32	240.0	8
	ecs.gn5-c28g1.14xlarge	880	56	224.0	8
	ecs.gn5-c8g1.14xlarge	3,520	56	480.0	8
d1	ecs.d1.2xlarge	4 * 5,500	8	32.0	3
	ecs.d1.4xlarge	8 * 5,500	16	64.0	8
	ecs.d1.6xlarge	12 * 5,500	24	96.0	8

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.d1-c8d3.8xlarge	12 * 5,500	32	128.0	8
	ecs.d1.8xlarge	16 * 5,500	32	128.0	8
	ecs.d1-c14d3.14xlarge	12 * 5,500	56	160.0	8
	ecs.d1.14xlarge	28 * 5,500	56	224.0	8
gn4	ecs.gn4-c4g1.xlarge	N/A	4	30.0	2
	ecs.gn4-c8g1.2xlarge	N/A	8	60.0	3
	ecs.gn4.8xlarge	N/A	32	48.0	8
	ecs.gn4-c4g1.2xlarge	N/A	8	60.0	3
	ecs.gn4-c8g1.4xlarge	N/A	16	60.0	8
	ecs.gn4.14xlarge	N/A	56	96.0	8
ga1	ecs.ga1.xlarge	1*87	4	10.0	2
	ecs.ga1.2xlarge	1*175	8	20.0	3
	ecs.ga1.4xlarge	1*350	16	40.0	8
	ecs.ga1.8xlarge	1*700	32	80.0	8
	ecs.ga1.14xlarge	1*1,400	56	160.0	8
se1ne	ecs.se1ne.large	N/A	2	16.0	1
	ecs.se1ne.xlarge	N/A	4	32.0	2

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.se1ne.2xlarge	N/A	8	64.0	3
	ecs.se1ne.4xlarge	N/A	16	128.0	8
	ecs.se1ne.8xlarge	N/A	32	256.0	8
	ecs.se1ne.14xlarge	N/A	56	480.0	8
sn2ne	ecs.sn2ne.large	N/A	2	8.0	1
	ecs.sn2ne.xlarge	N/A	4	16.0	2
	ecs.sn2ne.2xlarge	N/A	8	32.0	3
	ecs.sn2ne.4xlarge	N/A	16	64.0	8
	ecs.sn2ne.8xlarge	N/A	32	128.0	8
	ecs.sn2ne.14xlarge	N/A	56	224.0	8
sn1ne	ecs.sn1ne.large	N/A	2	4.0	1
	ecs.sn1ne.xlarge	N/A	4	8.0	2
	ecs.sn1ne.2xlarge	N/A	8	16.0	3
	ecs.sn1ne.4xlarge	N/A	16	32.0	8
	ecs.sn1ne.8xlarge	N/A	32	64.0	8
gn5i	ecs.gn5i-c2g1.large	N/A	2	8.0	1

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.gn5i-c4g1.xlarge	N/A	4	16.0	2
	ecs.gn5i-c8g1.2xlarge	N/A	8	32.0	2
	ecs.gn5i-c16g1.4xlarge	N/A	16	64.0	2
	ecs.gn5i-c28g1.14xlarge	N/A	56	224.0	2
g5	ecs.g5.large	N/A	2	8.0	2
	ecs.g5.xlarge	N/A	4	16.0	3
	ecs.g5.2xlarge	N/A	8	32.0	4
	ecs.g5.4xlarge	N/A	16	64.0	8
	ecs.g5.6xlarge	N/A	24	96.0	8
	ecs.g5.8xlarge	N/A	32	128.0	8
	ecs.g5.16xlarge	N/A	64	256.0	8
	ecs.g5.22xlarge	N/A	88	352.0	15
c5	ecs.c5.large	N/A	2	4.0	2
	ecs.c5.xlarge	N/A	4	8.0	3
	ecs.c5.2xlarge	N/A	8	16.0	4
	ecs.c5.4xlarge	N/A	16	32.0	8
	ecs.c5.6xlarge	N/A	24	48.0	8
	ecs.c5.8xlarge	N/A	32	64.0	8
	ecs.c5.16xlarge	N/A	64	128.0	8
r5	ecs.r5.large	N/A	2	16.0	2
	ecs.r5.xlarge	N/A	4	32.0	3
	ecs.r5.2xlarge	N/A	8	64.0	4

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.r5.4xlarge	N/A	16	128.0	8
	ecs.r5.6xlarge	N/A	24	192.0	8
	ecs.r5.8xlarge	N/A	32	256.0	8
	ecs.r5.16xlarge	N/A	64	512.0	8
	ecs.r5.22xlarge	N/A	88	704.0	15
se1	ecs.se1.large	N/A	2	16.0	2
	ecs.se1.xlarge	N/A	4	32.0	3
	ecs.se1.2xlarge	N/A	8	64.0	4
	ecs.se1.4xlarge	N/A	16	128.0	8
	ecs.se1.8xlarge	N/A	32	256.0	8
	ecs.se1.14xlarge	N/A	56	480.0	8
d1ne	ecs.d1ne.2xlarge	4 * 5,500	8	32.0	4
	ecs.d1ne.4xlarge	8 * 5,500	16	64.0	8
	ecs.d1ne.6xlarge	12 * 5,500	24	96.0	8
	ecs.d1ne.8xlarge	16 * 5,500	32	128.0	8
	ecs.d1ne.14xlarge	28 * 5,500	56	224.0	8
f3	ecs.f3-c16f1.4xlarge	N/A	16	64.0	8
	ecs.f3-c16f1.8xlarge	N/A	32	128.0	8
	ecs.f3-c16f1.16xlarge	N/A	64	256.0	16

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
ebmg5	ecs.ebmg5.24xlarge	N/A	96	384.0	32
i2	ecs.i2.xlarge	1 * 894	4	32.0	3
	ecs.i2.2xlarge	1 * 1,788	8	64.0	4
	ecs.i2.4xlarge	2 * 1,788	16	128.0	8
	ecs.i2.8xlarge	4 * 1,788	32	256.0	8
	ecs.i2.16xlarge	8 * 1,788	64	512.0	8
re5	ecs.re5.15xlarge	N/A	60	990.0	8
	ecs.re5.30xlarge	N/A	120	1,980.0	15
	ecs.re5.45xlarge	N/A	180	2,970.0	15

2.4.1.3 Instance UserData

As the basis of personalized customization of ECS instances, the UserData function of Alibaba Cloud allows you to customize the startup behaviors of an ECS instance and to pass data into an ECS instance.

UserData is mainly implemented via different types of user-defined scripts. Before this function was introduced, the initially started ECS instances can be thought as having the same environment and configuration. With UserData, enterprises or individuals can enter effective UserData as needed and the initially started instances have the configuration you need.

How to use it

- UserData-Scripts: suitable for users who need to initialize instances by running shell scripts, starting with `#!/bin/sh`. In practice, most of the users use this method to enter UserData and it is fit for relatively complex deployments.
- Cloud-Config: a unique format supported by cloud-init. With this method, the common personalized configuration is packed in a YAML file, thus implementing the common configuration more conveniently. The first line is `#cloud-config`, followed by an associative

array. This method provides such keys as ssh_authorized_keys, hostname, write_files, manage_etc_hosts, and so on.

Ideal for

- SSH authentication
- Updating and configuring software resources
- DNS configuration
- Installing and configuring applications

2.4.1.4 Instance lifecycle

The lifecycle of an ECS instance begins with creation and ends with release. During this process, an instance may undergo several status changes as explained in *Table 2-4: Inherent statuses of an instance lifecycle*.

Table 2-4: Inherent statuses of an instance lifecycle

Status	Status attribute	Description	Corresponding API status
Preparing	Transitory	An instance is in this status during its time of creation, but before running. If an instance is in this status for a long time, an exception occurs.	Pending
Starting	Transitory	An instance is in this status after it is started or restarted in the console or by using an API until it is running. If an instance is in this status for a long time, an exception occurs.	Starting
Running	Stable	The instance is running properly and can accommodat e your business needs.	Running
Stopping	Transitory	An instance is in this status after the stop operation is performed in the console or by using an API but before it actually stops. If an instance is in this status for a long time, an exception occurs.	Stopping
Stopped	Stable	The instance has been stopped properly. In this status , the instance cannot accommodate external services.	Stopped
Re-initializing	Transitory	An instance is in this status after the system disk and/ or data disk is re-initialized in the console or by using an API until it is running. If an instance is in this status for a long time, an exception occurs.	Stopped

Status	Status attribute	Description	Corresponding API status
Replacing System Disk	Transitory	An instance is in this status after the operating system is replaced in the console or by using an API until it is running. If an instance is in this status for a long time, an exception occurs.	Stopped

2.4.1.5 ECS Bare Metal Instance

ECS Bare Metal (EBM) Instance is a new type of computing product that features both elasticity of virtual machines as well as performance and characteristics of physical machines. As a product completely developed by Alibaba Cloud, EBM Instances are based on the next-generation virtualization technology.

Compared with the previous generation of virtualization technology, the next-generation virtualization technology is innovative in that it not only supports the common virtual cloud server but also completely supports the nested virtualization technology. It retains the resource elasticity of common cloud servers and adopts nested virtualization technology such that it keeps the user experience of physical machines intact.

Advantages

EBM Instances deliver value to customers via technological innovation. Specifically, EBM Instances have the following advantages:

- **Exclusive computing resources**

  As a cloud-based elastic computing product, the EBM Instances outshine the existing physical machines regarding performance and isolation and provides exclusive computing resources without virtualization performance overheads or feature loss. EBM Instances support 8/16/32/96-core CPU and ultra-high clock speed. Take an EBM Instance with 8 cores for example. It supports an ultra-high clock speed ranging from 3.7 to 4.1 GHz, providing better performance and responsiveness for gaming and financial industries than similar products.

- **Encrypted compute**

  For the sake of security, the EBM Instances use a chip-level trusted execution environment (Intel® SGX) in addition to the physical server isolation. This allows the encrypted data to be computed only in a safe and trusted environment, and provides improved security for the customer data on the cloud. This chip-level hardware security protection provides a "safe box

" for the data of cloud users and allows users to control the entire data encryption and key protection procedures.

- **Any Stack on Alibaba Cloud**

  An EBM Instance combines the performance strengths and complete features of physical machines with the ease-of-use and cost-effectiveness of cloud servers. It better meets your demands for high-performance computing and helps you build new hybrid clouds. Thanks to the flexibility and elasticity of virtual machines as well as all the strengths of physical machines , it has the re-virtualization ability. As a result, offline private clouds can be seamlessly migrated to Alibaba Cloud without performance overhead that may arise from nested virtualization, thus providing a new option for you to move business to the cloud.

- **Heterogeneous instruction set processor support**

  EBM Instances adopt the virtualization 2.0 technology completely developed by Alibaba Cloud and supports ARM and other instruction set processors at no cost.

Configuration

The following table describes the configuration of EBM instances:

Table 2-5: EBM configuration

Item	Description
CPU	EBMG5, general-purpose ECS Bare Metal Instance type family is supported currently.
Memory	Scalable from 32 GiB to 384 GiB as needed. For the sake of compute performance, the ratio of vCPU to memory is 1:2 or 1:4.
Storage	Startup is allowed from the virtual machine image or cloud disk, realizing delivery in seconds.
Network	VPC is supported. Interconnectivity is available with ECS instances, GPU instances and other cloud products. Meanwhile, the performance and stability is comparable with that of physical servers.
Image	ECS instance image can be used directly.
Security	The same security policy and flexibility as the existing ECS instances.

2.4.2 Block storage

Alibaba Cloud provides a wide variety of block storage products, including elastic block storage based on the distributed storage architecture and local storage based on the local disks of physical servers.

Details are as follows:

- *Elastic block storage* is a low-latency, persistent, and highly reliable random block-level storage service for ECS users. It adopts a triplicate distributed system to provide 99.9999999% data reliability for ECS instances. Elastic block storage can be created, resized, and released at any time.
- *Local storage*, also called local disks, are the disks attached to the physical servers (host machines) where ECS instances are hosted. They provide temporary block-level storage for instances, featuring low latency, high random IOPS, and high I/O throughput. They are designed for business scenarios requiring high storage I/O performance.

Block storage, OSS, vs. NAS

Currently, Alibaba Cloud provides three types of data storage products, namely block storage, Network Attached Storage (NAS) and Object Storage Service (OSS).

Table 2-6: Types of data storage products

Type	Feature	Ideal for
Block storage	A high-performance and low-latency block-level storage product for ECS instances. It supports random reads/writes. You can format the block storage and create a file system as with a hard disk.	It can be used for data storage in most of the common business scenarios.
OSS	You can treat it as a massive storage space, which is suitable for storing massive unstructured data, including images, short videos, and audios generated on the Internet. You can access the data stored in the OSS anytime and anywhere by using APIs.	Generally, OSS is used for such business scenarios as Internet business website construction, separation of dynamic and static resources, and CDN acceleration.

Type	Feature	Ideal for
NAS	Like OSS, it is suitable for storing massive unstructured data. However, you must access the data by using standard file access protocols, such as the Network File System (NFS) protocol for the Linux system and the Common Internet File System (CIFS) protocol for the Windows system. You can set the permissions to allow different clients to access the same file at the same time.	NAS is applicable to such business scenarios as file sharing across departments in a company, radio and television non-linear editing, high-performance computing, and Docker.

2.4.2.1 Performance

The following contents describe the key measures and performance of block storage products.

2.4.2.1.1 Elastic block storage

Elastic block storage includes cloud disks and shared block storage, whose performance is detailed in the following contents.

Cloud disks

Note:

the following data is obtained with standard tests and configuration.

Table 2-7: Performance data

Block storage	SSD cloud disk	Ultra cloud disk	Basic cloud disk
Capacity of a single disk	32,768 GiB	32,768 GiB	2,000 GiB
Max. IOPS	20,000	3,000	Several hundreds
Max. throughput	300 MBps	80 MBps	20 ~ 40 MBps

Formulas to calculate performance of a single disk	IOPS = min{30 * capacity, 20,000}
	Throughput = min{50 + 0.5 * capacity, 300} MBps
		IOPS = min{1,000 + 6 * capacity, 3,000}
		Throughput = min{50 + 0.1 * capacity, 80} MBps
			N/A

Block storage	SSD cloud disk	Ultra cloud disk	Basic cloud disk
Data reliability	99.9999999%	99.9999999%	99.9999999%
API name	cloud_ssd	cloud_efficiency	cloud
Typical scenarios	• I/O-intensive applications • Big and medium -sized relational databases • NoSQL databases	• Small and medium -sized relational databases • Large-scale development and testing • Web server log	Applications with occasional access requests or low I/O loads

Shared block storage

Table 2-8: Performance Data

Parameters	SSD shared block storage	Ultra shared block storage
Capacity	• Single disk: 32,768 GiB • All disks attached to an instance: up to 128 TiB	• Single disk: 32,768 GiB • All disks attached to an instance: up to 128 TiB
Max. random read/write IOPS	30,000	5,000
Max. sequential read/write throughput	512 MBps	160 MBps
Formulas to calculate performance of a single disk	IOPS = min{40 * capacity, 30, 000}	IOPS = min{1,000 + 6 * capacity, 5,000}
	Throughput = min{50 + 0.5 * capacity, 512} MBps	Throughput = min{50 + 0.15 * capacity, 160} MBps
Typical scenarios	• Oracle RAC • SQL Server • Failover clusters • High-availability of servers	• High-availability of servers • High-availability of development and testing databases

2.4.2.1.2 Local storage

For the performance of local disks, see *Local storage*.

2.4.2.1.3 Testing

To test the performance of block storage, different tools are used based on the operating system of ECS instances:

- Linux instances: DD, fio and sysbench can be used.
- Windows instances: fio and Iometer can be used.

The following contents describe how to perform the test by taking Linux instances and fio for example. Before the test, make sure the block storage is 4 KiB aligned.

Note:

Testing bare disks can obtain more accurate performance data, but damage the structure of the file system. For this reason, back up your data before testing. It is recommended that you perform the test on an ECS instance without data on its disks.

- Test random write IOPS

  ```
  fio -direct=1 -iodepth=128 -rw=randwrite -ioengine=libaio -bs=4k -size=1G -numjobs=1 -runtime=1000 -group_reporting -filename=iotest -name=Rand_Write_Testing
  ```

- Test random read IOPS

  ```
  fio -direct=1 -iodepth=128 -rw=randread -ioengine=libaio -bs=4k -size=1G -numjobs=1 -runtime=1000 -group_reporting -filename=iotest -name=Rand_Read_Testing
  ```

- Test write throughput

  ```
  fio -direct=1 -iodepth=64 -rw=write -ioengine=libaio -bs=64k -size=1G -numjobs=1 -runtime=1000 -group_reporting -filename=iotest -name=Write_PPS_Testing
  ```

- Test read throughput

  ```
  fio -direct=1 -iodepth=64 -rw=read -ioengine=libaio -bs=64k -size=1G -numjobs=1 -runtime=1000 -group_reporting -filename=iotest -name=Read_PPS_Testing
  ```

The following table describes the parameters of fio for the above test.

Parameters	Description
-direct=1	Ignore I/O buffer when testing. Data is written directly.
-rw=randwrite	Read and write policies. Available options: randread (random read), randwrite (random write), read (sequential read), write (sequential write), and randrw (random read and write).

Parameters	Description
-ioengine=libaio	Use libaio as the testing method (Linux AIO, Asynchronous I/O). An application can use I/O in two ways: synchronous and asynchronous. Synchronous I/O only sends out one I/O request each time, and returns only after the kernel is completed. In this case, the iodepth is always less than 1 for a single job, but can be resolved by multiple concurrent jobs. Usually 16~32 concurrent jobs can fill up the iodepth. The asynchronous method uses libaio to submit a batch of I/O requests each time, thus reduces interaction times, and makes interaction more effective.
-bs=4k	The size of each block for one I/O is 4k. If not specified, the default value 4k is used.
-size=1G	The size of the testing file is 1 GiB.
-numjobs=1	The number of testing jobs is 1.
-runtime=1000	Testing time is 1,000 seconds. If not specified, the test will go on with the value specified for -size, and write data in -bs each time.
-group_reporting	The display mode of testing results. Group_reporting gives the statistics of each job, instead of showing statistics by different jobs.
-filename=iotest	The output path and name of the test files. After the testing, be sure to remove the relevant files to free up the disk space.
-name=Rand_Write_Testing	The name of the testing task.

2.4.2.2 Elastic block storage

Based on whether they can be attached to multiple ECS instances, elastic block storage products fall into:

- **Cloud disk**: One cloud disk can only be attached to one ECS instance in the same zone.
- **Shared block storage**: One shared block storage can be attached to 4 ECS instances in the same zone at the same time.

2.4.2.2.1 Cloud disks

Cloud disks can be categorized in two ways:

- **By performance**

  Cloud disks fall into basic cloud disks, ultra cloud disks and SSD cloud disks.

  - Basic cloud disks are intended for scenarios of low I/O loads, providing hundreds of IOPS for ECS instances.
  - Ultra cloud disks are intended for scenarios of medium I/O loads, providing up to 3,000 random IOPS for ECS instances.
  - SSD cloud disks are intended for I/O-intensive applications, providing stable high random IOPS.
- **By uses**

  Cloud disks fall into system disks and data disks.

  - System disks cannot be accessed in a shared mode. They are created and released along with instances and have a lifecycle that is the same as their ECS instance.
  - Data disks can be created either along with ECS instances or separately, and cannot be accessed in a shared mode. If created along with ECS instances, they have a lifecycle that is the same as their ECS instance and are released along with instances. If created separately, they can be released separately or along with ECS instances depending on your configuration. The capacity of a data disk is determined by the disk type. For details, please refer to *Performance*.

2.4.2.2.2 Shared block storage

Shared block storage is a block level data storage service with high level of concurrency, performance, and reliability. It supports concurrent reads/writes to multiple ECS instances. It delivers the data reliability of up to 99.9999999%. One shared block storage can be attached to 4 ECS instances at the same time.

Shared block storage can only be used as data disks and created separately. Shared access is allowed. Shared block storage can be configured to be released with the ECS instances.

Based on their performance, the shared block storage can be divided into:

- **SSD shared block storage**: It adopts SSD as the storage medium to provide stable and high-performance storage with enhanced random I/O and data reliability.
- **Ultra shared block storage**: It adopts the hybrid media of SSD and HDD as the storage media.

When used as data disks, shared block storage shares the data disk quota with cloud disks, that is , up to 16 data disks can be attached to one ECS instance.

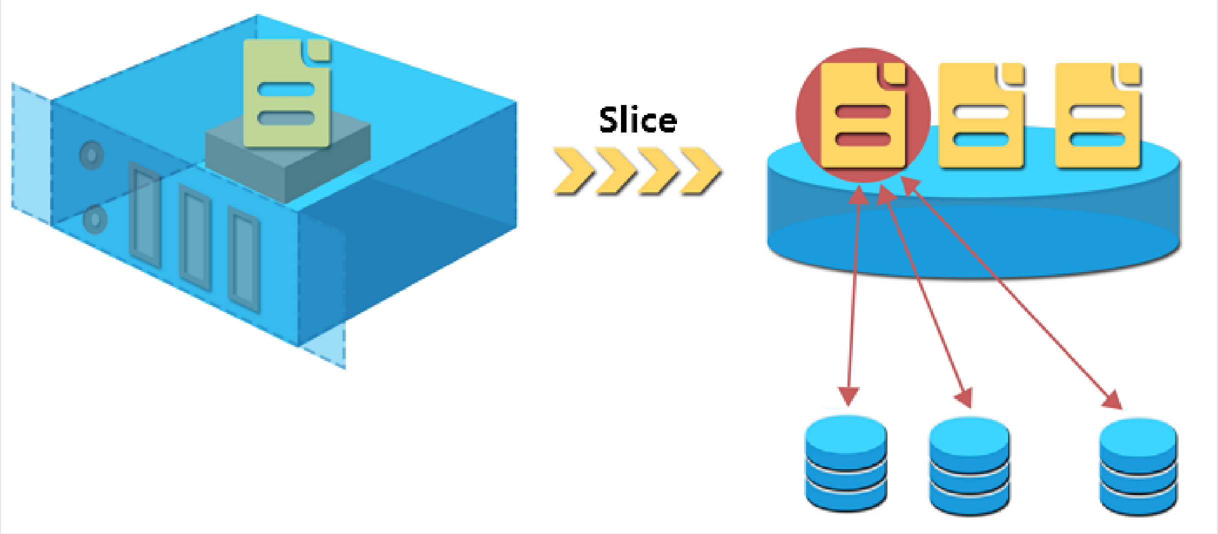
2.4.2.2.3 Triplicate technology

The Alibaba Cloud Distributed File System provides stable and efficient data access and reliability for ECS.

Chunks

ECS users' reads/writes to virtual disks are mapped to the reads/writes to the files on the file platform of Alibaba Cloud. The Distributed File System of Alibaba Cloud uses a flat design in which a linear address space is divided into slices, also called chunks. Each chunk has three copies stored on different server nodes on different racks, thus guaranteeing data reliability.

Figure 2-2: Triple replication of data



How triplicate technology works

Triplicate technology involves three key components: Master, Chunk Server, and Client. To demonstrate how triplicate technology works, in this example, the write operation of an ECS user undergoes several conversions before being executed by the Client. The process is as follows:

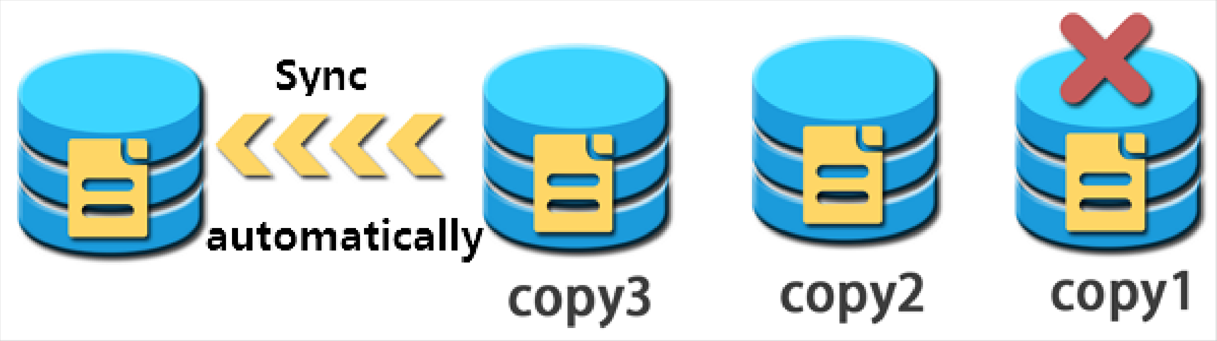
1. The Client determines the location of a chunk corresponding to one of your write operations.
2. The Client sends a request to the Master to query the storage locations (that is, the Chunk Servers) of the three copies of the chunk.
3. The Client sends write requests to the corresponding three Chunk Servers according to the results returned from the Master.
4. The Client returns a message to the user indicating whether the operation was successful.

The distribution strategy of the Master takes into account such factors as the disk usage of all the Chunk Servers in a rack, how they are distributed in different racks, availability of power supply and machine workloads, thereby guaranteeing that all the copies of a chunk are distributed on different Chunk Servers on different racks. This approach effectively reduces the potential of total data loss caused by failure of a Chunk Server or a rack.

Data protection

If a system failure occurs because of a corrupted node or hard drive failure, some chunks may lose one or more of the three valid chunk copies associated with them. If this occurs and triplicate technology is enabled, the Master replicates data between Chunk Servers to restore the missing chunk copies across different nodes.

Figure 2-3: Auto sync. of data



As described above, whenever users add, modify or delete data on the cloud disks, their operations are synchronized to the three copies. By doing so, the reliability and consistency of users' data is guaranteed.

For data loss in ECS instances caused by virus, operational mistakes or hackers, solutions include backup, snapshots, and so on. However, there is no single technology that solves all the issues and it is important to take appropriate measures to protect your data based on the actual situation.

2.4.2.3 ECS disk encryption

As a simple and secure encryption method, ECS disk encryption encrypts newly created cloud disks. You do not have to create, maintain, or protect your own key management infrastructure, nor change any of your existing applications or O&M processes. In addition, no extra encryption /decryption operations are required, so your business is not impacted by the disk encryption function.

After an encrypted cloud disk is created and attached to an ECS instance, the data in the following list can be encrypted:

- Data on the cloud disk.
- Data transmitted between the cloud disk and the instance. However, data in the instance operating system is not encrypted.
- All snapshots created from the encrypted cloud disk, which are called encrypted snapshots.

Encryption and decryption are performed on the host that runs the ECS instance, so the data transmitted from the ECS instance to the cloud disk is encrypted.

ECS disk encryption supports all available cloud disks (basic cloud disks, ultra cloud disks, and SSD cloud disks) and shared block storage (ultra and SSD) in a VPC.

2.4.2.4 Local storage

Local storage, also called local disks, refers to the disks attached to the physical servers (host machines) where ECS instances are hosted. They provide temporary block level storage for instances, featuring low latency, high random IOPS, and high I/O throughput. They are designed for business scenarios requiring high storage I/O performance.

Because a local disk is attached to a single physical server, the data reliability depends on the reliability of the physical server, which is subject to the single point failure. We recommend that you implement data redundancy at the application layer to guarantee data availability.

Note:

Using a local disk for data storage comes with the risk of losing your data in some cases, such as when the host machine is down. Therefore, never store any business data that requires long-term persistence on a local disk. If no data reliability architecture is available for your application, we strongly recommend that you build your ECS with cloud disks or shared block storage.

Categories

Currently, Alibaba Cloud provides two types of local disks:

- **Local NVMe SSD**: This disk is used together with instances of the following type families: gn5 and gal.
- **Local SATA HDD**: This disk is used together with instances of the d1ne and d1 type families. It is applicable to the Internet, finance, and other allied industries that require big data computing and storage analysis for massive data storage and offline computing. It fully meets the needs

of distributed computing business models represented by Hadoop regarding instance storage performance, capacity, and Intranet bandwidth.

SATA HDD

The following table lists the performance of local SATA HDD of a d1ne or d1 ECS instance.

Table 2-9: Performance

Parameters	SATA HDD
Capacity	• Single disk: 5,500 GiB • Total capacity per instance: 154,000 GiB
Throughput	• Single disk: 190 MBps • Total throughput per instance: 5,320 MBps
Access latency	In milliseconds

2.4.3 Images

An image is a running environment template for ECS instances. It includes an operating system and in some cases, pre-installed software. An image file is like a duplicate that contains all the data of one or more disks. For ECS, such disks can be a single system disk or the combination of a 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 various types of images for you to easily access image resources.

Table 2-10: Image types

Type	Description
Public image	Public images officially provided by Alibaba Cloud support nearly all mainstream Windows and Linux versions. Including: • Windows • CentOS • CoreOS • Debian • Gentoo • FreeBSD • OpenSUSE

Type	Description
	• SUSE Linux • Ubuntu
Custom image	Custom images are created based on your existing physical server, virtual machine, or cloud host. These images can meet your personalized needs flexibly.

How to obtain an image

You can obtain an image for your ECS instance by:

- Creating a custom image based on an existing ECS instance.
- Choosing an image shared from other accounts.
- Importing a local image file into an ECS cluster to generate a custom image.
- Copying a custom image to other regions to maintain a consistent deployment of environment and application across multiple regions.

Image format

Currently, ECS supports images in VHD and RAW formats. You must convert other formats into VHD or RAW to use them in ECS. For how to convert a format, please refer to **How to convert the image format** in *Cite LeftECS User GuideCite Right*.

2.4.4 Snapshots

A snapshot is a copy of data on a disk created at a specific point in time. In reality, you may have the following needs:

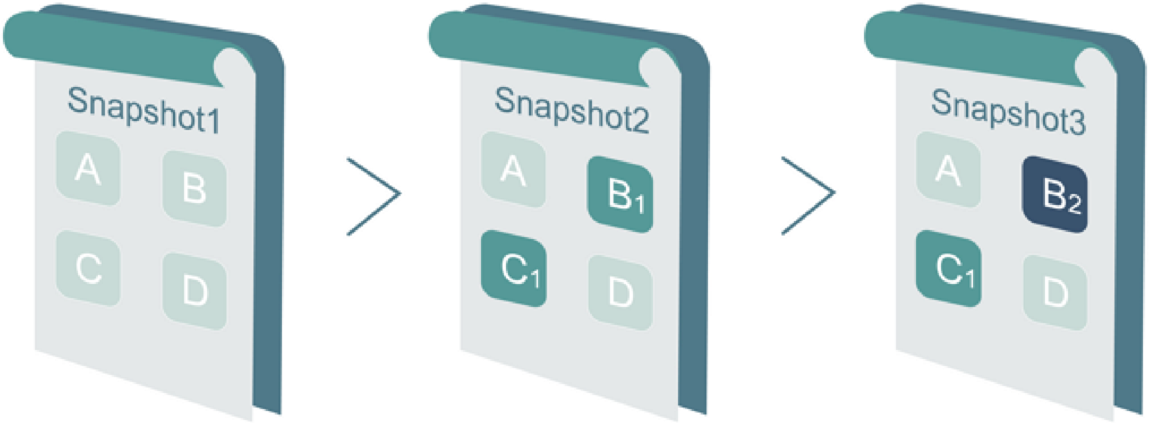
- When writing or storing data on multiple disks, you want to use snapshot data from one disk as the bassis for other disks.
- While cloud disks represent a secure way to store data, the data on them may be subject to errors (for example, data errors due to application errors or malicious read/write by hackers), which requires other mechanism to safeguard your data. For this reason, you may want to use snapshots to restore data to a previous point in time even if cloud disks are used.

2.4.4.1 Incremental snapshot mechanism

A snapshot is a copy of data in a certain point in time. You can create snapshots as scheduled to guarantee business continuity.

Snapshots are created in an incremental mode, that is, only data changes between two snapshots are copied. *Figure 2-4: Schematic diagram of snapshot* shows the incremental snapshot process.

Figure 2-4: Schematic diagram of snapshot



As shown above, Snapshot1, Snapshot2, and Snapshot3 are the first, second, and third snapshot of a disk respectively. During the snapshot creation process, the Distributed File System checks the disk data by blocks. Only the blocks with changed data are copied to the snapshot. In the preceding figure:

- In Snapshot 1, all data on the disk is copied because it is the first disk snapshot.
- Snapshot 2 only copies the changed data blocks B1 and C1. Data blocks A and D only reference their counterparts in Snapshot 1.
- Snapshot 3 copies the changed data block B2 but references data blocks A and D from Snapshot 1 and data block C1 from Snapshot 2.
- When you roll back the disk to Snapshot 3, blocks A, B2, C1, and D are copied to the disk to replicate Snapshot 3.
- When you delete Snapshot 2, block B1 is deleted, but block C1 is retained because blocks that are referenced by other snapshots cannot be deleted. When you roll back a disk to Snapshot 3 , block C1 is recovered.

Note:

Snapshots are stored in the Object Storage Service (OSS), but are invisible to users. They do not consume the bucket space in OSS. Snapshot operations can only be performed by using the ECS console or APIs.

2.4.4.2 ECS Snapshot 2.0

Built on original basic snapshot features, ECS Snapshot 2.0 data backup service provides a higher snapshot quota and more flexible automatic snapshot policies, further reducing its impact on business I/O. The features of ECS Snapshot 2.0 are described in the following table.

Table 2-11: Snapshot 2.0 vs. original Snapshot

Feature	Original snapshot specifications	Snapshot 2.0 specifications	Benefits of using ECS Snapshot 2.0	Comments
Snapshot quota	Number of disks * 6 + 6.	64 snapshots for each disk.	Longer protection circle; smaller protection granularity.	• Snapshot backup of a data disk for non-core business occurs at 00:00 every day. The backup data is retained for over two months. • Snapshot backup of a data disk for core business occurs every four hours. The backup data is retained for over 10 days.
Automatic snapshot policy	Triggered once daily by default. It cannot be modified.	Customizable weekly snapshot day, time of day, and snapshot retention period. You can query the quantity and details of disks associated with an automatic snapshot policy.	More flexible protection policy.	• You can take snapshots on the hour and several times in a day. • You can choose any day of the week as the recurring day for taking snapshots. • You can specify the snapshot retention period or choose to retain it permanently (when the maximum number of automatic snapshots is reached, the oldest automatic snapshot is automatically deleted).
Implementation	COW (Copy-On-Write).	ROW (Redirect-On-Write).	Mitigated performance impact of the	No interruption to your business, allowing snapshots to be taken at any time.

Feature	Original snapshot specifications	Snapshot 2.0 specifications	Benefits of using ECS Snapshot 2.0	Comments
			snapshot task on business I/O writes.	

2.4.4.3 ECS Snapshot 2.0 vs. traditional storage products

Alibaba Cloud ECS Snapshot 2.0 has many advantages as compared with the snapshot feature of traditional storage products, as described in the following table.

Table 2-12: Comparison of technical advantages

Item	ECS Snapshot 2.0	Traditional snapshot products
Capacity	Unlimited capacity that meets the requirements of protecting ultra-large-scale business data.	Limited capacity, often determined by the initial storage device capacity, which meets the requirements of protecting the core business data.

2.4.5 Deployment sets

You may require higher reliability and performance when buying multiple ECS instances in the same zone, for example:

- **Improve business reliability**

  To avoid interrupting business when a physical host, rack or switch malfunctions, you may want the same instance to be distributed on different hosts, racks or switches.

- **Improve business network performance**

  There are many interactions among instances in some business scenarios. You may want to have a lower network latency and higher network bandwidth. In this case, it is necessary to put instances together on one switch to meet your needs.

ECS provides deployment sets for you to perceive the physical topology of hosts, racks and switches, and to choose appropriate deployment policies according to your business types to improve overall business reliability and performance.

Deployment granularities and policies

- **Deployment granularities**
  - Host: means that the minimum scheduling granularity is a physical server. It is also the default value.
  - Rack: means that the minimum scheduling granularity is a rack.
  - Switch: means that the minimum scheduling granularity is a network switch.
- **Deployment policies**
  - LooseAggregation
  - StrictAggregation
  - LooseDispersion
  - StrictDispersion

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

For deployment policies and business scenarios that correspond to each deployment granularities, see the *Table 2-13: Granularities and policies*.

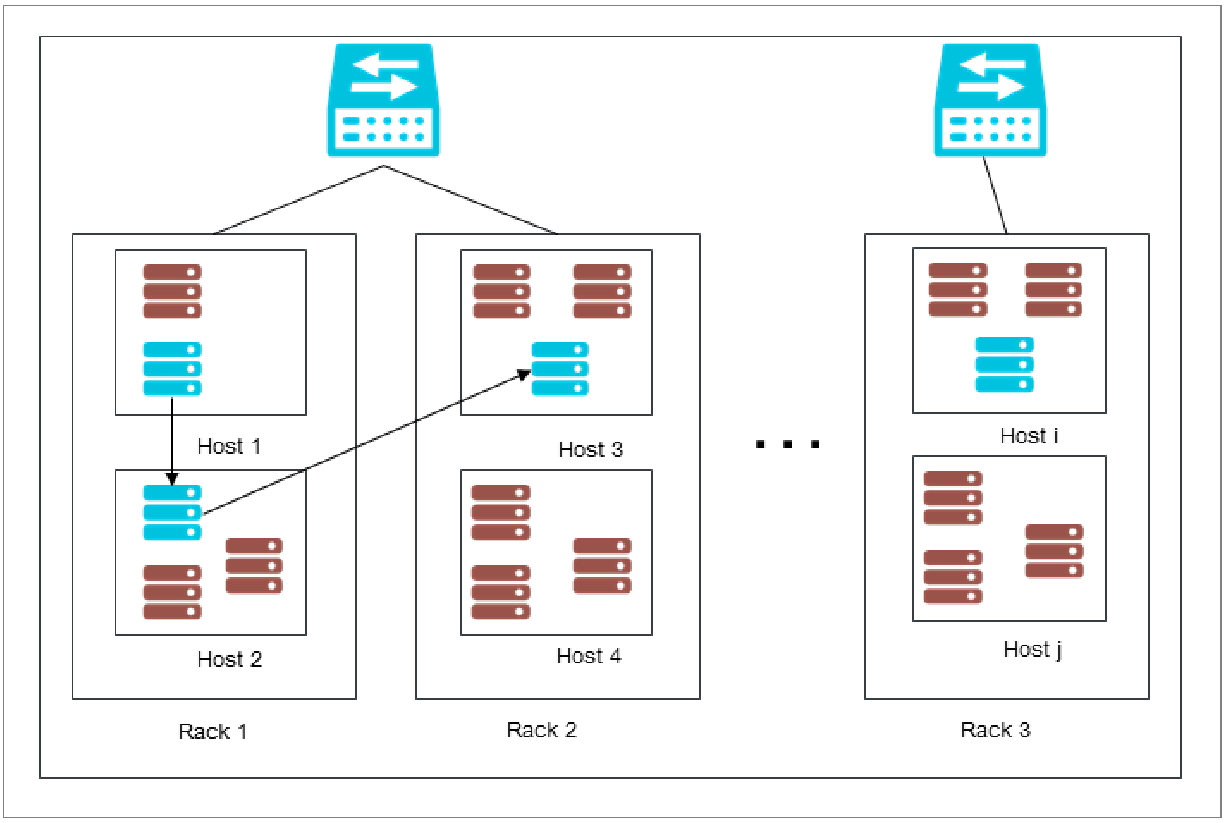
Table 2-13: Granularities and policies

Deployment granularity	Deployment policy	Business scenario
Host	StrictDispersion	General
	LooseDispersion	
Rack	Strict Distribution	Big data, Databases
	LooseDispersion	Game clients
Switch	StrictDispersion	VPN
	LooseDispersion	Game clients
	StrictAggregation	Big data, Databases
	LooseAggregation	Game clients

Example

The figure below shows a typical example that improves business reliability through deployment sets. The three ECS instances of the user are distributed on three physical hosts on at least two racks.

Figure 2-5: Example



Note:

For specific APIs related to deployment sets, see **Deployment sets** in *Cite LeftECS User GuideCite Right*.

2.4.6 Network and security

2.4.6.1 IP address of a VPC

This section introduces the supported types of IP addresses and specific application scenarios.

IP address types

ECS instances have the following two types of IP address.

- **Private IP**

A private IP is assigned when you create an ECS instance according to the VPC and switch segment that the instance belongs to.

- **Elastic IP (EIP)**

  An EIP is a public IP address that you can apply for separately.

Application scenarios

- **Private IP**: A private IP is used to access Intranet. You can directly configure the private IP when you create an ECS instance.

  Note:

  The system automatically assigns a private IP if it is not configured.

- **Elastic IP (EIP)**: An EIP is used to access Internet. You need to separately associate an EIP with the ECS instance that you have created. For specific operations, see **Elastic Internet IP** in *VPC User Guide*. An EIP can be applied for separately and kept for a long period. You can bind/unbind it from an instance, delete it separately or change its bandwidth.

2.4.6.2 Elastic network interface

This section briefly introduces the concept and application scenarios of elastic network interface.

Elastic network interface (ENI), also known as the auxiliary network interface, is a kind of virtual network interface that can be attached to ECS instances in a VPC. It can help you build highly available clusters, and implement low-cost failover and fine network management.

ENIs are applicable for the three scenarios:

- **Build highly available clusters**

  Meet the highly available architecture's demands for a single instance with multiple network interfaces.

- **Low-cost failover**

  By separating ENIs from an ECS instance and attaching them to another, you can rapidly migrate business traffic on a malfunctioning instance to a standby instance to resume the service.

- **Fine network management**

  An instance can have multiple ENIs such as ENIs for internal management and those for Internet business access, thus separating management data from business data. You can configure accurate security group rules for each ENI to ensure that their traffic is under secure access control.

ENI properties

The table below shows the ENI information.

Table 2-14: Property description

Property	Number
Master private IP address	1
MAC address	1
Security group	1 ~ 5
Description	1
Name of network interface	1

Limitations

ENIs can be created and then attached to or detached from instances. However, ENIs have the following limitations:

- For one account, the maximum number of ENIs that can be created in one region is 100.
- ECS instances and ENIs must be in the same zone of the same VPC, but can belong to different switches.
- For instance types that support ENIs and the number of ENIs that they support, see *Instance types*.
- The instance bandwidth will not increase by attaching multiple ENIs.

Note:

The instance bandwidth is determined by instance type.

2.4.6.3 Intranet

Currently, Alibaba Cloud servers communicate through the Intranet using 1 Gbit/s shared bandwidth for non-I/O optimized instances, and 10 Gbit/s shared bandwidth for I/O optimized instances. However, because the servers communicate over a shared network, the bandwidth may fluctuate.

Note:

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

If you need to transmit data between two ECS instances in the same region, Intranet communication is recommended. Intranet communication is also recommended for RDS, Server Load Balancer, and OSS as they communicate with the 1 Gbit/s shared bandwidth in Intranet as well.

Currently, RDS, Server Load Balancer, and OSS can communicate with ECS through Intranet directly so long as they are in the same region.

For ECS instances in a **VPC** of Intranet:

- For instances in the same region and VPC and under the same account, Intranet communication is enabled by default if they are in the same security group. If instances are in different security groups, Intranet communication can be enabled via authorization among security groups.
- For instances under the same account and region while in different VPCs, Intranet communication can be enabled via Express Connect.
- Private IP addresses of instances can be modified or replaced.
- Private and public IP addresses of instances do not support virtual IP (VIP) configuration.
- Instances of different network types cannot communicate with each other in Intranet.

2.4.6.4 Security group rules

Security group rules can allow or deny inbound or outbound traffic of the Internet and Intranet for ECS instances.

You can authorize or cancel security group rules at any time. Changes to security group rules are automatically applied to ECS instances associated with security groups.

When configuring security group rules, make sure the rules are concise. If you associate an instance with multiple security groups, hundreds of rules may apply to the instance, which may cause connection errors when you access the instance.

2.5 Application scenarios

ECS is a highly flexible solution since it can be used either independently as a simple web server, or with other Alibaba Cloud products, such as OSS and CDN, to provide advanced multimedia solutions. ECS can be used in the following scenarios:

Corporate websites and simple web applications

In the initial stage, corporate websites have low traffic volumes and require only low-configuration ECS instances to run applications, databases, storage files, and other resources. As your business expands, you can upgrade the ECS configuration and increase the number of ECS instances at any time. You no longer need to worry about insufficient resources during peak traffic.

Multimedia and large-traffic apps or websites

ECS can be used with OSS to store static images, videos, and downloaded packages, reducing storage fees. In addition, ECS can be used with CDN or Server Load Balancer to greatly reduce response time and bandwidth fees, thus improving availability.

Apps or websites with significant traffic fluctuations

Some applications like the 12306 website may encounter large traffic fluctuations within a short period. When ECS is used with Auto Scaling, the number of ECS instances is automatically adjusted based on traffic. This feature allows you to meet resource requirements at a low cost. ECS can be used with Server Load Balancer to implement a high availability architecture.

Databases

ECS supports databases with high I/O requirements. An I/O-optimized ECS instance with high configuration can be used with an SSD cloud disk to deliver high I/O concurrency and higher data reliability. Alternatively, multiple I/O-optimized ECS instances with lower configuration can be used with Server Load Balancer to build a highly available architecture.

2.6 Usage limitations

ECS has the following limitations:

- For a cloud server with 4 GiB or more RAM, select the 64-bit operating system (the 32-bit operating system has the 4GiB RAM addressing limitation).
- Windows 32-bit Operating System supports up to 4 vCPUs.
- Windows does not support instance types that have more than 64 vCPUs.
- ECS does not support virtual application installation or re-virtualization (for example, VMware).

- ECS does not support sound card applications (currently, only GPU instances support analog sound card) or directly loading external hardware device, such as hardware dongle, USB memory, external hard disk and bank U key.
- ECS does not support multicast protocol. If multicasting services are required, we recommend that you use unicast point-to-point method.

Besides the preceding limitations, others are mentioned in the following table.

Table 2-15: Other limitations

Type	Limitation description
Instance type	For specific limitations, see *Instance type families* and *Instance types*.
Block storage	**Type limitation** • Quota of system disks for one ECS instance: 1. • Quota of data disks for one ECS instance: 16. • Instances to which one shared block storage can be attached: 4. • Capacity of the system disk: 40 GiB ~ 500 GiB. • Capacity of one Basic Cloud Disk: 5 GiB ~ 2,000 GiB. • Capacity of one SSD Cloud Disk: 20 GiB ~ 32,768 GiB. • Capacity of one Ultra Cloud Disk: 20 GiB ~ 32,768 GiB. • Total capacity of one Ultra Block Storage: 32,768 GiB. **Usage limitations** • Only data disks can be encrypted, while system disks cannot. • Existing non-encrypted disks cannot be directly converted into encrypted disks. • Existing encrypted disks cannot be converted into non-encrypted disks either. • Snapshots generated from existing non-encrypted disks cannot be directly converted into encrypted snapshots. • Encrypted snapshots cannot be converted into non-encrypted snapshots either. • Images with encrypted snapshots cannot be shared. • Images with encrypted snapshots cannot be exported.
Quota of snapshots	Number of disks × 64.
Images	• Quota of accounts to share one custom image: 50. • Requirements of images for instance types: 32-bit images are not supported on an instance with 4 GiB or more RAM.

Type	Limitation description
Security groups	• The number of instances within one security group cannot exceed 1,000. If more than 1,000 instances need to access each other in Intranet, they can be divided into several security groups and access each other via mutual authorization. • Each instance can join up to 5 security groups. • Each account can have up to 100 security groups. • Each security group can have up to 100 security group rules. • Adjustment operation on security group will not interrupt your service continuity. • Security groups are stateful. If data packets are permitted in the outbound direction, so are the packets in the inbound direction.
ENI	For the number of ENIs that can be bound to different instance type families, see *Instance types*.
Instance UserData	Currently, the ECS UserData feature only supports VPC+I/O optimized instance. Besides, cloud-init should be installed in the image as UserData depends on the cloud-init service. For details, see **Install cloud-init** in *Cite LeftECS User GuideCite Right*.

2.7 Basic concepts

Cloud server

A simple and efficient cloud computing service with elastic processing capacity that is supported in Linux, Windows, and other operating systems.

Instance

An independent virtual machine that includes basic cloud computing components such as CPU, memory, operating system, bandwidth, disks, and so on.

Security group

A security group is a virtual firewall that has such functions as detecting the state, filtering packets , and setting up network access control for one or more cloud servers. Instances within one security group are interconnected. Instances between different security groups can only access each other through authorization between the two security groups.

Image

A running environment template for ECS instances. It generally includes an operating system and preinstalled software. There are three types of images: public images, custom images, and shared 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 time point. Snapshots are classified into Auto Snapshot and Manual Snapshot.

Cloud disk

A kind of indpendent disk that can be attached to any ECS instance in the same region and zone. Cloud disks are classified into Ultra Cloud Disk, SSD Cloud Disk and Basic Cloud Disk according to their perfomances.

Alibaba Cloud Block Storage (Block Storage)

A low-latency, persistent, high-reliability, block-level, random storage for ECS instances.

Throughput

The amount of data successfully transmitted through a network, device, port, virtual circuit, or other facilities within a unit time.

Performance Testing Service (PTS)

A world-class, powerful testing platform that simulates real-world business scenarios involving massive users to observe real world capabilities and identify limitations.

Alibaba Cloud Virtual Private Cloud (VPC)

An Alibaba Cloud Virtual Private Cloud (VPC) is a private network built and customized based on Alibaba Cloud. Full logical isolation is achieved between Alibaba VPCs. Users can create and manage cloud product instances, such as ECS, Intranet Server Load Balancer, and RDS in their own VPCs.

Private IP address

A connection address used to access the host on a private network.

GPU cloud server

Computing service based on GPU applications that is applicable for video decoding, graphics rendering, deep learning, scientific computing, and other scenarios. GPU cloud server features real-time, high-speed, concurrent computing and powerful floating-point computing capacity.

3 Object Storage Service (OSS)

3.1 What is OSS

Alibaba Cloud Object Storage Service (OSS) is a storage service that enables you to store, back up, and archive any amount of data in the cloud. OSS is a cost-effective, highly secure, and highly reliable cloud storage solution. It uses RESTful APIs and is designed for 99.999999999% (11 nines) durability and 99.99% availability. Using OSS, you can store and retrieve any type of data at any time, from anywhere on the web.

You can use API and SDK interfaces provided by Alibaba Cloud or OSS migration tools to transfer massive amounts of data into or out of Alibaba Cloud OSS. You can use the Standard storage class of OSS to store image, audio, and video files for apps and large websites. You can use the Infrequent Access (IA) or Archive storage class as a low-cost solution for backup and archiving of infrequently accessed data.

3.2 Benefits

Benefits of OSS over traditional storage

Item	OSS	User-created server storage
Reliability	• 99.9% service availability. • Automatic scaling without affecting external services. • 99.99999999% data persistence. • Automatic redundant data backup.	• Limited by hardware reliability. Traditional storage is prone to errors. If a disk has a bad sector, data may be irretrievably lost. • Manual data recovery is complex, requiring time and effort.
Security	• Provides enterprise-grade, multilevel security. • Multi-user resource isolation mechanisms; supports disaster recovery within the same region. • Provides authentication and authorization mechanisms, as well as whitelist, anti-leeching, and RAM account features.	• Cleaning equipment and black hole equipment must be purchased separately. • Security must be implemented independently.
Data processing capabilities	Provides an image processing function.	Server storage must be purchased and deployed separately.

More benefits of OSS

- Convenient and fast
  - Provides standard RESTful APIs, a wide range of SDKs, client tools, and a console. You can easily upload, download, retrieve, and manage massive amounts of data for websites and applications just as if you were using regular files in Windows.
  - The number and size of files are not limited. You can resize the storage space based on your needs, avoiding the storage scaling problems of traditional hardware.
  - Supports simultaneous stream writing and reading. Perfect for business scenarios where videos and other large files must be read and written simultaneously.
  - Supports lifecycle management. You can delete expired data in batches.
- Powerful and flexible security

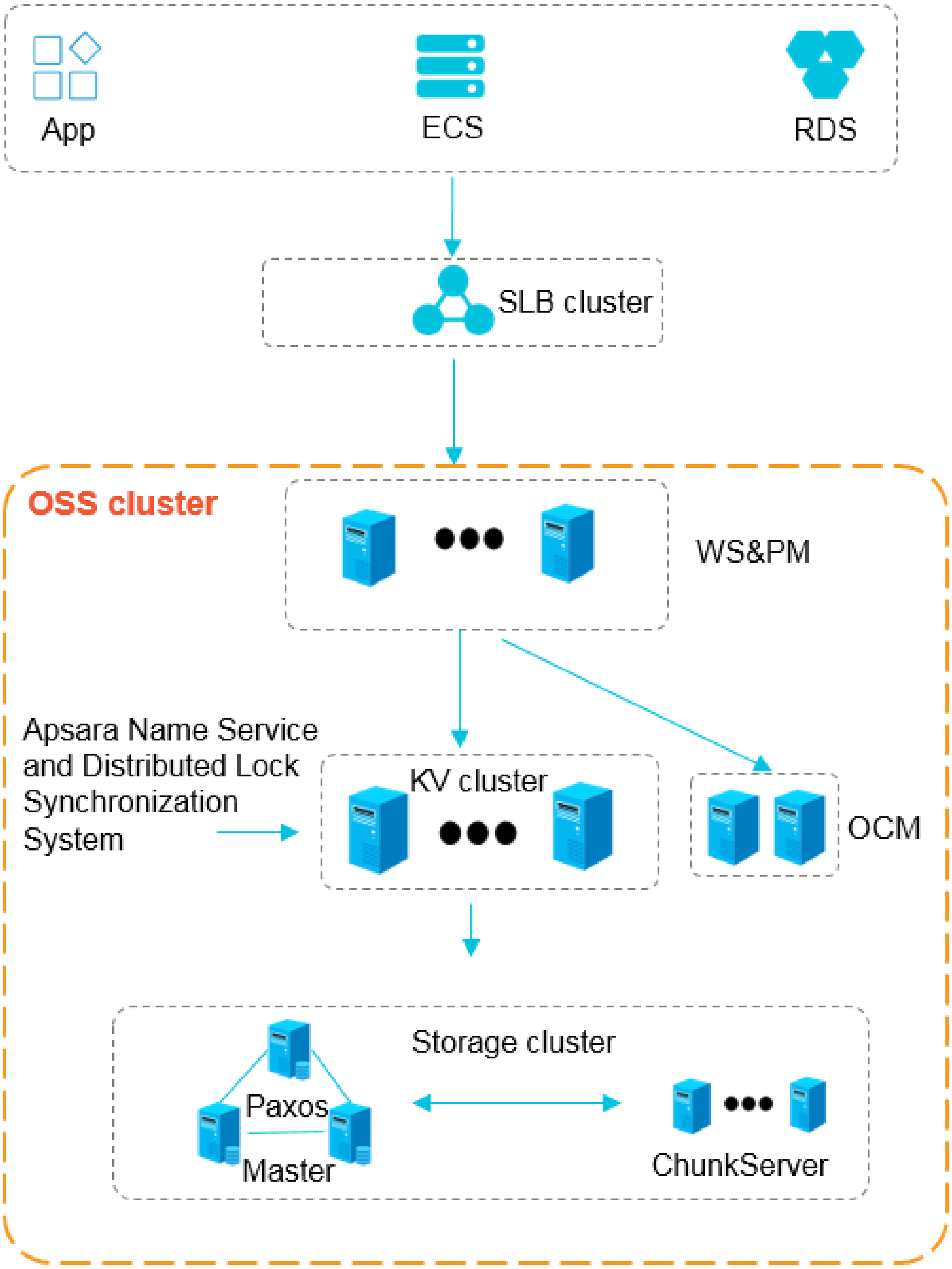
  Flexible authentication and authorization. OSS provides STS and URL authentication and authorization, as well as whitelist, anti-leeching, and RAM account features.
- Rich image service

  OSS supports Image Service which enables format conversion, thumbnails, cropping, watermarks, scaling, and other operations for a wide variety of file formats including JPG, PNG , BMP, GIF, WEBP, and TIFF.

3.3 Architecture

Object Storage Service (OSS) is a storage solution built on the Alibaba Cloud Apsara platform. It is based on infrastructure such as the Apsara distributed file system and distributed job scheduling system, and provides distributed scheduling, high-speed networks, and distributed storage features. *Figure 3-1: Architecture of OSS* shows the architecture of OSS.

Figure 3-1: Architecture of OSS



- WS&PM protocol layer: receives users' requests sent through the RESTful protocol and performing authentication. If authentication succeeds, users' requests are forwarded to the key-value engine for further processing. If authentication fails, an error message is returned.

- KV cluster: processes structured data, including reading and writing data based on the Key (object name). This layer also supports large-scale concurrent requests. When the running physical location of a service is changed due to changes from the service cluster, this layer can relocate the access point of the service.
- Storage cluster: Metadata is stored on the Masters, and the distributed message consistency protocol (Paxos) is adopted between Masters to ensure metadata consistency. In this way, efficient distributed file storage and access are achieved.

3.4 Functions

Table 8-1 describes key OSS functions:

Table 3-1: OSS functions

Category	Function	Description
Bucket	Create a bucket	Before uploading any file to OSS, you must create a bucket to store files.
	Delete a bucket	If you no longer need a bucket, delete it to avoid further fees.
	Modify read and write permissions for a bucket	OSS provides an Access Control List (ACL) for permission control. You can configure an ACL when creating a bucket and modify the ACL after creating the bucket.
	Set static website hosting	You can convert a bucket configuration to static website hosting mode and then access the static website through the bucket domain.
	Set anti-leech	To reduce extra fees due to undesired hotlinking of data on OSS,OSS supports anti-leeching based on the referer field in the HTTP header.
	Manage CORS	OSS provides Cross-Origin Resource Sharing (CORS) settings in the HTML5 protocol to help you achieve cross-origin access.
	Set the lifecycle	You can define and manage the lifecycle of some or all objects in the bucket. Lifecycle settings include operations such as multiple files management and automatic fragment deletion.
Object (file)	Upload files	You can upload all types of files to a bucket.

Category	Function	Description
	Create a folder	You can manage OSS folders the same way you manage Windows folders.
	Search for files	You can search for files with the same name prefix in a bucket or folder.
	Get file access addresses	You can use file access addresses to share and download files.
	Delete files	You can delete a single file or delete multiple files.
	Delete folders	You can delete a folder or delete multiple folders.
	Modify read and write permissions for a file	You can configure an ACL when uploading a file and modify the ACL after uploading the file.
	Manage parts	You can delete some or all part files in a bucket.
Image service	Image Service	You can perform operations such as format conversion , cropping, scaling, rotation, watermark, and style encapsulation on images stored on OSS.
OSS VPC access control	Single Tunnel	You can create single tunnels to access resources stored on OSS through VPC.
API	API	Provides RESTful API operations supported by OSS and related examples.
SDK	SDK	Provides SDK development operations and examples in mainstream languages.

3.5 Scenarios

OSS is applicable for the following scenarios:

Massive-scale storage for image and audio/video applications

OSS can be used to store massive-scale data, such as images, audios, videos, and logs. It supports various devices and direct data read/write to and from OSS by websites and apps. OSS supports file upload and streaming upload.

Static/dynamic resource separation for web pages and apps

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

Local data archiving

You can transfer internal data which needs to be locally archived for a long period to OSS, which is low in cost and high in availability.

3.6 Limits

Restricted item	Description
Bucket	• You can create a maximum of 10 buckets. • The name and region of a bucket cannot be modified after it is created.
File uploading	• The size of each file uploaded by console upload, simple upload, form upload, and append upload cannot be greater than 5 GB. To upload a file greater than 5 GB , you must use multipart upload. The size of each file uploaded by multipart upload cannot be greater than 48.8 TB. • You can upload files with same names but the existing files are overwritten.
File deleting	• Deleted files cannot be restored. • You can delete up to 50 files in batches in the console. To delete more files in batches, you must use APIs or SDKs.
Lifecycle	You can configure with up to 1,000 lifecycle rules for each bucket.
Image processing	• For the original image: — Only jpg, png, bmp, gif, webp, and tiff formats are supported. — File size cannot exceed 20 MB. — For the image rotation, the width or height of the image cannot exceed 4096. • For a thumbnail: — The product of the width and height cannot exceed 4096 x 4096. — The length of each side cannot exceed 4096.

3.7 Glossary

This section introduces the basic concepts of OSS.

Object

A discrete unit of data (sometimes known as a file).

An object is composed of:

- Metadata, known as Object Meta, which is a key-value pair that expresses the object’s attributes, such as its last modification time and size, as well as user-defined information.
- User data
  - A unique object name, known as a Key.

The size of an object varies with the upload method. Multipart Upload supports objects of up to 48. 8 TB. Other upload methods only support objects of up to 5 GB.

An object’s lifecycle starts from when it has been successfully uploaded, and ends when it has been deleted. During an object’s lifecycle, its information cannot be changed. If you upload an object with a duplicate name in a bucket, it will overwrite the existing one. Therefore, OSS does not allow users to modify only part of an object.

OSS provides an Append Upload function, which allows you to continually append data to the end of an object.

Bucket

A virtual division of object storage that, unlike file systems, manages objects in a flat structure.

Bucket properties are as follows:

- All objects must belong to a bucket. During an object’s lifecycle it remains directly affiliated with the corresponding bucket.
- A user can have multiple buckets, with each bucket able to contain an unlimited number of objects.
- You can set and modify the attributes of a bucket for region and object access control and object lifecycle management. The attributes apply to all objects in the bucket.
- You can create different buckets to perform different management functions.
- The name of a bucket must be globally unique within the OSS. Once a bucket name is created , it cannot be changed.

High consistency

In OSS, object operations are atomic, that is, operations must either succeed or fail without an intermediate status. After you upload an object, OSS ensures that it is complete. OSS will not return a partial success response when uploading objects.

Object operations in OSS are highly consistent. Once you receive a successful upload (PUT) response, this object can be read immediately, and the data is already available in triplicate. The same concept applies to delete operations. Once you delete an object, that object no longer exists .

This high-consistency feature facilitates your architectural design. The logic of OSS usage is the same as that of a traditional storage device: modifications are immediately committed and users do not have to consider consistency issues.

Comparison between OSS and file system

OSS is a distributed object storage service that uses a Key-Value pair format, whereas a file system uses a tree-type index structure of directories that contain files. In OSS, you can retrieve object content based on unique object names (Keys). In file systems, you can retrieve files based on their location in a directory.

The benefit of OSS is that it supports massive concurrent access volumes, which means large volumes of unstructured data (such as images, videos, and documents) can be stored and retrieved without excessive use of resources. The benefit of a file system is that folder operations such as renaming, moving, and deleting directories (which means renaming, moving, and deleting data) is considerably easier as data does not need to be copied and replaced.

The limitation of OSS is that saved objects cannot be modified. If an object needs modification , the entire object must be uploaded again to make the modification take effect. One exception is through using the append object operation, whereby users call a specific API that, allows a generated object be of a different type than normally uploaded objects. The limitations of a file system are that system performance is limited to a single device, and the more files and directories that are created in the system, the more resources are consumed, and the lengthier user processes become.

4 Table Store

4.1 What is Table Store

Table Store is a NoSQL database service built on Alibaba Cloud’s Apsara distributed file system that can store and access massive structured data in real time.

Table Store allows users to:

- Organize data into instances and tables that can seamlessly scale using data partitioning and load balancing.
- Shield applications from faults and errors that occur on the underlying hardware platform, providing fast recovery capability and high service availability.
- Manage data with multiple backups using solid state disks (SSDs), enabling quick data access and high data reliability.

4.2 Benefits

Scalability

- Dynamic adjustment of reserved read/write throughput

  When creating a table, you can configure the reserved read/write throughput for an application based on business access conditions. Table Store schedules and reserves resources based on the table's reserved read/write throughput. Then, Table Store can dynamically adjust the table’s reserved read/write throughput based on the application.
- Unlimited capacity

  There is no limit to the amount of data stored in Table Store tables. If a table size increases, Table Store will adjust the data partitions for immediate storage space allocation to the increased table.

Data reliability

Table Store stores multiple data copies across different servers in different racks. If one backup fails, servers with copied data will immediately restore services, achieving zero data loss.

High availability

Through automatic failure detection and data migration, Table Store shields applications from both hardware and network-related faults to deliver high availability.

Ease of management

Table Store automatically manages complex tasks, such as the management of data partitions, software/hardware upgrades, configuration updates, and cluster resizing, allowing you to focus on growing your business.

Access security

Table Store performs identity authentication for each application request, preventing unauthorized data access and ensuring data access security.

High consistency

Table Store ensures high consistency of writing data. Once a successful result is returned for a write operation, applications can read the latest data.

Flexible data models

Table Store tables do not require a fixed format. The column numbers of each row, and the value types in columns of the same name but different rows, can be different. Furthermore, Table Store supports multiple data types, such as Integer, Boolean, Double, String, and Binary.

Pay-As-You-Go

Table Store only charges fees based on the actual resources you have reserved and used.

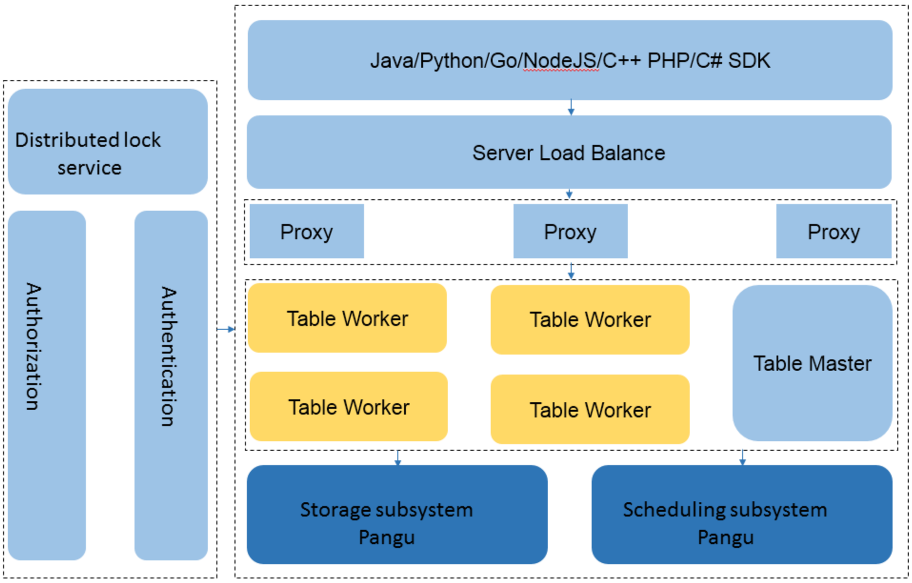
Monitoring integration

The Table Store console provides real-time monitoring information, including the requests number per second and the average response latency.

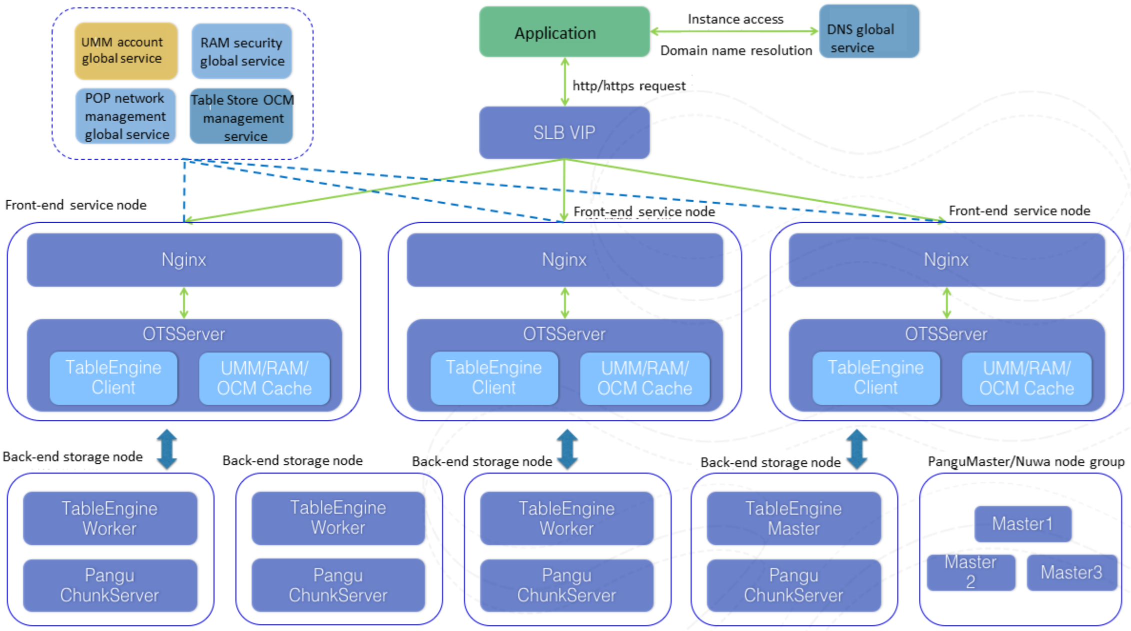
4.3 Architecture

The architecture of Table Store is very much similar to that of Google's BigTable. The log-structured merge-tree (LSM) storage engine is used to provide extremely high write performance. In addition, the primary key-based single-row query and range query can deliver stable and predictable performance, without being impacted by data size and access concurrency.

The following figure shows the underlying architecture of Table Store.



The following figure shows the detailed architecture of Table Store.



- The uppermost layer is the protocol access layer. Server Load Balancer distributes user requests to different proxy nodes. After receiving the requests sent by the user through the RESTful protocol, the proxy nodes perform authentication. If the authentication succeeds, the user requests are sent to the corresponding data engine for processing based on the value of the first primary key; if the authentication fails, an error message is returned to the user.

- Table Worker is located at the data engine layer and processes structured data, including reading and writing data based on the primary key. It also supports large-scale concurrent requests.
- The bottommost layer is the persistent storage layer, at which the large-scale distributed file system is deployed. Metadata is stored on the master nodes. A distributed message consistency protocol (Paxos) is adopted between master nodes to ensure the consistency of metadata. In this way, efficient distributed file storage and access are achieved to ensure that data has three copies in the system and can be recovered from any software or hardware faults. The design of Table Store provides a minimum of 99.9% availability and 99.99999999% data reliability.

4.4 Features

Data partition and load balancing

The first column of primary keys in each row of a table is referred to as a partition key. Based on the partition key range, the system partitions a table into multiple shards that are evenly dispatched to different storage nodes. If data in a partition increases and exceeds a threshold, the partition is automatically divided into two small partitions to where the access load is distributed. By spreading the access load across different nodes, the solution can achieve linear expansion of the data size and access pressure of a table. Since the storage is shared, data partitions become logic structures of data. Therefore, splitting of a partition does not involve actual data migration, and the theoretical delay caused by load balancing on the partition is hundreds of milliseconds.

Automatic recovery upon a single node failure

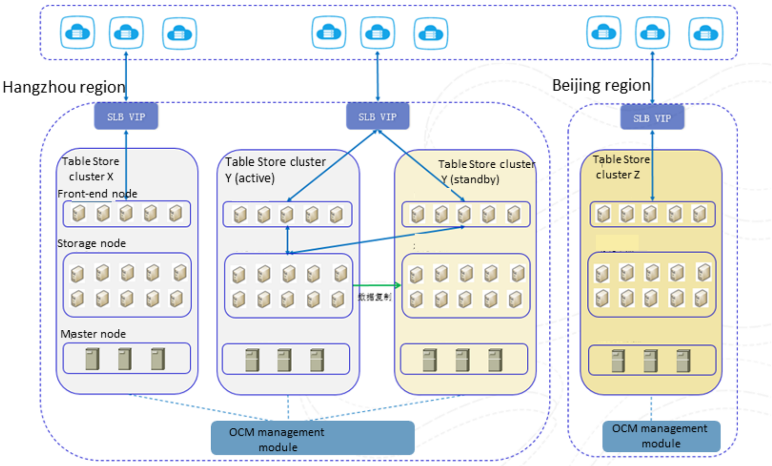
In the storage engine of Table Store, each node serves data partitions in different tables, and partition distribution and dispatching are managed by a master node that monitors the health of each service node as well. If a service node fails, the master node dispatches data partitions originally allocated to this faulty node to other healthy nodes. Migration of data partition is essentially a logical operation, and the distributed file system is deployed at the underlying layer, so the actual data is not migrated. Therefore, services can rapidly recover in case of a single node failure.

Intra-city and remote disaster tolerance

To meet business security and availability requirements, Table Store provides instance-based intra-city and remote active/standby disaster tolerance. Operations performed on a table on an active instance, including insertion, update, and deletion, are asynchronously replicated to the table with the same name on the standby instance. The duration of data synchronization between active and standby instances depends on the network environment of the active/standby cluster . In an ideal network environment, the delay is only several milliseconds. Therefore, before the manual switchover, resource access to the active cluster must be stopped. After all data is backed up, services can be switched over to the standby cluster. After the active/standby switchover, the switchover cannot be performed again within one hour. In addition, the original cluster data must be cleared and the standby cluster information must be set again.

In intra-city active/standby cluster scenarios, applications' domain names remain unchanged when they access Table Store in the active and standby clusters. That means the applications do not need to be changed after the switchover. In remote active/standby cluster scenarios, the service domain names of the active/standby clusters differ. After switchover, applications' domain names need to be changed.

The following figure shows the multi-cluster architecture for active/standby cluster disaster tolerance, in which the Table Store cluster Y (active) and Table Store cluster Y (standby) are used for disaster tolerance.



In the single-cluster architecture, Table Store keeps three copies for all data and ensures that the results are returned to users after all copies are written to the disk for high reliability of each cluster.

During intra-city and remote disaster tolerance, the service node of Table Store asynchronously sends data to the front-end service node of the standby cluster in replication mode based on the data writing time. After receiving the replication request, the front-end service node writes the data to the standby cluster by following the normal writing process.

In this disaster tolerance mode, data synchronization and switchover are the most important activities. When the active and standby clusters are initially built, data synchronization includes full synchronization and real-time incremental synchronization. After switchover, the standby cluster starts to provide services and re-establishes the synchronization relationship after the active cluster recovers. Finally, the services can be switched back to the active cluster.

Note that data synchronization here is an asynchronous process. A piece of data is returned to the client after it is successfully written to the active cluster. At this time, data may not be synchronized to the standby cluster. Asynchronous synchronization may cause inconsistency of some data . Therefore, service personnel must understand the impact of such data inconsistency and take corresponding measures.

Specifically, assume that some data on the active cluster is not synchronized to the standby cluster yet when active/standby switchover is implemented upon a failure. In this case:

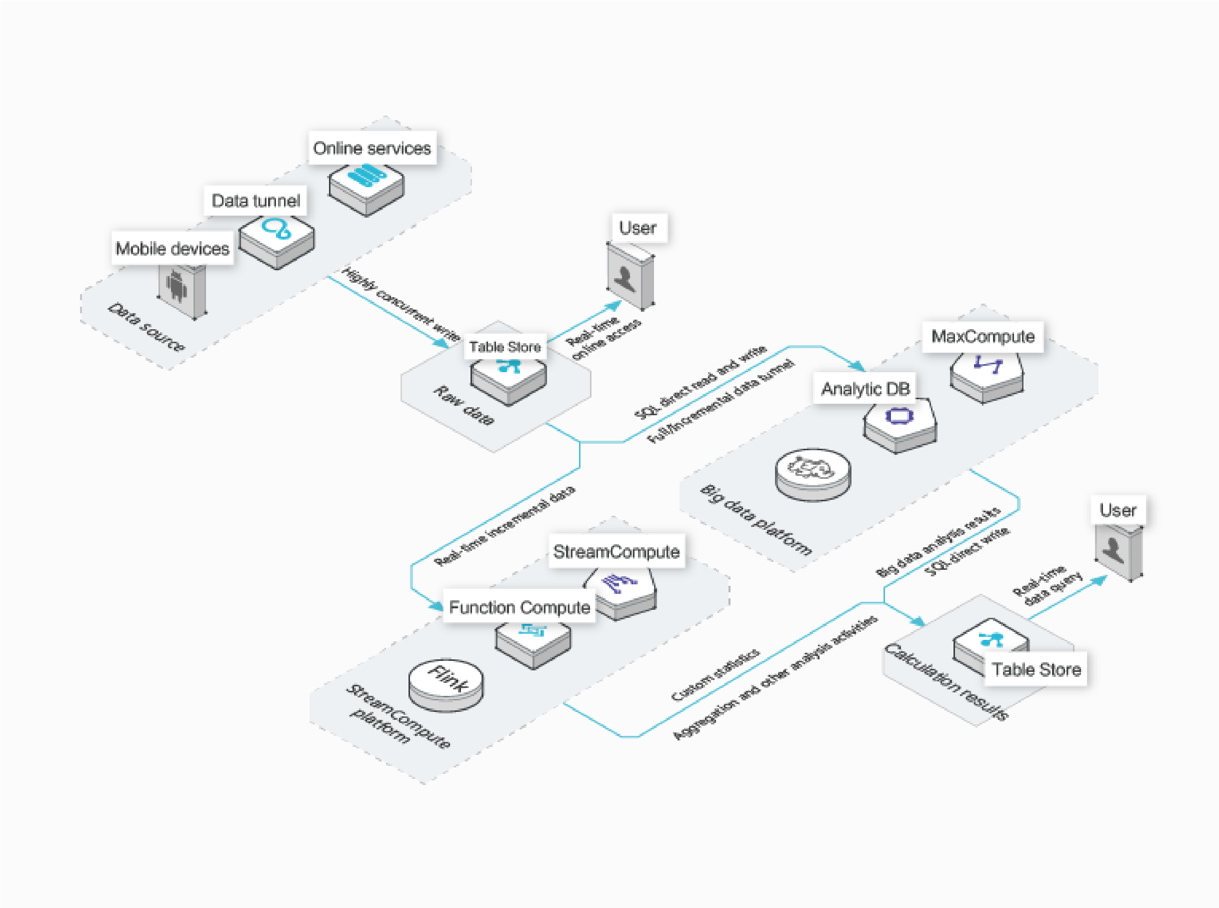
- The original active cluster has all data written before the switchover, including the data that is not synchronized to the standby cluster before the switchover.
- Except the data that is not synchronized before the switchover, the new active cluster (that is, the original standby cluster) has all written data.

In normal network conditions, Table Store active/standby clusters provide the recovery point objective (RPO) of one minute and the recovery time objective (RTO) of five minutes.

4.5 Scenarios

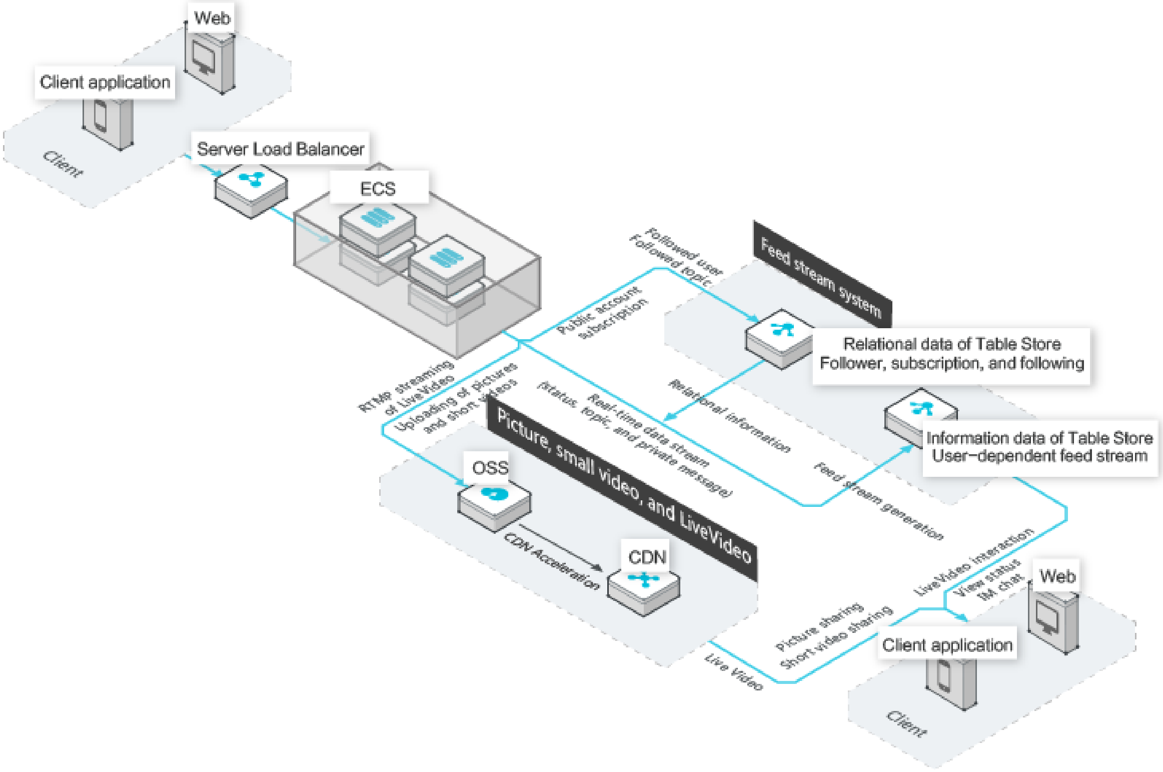
Big data storage and analysis

Table Store provides low-cost, low-latency, and high-concurrency storage and online access of high volumes of data. In addition, Table Store provides incremental and full data tunnels, and also SQL direct read and write on big data analysis platforms, such as MaxCompute. An efficient incremental streaming read interface is provided for easy computing of real-time data streams.



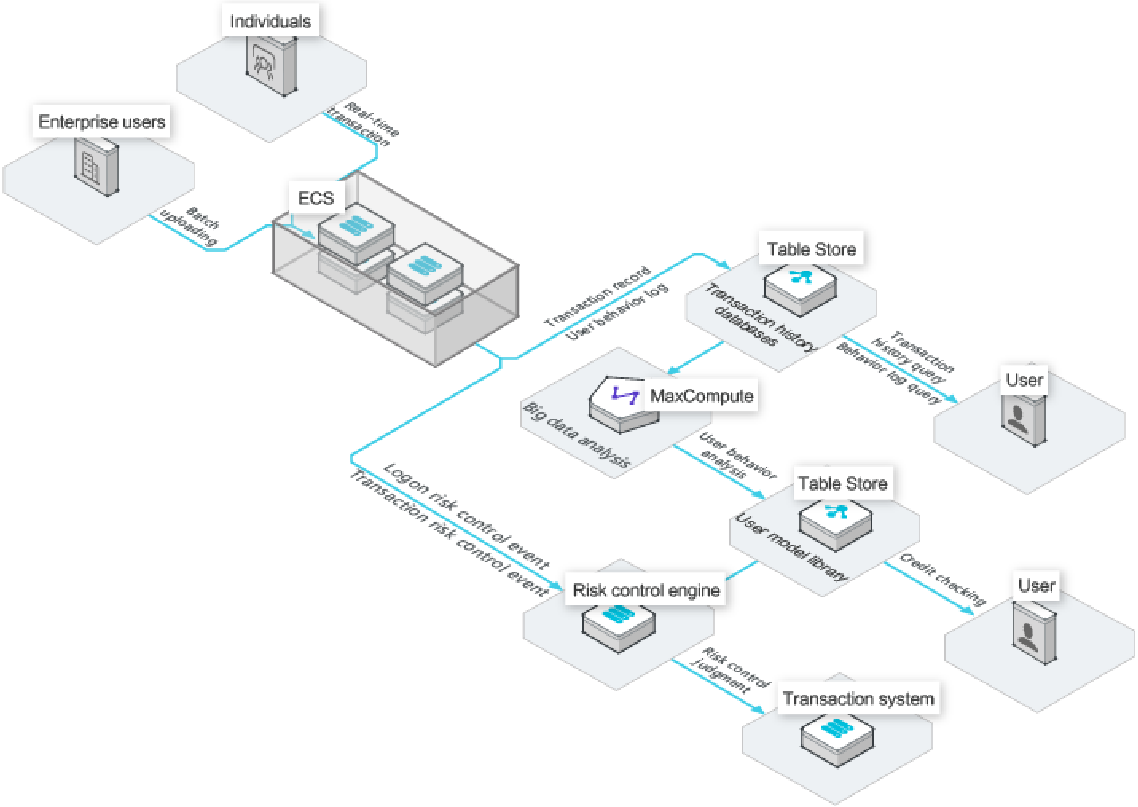
Social feed stream storage

Table Store can store massive volumes of social information produced by interactions between people, including IM chats, comments, threads, and likes. It stores images and videos on OSS and, with CDN acceleration, provides optimal user experience. Table Store can meets the needs of applications that feature significant traffic fluctuations and high concurrency when low latency is required.



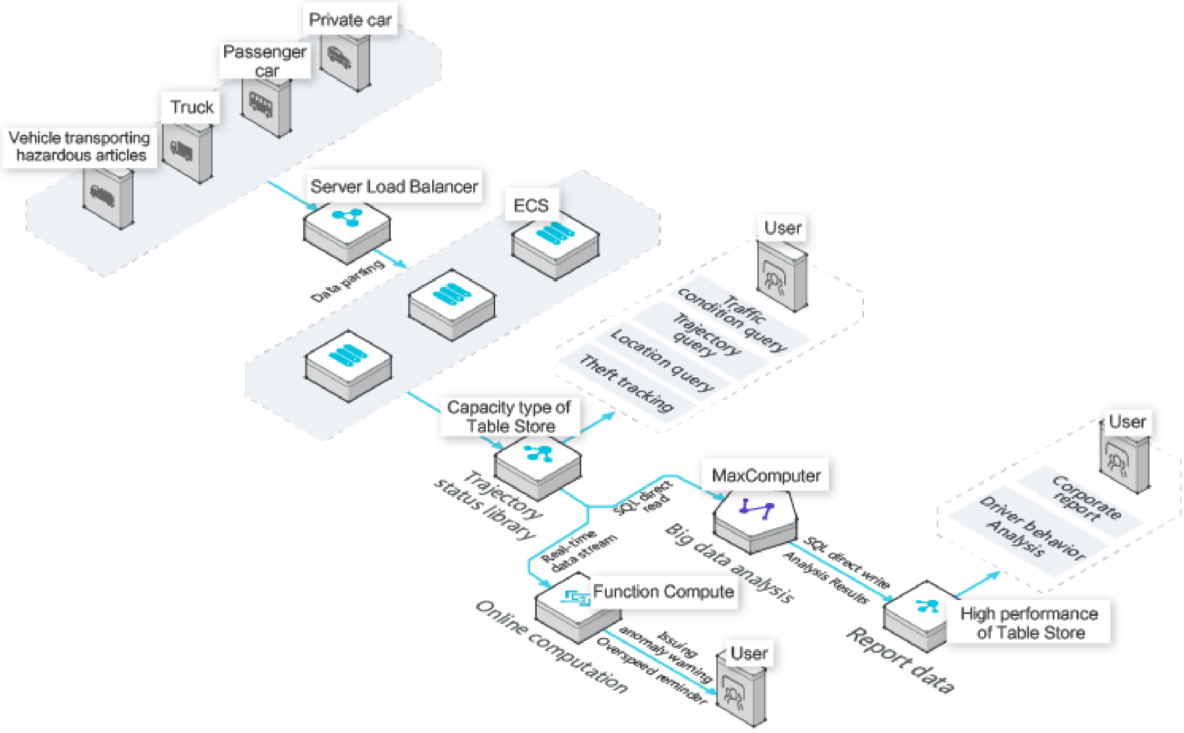
Financial risk control

The advantages of Table Store, such as low latency and high concurrency, optimize the risk control system, allowing you to strictly control over transaction risks. Flexible data structures enable fast iteration of business models as market needs shift.



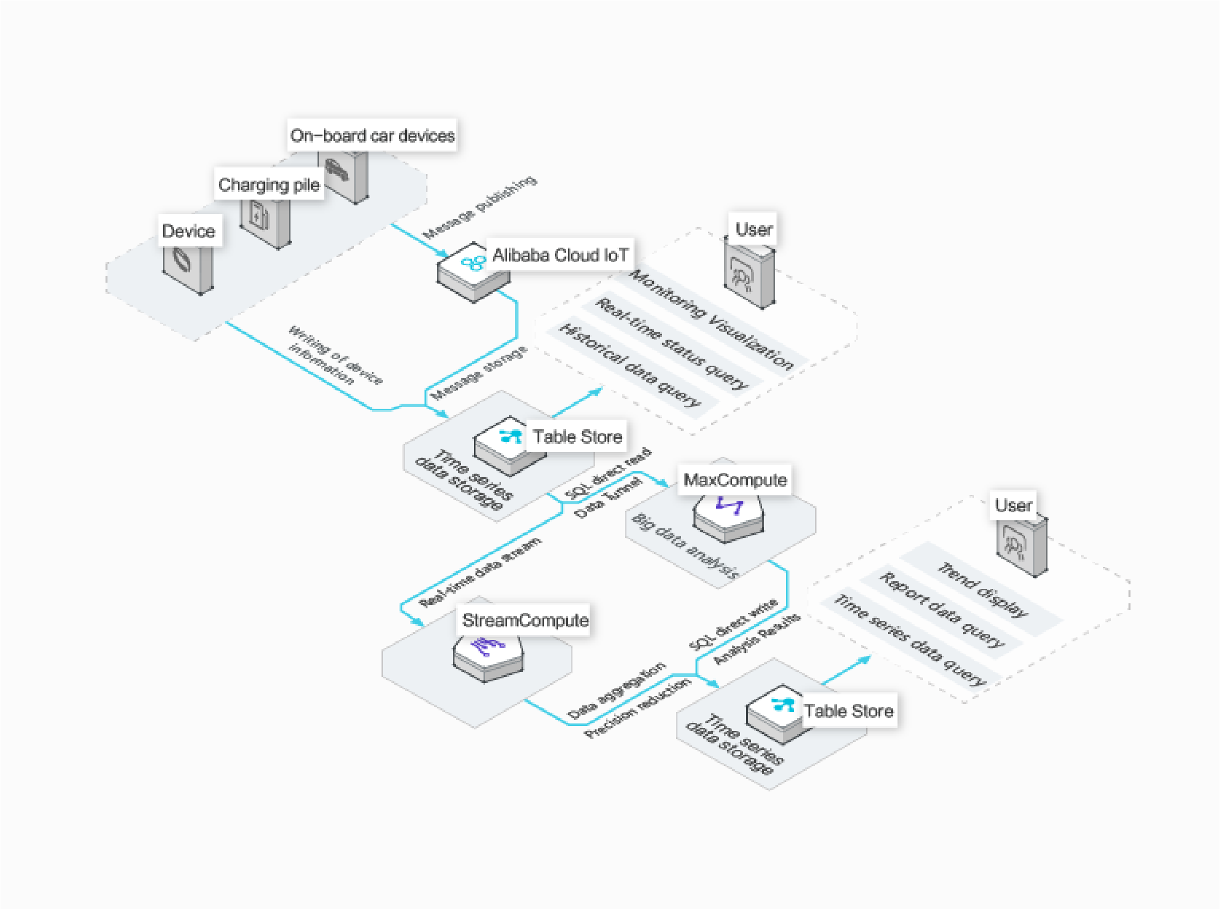
IoV data storage

A single table can store petabytes of data without distributing data in separate databases and tables, which simplifies the business logic. The schema-free data model enables easy access to the monitoring data of different vehicle-mounted devices. Table Store can seamlessly integrate with multiple big data analysis platforms and real-time computing services for ease of real-time online query and business report analysis of your vehicular fleet.



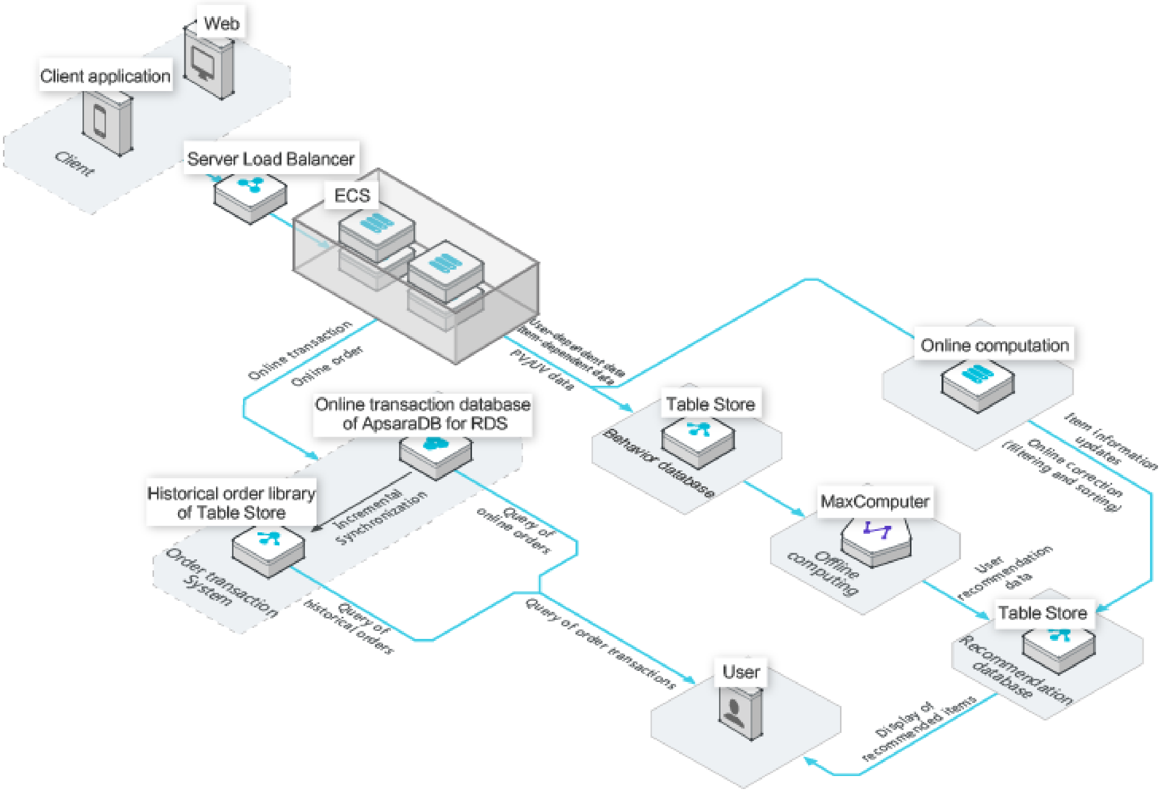
IoT time series data storage

With a single table capable of storing petabytes of data and processing thousands of queries per second (QPS), Table Store makes it easy to store the time series data of IoT devices and monitoring systems. The big data analysis SQL direct read function and the efficient incrementa l streaming read interface provide an easy way of offline data analysis and real-time streaming computing.



E-commerce recommendation

Table Store makes it possible for you to deal with data volumes and access performance with ease when handling a large number of historical transaction orders. Combined with MaxCompute , Table Store enables precision marketing, and elastic resource storage, so that you can easily cope with peak traffic during certain shopping periodshours when all customers go online.



4.6 Limits

Table 4-1: Limits

Item	Limits	Description
Number of instances created under an Alibaba Cloud user account	10	If you want to raise the limit, contact the administrator.
Number of tables in an instance	64	If you want to raise the limit, contact the administrator.
Instance name length	3-16 Bytes	Can contain letters, numbers, and hyphens (-). It must begin with a letter. It cannot end with a hyphen.
Table name length	1-255 Bytes	Can contain letters, numbers, and underscores (_). It must begin with a letter or underscore.

Item	Limits	Description
Column name length	1-255 Bytes	Can contain letters, numbers, and underscores (_). It must begin with a letter or underscore.
Number of primary key columns	1-4	There must be at least one column.
Size of String type primary key column values	1 KB	Applies to each primary key column of the String type.
Size of String type attribute column values	2 MB	Applies to each attribute column of the String type.
Size of Binary type primary key column values	1 KB	Applies to each primary key column of the Binary type.
Size of Binary type attribute column values	2 MB	Applies to each attribute column of the Binary type.
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 columns	The number of attribute columns written by one PutRow, UpdateRow, or BatchWrite Row request in a single row.
Data size of a single row	Unlimited	The total size of all column names, and column value data, for a single row is unlimited.
Number of columns in a read request's columns_to_get parameter	128	The maximum number of columns obtained in a row of data in the read request.
Number of UpdateTable operations for a single table	• Raise: Unlimited • Lower: Unlimited	The number of read/write operations within a single calendar day (from 00:00:00 to 00:00:00 of the next day in UTC time).
UpdateTable frequency for a single table	Maximum of one update every 2 minutes	The reserved read/write throughput for a single table cannot be adjusted beyond the frequency of once every 2 minutes.
The number of rows read by one BatchGetRow request	100	N/A
The number of rows written by one BatchWriteRow request	200	N/A

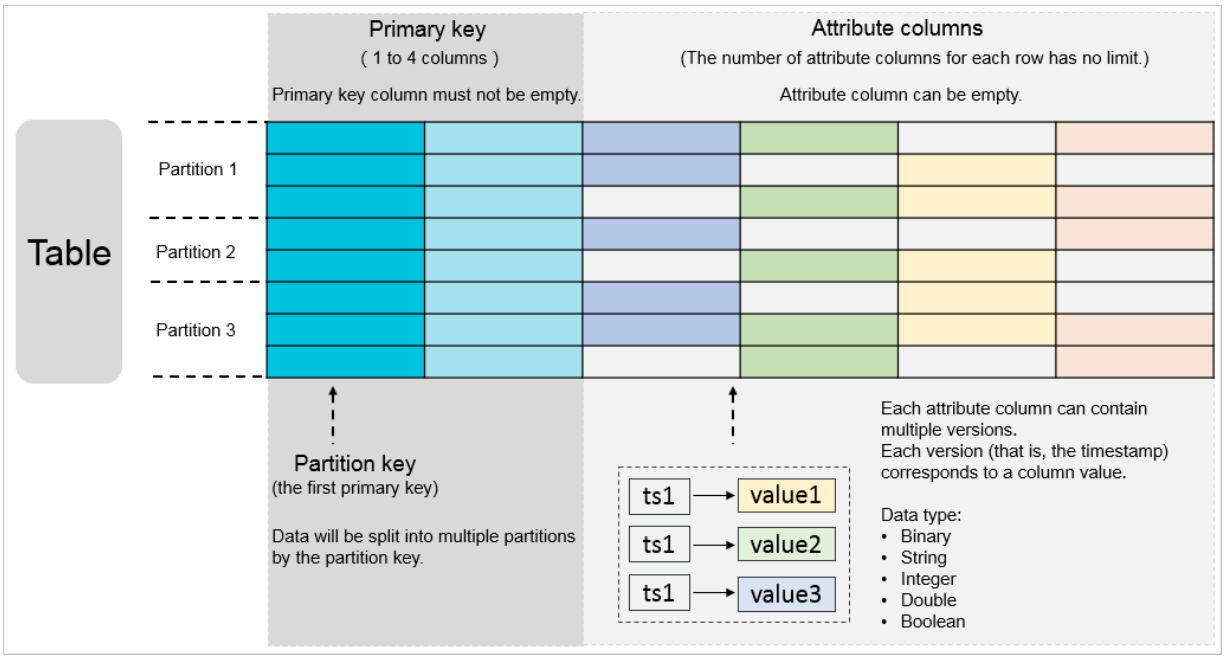
Item	Limits	Description
Data size of one BatchWriteRow request	4 MB	N/A
Data returned by one GetRange operation	5,000 rows or 4 MB	The data returned by a single operation cannot exceed 5000 rows or 4 MB. Otherwise, you must read the excessive data with a returned token.
The data size of an HTTP Request Body	5 MB	N/A

4.7 Terms

4.7.1 Data model

The data model of Table Store is described by Table, Row, Primary Key, and Attribute, and is shown in *Figure 4-1: Data model diagram*.

Figure 4-1: Data model diagram


Primary key
(1 to 4 columns)
Primary key column must not be empty.
Attribute columns
(The number of attribute columns for each row has no limit.)
Attribute column can be empty.
Table
Partition 1
Partition 2
Partition 3
Partition key
(the first primary key)
Data will be split into multiple partitions by the partition key.
ts1
value1
ts1
value2
ts1
value3
Each attribute column can contain multiple versions.
Each version (that is, the timestamp) corresponds to a column value.
Data type:
Binary
String
Integer
Double
Boolean

- A table is a set of rows, and a row consists of the *primary key and attribute*.
- The primary key and attribute columns consist of names and values.
- All rows in a table must contain primary key columns with the same number and name. However, the number, name, and data type of attribute columns within the rows can vary.

- Each attribute column can contain multiple versions, and each version (that is, the timestamp) corresponds to a column value, which is different from that of a primary key column.

Note:

Timestamp is the sum of milliseconds counted from 1970-01-01 00:00:00 UTC to the time when data is written.

The following example illustrates two rows in a table. The **`ID`** column is the primary key column.

Table 4-2: Example

ID	Type	ISBN	PageCount	Length
‘4776’	timestamp = 1466676354000, value = ‘Book’	timestamp = 1466676354000, value = ‘123*45678912345’	timestamp = 1466676354000, value = 666	-
‘6555’	timestamp = 1466676354000, value = ‘Music’	-	-	timestamp = 1466676354000, value = ‘400’; timestamp = 1466762754000, value = ‘500’

The meaning of the preceding table is as follows:

- **`ID`** is the primary key of the table. Rows with the ID of **`‘4776’`** and **`‘6555’`** have different attributes and can be stored in the same table.
- The attribute column **`Type`** of the row with ID **`‘4776’`** only has one version. The version is **`1466676354000`** and the data is **`‘Book’`**.
- The attribute column **`Length`** of the row with ID **`‘6555’`** has two versions. The data of version **`1466676354000`** is **`‘400’`** and the data of version **`1466762754000`** is **`‘500’`**.

4.7.2 Max Versions

Max Versions is a table attribute used to indicate the maximum number of data versions in each attribute column of a data table. When the number of data versions in an attribute column exceeds the value of Max Versions, the earliest versions will be deleted asynchronously.

After creating a table, you can use the **`UpdateTable`** operation to dynamically change the value of Max Versions.

When dynamically modifying the value of Max Versions, the following restrictions may apply:

- Data that exceeds the value of Max Versions is invalid and cannot be viewed or accessed, even though it is not deleted.
- If you reduce the value of Max Versions, data that exceeds the new value will be deleted asynchronously.
- If you increase the value of Max Versions, data that exceeds the old value, but has not been deleted, remains accessible.

4.7.3 Time To Live

Time To Live (TTL) is a table attribute measured in seconds that indicates the validity period of data. To save data storage space and reduce storage costs, Table Store runs in the background and automatically clears any data that exceeds the TTL. TTL usage is described as follows:

- Set TTL when creating a table. If you do not want data to expire, set TTL to **`-1`**.
- After creating a table, you can use the **`UpdateTable`** operation to dynamically change the value of TTL.

For example, for a table for which the TTL is 86400 seconds (one day), all attribute columns with versions earlier than 1468944000000 (divided by 1000 and converted to seconds to get 2016-07-20 00:00:00 UTC) expires at 2016-07-21 00:00:00 UTC and will be automatically cleared.

Expired data will be cleared asynchronously. When dynamically modifying the value of TTL, the following restrictions may apply:

- Data with a date that exceeds the value of TTL is invalid and cannot be viewed or accessed, even though it is not deleted.
- If you reduce the value of TTL, data with a date that exceeds the later value will be deleted asynchronously.
- If you increase the value of TTL, data with a date that exceeds the earlier value, but has not been deleted, remains accessible.

4.7.4 Max Version Offset

Max Version Offset is a table attribute measured in seconds. To prevent the writing of unexpected data, the server checks the attribute columns’ versions when processing writing requests. Writing data to a specified row fails if the row has an attribute column in which:

- Its version is earlier than the current writing time minus the value of Max Version Offset.

- Its version is later than or equal to the current writing time plus the value of Max Version Offset.

The valid version range for attribute columns is **`[Data written time – the value of Max Version Offset, Data written time + the value of Max Version Offset)`**. The data written time is the sum of the seconds counted from 1970-01-01 00:00:00 UTC to the time when data is written. Versions of the attribute columns (expressed in milliseconds) must, after being divided by 1000 and converted to seconds, fall into the valid version range.

For example, for a table for which the valid version range is 86400 seconds (one day), then at 2016-07-21 00:00:00 UTC, only data with versions later than 1468944000000 (converted to seconds to get 2016-07-20 00:00:00 UTC) but earlier than 1469116800000 (converted to seconds to get 2016-07-22 00:00:00 UTC) can be written to the table. If a row has an attribute column whose version is 1468943999000 (converted to seconds to get 2016-07-19 23:59:59 UTC, which is less than a day), the data cannot be written to the row.

Max Version Offset must be set to a non-zero value based on the number of seconds during the period from 1970-01-01 00:00:00 UTC to the current time.

If you do not set the value of Max Version Offset when creating a table, the table uses the default value **`86400`**.

After creating a table, you can use the **`UpdateTable`** operation to dynamically change the value.

4.7.5 Primary key and attribute

Primary key

A primary key is the unique identifier of each row in a table. It consists of 1 to 4 attribute columns.

When a table is created, the primary key must be defined. To define a primary key, the column name and data type of each primary key column, and the fixed sequence of the primary key columns, must be provided.

Data types of the primary key column can only be String, Integer, and Binary. For a primary key column of String or Binary data type, the size of the column value cannot exceed 1 KB.

Partition key

The first primary key column is also called a partition key. Table Store checks the range where the partition key value of each row is located, and automatically allocates the data in this row to the corresponding partition and machine to achieve load balancing.

Rows with the same partition key value belong to the same partition. A partition may contain multiple partition key values. A partition key is the smallest partition unit, which means the data of the same partition key value cannot be split. To avoid a partition becoming too large to be split, we recommend that the total data volume of all rows under a single partition key value does not exceed 10 GB.

To improve load balancing, Table Store splits and merges partitions according to specific rules. This process is automated without application intervention.

Attribute

Data of rows are stored in the attribute columns. There is no limit to the number of attribute columns for each row.

Version

When writing data, you can specify the attribute columns' versions. If you do not specify any versions, the server generates the versions of the attribute columns based on the current timestamp (expressed by the number of milliseconds that have elapsed since 1970-01-01 00:00:00 UTC). You can specify the maximum number of versions per column, or the version range, to limit the data that can be read from each row.

If the number of versions written to an attribute column exceeds the value of Max Versions, the data of earlier versions is discarded so that the number of remaining versions is equal to the Max Versions value.

A version number also indicates the time when data is generated. The version is expressed by the number of milliseconds that have elapsed since 1970-01-01 00:00:00 UTC. For a table where the TTL is set to 86400 (one day), the data in the attribute column where the version is 1468944000000 (2016-07-20 00:00:00 UTC) will expire at 2016-07-21 00:00:00 UTC and be automatically deleted.

Note:

- During TTL comparison and Max Version Offset calculation, versions in milliseconds must be converted to seconds by dividing by 1000.
- If versions are determined by the server, data will be cleared after the time (in seconds) indicated when TTL has elapsed since the data was written.

- To prevent invalid written data, the system denies the writing of expired data. For example, at 2016-07-21 00:00:00, the data in which the version is earlier than 1468944000000 (2016-07-20 00:00:00 UTC) cannot be written to a data table where the TTL is 86400.
- To prevent writing errors, the system requires that the attribute column version to which the data is written is within range of `[Data written time - The value of Max Version Offset, Data written time + The value of Max Version Offset)`.

Column naming conventions

The primary key and attribute columns obey the following naming conventions:

- Each name can be 1 to 255 characters in length.
- Letters, numbers, and underscores (_) are permitted.
- The name must start with a letter or underscore (_).
- All characters are case sensitive.

Data types of column values

Table Store supports five data types of column values, as shown in *Table 4-3: Data type of column values*

Table 4-3: Data type of column values

Data type	Definition	Permitted as primary key	Size limitation
String	UTF-8, or empty	Yes	For a primary key column, maximum of 1 KB. For an attribute column, see *Limits*.
Integer	64 bits	Yes	8 Bytes.
Double	64 bits	No	8 Bytes.
Boolean	True/False	No	1 Byte.
Binary	Defined, or empty	Yes	For a primary key column, maximum of 1 KB. For an attribute column, see *Limits*.

4.7.6 Read/write throughput

The read/write throughput is measured by read/write capacity units (CUs), which is the smallest billing unit for the data read/write operations.

One read CU indicates that 4 KB data is read from the table, and one write CU indicates that 4 KB data is written into the table. Data smaller than 4 KB during the operation will be rounded up to an integer. For example, writing 7.6 KB data will consume two write CUs, and reading 0.1 KB data will consume one read CU.

When applications use an API to perform Table Store read/write operations, the corresponding amount of read/write CUs will be consumed.

Reserved read/write throughput

The reserved read/write throughput:

- Is an attribute of a table.

  When creating a table, the application specifies the read/write throughput reserved for the table . Configuring the reserved read/write throughput does not affect the table's access performance , or service capability.
- Can be set to zero.

  If the reserved read/write throughput is greater than zero, Table Store assigns and reserves enough resources for the table according to this configuration to ensure low resource costs. The reserved read/write throughput of a table can be dynamically changed using the `UpdateTable` operation.

For a non-zero reserved read/write throughput, your use of Table Store is billed even if there are no read and write requests. To ensure billing accuracy, Table Store limits the maximum reserved read/write throughput to 5,000 CUs per table (neither read throughput, nor write throughput, can exceed 5,000 CUs). If you require more than 5,000 CUs of reserved read/write throughput for a single table, contact the administrator to increase the throughput.

The reserved read/write throughput of a non-existent table is regarded as zero. To access a non-existent table, one additional read CU or one additional write CU will be consumed depending on the actual operation.

Additional read/write throughput

The additional read/write throughput refers to the portion of the actual consumed read/write throughput that exceeds the reserved read/write throughput. Its statistical period is one second. For example, if the reserved read throughput of a table is set to 100 CUs and, within one second, the read operation consumes 120 CUs, then the additional read throughput consumed within the second is 20 CUs.

Note:

Because it is difficult to accurately reserve resources based on the additional read/write throughput, in extreme situations, Table Store may return an error `OTSCapacityUnitExhausted` to an application when an access to a single partition key consumes 10,000 CUs per second. In this case, policies, such as backoff retry, are used to reduce the frequency of access to the table.

4.7.7 Instance

Overview

An instance is a logical entity in Table Store used to manage tables, that is, it is the basic unit of Table Store's management and is used to create and manage tables within a Table Store instance . An instance is managed through the Table Store console. Table Store implements access control and resource metering at the instance level.

You can create different instances for different businesses to manage their respective tables. You can also create multiple instances for one business based on different development, testing, and production purposes.

Table Store allows one Alibaba Cloud account to create up to 10 instances, and up to 64 tables can be created within each instance. If more instances or tables are needed, contact the administrator.

Naming conventions

The name of each instance must:

- Be unique within each service region. If the instances are in different service regions, you can create instances of the same name.
- Be 3 Bytes to 16 Bytes in length.
- Contain only letters, numbers, and hyphens (-).
- Start with a letter.
- Not end with a hyphen (-).

Instance type

Table Store supports two instance types: high-performance instance and capacity instance.

Note:

An instance type cannot be modified once the instance is created.

The two instance types provide the same functions and support petabyte-sized data volumes for a single table, however, they differ in application scenarios.

- High-performance instance

  High-performance instances support millions of read-write transactions per second (TPS) with a 1 ms average latency of read and write operations per row. High-performance instances are suitable for scenarios requiring high read and write performance and concurrency, such as gaming, financial risk control, social networking apps, recommendation systems, and public opinion monitoring.
- Capacity instance

  Capacity instances provide write throughput and write performance comparable to that of the high-performance instances. However, the capacity instances do not match the read performance and concurrency of high-performance instances. Capacity instances are suitable for services with high write frequency, but low read frequency, as well as services with high affordability, but relatively low performance requirements. This includes access to log monitoring data, Internet of Vehicles (IoV) data, device data, time sequence data, and logistic data.

  Note:

  Capacity instances do not support reserved read/write throughput. All reads and writes are billed based on the additional read/write throughput.

5 Network Attached Storage (NAS)

5.1 What is NAS

Alibaba Cloud Network Attached Storage (NAS) is a highly reliable, highly available file storage service for Alibaba Cloud ECS, E-HPC, and Container Service. The service features a distributed file system with unlimited capacity and performance scaling ability. It supports a single namespace and allows multiple user access. Additionally, standard file access protocols are supported. You do not need to modify your application to use the service.

After creating a NAS file system and a mount point, you can mount the file system on multiple compute nodes (for example, ECS, E-HPC, and Container Service) using the NFS protocol, and use POSIX interfaces to access the file system. The same file system can be mounted on multiple compute nodes to share files and directories.

5.2 Benefits

Multiple access

The same file system can be mounted on multiple computing nodes, significantly reducing data copying and synchronization costs.

High reliability

In comparison to a self-built NAS, Alibaba Cloud NAS provides a data reliability of 99.99999999 %, saves substantial maintenance costs, and reduces data security risks.

Elastic scaling

Each file system has a maximum capacity of 10 PB, which can be scaled on demand effortlessly.

High performance

The throughput of each file system scales linearly with the storage capacity, substantially reducing the costs of purchasing a high-end NAS device.

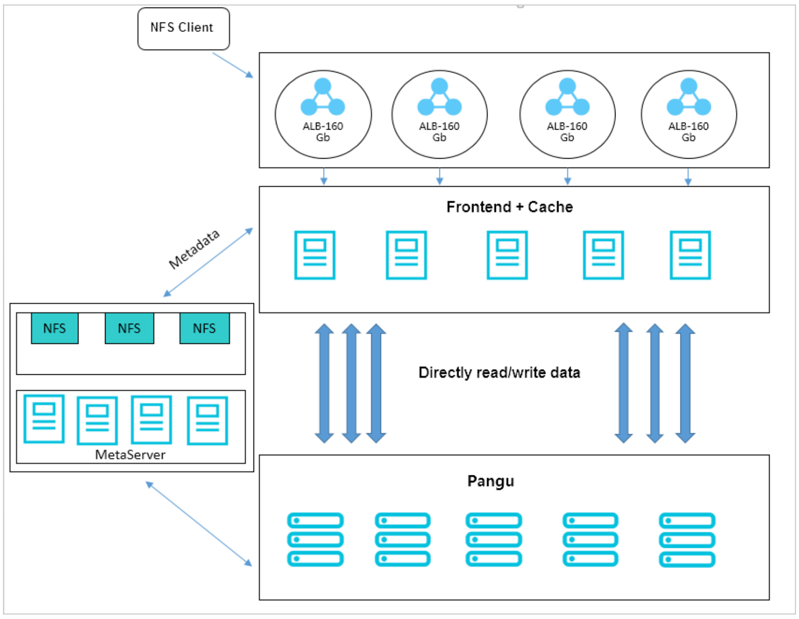
Ease of use

Alibaba Cloud NAS supports the NFSv3 and NFSv4 protocols. Compute nodes such as ECS instances, E-HPC, and Docker clusters can access the file system by using standard POSIX interfaces.

5.3 Architecture

Based on Apsara distributed file system, Alibaba Cloud NAS stores and distributes three copies of each data file on multiple storage nodes. The frontend nodes receive connection requests from NFS clients. Deployed in a distributed fashion, these nodes are stateless with cache feature, and ensure frontend high availability. The metadata of the file system is stored on MetaServers. I/O requests from the client can directly access user data stored on backend nodes after obtaining the metadata of the file system from MetaServers.

Both the frontend and backend can expand elastically as demand changes, ensuring high availability, high throughput, and low latency.



5.4 Features

Seamless integration

Alibaba Cloud NAS supports the NFSv3 and NFSv4 protocols and provides standard file system semantics for data access. Most mainstream applications and tasks can be seamlessly integrated with the service without any modifications.

Shared access

Multiple compute nodes can simultaneously access the same file system, allowing applications deployed across multiple ECS instances, E-HPC or Docker clusters to access the same data source.

Elastic scaling

Each file system has a maximum capacity of 10 PB, providing great scalability to meet your needs .

Security control

Multiple security mechanisms are implemented to guarantee system data security, including network isolation (VPC) and user isolation (classic network), standard access control, permission groups, and account authorization.

Linearly scalable performance

Alibaba Cloud NAS is characterized by high throughput, high IOPS, and low latency. Performance and capacity can be linearly scaled to meet real-time demands.

5.5 Scenarios

SLB shared storage and high availability

When your SLB instances have been connected to multiple ECS instances, we recommend that the applications deployed on these ECS instances store their data on a shared NAS file system, achieving data sharing and high availability of these SLB instances.

File sharing

When multiple employees need shared access to the same files, we recommend that the administrator can create a NAS file system and provide specific users or groups of users with different file or directory permissions.

Data backup

You can back up local data to a NAS file system, which provides a standard file access interface for data access. We recommend that you use NAS to back up data from the equipment room.

Server log sharing

The server logs of applications deployed on multiple compute nodes can be stored on a shared NAS file system. We recommend that you use NAS to store the server logs, making it easy to perform log analysis in the future.

5.6 Limits

- NAS currently supports NFSv3 and NFSv4 protocols.

- The following table describes attributes not supported by NFSv4.0 and NFSv4.1 and corresponding error messages displayed on the client:

Protocol	Attribute not supported	Error message
NFSv4.0	FATTR4_MIMETYPE，FATTR4_QUOTA_AVAIL_HARD，FATTR4_QUOTA_AVAIL_SOFT，FATTR4_QUOTA_USED，FATTR4_TIME_BACKUP，FATTR4_TIME_CREATE	NFS4ERR_ATTRNOTSUPP
NFSv4.1	FATTR4_DIR_NOTIF_DELAY，FATTR4_DIRENT_NOTIF_DELAY，FATTR4_DACL，FATTR4_SACL，FATTR4_CHANGE_POLICY，FATTR4_FS_STATUS，FATTR4_LAYOUT_HINT，FATTR4_LAYOUT_TYPES，FATTR4_LAYOUT_ALIGNMENT，FATTR4_FS_LOCATIONS_INFO，FATTR4_MDSTHRESHOLD，FATTR4_RETENTION_GET，FATTR4_RETENTION_SET，FATTR4_RETENTEVT_GET，FATTR4_RETENTEVT_SET，FATTR4_RETENTION_HOLD，FATTR4_MODE_SET_MASKED，FATTR4_FS_CHARSET_CAP	NFS4ERR_ATTRNOTSUPP

- In addition, the following OPs are not supported by NFSv4: OP_DELEGPURGE, OP_DELEGRETURN, and NFS4_OP_OPENATTR, which result in the NFS4ERR_NOTSUPP error message displayed on the client:
- Currently, NFSv4 does not support the Delegation function.
- About UID and GID:

- For NFSv3 protocol, if the UID or GID of a file is configured in the Linux local account, the user name or group name of the file is displayed based on the mapping relationship between UID/GID and user name/group name. If the UID or GID of a file is not configured in the Linux local account, the UID or GID is directly displayed.
- For NFSv4 protocol, if the version of the Linux kernel is earlier than 3.0, the UIDs and GIDs of all files are displayed as nobody. If the version of the Linux kernel is later than 3.0, the UID and GID of a file are displayed in the same way as that in NFSv3.

> **Note:**
> If you use the NFSv4 protocol to mount file systems and the version of the Linux kernel is earlier than 3.0, we suggest you do not change the owner or group of a file or directory. Otherwise, the UID and GID of the file or directory are changed to nobody.

- A single file system can be mounted and accessed by 10,000 computing nodes at the same time.

5.7 Glossary

Mount point

A mount point is the access address of a NAS file system in a VPC or classic network. Each mount point corresponds to a domain name. You need to specify the domain name of the mount point to mount the corresponding NAS file system to a local target.

Permission group

The permission group functions as a whitelist in NAS. You can add rules to provide IP addresses or network segments with different permissions to access the file system.

> **Note:**
> Each mount point must have a specified permission group.

Authorized object

An authorized object is an attribute of the permission group rule and specifies the IP address or network segment to which the permission group rule is applied. In a VPC, an authorized object can be a single IP address or a network segment. In a classic network, an authorized object can be a single IP address only, generally the intranet IP address of an ECS instance.

6 Relational Database Service (RDS)

6.1 What is ApsaraDB for RDS?

Alibaba Cloud ApsaraDB for Relational Database Service (RDS) is a stable, reliable, and auto-scaling online database service. Based on Alibaba Cloud's distributed file system and high-performance storage, ApsaraDB provides a complete set of solutions for disaster tolerance, backup, recovery, monitoring, and migration to free you from worries about database O&M.

ApsaraDB for MySQL

Based on Alibaba Cloud's MySQL source code branch, ApsaraDB for MySQL has proven to have excellent performance and throughput. It has withstood the massive data traffic and large number of concurrent users during many November 11 shopping festivals. ApsaraDB for MySQL also provides a range of advanced functions such as optimized read/write splitting, data compression, and intelligent optimization.

MySQL is the world's most popular open source database. It is used in a variety of applications and is an important part of LAMP, a combination of open source software (Linux+Apache+MySQL+Perl/PHP/Python).

Two popular Web 2.0-era technologies, BBS software system Discuz! and the blogging platform – WordPress, are built on the MySQL-based architecture. In the Web 3.0 era, leading Internet companies such as Alibaba, Facebook, and Google have all taken advantage of the flexibility of MySQL to build their mature database clusters.

ApsaraDB for SQL Server

SQL Server is one of the first commercial databases and is an important part of the Windows platform (IIS + .NET + SQL Server), with support for a wide range of enterprise applications. The SQL Server Management Studio software comes with a rich set of built-in graphical tools and script editors. You can quickly get started with a variety of database operations through a visual interface.

ApsaraDB for SQL Server provides strong support for a variety of enterprise applications powered by the high-availability architecture and the ability to recover to any point in time. It also covers Microsoft’s licensing fee.

ApsaraDB for PostgreSQL

PostgreSQL is the world's most advanced open source database. As the forerunner among academic relational database management systems, PostgreSQL excels for its full compliance with SQL specifications and robust support for a diverse range of data formats such as JSON, IP, and geometric data, which are not supported by most commercial databases.

In addition to excellent support for features such as transactions, subqueries, Multi-Version Concurrency Control (MVCC), and data integrity check, ApsaraDB for PostgreSQL integrates a series of important functions including high availability, backup, and recovery that help ease your O&M burden.

ApsaraDB for PPAS

Postgres Plus Advanced Server (PPAS) is a stable, secure, and scalable enterprise-class relational database. Based on PostgreSQL, the world's most advanced open source database, PPAS brings enhancements in terms of performance, application solutions, and compatibility. It also provides the capability of directly running Oracle applications. You can run enterprise-class applications on PPAS stably and obtain cost-effective services.

ApsaraDB for PPAS provides account management, resource monitoring, backup, recovery, and security control, and more functions, and is continuously updated and improved.

6.2 Benefits

6.2.1 Easy-to-use

Out-of-the-box experience

You can specify RDS instance types through the APIs. After the order is created, RDS generates the specified instance immediately.

On-demand upgrade

Along with changes in the database load and data storage capacity, you can flexibly adjust the instance types, and RDS will not interrupt the data link service during the upgrade.

Transparent and compatible

The use method of RDS is the same as that of the native database engine. You can get started easily without further learning. In addition, RDS is compatible with your current programs and tools . Data can be migrated to RDS using a data import and export tool with minimal labor required.

Convenient management

Alibaba Cloud is responsible for ensuring the normal operation of RDS through daily maintenance and management, such as hardware/software fault processing and database patch updates. You can also independently perform database addition, deletion, restart, backup, recovery and other management operations through the Apsara Stack console.

6.2.2 High performance

Parameter optimization

Alibaba Cloud has gathered top database experts in China, and the parameters of all RDS instances have been obtained from many years' experience in production and optimization. The professional database administrators continuously optimizes RDS instances over their life cycles to ensure that RDS is running at optimal performance.

SQL optimization

Based on your application scenario, RDS will lock low-efficiency SQL statements and provide recommendations for optimizing your business code.

High-end backend hardware

All servers used by RDS have been evaluated by multiple parties to ensure the exceptional performance and stability.

6.2.3 High security

DDoS attack

> **Note:**
> This function requires activation of Alibaba Cloud security products.

When you access your RDS instance from the Internet, the instance may suffer from DDoS attacks. If this occurs, the RDS security system initiates traffic cleaning. If the attacks reach the blackhole threshold or the traffic cleaning fails, blackhole filtering will be triggered.

The following describes how and when traffic cleaning and blackhole filtering are triggered:

- Traffic cleaning

  This applies only to inbound traffic from the Internet. During this process, the RDS instance can be normally accessed.

The system automatically triggers and ends traffic cleaning. Traffic cleaning is triggered if a single RDS instance meets any of the following conditions:

- Packets Per Second (PPS) reaches 30,000.
- Bits Per Second (BPS) reaches 180 Mbit/s.
- The number of concurrent connections created per second reaches 10,000.
- The number of concurrent active connections reaches 10,000.
- The number of concurrent inactive connections reaches 10,000.

- Blackhole filtering

  This only applies to inbound traffic from the Internet. During this process, the RDS instance cannot be accessed from the Internet, and its applications typically become unavailable. Blackhole filtering provides a way to ensure the availability of the overall RDS service.

  Blackhole filtering will be triggered if the following conditions are met:

  - BPS reaches 2 Gbit/s.
  - Traffic cleaning fails.

  The blackhole will end automatically 2.5 hours after being triggered.

Note:

We recommend that you access your RDS instance over an intranet to prevent the risk of DDoS attacks.

Access control policy

You can define the IP addresses that are allowed to access RDS. Access from unspecified IP addresses will be denied.

Each account can only view and operate its own database.

System security

RDS is protected by multiple firewall layers that can effectively block a variety of malicious attacks and ensure data security.

Direct logon to the RDS server is not allowed. Only the ports required by the specific database services are open.

The RDS server cannot initiate an external connection. It can only accept access requests.

6.2.4 High availability

Hot standby

RDS adopts a hot standby architecture. If the master server fails, services fail over in seconds. The entire failover process is transparent to applications.

Multi-copy redundancy

The data on the RDS server is built on RAID, and data backups are stored on OSS.

Data backup

RDS provides an automatic backup mechanism. You can set a backup schedule or initiate a temporary backup at any time based on the business characteristics.

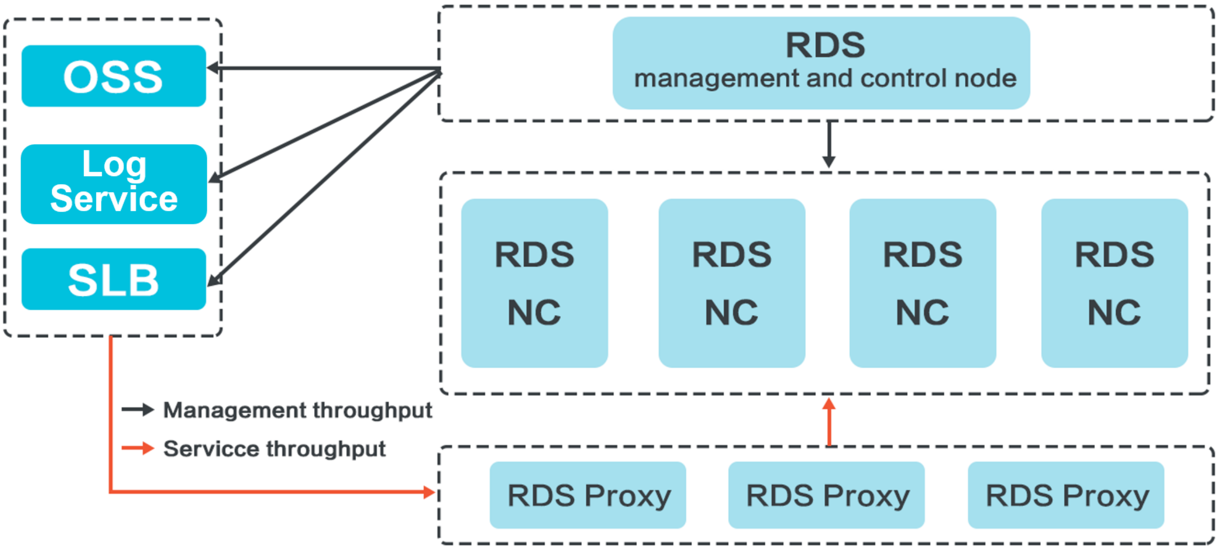
Data recovery

Data can be recovered from the backup set and for the specified time point. Generally, you can restore data at any time point within 7 days to a temporary RDS instance. After the data is verified , the data can be migrated back to the master RDS instance.

6.3 Architecture

The RDS system architecture is as follows.

Figure 6-1: RDS system architecture



6.4 Features

6.4.1 Data link service

Alibaba Cloud ApsaraDB provides all of the data link services, including DNS, Server Load Balancer (SLB), and Proxy. Since RDS uses native DB engines, and database operations are highly similar across engines, there is essentially no learning cost for users who are familiar with the engines. Besides, ApsaraDB provides a Database Management System (DMS), which greatly facilitates your access to and use of the database.

DNS

The DNS module supports the dynamic resolution of domain names to IP addresses, to prevent IP address changes from affecting the performance of your RDS instance.. After its domain name has been configured in the connection pool, an RDS instance can continue to be accessed even if its IP address changes.

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

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 configured in the connection pool is **`test.rds.aliyun.com`**, the instance can still be accessed. However, if the IP address configured in the connection pool is **`10.10.10.1`**, the instance will no longer be accessible.

SLB

The SLB module provides instance IP addresses (including both intranet and Internet IP addresses) to prevent physical server changes from affecting the performance of your RDS instance.

For example, the intranet IP address of an RDS instance is **`10.1.1.1`**, and the corresponding Proxy or DB Engine runs on **`192.168.0.1`**. Normally, the SLB module redirects all traffic destined for **`10.1.1.1`** to **`192.168.0.1`**. If **`192.168.0.1`** fails, another server in hot standby status with the IP address of **`192.168.0.2`** takes over for **`192.168.0.1`**. 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 its services normally.

Proxy

The Proxy module provides a number of functions including data routing, traffic detection, and session holding.

- Data routing: This supports distributed complex query aggregation for big data and provides the corresponding capacity management.
- Traffic detection: This reduces SQL injection risks and supports SQL log backtracking when necessary.
- Session holding: This prevents database connection interruptions if any fault occurs.

DMS

DMS is a web service designed to access and manage cloud data. DMS provides such functions as data management, object management, data stream management, and instance management.

The DMS supports data sources such as MySQL, SQL Server, PostgreSQL and PPAS.

6.4.2 High-availability service

The high-availability service consists of several modules including the Detection, Repair, and Notice modules. These modules guarantee the availability of the data link services and process internal database exceptions.

In addition, RDS can improve the performance of its high-availability service by migrating instances to a region that supports multi-zone instances and by adopting appropriate high-availability policies.

Detection

The Detection module checks whether the master and slave nodes of the DB Engine are providing their services normally. The HA node uses heartbeat information, acquired at an interval of 8 to 10 seconds, to determine the health status of the master node. This information, combined with the health status of the slave node and heartbeat information from other HA nodes, allows the Detection module to eliminate any risk of false judgements caused by exceptions such as network jitter. As a result, a failover can be completed within 30 seconds.

Repair

The Repair module maintains the replication relationship between the master and slave nodes of the DB Engine. It can also repair any errors that may occur at either node.

For example:

- Automatic restoration of master/slave replication in case of an unexpected disconnection
- Automatic repair of table-level damage to either node
- On-site saving and automatic repair in case of crashes

Notice

The Notice module notifies the SLB or Proxy of status changes to the master and slave nodes to ensure that you always access the correct node.

For example, the Detection module discovers that the master node has an exception and instructs the Repair module to fix it. If the Repair module fails to resolve the problem, the Notice module will be informed of this information. The Notice module then forwards the failover request to the SLB or Proxy module, which begins to redirect all traffic to the slave node. At the same time, the Repair module creates a new slave node on another physical server and synchronizes this change back to the Detection module. The Detection module starts to recheck the health status of the instance.

Multi-zone

A multi-zone is a physical area that is formed by combining multiple individual zones within the same region. Multi-zone RDS instances can withstand higher level disasters than single-zone ones.

For example, a single-zone RDS instance can withstand server and rack failures, while a multi-zone RDS instance can withstand the failure of an entire equipment room.

There is currently no extra charge for multi-zone RDS instances. Users in a region where multi-zone has been enabled can buy multi-zone RDS instances directly or convert single-zone RDS instances into multi-zone ones by using inter-zone migration.

Note:

Certain network latency may exist between multiple zones. As a result, when a multi-zone RDS instance uses a semi-synchronous data replication solution, its response time to any individual update may be longer than that of a single-zone instance. In this case, the best way to improve overall throughput is to increase concurrency.

High-availability policy

The high-availability policies use a combination of service priorities and data replication modes to meet the needs of your business.

There are two service priorities:

- Recovery Time Objective (RTO) priority: The database must restore services as soon as possible, thst is, the longest available time. This is suitable for users who require their databases to provide uninterrupted online services.
- Recovery Point Objective (RPO) priority: The database must protect the reliability of the data as much as possible, that is, the least amount of data lost. This is suitable for users whose has higher preference to data consistency.

There are three data replication modes:

- Asynchronous replication (Async): When an application initiates an update request, which may include addition, deletion, or modification operations, the master node responds to the application immediately after completing an operation, and then the master node replicates data to the slave node asynchronously. This means that the operation of the master database is not affected if the slave node is unavailable, but it does make it possible for data inconsistencies to occur between the master and slave nodes if the master node is unavailable.
- Synchronous replication (Sync): When an application initiates an update request, which may include addition, deletion, or modification operations, the master node replicates the data to the slave node immediately after completing an operation and waits for the slave node to return a success message before it responds to the application. Since the master node replicates data to the slave node synchronously, unavailability of the slave node will affect the operation on the master node, but unavailability of the master node will not cause data inconsistency.
- Semi-synchronous replication (Semi-sync): Normally, data is replicated in Sync mode. If an exception occurs (unavailability of the slave node or a network exception between the two nodes) when the master node replicates data to the slave node, the master node will suspend response to the application until the Sync replication times out and degrades to Async replication. If the application is allowed to update data in such situation, unavailability of the master node will cause data inconsistency. When data replication between the two nodes resumes normal, because the slave node or network connection is recovered, data replication mode will change from Async to Sync.

You can select different combination modes of service priorities and data replication modes to improve availability according to the business characteristics. *Table 6-1: Characteristics of different combinations* describes the characteristics of different combinations.

Table 6-1: Characteristics of different combinations

Engine	Service priority	Data replication mode	Combination characteristics
MySQL 5.6	RPO	Async	If the master node fails, services fail over after the slave node applies all of the relay logs. If the slave node fails, application operations on the master node are not affected. The data on the master node will be synchronized after the slave node recovers.
MySQL 5.6	RTO	Semi-sync	If the master node fails and data replication has not degraded, RDS will immediately trigger a failover and direct traffic to the slave node because data consistency has been guaranteed. If the slave node fails, application operations on the master node will time out, and data replication will degrade to Async replication. After the slave node recovers and the data on the master node is synchronized completely, data replication will return to synchronous replication. If the master node fails in case of data inconsistency between the master and slave node and the data replication mode has degraded to asynchronous replication, services will fail over after the slave node applies all of the relay logs.
MySQL 5.6	RPO	Semi-sync	If the master node fails and data replication has not degraded, RDS will immediately trigger a failover and direct traffic to the slave node because data consistency has been guaranteed. If the slave node fails, application operations on the master node will time out, and data replication will degrade to asynchronous replication. When the slave node can obtain information from the master node again (the slave node or network connection recovers), data replication will return to synchronous replication. If the master node fails in case of data inconsistency between the master and slave node and the data difference on the slave node cannot be supplemented, you can obtain the time of the slave

Engine	Service priority	Data replication mode	Combination characteristics
			node through the APIs and decide the failover time and data supplementation method.

6.4.3 Backup and recovery service

This service supports the offline backup, dump, and recovery of data.

Backup

The Backup module compresses and uploads the data and logs on both the master and slave nodes. RDS uploads this data to OSS by default, but the backup files can also be dumped to a cheaper and more persistent Archive Storage. Normally, backup is initiated on the slave node so as to not affect the services on the master node. However, if the slave node is unavailable or damaged, the Backup module will initiate backup on the master node.

Recovery

The Recovery module restores backup files stored on OSS to the target node.

- Master node rollback: This function rolls back the master node to the status of a specified time point in the case of faulty operation.
- Slave node repair: This function creates a new slave node to reduce risks in the case of an irreparable fault occurred to the slave node.
- Read-only instance creation: A read-only instance is created from a backup.

Storage

The Storage module uploads, transfers, and downloads backup files. Currently, all backup data is uploaded to OSS for storage, and you can obtain temporary links to download data as needed. In certain scenarios, the Storage module also supports dumping backup files from OSS to Archive Storage for inexpensive and persistent offline storage.

6.4.4 Monitoring service

ApsaraDB provides multi-level monitoring services across the physical, network, and application layers to ensure business availability.

Service

The Service module tracks service-level status. It monitors whether SLB, OSS, Archive Storage , Log Service, and other cloud products on which RDS depends are normal, including their functions and response time. It also uses logs to determine whether the internal RDS services are operating normally.

Network

The Network module tracks status at the network layer. It monitors the connectivity between ECS and RDS and between RDS physical servers, as well as the packet loss rates on the router and switch.

OS

The OS module tracks status at the hardware and OS kernel layer, including:

- Hardware overhaul: It constantly checks the operational status of the CPU, memory, main board, and storage, pre-judges whether a fault will occur, and automatically submits a repair report in advance.
- OS kernel monitoring: It tracks all database calls and uses kernel status to analyze the reasons for slowdowns or call errors.

Instance

The Instance module collects RDS instance-level information, including:

- Available instance information
- Instance capacity and performance indicators
- Instance SQL execution records

6.4.5 Scheduling service

The scheduling service consists of the Resource and Version modules. It mainly allocates resources and manages instance versions.

Resource

The Resource module allocates and integrates the underlying RDS resources, which means instance enabling and migration to you. For example, when you create an instance through the RDS console or APIs, the Resource module will calculate the most suitable physical server to carry traffic. This module also allocates and integrates the underlying resources required for the inter-zone migration of RDS instances. After lengthy instance creation, deletion, and migration operations, the Resource module calculates the degree of resource fragmentation in a zone and initiates resource integration regularly to improve the service carrying capacity of the zone.

Version

The Version module is responsible for version upgrades of RDS instances. For example:

- Major version upgrade: Upgrading MySQL 5.1 to MySQL 5.5, MySQL 5.5 to MySQL 5.6, and so on.
- Minor version upgrade: Fixing bugs in the MySQL source code.

6.4.6 Migration service

The migration service helps you migrate data from a local database to ApsaraDB, or migrate data from an ApsaraDB instance to another. ApsaraDB provides a Data Transmission Service (DTS) tool to facilitate quick database migration.

DTS

DTS is a cloud data transfer service for efficient data migration from local databases to RDS, as well as from an RDS instance to another. Currently, DTS supports three types of databases: MySQL, SQL Server, and PostgreSQL.

DTS provides three migration modes: structure migration, full migration, and incremental migration .

- Structure migration

  DTS migrates the structure definitions of migration objects to the target instance. Currently, tables, views, triggers, stored procedures, and stored functions can be migrated in this mode.
- Full migration

  DTS migrates all existing data of migration objects in source databases to the target instance.

  Note:

  To ensure data consistency, non-transaction tables that do not have a Primary Key will be locked during full migration. Locked tables cannot be written to, and the lock duration depends on the volume of data in the tables. The locks are released only after the non-transaction tables without a Primary Key have been fully migrated.
- Incremental migration

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

Note:

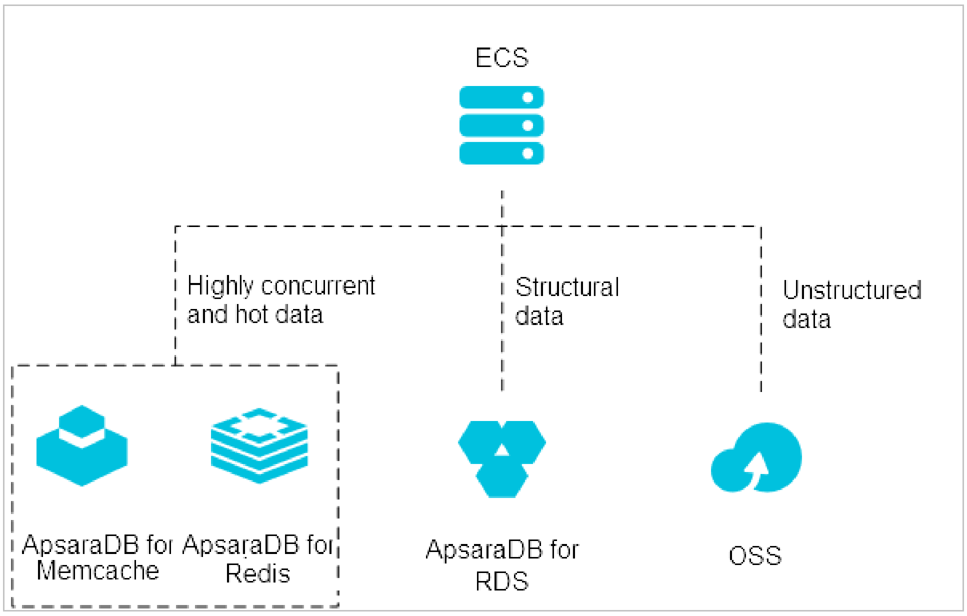
If a DDL operation is performed during data migration, structure changes will not be synchronized to the target instance.

6.5 Scenarios

6.5.1 Diversified data storage

RDS supports diversified storage extension through ApsaraDB for Memcache, ApsaraDB for Redis, OSS, and other storage products, as shown in *Figure 6-2: Diversified data storage*.

Figure 6-2: Diversified data storage



Cache data persistence

RDS can be used together with Memcache and Redis to form a high-throughput and low-latency storage solution. Compared with RDS, the ApsaraDB cache products has two benefits:

- Higher response speed: The request delay of Memcache and Redis is usually within several milliseconds.
- Higher Queries Per Second (QPS)

Multi-structure data storage

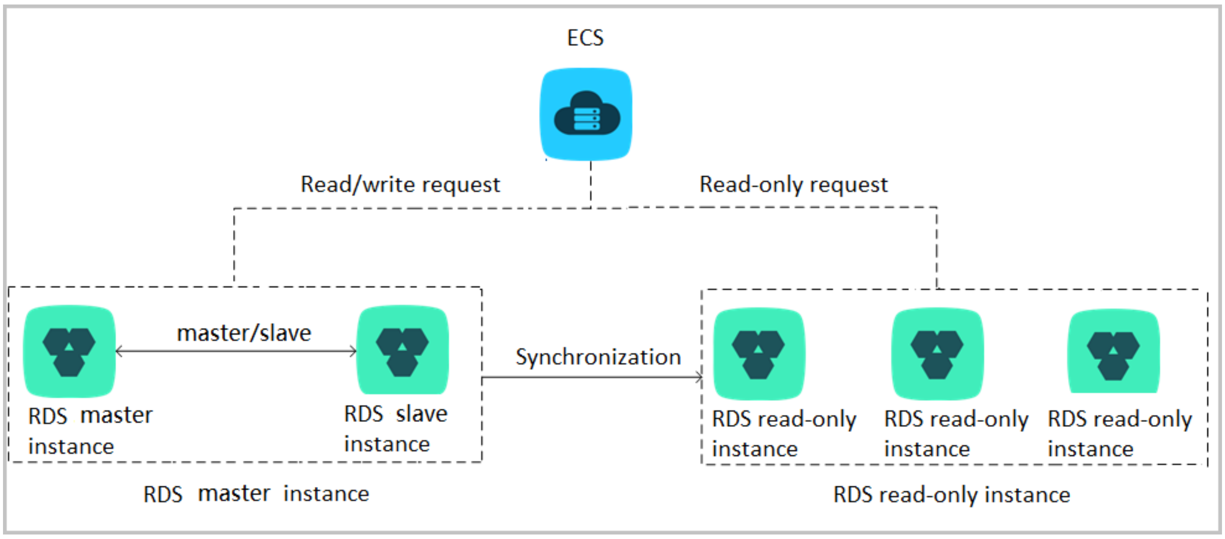
OSS is an Alibaba Cloud storage service that features massive capacity, robust security, low cost, and high reliability. RDS and OSS can work together to form multiple types of data storage solutions. For example, when RDS and OSS are used in a forum, resources such as the images of registered users and those posted on the forum can be stored in OSS, reducing the storage pressure on RDS.

6.5.2 Read/write splitting

RDS for MySQL allows read-only instances to be directly attached to the master instance in order to distribute the read pressure on the master instance.

The master instance and read-only instances of RDS for MySQL have their own connection addresses. After you enable the read/write splitting function, RDS will offers an extra read/write splitting address, which associate the master instance and its read/write splitting instances to achieve automatic read/write splitting. Applications only need to connect to the same read/write splitting address to perform read and write operations. The read/write splitting module automatically sends write requests to the master instance and read requests to each read-only instance based on specified weights you have set. You can simply keep strengthening RDS's processing ability without any changes to applications by adding the number of read-only instances, as shown in *Figure 6-3: Read/write splitting*.

Figure 6-3: Read/write splitting

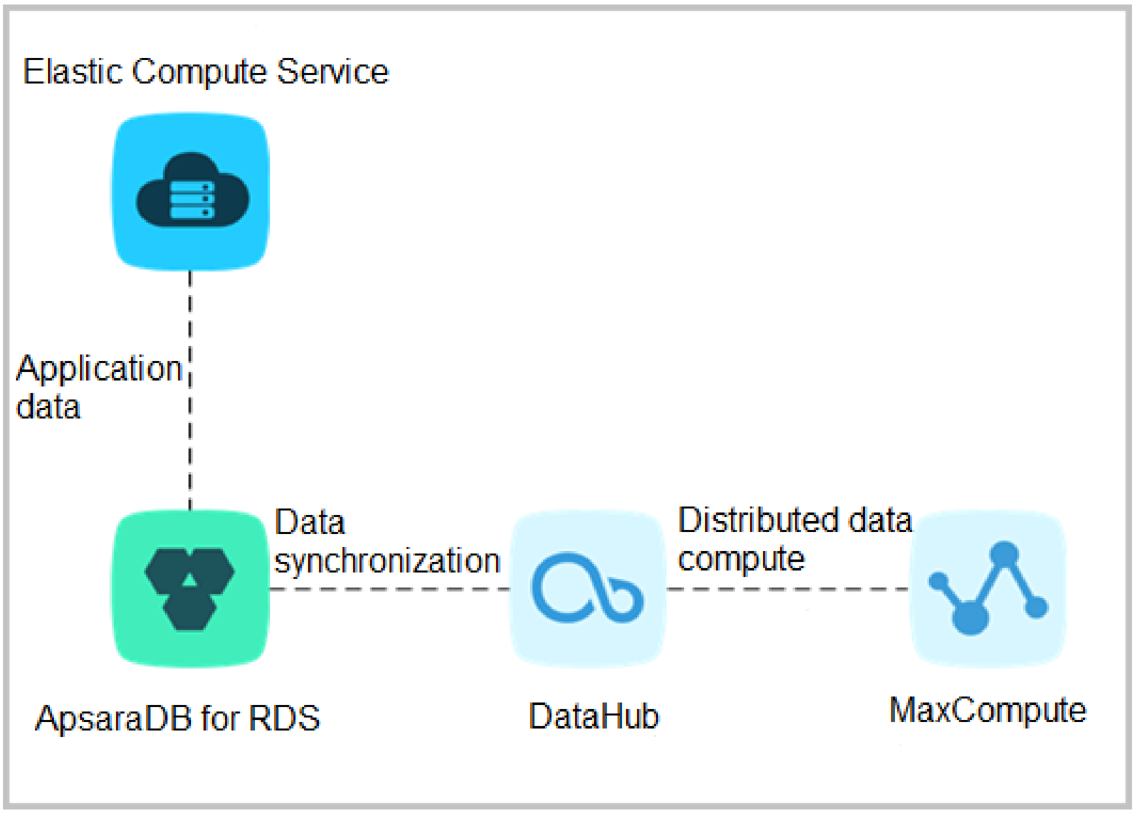


6.5.3 Big data analysis

Alibaba Cloud provides MaxCompute (formerly known as ODPS) for storage and processing of massive amounts of structured data. The service offers mass data warehouse solutions and analytical modeling services for big data.

You can import data from an RDS instance into a MaxCompute instance through Data IDE to achieve large-scale data computing, as shown in *Figure 6-4: Big data analysis*.

Figure 6-4: Big data analysis



6.6 Limits

6.6.1 Restrictions on MySQL

To guarantee the stability and security of ApsaraDB for MySQL, certain restrictions apply to the database and management properties, as shown in *Table 6-2: Restrictions on MySQL*.

Table 6-2: Restrictions on MySQL

Operation	Restriction
Parameter modification	The RDS console or open APIs must be used to modify database parameters. However, some parameters cannot be modified. For more information, see *Set parameters*.
Root permission	Root or sa permission is not provided.
Backup	• Command lines or graphical interfaces can be used to perform logical backup. • For physical backups, the RDS console or APIs must be used.
Restoration	• Command lines or graphical interfaces can be used to perform logical restoration.

Operation	Restriction
	• For physical backups, the RDS console or APIs must be used.
Migration	• Command lines or graphical interfaces can be used to perform logical import. • You can use MySQL command line tool or Data Transmission Service (DTS) to perform data migration.
MySQL storage engine	• Currently, only InnoDB and TokuDB are supported. Due to defects inherent to the MyISAM engine, data may be lost . Therefore, any MyISAM table of a new instance will be automatically converted to an InnoDB table. • The InnoDB storage engine is recommended for better performance and higher security. • The Memory engine is not supported. Any Memory table of a new instance will be automatically converted to an InnoDB table.
Replication	MySQL supports dual-node clusters based on a master/slave replication architecture without manual setup. The slave instances in this replication architecture are not publicly available . You cannot access them directly.
RDS instance restart	Instances must be restarted through the RDS console or APIs.
User, password, and database management	By default, MySQL uses the RDS console to manage users, passwords, and databases. For example, RDS for MySQL allows you to create or delete an instance, modify permissions, and change passwords. Additionally, RDS for MySQL allows you to create a master account to manage users, passwords, and databases.
Common account	• Does not allow customizable authorization. • The account management and database management interfaces are provided on the RDS console. • Instances that can create common accounts can also create master accounts.
Master account	• Allows customizable authorization. • The account management and database management interfaces are not provided on the RDS console. To manage accounts and databases, use SQL statements or DMS. • The master account cannot be rolled back to a common account.

6.6.2 Restrictions on SQL Server

RDS for SQL Server provides instances with accompanying licenses only. After an instance is created, it is granted a Microsoft SQL Server Enterprise Edition license. It does not allow users to bring their own licenses. Furthermore, to ensure instance stability and security, SQL Server has following restrictions:

Table 6-3: Restrictions on SQL Server

Feature	Description
Number of databases	50
Number of database accounts	500
User, login, or database creation	Supported
Database-level DDL trigger	Limited
Granting permission within databases	Limited
Thread killing permission	Supported
Linked server	Limited
Distributed transaction	Limited
SQL Profiler	Limited
Optimization consultant	Limited
Change data capture	Limited
Change tracking	Supported
Windows domain account login	Limited
Email	Limited
SQL Server Integration Services (SSIS)	Limited
SQL Server Analysis Services (SSAS)	Limited
SQL Server Reporting Services (SSRS)	Limited
R language service	Limited
Common language runtime (CLR)	Limited
Asynchronous messaging	Limited
Replication	Limited
Policy management	Limited

6.6.3 Restrictions on PostgreSQL

To guarantee instance stability and security, there are some restrictions on ApsaraDB for MySQL.

Table 6-4: Restrictions on PostgreSQL

Operation	Restriction
Database parameter modification	Not supported.
Root permission of databases	Administrator permissions cannot be provided to users.
Database backup	Data can be backed up only through **pg_dump**.
Data migration	Only the data backed up through **pg_dump** can be restored through **psql**.
Database replication	• The system automatically builds HA databases based on PostgreSQL streaming replication. • PostgreSQL standby nodes are not visible to users, and cannot be accessed directly.
RDS instance restart	RDS instances must be restarted on the RDS console or through APIs.
Network management	If instances are used in safe mode, net.ipv4.tcp_timestamps cannot be enabled in SNAT mode.

6.6.4 Restrictions on PPAS

To guarantee instance stability and security, there are some restrictions on ApsaraDB for MySQL.

Table 6-5: Restrictions on PPAS

Operation	Restriction
Database parameter modification	Not supported.
Root permission of databases	Administrator permissions cannot be provided to users.
Database backup	Data can be backed up only through the **pg_dump** command.
Data migration	Only the data backed up through **pg_dump** can be restored through **psql**.

Operation	Restriction
Database replication	• The system automatically builds HA databases based on PPAS streaming replication. • PPAS standby nodes are not visible to users, and cannot be accessed directly.
RDS instance restart	RDS instances must be restarted on the RDS console or through APIs.
Network Management	If instances are used in safe mode, net.ipv4.tcp_timestamps cannot be enabled in SNAT mode.

6.7 Concepts

Concept	Description
Region	A region indicates the geographic location where your RDS instance resides. You must specify a region when creating an RDS instance, and you cannot change the region after the instance is created. RDS must be used together with ECS and supports access only from the intranet . Therefore, the RDS instance must be in the same region as the ECS instance.
Zone	A region contains one or more zones. A zone is a physical area with independent power supply and networks. Zones of a region are connected through the intranet, but faults are isolated between the zones. A single-zone instance indicates that the master and slave nodes of the instance are all in the same zone. Network latency is shorter if the ECS and RDS instances are in the same zone rather than different zones.
Instance	An RDS instance is the basic unit of RDS services. An instance is the operating environment of AprasaDB for RDS and works as an independent process on a host. You can create, modify, or delete an RDS instance on the RDS console. An instance is independent from another, and resources of an instance are isolated from those of another. There are no resource preemption problems, such as CPU, memory, or I/O preemption, among instances. Each instance has its own characteristics, such as database type and version, and RDS use parameters to control instance behavior.
Memory	Memory indicates the maximum memory space an RDS instance can use.

Concept	Description
Disk capacity	Disk capacity indicates the size of the disk you decide to create when creating an RDS instance. The disk capacity an instance occupies includes data as well as the space required for proper instance running , such as the space occupied by system databases, database rollback logs, redo logs, and indexes. Ensure that your RDS instance has sufficient disk capacity to store data; otherwise, your instance might be locked. If your instance is locked, you can expand the disk capacity to unlock the instance.
IOPS	IOPS indicates the maximum number of read/write operations performed on block devices per second at a 4-KB granularity.
Number of CPU cores	The CPU core indicates the highest computing capability of an instance . A CPU core has the computing capability of no lower than 2.3 GHz Hyper-Threading (of the Intel Xeon series).
Number of onnections	Number of connections indicates the number of TCP connections between the client and the RDS instance. If the client uses a connection pool, the connections are persistent connections. Otherwise, the connections are short-lived connections.

7 ApsaraDB for Redis

7.1 What is ApsaraDB for Redis

Alibaba Cloud ApsaraDB for Redis is an online Key-Value storage service compatible with the open-source Redis protocol. ApsaraDB for Redis supports many data types including String, List, Set, SortedSet, and Hash, and provides advanced functions such as Transactions and Pub/Sub. Using memory+hard disk storage, ApsaraDB for Redis meets your data persistence requirements, while providing high-speed data read/write capability.

In addition, ApsaraDB for Redis is used as a cloud computing service, with hardware and data deployed on the cloud, supported by comprehensive infrastructure planning, network security protection, and system maintenance services. This service enables you to focus fully on business innovation.

7.2 Benefits

Cluster functions

- The cluster function supports ultra-high capacity and performance. Cluster instances provide 128 GB or larger capacity, meeting requirements for large capacity and high performance.
- Master-slave dual-node instances are of 64 GB or smaller capacity, meeting average users' requirements for capacity and performance.

Elastic resizing

- One-click storage resizing: You can use the console to adjust the storage capacity of your instances as needed.
- Online resizing with no service interruption: You can adjust the instance capacity online without suspending your services or affecting your business.

Resource isolation

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

Data security

- Persistent data storage: With memory plus hard disk storage, ApsaraDB for Redis provides high-speed data read/write capability and meets the data persistence requirements.
- Master-slave dual-backup for data: All data on the master node has a backup copy on the slave node.

- Password authentication is required for secure and reliable access.
- Data transmission encryption: Secure Sockets Layer (SSL) and Transport Layer Security (TLS) are supported for data transmission security.

High availability

- Each instance has a master node and a slave node: This prevents service interruption caused by SPOF.
- Automatic detection and recovery of hardware failure: This feature can automatically detect hardware failures and fail over to the slave node, restoring service in a matter of seconds.

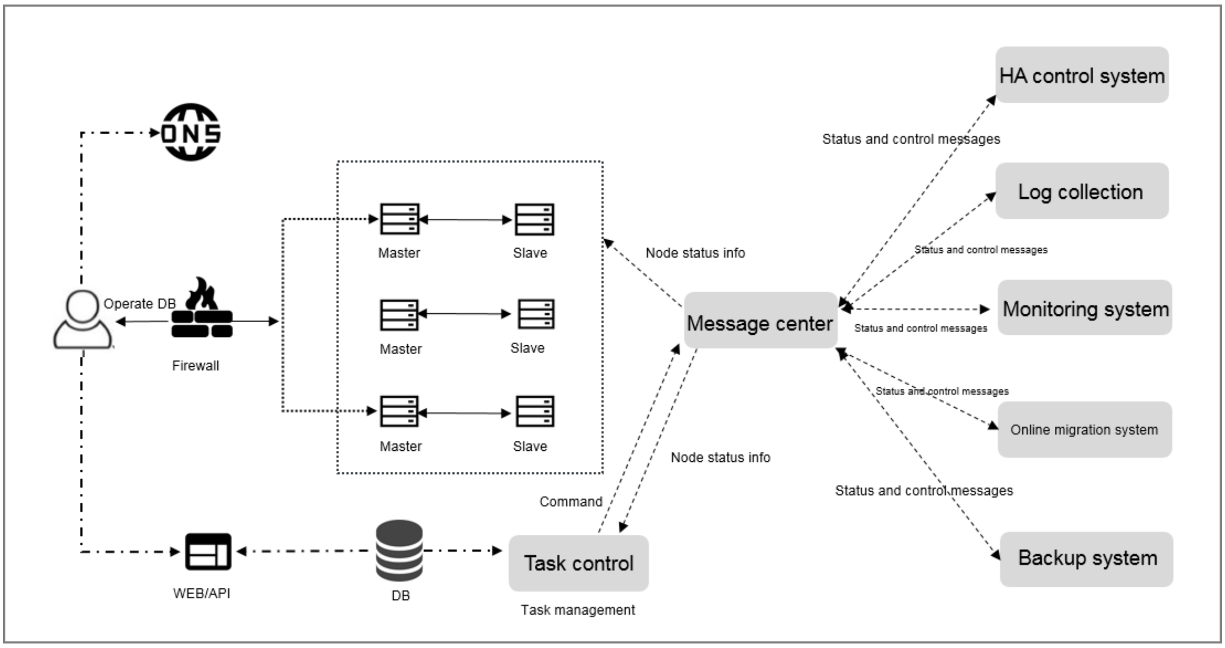
Easy to use

- Out-of-the-box service: This product requires no setup or installation and can be used right after purchase for quick and convenient business deployment.
- Compatible with open-source Redis: This product is compatible with Redis commands, and any Redis client can easily establish a connection with ApsaraDB for Redis to perform data operations.

7.3 Architecture

Figure 7-1: Architecture diagram shows the architecture of ApsaraDB for Redis.

Figure 7-1: Architecture diagram



ApsaraDB for Redis automatically constructs a master-slave dual-node architecture for you.

- **HA Control system**

  A high-availability detection module used to detect and monitor the operating status of ApsaraDB for Redis instances. If this module determines that a master node is unavailable, it will switch over to the slave node to ensure high availability of the ApsaraDB for Redis instances.

- **Log collection**

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

- **Monitoring system**

  This module collects performance monitoring information of ApsaraDB for Redis instances, including basic group monitoring, keys group monitoring, and String group monitoring.

- **Online migration system**

  When the physical server that runs an instance fails, the online migration system will recreate an instance based on the backup files in the backup system. This ensures that the business is not affected.

- **Backup system**

  This module generates and stores the backup files of ApsaraDB for Redis instances on the OSS system. At present, the backup system allows you to customize the backup settings and temporary backup configuration. The backup files are retained for seven days.

- **Task control**

  ApsaraDB for Redis instances support various management and control tasks, including instance creation, configuration changes, and instance backup. The task control module flexibly controls tasks and executes task tracking and error management based on the commands you give.

7.4 Features

High-availability technologies for smooth service running

Redis synchronizes data between the master and slave nodes in real time. If the master node fails , services are automatically switched over to the salve node. During the switchover, services are not affected, and therefore high availability is guaranteed.

A cluster instance is deployed based on a distributed high-availability architecture. Each node of the cluster has a master node and a slave node for automatic failovers. Therefore, high availability of services is guaranteed.

One-click backup and recovery and customizable backup policies

You can back up data manually, and customize automatic backup policies. Redis automatically reserves backup data of the past seven days, and supports one-click recovery, thereby minimizing the impact caused by incorrect operations.

Various network protection measures for data security

- TCP-layer network isolation provided by VPCs
- Monitoring and defense against DDoS attacks
- Up to 10000 whitelisted IP addresses for preventing unauthorized access

Optimized kernel for preventing attacks that exploit vulnerabilities

The Alibaba Cloud expert team performs in-depth kernel optimization for Redis source code. This effectively prevents memory overflows and fixes security vulnerabilities, offering you secure services.

Elastic scaling for higher capacity and better performance

ApsaraDB for Redis supports various types of memory specifications, allowing you to upgrade memory specifications as your service grows. The cluster version also allows elastic scaling of database storage space and throughput. This resolves the QPS performance bottleneck and supports millions of read/write operations per second.

Various instance specifications for flexible configuration changes

Redis supports the single-node cache architecture and dual-node storage architecture and allows configuration changes for different service scenarios.

Real-time monitoring and alarming for the instance status

Redis provides monitoring information about the CPU usage, number of connections, and disk space usage, as well as the alarm reporting function, so that you are fully aware of the instance status.

Graphical O&M platform for simplified and convenient O&M operations

You can clone instances, back up data, recover data, or other operations, with one click on the O&M platform.

Automatic database kernel upgrade for software defect prevention

Redis automatically upgrades itself to fix defects, freeing you from routine version management.

User-defined parameter settings for customization

You can set Redis parameters to fully utilize system resources.

7.4.1 Specifications and performance

Note:

The maximum bandwidth is the sum of incoming and outgoing traffic.

Standard-dual copy

Table 7-1: Standard package

Size (GB)	Max connections	Max intranet bandwidth (Mbit/s)	CPU core(s) (Relative)	Description
1 GB master/slave	10,000	10	1	Master-slave dual-node instance
2 GB master/slave	10,000	16	1	Master-slave dual-node instance
4 GB master/slave	10,000	24	1	Master-slave dual-node instance
8 GB master/slave	10,000	24	1	Master-slave dual-node instance
16 GB master/slave	10,000	32	1	Master-slave dual-node instance
32 GB master/slave	10,000	32	1	Master-slave dual-node instance

Table 7-2: Customized package

Size (GB)	Max connections	Max intranet bandwidth (Mbit/s)	CPU core(s) (Relative)	Description
1 GB master /slave (advanced)	20,000	48	1	Master-slave dual-node instance
2 GB master /slave (advanced)	20,000	48	1	Master-slave dual-node instance
4 GB master /slave (advanced)	20,000	48	1	Master-slave dual-node instance
8 GB master /slave (advanced)	20,000	48	1	Master-slave dual-node instance
16 GB master /slave (advanced)	20,000	48	1	Master-slave dual-node instance
32 GB master /slave (advanced)	20,000	48	1	Master-slave dual-node instance

Cluster

Size (GB)	Max connections	Max intranet bandwidth (Mbit/s)	CPU core(s) (Relative)	Description
16 GB cluster	80,000	384	8	High-performance cluster instance
32 GB cluster	80,000	384	8	High-performance cluster instance

Size (GB)	Max connections	Max intranet bandwidth (Mbit/s)	CPU core(s) (Relative)	Description
64 GB cluster	80,000	384	8	High-performance cluster instance
128 GB cluster	160,000	768	8	High-performance cluster instance
256 GB cluster	160,000	768	8	High-performance cluster instance

QPS performance reference

Table 7-3: QPS performance reference

Size (GB)	Max connections	Max intranet bandwidth (Mbps)	CPU core(s)	QPS reference value
8	10,000	24	1	80,000

Note:

The QPS of non-cluster instances ranges from 80,000 to 100,000, and that of cluster instances is the product of the number of nodes and the range (80,000 to 100,000).

Test scenario

Figure 7-2: Network topology

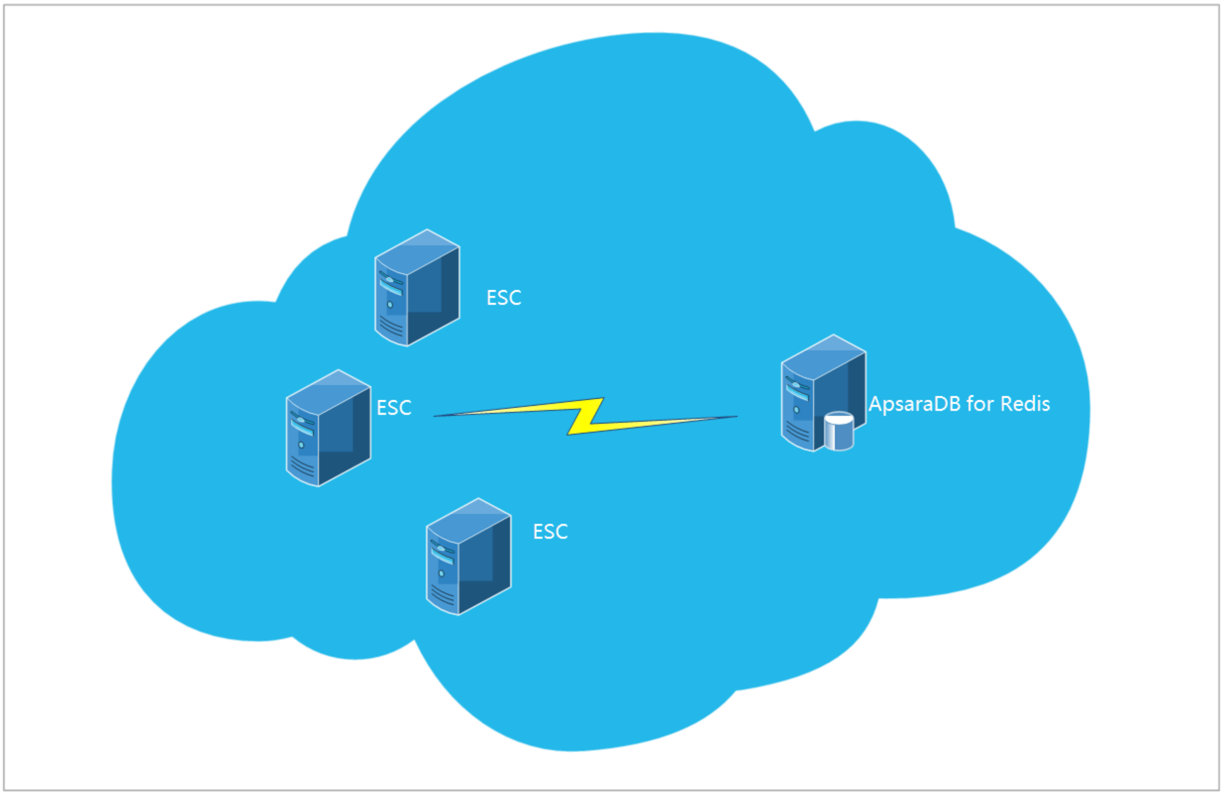


Table 7-4: Specification of cloud host

OS	CPU (number of cores)	Memory	Zone	Number of Hosts
Ubuntu 14.04 64-bit	1	2,048 MB	China South 1	3

1. Download the source code package for redis-2.8.19 to three ECS instances.

```
$ wget http://download.redis.io/releases/redis-2.8.19.tar.gz
$ tar xzf redis-2.8.19.tar.gz
$ cd redis-2.8.19
$ make
$ make install
```

2. Run the following command on the three ECS instances:

```
redis-benchmark -h ***********.m.cnsza.kvstore.aliyuncs.com -p 6379 -a
 password -t set -c 50 -d 128 -n 25000000 -r 5000000
```

3. Summarize the testing data from the three ECS instances. The QPS is the total for the preceding three servers.

7.5 Scenarios

Gaming industry applications

Game companies can use ApsaraDB for Redis as an important part of their deployment architecture.

- **Scenario 1: Using ApsaraDB for Redis for data storage**

  Game deployment architecture is relatively simple. With the main program deployed on ECS, all business data are stored in Redis as a persistent database. ApsaraDB for Redis supports persistence function, with master-slave dual-node redundant data storage.

- **Scenario 2: Using ApsaraDB for Redis as a cache to accelerate application access**

  Using Redis as a cache layer will accelerate application access. Data are stored in a backend database (RDS).

  Reliability is critical to ApsaraDB for Redis services. Once a ApsaraDB for Redis service becomes unavailable, business access may overload the backend database. ApsaraDB for Redis uses a hot standby high-availability architecture to ensure extremely high service reliability. The master node provides external services. If this node fails, the system will automatically set up the standby node to take over the services. The entire failover process is completely transparent to users.

Live video applications

Live video services are often highly reliant on ApsaraDB for Redis to store user data and friends interaction information.

- **Dual-node hot standby ensures high availability**

  ApsaraDB for Redis provides the hot-standby mode to maximize service availability.

- **Cluster version solves the performance bottleneck**

  ApsaraDB for Redis provides cluster version instances to break through the performance bottleneck of Redis single-thread mechanism. This approach can effectively cope with spikes in live video broadcast traffic and meet high performance requirements.

- **Easy resizing helps cope with business peaks**

ApsaraDB for Redis can support one-click resizing. The entire upgrade process is fully transparent to you and helps you easily cope with traffic bursts.

E-commerce industry applications

In the e-commerce industry, Redis is extensively used, mostly for item display, shopping recommendations, and other modules.

- **Scenario 1: Seckill-type shopping systems**

  During large-scale seckill promotions, a shopping system will be overwhelmed by traffic, which far exceeds the Read/Write capability of common databases.

  The persistence function supported by ApsaraDB for Redis allows you to directly use Redis as a database system.

- **Scenario 2: Inventory system with a counter**

  In such a system, the underlying architecture usually keeps actual data in RDS and count information in database fields. ApsaraDB for Redis reads the counts information while ApsaraDB for RDS stores the count information. In this scenario, ApsaraDB for Redis is deployed on a physical machine, with an underlying architecture based on SSD high-performance storage that can provide high-level data reading capabilities.

7.6 Limits

Item	Description
List data type	The number of lists is not restricted. The size of single element cannot exceed 512 MB. We recommend that one lists contain no more than 8192 elements, and the maximum value length cannot exceed 1 MB.
Set data type	The number of sets is not restricted. The size of single element cannot exceed 512 MB. We recommend that one set contain no more than 8192 elements, and the maximum value length cannot exceed 1 MB.
SortedSet data type	The number of SortedSets is not restricted. The size of single element cannot exceed 512 MB. We recommend that one SortedSet contain no more than 8192 elements, and the maximum value length cannot exceed 1 MB.
Hash data type	The number of fields is not restricted. The size of single element cannot exceed 512 MB. We recommend that one field contain no more than 8192 elements, and the maximum value length cannot exceed 1 MB.
Restriction on the database number	Each instance supports 256 databases.

Item	Description
Redis commands supported	For more information, see section **Supported Redis commands** in *Cite LeftApsaraDB for Redis User GuideCite Right*.
Monitoring alert	ApsaraDB for Redis does not provide the capacity alert function. You can configure this function on CloudMonitor. We recommend that you set alert for the following metrics: instance fault, instance master-slave switchover, connection usage, failed operation count, capacity usage, write bandwidth usage, and read bandwidth usage.
Expired data deletion policy	- Active expiration: The system periodically detects and deletes expired keys in the background.- Passive expiration: The system deletes expired keys when users access keys.
Idle connection recovery mechanism	Idle Refis connection is not automatically recovered by the server, and must be managed by the user.
Data persistence policy	AOF_FSYNC_EVERYSEC is enabled , and fysnc is performed every second.

7.7 Glossary

Redis

ApsaraDB for Redis is a high-performance Key-Value storage system (cache and store) released in compliance with the BSD open-source protocol.

Instance ID

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

ApsaraDB for Redis limits connection quantities, bandwidth, CPU specifications, and other parameters based on the capacity specifications of individual instances. On the console, you can view the list of IDs of the instances you have purchased. There are two types of ApsaraDB for Redis instances: master-slave dual-node instances and high-performance cluster instances.

Master-slave dual-node instances

This is an ApsaraDB for Redis instance that adopts a master-slave architecture. Master-slave dual-node instances are limited in terms of capacity and performance.

High-performance cluster instances

This is an ApsaraDB for Redis instance that adopts a scalable cluster architecture. Cluster instances have better scalability and performance, but are functionally limited to a certain extent.

Connection address

This is the host address used to connect to ApsaraDB for Redis. It is displayed as a domain name, and can be found at **Instance Information** > **Connection Information**.

Connection password

This is the password used to connect to ApsaraDB for Redis. The password format is `Instance ID:custom password`. For example, if you set the password as 1234 when you make the purchase and the allocated instance ID is xxxx, then the connection password will be `xxxx:1234`.

Eviction policy

This is consistent with the Redis eviction policy.

DB

This is the abbreviation of Redis database. Each ApsaraDB for Redis instance supports 256 DBs. By default, data is written to DB 0.

8 ApsaraDB for MongoDB

8.1 What is ApsaraDB for MongoDB

ApsaraDB for MongoDB is a dedicated high-performance distributed data storage service, which is fully compatible with MongoDB protocol and can provide stable, reliable, and auto scaling database service. It offers a full range of database solutions, such as disaster tolerance, backup, recovery, monitoring, and alarms.

ApsaraDB for MongoDB offers the following basic features:

- Automatically creates a three-replica MongoDB replica set for users to use. This encapsulates advanced functions such as disaster tolerance switchover and failover, and the whole process is completely transparent to the users.
- Provides cluster version instances based on multiple replica sets (with each replica set having three replicas), so you can easily scale the read/write performance and conveniently build a MongoDB distributed database system.
- Supports one-click database backup and recovery. Users can perform conventional database backup and database rollback with a single click on the console.
- Provides up to 20 performance metrics for monitoring and alarm functions, giving you a full view of database performance.
- Provides visual data management tools, making O&M more convenient.

8.2 Benefits

High availability

- The three-node replica set high-availability architecture delivers extremely high service availability.

  The ApsaraDB for MongoDB service uses a three-node replica set high-availability architecture. The three data nodes are located on different physical servers and synchronize data automatically . The primary and secondary nodes provide services. When the primary node fails, the system automatically elects a new primary node. When the secondary node is unavailable, the standby node takes over the services.
- Automatic backup and one-click recovery can resolve over 99.99% of system failures.

  The data is automatically backed up and uploaded to the Object Storage Service (OSS) every day, simultaneously improving data disaster recovery capabilities while effectively reducing the consumption of disk space. The backup files can restore the instance data to the original instance. This effectively prevents irreversible effects on service data caused by incorrect operations or other reasons.

High security

The multilevel security defense system can protect your network against over 90% of network attacks.

- Anti-DDoS protection: provides real-time monitoring at the network entry point. When high-traffic attacks are identified, their source IP addresses will be cleaned. In case cleaning is ineffective, the black hole mechanism will be triggered.
- IP whitelist configuration: supports the configuration of up to 1,000 server IP addresses which are allowed to connect to MongoDB instances, directly controlling risks at the source.

Ease of use

Sound performance monitoring will take over more than 60% of your O&M workload.

The product monitors instance information in real time, such as CPU utilization, IOPS, connection s, and disk space and reports alarms. This helps keep you updated on instance statuses at all times.

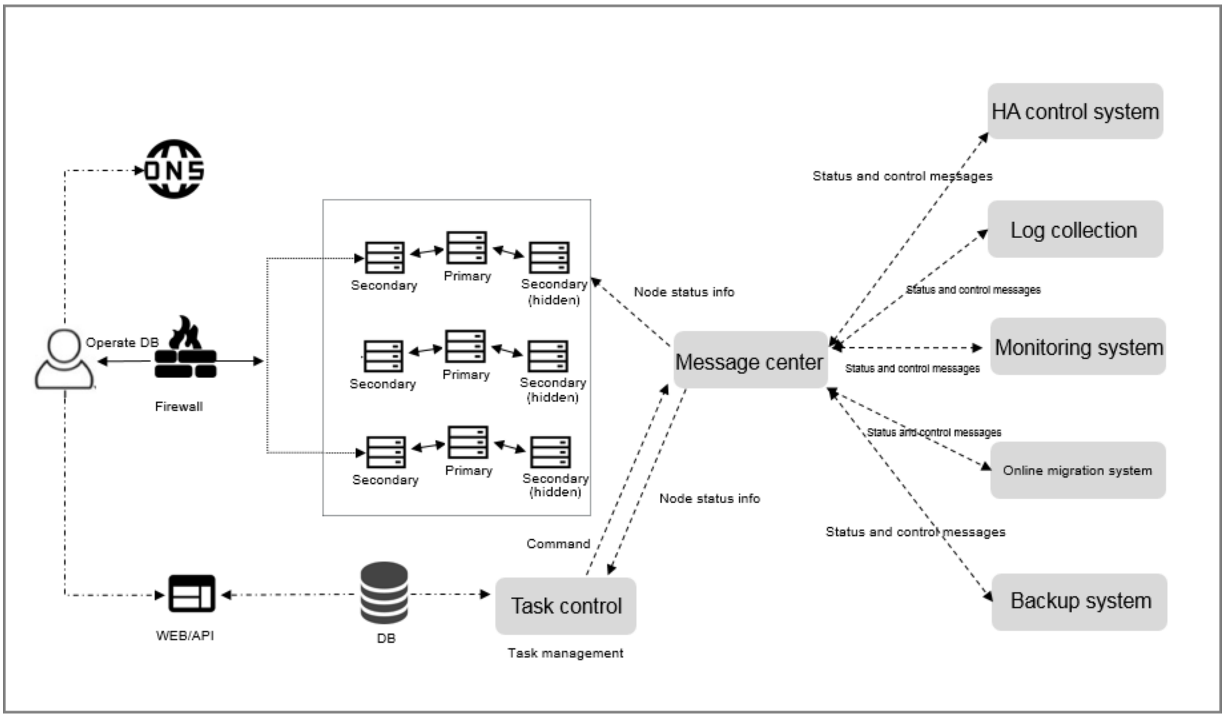
Scalability

The replica sets can be elastically resized.

ApsaraDB for MongoDB supports three-node replica sets to allow elastic resizing. You can change the configuration of your instance if the current configuration is too high or cannot meet the performance requirements of your application. The configuration change process is completely transparent and will not affect your services.

8.3 Architecture

ApsaraDB for MongoDB automatically creates a three-node replica set for you to use. You can directly operate on one primary node and one secondary node. The following figure shows the system architecture:



- HA control system: Instance high-availability detection modules are used to detect and monitor the operating status of MongoDB instances. If the system determines that the primary node instance is unavailable, it will switch over to the standby node, to ensure the high availability of MongoDB instances.
- Log collection: This process collects MongoDB operating condition logs, including instance slow query logs and RAM logs.
- Monitoring system: This system collects MongoDB instance performance monitoring information, including basic metrics, disk capacity, network requests, operation counts, and other core information.
- Online migration system: When the physical server that runs an instance fails, the online migration system will re-build an instance based on the backup files in the backup system to avoid service interruptions.
- Backup system: This system backs up MongoDB instances and stores the generated backup files on the OSS system. At present, the MongoDB backup system allows users to customize the backup settings and temporary backup configuration. Files are retained for 7 days.
- Task control: ApsaraDB for MongoDB instances support various management and control tasks, including instance creation, configuration changes, and instance backup. The task system flexibly controls tasks and executes task tracking and error management based on the commands you give.

8.4 Features

Flexible architecture

ApsaraDB for MongoDB automatically creates a three-node replica set for you. You can directly operate on the primary node and a secondary node. If the primary node fails, you can switch over services to the secondary node, thereby ensuing high availability of MongoDB instances.

Elastic capacity expansion

- One-click storage capacity expansion: You can adjust the instance storage capacity on the MongoDB console based on service requirements.
- Online capacity expansion without service interruption: You can adjust the instance storage capacity when services arerunning. This will not affect services.

Data security

- Automatic backup: ApsaraDB for MongoDB allows you to set backup cycles. You can flexibly configure backup start times according to your service off-peak times. The backup files are retained for free for up to 7 days.
- Temporary backup: You can initiate temporary backup as required. The backup files are retained for free for up to 7 days.
- Data recovery: Using backup files, you can directly overwrite existing data and restore an instance to a previous state.
- Backup file download: ApsaraDB for MongoDB retains your backup files for free for up to 7 days. During this period, you can log onto the console and download the backup files to your local device.
- Creating instances from backup sets: On the console, you can create an instance from backup files with a single click, for fast deployment.
- Multi-layer network security protection:
  - TCP-layer network isolation provided by VPCs
  - Monitoring and defense against DDoS attacks
  - Up to 10000 whitelisted IP addresses for preventing authorized access

Intelligent O&M

- Monitoring platform: Redis provides monitoring information about the CPU usage, number of connections, and disk space usage, as well as the alarm reporting function, so that you are fully aware of the instance status.

- Graphical O&M platform: You can clone instances, back up data, recover data, or other operations, with one click on the O&M platform.
- Database kernel version management: Redis automatically upgrades itself to fix defects, freeing you from routine version management.

8.5 Scenarios

- **Read/Write splitting**

  The ApsaraDB for MongoDB service uses a three-node replica set high-availability architecture . The three data nodes are located on different physical servers and automatically synchronize data. The primary and secondary nodes provide service. The two nodes provide independent domain names and, with the MongoDB driver, can independently allocate read pressure.
- **Service flexibility**

  Because MongoDB uses a No-Schema method, it is very suitable for businesses in their initial stages, as it avoids the need to change table structures. By storing fixed, structured data in RDS, flexible business data in MongoDB, and frequently accessed data in KVStore for Memcache or KVStore for Redis, you can achieve efficient data storage and reduce investment costs.
- **Mobile applications**

  ApsaraDB for MongoDB supports two-dimensional space indexes, so it provides great support for location-based mobile app services. At the same time, the dynamic storage method of MongoDB is especially suitable for storing heterogeneous data from multiple systems thus satisfying the needs of mobile apps.
- **IoT applications**

  ApsaraDB for MongoDB provides an asynchronous data writing function. It can provide memory database performance that is effective for special scenarios such as IoT high concurrency writing. At the same time, MongoDB map-reduce function can perform aggregated analysis on large data volumes.

  ApsaraDB for MongoDB supports cluster versions, so it can dynamically add mongos and shard components and resize their configurations, allowing unlimited performance and storage space scalability. This is well-suited for IoT scenarios with massive data volumes and high concurrency and performance requirements.
- **Core log systems**

In asynchronous disk scenarios, ApsaraDB for MongoDB can provide excellent plugin performance and has memory database processing capabilities. MongoDB provides a secondary index function, to meet the need for dynamic queries. It can use the map-reduce aggregate framework to perform multidimensional data analysis.

8.6 Limits

Operation	Limit
Database replication	Currently, you cannot manually build a secondary node.
Database restart	You can restart the instance only through the console.

8.7 Glossary

Concept	Explanation
Region	Region refers to the geographical location of the server for a user-purchased MongoDB instance. You can specify the region when activating the MongoDB instance. For now, the region cannot be modified after the instance has been purchased. When purchasing a MongoDB instance, you must use it with an Alibaba Cloud ECS instance. MongoDB only supports intranet access, so the selected region must be the same as that of the ECS instance.
Zone	Zone refers to the physical zones with separate power supplies and networks in the same region. Intranet communication can take place between zones, but network latency is lower within a zone. Fault isolation can be performed between zones. Single-zone refers to the case where the three nodes in the MongoDB instance replica set are located in the same zone. If the ECS and MongoDB instances are deployed in the same zone, the network latency will be lower.
Instance	A MongoDB instance, or simply an instance, is the basic unit of the MongoDB service purchased by users. The instance is the operating environment for ApsaraDB for MongoDB and exists as a separate process on the host. Users can use the console to create, modify, and delete MongoDB instances. Instances are mutually independent and their resources are isolated. They do not compete for CPU, memory, IO, and other resources. Each instance has its own features, such as database type and version. The system has corresponding parameters to control instance behavior.
Memory	The maximum memory that can be used by an ApsaraDB for MongoDB instance.

Concept	Explanation
Disk capacity	Disk capacity is the size of the disk which the user selects when purchasing the MongoDB instance. The disk capacity occupied by the instance includes set data and the space required for normal instance operation, such as the system database, database rollback log, redo log, and indexing. Ensure that the disk capacity is sufficient for the MongoDB instance to store data, otherwise, the instance may be locked. If insufficie nt disk space causes the instance to be locked, the user can purchase a larger disk to unlock the instance.
IOPS	Measured in units of 4KB, IOPS is the maximum number of block device reads/writes per second.
CPU core	The maximum computing power of the instance. One core CPU has a minimum of 2.3 GHz hyperthreading (Intel Xeon series Hyper-Threading) computing power.
Connections	The number of TCP connections between clients and the MongoDB instance. If the client uses a connection pool, the connections between the client and MongoDB instance will be persistent. Otherwise, they will be short connections.
Cluster Version	ApsaraDB for MongoDB supports cluster versions. You can purchase multiple mongos and shard nodes and combine them with a single ConfigServer to form a cluster version. This allows you to easily create a MongoDB distributed database system.
Mongos	Mongos are MongoDB’s cluster request portals. All requests must be coordinated through mongos which act as request distribution centers. They are responsible for forwarding data requests to the corresponding shard servers. You can use multiple mongos as request portals, so that, if one goes offline, MongoDB requests can still be processed.
Configserver	The configserver stores all database metadata (route, shard, etc.) configuration. Mongos themselves do not store shard servers and data routing information, but only cached those information in the memory. When mongos are started for the first time or shut down and then restarted , they load configuration information from the configserver. If the configserv er information changes, all mongos are notified to update their statuses. This way, the mongos always have the correct routing information. The configserver stores shard route metadata. As there are high requirements for service availability and data reliability, ApsaraDB for MongoDB uses three-node replica sets to comprehensively ensure the reliability of the configserver's services.

9 ApsaraDB for Memcache

9.1 What is ApsaraDB for Memcache

ApsaraDB for Memcache is a memory-based cache service that supports high-speed access to large volumes of small data. ApsaraDB for Memcache can greatly cut down the back-end storage load and speed up the response of websites and applications.

ApsaraDB for Memcache supports the key-value data structure and can communicate with clients compatible with the Memcached protocol.

ApsaraDB for Memcache supports out-of-the-box deployment. It also relieves the database load for dynamic web applications through the cache service, thus improving the overall response speed of the website.

Like local self-built Memcached databases, ApsaraDB for Memcache is also compatible with the Memcached protocol and user environments, and you can use ApsaraDB for Memcache directly. The difference is that the hardware and data of ApsaraDB for Memcache are deployed in the cloud, providing complete infrastructure, network security, and system maintenance services.

9.2 Benefits

Ease of use

- Out-of-the-box service: Out-of-the-box use is immediately available after purchase, facilitating fast business deployment.
- Compatible with open-source Memcache: Compatible with Memcached Binary Protocol. All clients complying with the protocol (binary SASL) can connect to ApsaraDB for Memcache.
- Visualized management and monitoring panel: The console provides multiple monitoring metrics to help you manage Memcache instances.

Cluster features

Super large capacity and super high performance: The default cluster output utilizes super large cluster instances to meet demands for large capacity and high performance.

Elastic resizing

- One-click resizing of storage capacity: You can adjust the storage capacity of an instance in the console based on business requirements.

- Uninterrupted services during online expansion: Instance storage capacity can be adjusted online without the need to stop services, with no impact to your own business.

Resource isolation

Instance-level resource isolation to better secure stability of a single user service.

Safe and reliable

- Password authentication is supported to ensure safe and reliable access.
- Persistent data storage: Memory plus hard disk storage meets data persistence demands while providing high-speed data reading/writing.

Seconds-level monitoring

- ApsaraDB for Memcache provides minute-level monitoring of historical data at the engine and resource levels.
- ApsaraDB for Memcache provides monitoring information on various data structures and interfaces to clearly display access information, so that you can fully understand the situation and effectively use ApsaraDB for Memcache.

High availability

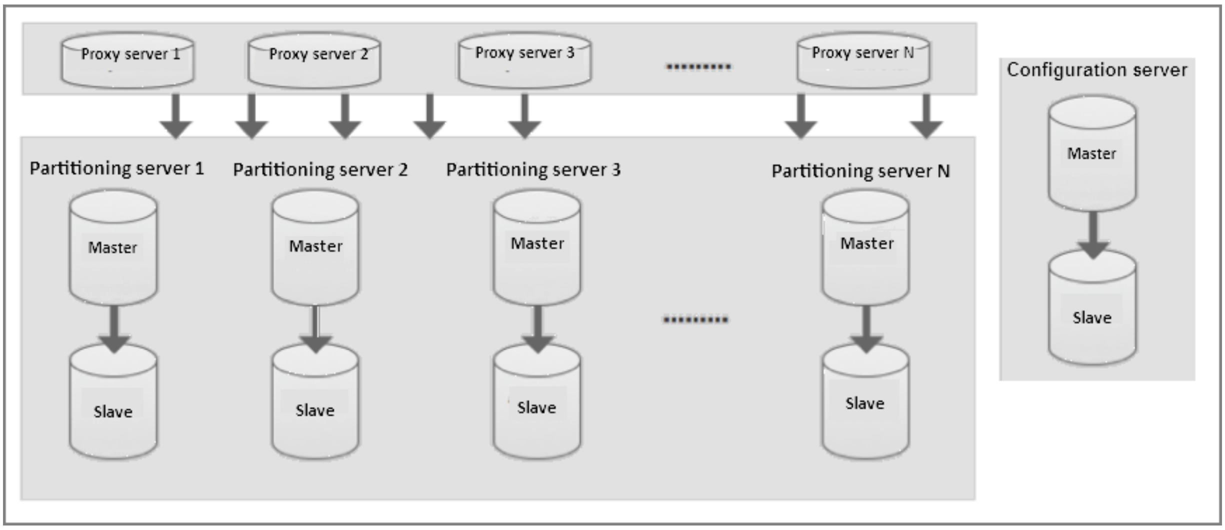
- Each instance has two nodes - a master node and a slave node - to avoid service interruptions because of single point of failure (SPOF).
- Automatic detection and recovery of hardware faults: ApsaraDB for Memcache enables automatic detection of hardware faults and can switchover within seconds to recover services.

9.3 Architecture

ApsaraDB for Memcache uses a cluster-based architecture. It is embedded with data partitioning and reading algorithms. The whole process is transparent to users, saving development and O&M troubles. Each partition node uses master-slave architecture to ensure high availability of services.

ApsaraDB for Memcache is comprised of three components, namely, the proxy server (service proxy), the partitioning server, and the configuration server.

Figure 9-1: Memcache architecture



Proxy server

Single-noded. A cluster structure may contain multiple proxies. The system automatically implements load balancing and fail-over for proxies.

Partitioning server

Each partitioning server is in a dual-copy high-availability architecture. The system automatica lly implements the master-slave switchover in case of a fault in the master node to ensure high availability of services.

Configuration server

The server is used to store cluster configuration information and partitioning policies. It adopts dual-copy architecture to ensure high availability.

Note:

- The number and configuration of the three components are specified by the system at purchase and are not customizable. Specification details are as follows:

Specification	Number of proxies	Number of partitioning server	Memory size of single partitioning server
1 GB	1	1	1 GB
2 GB	1	1	2 GB
4 GB	1	1	4 GB

Specification	Number of proxies	Number of partitioning server	Memory size of single partitioning server
8 GB	1	1	8 GB
16 GB	2	2	8 GB
32 GB	4	4	8 GB
64 GB	8	8	8 GB
128 GB	16	16	8 GB
256 GB	16	16	16 GB
512 GB	32	32	16 GB

- A Memcache cluster exposes a uniform domain for access. You can visit this domain for normal access to and data operations on Memcache. The proxy server, the partitioning server , and the configuration server do not provide domain access and you cannot directly access them for operations.

9.4 Specifications

ApsaraDB for Memcache uses a cluster-based architecture and its specifications are defined as follows.

Specification	CPU processing capability	Number of nodes	Maximum number of connections	Maximum intranet bandwidth
1 GB	Single-core	1	10,000	10
2 GB	Single-core	1	10,000	16
4 GB	Single-core	1	10,000	24
8 GB	Single-core	1	10,000	24
16 GB	Dual-core	2	10,000	96
32 GB	4-core	4	40000	192
64 GB	8-core	8	80000	384
128 GB	16-core	16	160000	768
256 GB	16-core	16	160000	768
512 GB	32-core	32	320000	1,536

Note:

512 GB type instances are not directly available now. You need to submit a *ticket* to apply for activation.

9.5 Features

Distributed architecture, freeing businesses from the impact of single point of failure (SPOF) events

- ApsaraDB for Memcache uses a distributed cluster architecture. Each node is composed of two servers for hot backup and is capable of automatic disaster tolerance and fail-over.
- Multiple types are available to cope with different business stresses with unlimited database performance expansion.
- ApsaraDB for Memcache supports data persistence and backup recovery policies to effectively secure data reliability and avoid the impact of huge stress to the backend databases when the cache becomes invalid because of physical node faults.

A multi-level security defense system to resist more than 90% of network attacks

- DDoS defense: Real-time monitoring at the entry point of network. The source IP address will be cleaned in the event of high-traffic attacks. If the cleaning turns out ineffective, the traffic will be redirected to a black hole.
- IP address white list mechanism: A maximum of 100 server IP addresses can be configured in the white list for accessing an instance, directly putting risks under control at the source.
- VPC virtual network: ApsaraDB for Memcache is fully connected to VPC and you can build an isolated network environment based on Alibaba Cloud.
- SASL authentication: SASL-enabled user identify authentication to safeguard data access security.

Improved tools to share your O&M workload for cache databases

- Monitoring and alarming: Real-time monitoring and alarming on instance information such as CPU utilization, IOPS, connections and disk space is provided so that you can see the status of the instance at all times.
- Data management: Visualized data management tools are available for you to easily handle data operations.

- Source code and distributed maintenance: Professional database kernel experts offer maintenance services to save your efforts for maintaining Memcache source code and distributed algorithms.

Specifications

ApsaraDB for Memcache uses a cluster-based architecture and its specifications are defined as follows.

Specification	CPU processing capability	Number of nodes	Maximum number of connections	Maximum intranet bandwidth
1 GB	Single-core	1	10,000	10
2 GB	Single-core	1	10,000	16
4 GB	Single-core	1	10,000	24
8 GB	Single-core	1	10,000	24
16 GB	Dual-core	2	10,000	96
32 GB	4-core	4	40000	192
64 GB	8-core	8	80000	384
128 GB	16-core	16	160000	768
256 GB	16-core	16	160000	768
512 GB	32-core	32	320000	1,536

9.6 Scenarios

Frequently-accessed businesses

Businesses such as social networks, e-businesses, games and advertisements, can store frequently-accessed data in ApsaraDB for Memcache and the underlying data in RDS.

Large promotion businesses

Large promotion or flash sales systems are usually under high access pressure. The average database simply cannot handle this amount of read stress, but ApsaraDB for Memcache can be a viable alternative.

Inventory system with a counter

ApsaraDB for RDS and ApsaraDB for Memcache can be used in combination. RDS stores the specific data information, while the database fields store the specific counter statistics. ApsaraDB for Memcache reads the statistics, while RDS stores the statistics.

Data analysis businesses

ApsaraDB for Memcache can be used in combination with open data processing service MaxCompute. It implements distributed analysis and processing of big data, which is suitable for big data processing scenarios such as business analysis and data mining. Data Integration service allows you to synchronize data between ApsaraDB for Memcache and MaxCompute on your own, simplifying data operations.

9.7 Limits

- ApsaraDB for Memcache only supports the key-value format data and does not support complex data types such as arrays, maps, and lists. Therefore, ApsaraDB for Memcache is not suitable for storing complex data types.
- ApsaraDB for Memcache's data is stored in the memory, and it is not guaranteed that the cache data will not be lost. Therefore, ApsaraDB for Memcache is not suitable for storing data which requires high consistency.
- ApsaraDB for Memcache supports a maximum of 1 KB and 1 MB respectively for the key size and value size of a single piece of cached data. ApsaraDB for Memcache is not suitable for storing sizable data.
- ApsaraDB for Memcache does not support transactions. Therefore, ApsaraDB for Memcache is not suitable for storing data with transaction requirements. Data with transaction requirements should be written directly to the database.
- When data access are evenly distributed, and there is no obvious hotspot or less popular data, a large number of access requests cannot hit the cache data in ApsaraDB for Memcache. Therefore, ApsaraDB for Memcache does not effectively function as the database cache. You should give full consideration to the data access requirements of the business model when selecting the database cache.

9.8 Glossary

Memcached

Memcached is a high-performance distributed caching system for memory objects. For the official introduction of Memcached, see *here*. ApsaraDB for Memcache is compatible with Memcached binary protocol and text protocol.

Instance ID

An instance corresponds to a user space. It is the basic unit for using ApsaraDB for Memcache. ApsaraDB for Memcache imposes different QPS and traffic limits on single instances of different capacity specifications. You can view the instance ID list on the console.

Connection address

The host address used for connecting to ApsaraDB for Memcache is displayed in the form of domain names. You can query the connection address in **Instance Information** > **Basic Information** > **Instance Details** > **Intranet Address** .

Connection password

The password used for connecting to ApsaraDB for Memcache. You can set the password at purchase, or reset the password after purchase.

Hit rate

Hit rate=Number of successful reads by the user/number of reads by the user.

Password-free access

Allows you access to the corresponding ApsaraDB for Memcache on an authorized ECS without a password. For more information, see Password-free Access chapter in the user guide of ApsaraDB for Memcache.

SASL

SASL is short for Simple Authentication and Security Layer. It is a mechanism that expands the verification capability of the client/server (C/S) mode. From version 1.4.3 Memcached supports SASL authentication. ApsaraDB for Memcache also uses SASL as the authentication mechanism because ApsaraDB for Memcache is shared by multiple tenants. Essentially, SASL uses passwords to ensure security of cached data. You are recommended to use a strong password and change the password periodically. ApsaraDB for Memcache will automatically perform an authentication every 60 seconds.

9.9 ApsaraDB for Memcache updates

Background

ApsaraDB for Memcache (original version) uses distributed cache architecture, but does not provide data reliability protection. When a fault occurs in a service node, although service reliability is protected, users need to push Memcache system after data loss by themselves due to the lack of a data persistence policy. This causes significant inconvenience.

To provide better services to customers, Alibaba Cloud database team has updated the ApsaraDB for Memcache product (released on May 10, 2017), which provides two servers for hot backup, data persistence, backup recovery and other advanced features, while also protecting service reliability, and providing a full set of database solutions such as disaster tolerance, recovery, monitoring, and migration to users.

Product form comparison

Modules	New Memcache	Old Memcache
Distributed architecture	Supported	Supported
Data persistence protection	Supported	Not supported
Two servers for hot backup architecture	Supported	Not supported
Memcache protocol compatibility	Completely compatible	Completely compatible

Sales mode comparison

The sale mode is more flexible, supporting subscription and pay-as-you-go.

Modules	New Memcache	Old Memcache
New subscription purchase	Supported	Not supported
Subscription upgrade	Supported	Not supported
Subscription renewal	Supported	Not supported
Subscription renewal and configuration change	Supported	Not supported
Auto renewal for subscription	Supported	Not supported
New pay-as-you-go purchase	Supported	Supported
Change pay-as-you-go configurations	Supported	Supported

Modules	New Memcache	Old Memcache
Release pay-as-you-go	Supported	Supported
Switch from pay-as-you-go to subscription	Supported	Not supported

Sales region support

Supports broader sale regions, including all international regions and regions in China.

Table 9-1: International region list

Region	Zone	New Memcache	Old Memcache
Asia Pacific (Singapore)	Asia Pacific 1 zone A	Supported	Not supported
Japan	Japan zone A	Supported	Not supported
Germany (Frankfurt)	Germany zone A	Supported	Not supported
Asia Pacific SE 2 (Sydney)	Australia zone A	Supported	Not supported
Hong Kong	Hong Kong zone C	Supported	Not supported
US East	US East 1 zone A	Supported	Not supported
Silicon Valley, USA	US West 1 zone B	Supported	Supported

Table 9-2: Chinese domestic region list

Region	Zone	New Memcache	Old Memcache
China East 1	China East 1 zone B	Supported	Supported
	China East 1 zone D	Supported	Supported
	China East 1 zone E	Supported	Supported
China East 2	China East 2 zone A	Supported	Supported
	China East 2 zone B	Supported	Supported
China South 1	China South 1 zone A	Supported	Supported
	China South 1 zone B	Supported	Supported
China North 1	China North 1 zone B	Supported	Supported
China North 2	China North 2 zone A	Supported	Supported
	China North 2 zone B	Supported	Supported

Region	Zone	New Memcache	Old Memcache
	China North 2 zone C	Supported	Supported

Function module comparison

Function coverage is comprehensively improved, with more focus on advanced database functions.

Function type	Features	New Memcache	Old Memcache
Backup recovery	Full backup	Supported	Not supported
	Backup recovery	Supported	Not supported
	Clone instance	Supported	Not supported
	Data flow operation	It is expected that DMS will support graphics by the end of June	Brief command line
Monitoring alarm	Resource monitoring	Supported	Supported
	Resource alarms	Supported by the end of May	Supported
Data security	White lists	Supported	Supported
	VPC support	Supported	Supported
	Password free access for white list	Supported	Supported
	Multiple Memcache instances support password free access	Supported	Not Supported

Related FAQs

- Q: How can I manage and change the configuration for the old version ApsaraDB for Memcache instance?

  A: You can continue managing old-version ApsaraDB for Memcache instances in Alibaba Cloud console. Activated old-version instances can be managed normally, changed in configuration and released.
- Q: How can I purchase another old-version ApsaraDB for Memcache instance?

A: The old-version ApsaraDB for Memcache instances cannot be purchased now. You may only purchase a new-version ApsaraDB for Memcache instance.

Q: Alibaba Cloud is expected to launch single-node ApsaraDB for Memcache instances in August 2017. By that time, the single-node instances will have the same prices as the old-version ApsaraDB for Memcache. You can flexibly choose single-node or dual-node instances to meet diverse business requirements.

- Q: How can I upgrade an old-version ApsaraDB for Memcache instance to a new version?

  A: Now ApsaraDB for Memcache does not support one-click upgrade from the old-version to the new version. To do this, you need to purchase a new-version instance, cache data manually, point your apps to the domain name of the new version instance and then release the old-version instance.

10 Server Load Balancer (SLB)

10.1 What is Server Load Balancer

Load Balancer is a traffic distribution control service that distributes the incoming traffic among multiple Elastic Compute Service (ECS) instances according to the configured forwarding rules. It expands the service capabilities of the application and increases the availability of the application.

By setting a virtual service IP address, Server Load Balancer virtualizes the ECS instances located in the same region into a high-performing and high-available application service pool. Client requests are distributed to the ECS instances in the cloud server pool according to the defined forwarding rules.

Server Load Balancer checks the health status of the ECS instances in the cloud server pool and automatically isolates any ECS instances with an abnormal status. This eliminates the single point of failure (SPOF) of an ECS instance and improves the overall service capability. Additionally, Server Load Balancer also provides the capability of defending DDoS attacks, which enhances security of the application.

Components

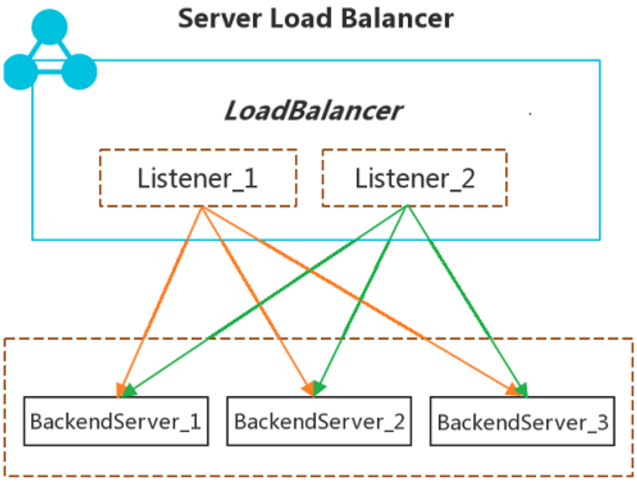
Server Load Balancer consists of the following components:

- **Server Load Balancer instances**: A Server Load Balancer instance is a running load balancing service that receives and distributes the incoming traffic to the backend servers.

  To use the Server Load Balancer service, you must create a Server Load Balancer instance with at least one listener and two ECS instances configured.
- **Listeners**: A listener checks the client requests and forwards the requests to the backend servers. It also performs health check on the backend servers.
- **Backend servers**: Backend servers are the ECS instances added to a Server Load Balancer instance to process the distributed requests. You can group the ECS instances hosting different applications or functioning different roles into different server groups.

As shown in the following figure, after the Server Load Balancer instance receives a client request , the listener forwards the request to the corresponding backend ECS instances according to the configured listening rules.

Figure 10-1: SLB components



10.2 Benefits

High availability

Server Load Balancer is designed to work in the full-redundancy mode without SPOF. Server Load Balancer supports local and cross-region disaster tolerance. When using together with DNS, service availability is up to 99.95%.

Scalability

Server Load Balancer can flexibly scale its service based on the application load without interrupting external services during traffic fluctuation.

Low cost

Server Load Balancer is 60% more cost-efficient than traditional hardware load-balancing systems without generating any O&M cost.

Security

Combined with Alibaba Cloud Security, Server Load Balancer can defend against up to 5 Gbps DDoS attacks, such as HTTP flood and SYN flood attacks.

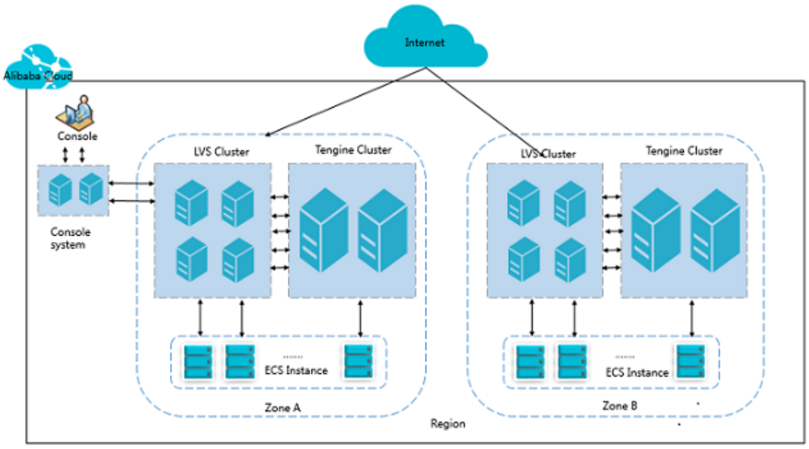
10.3 Architecture

Server Load Balancer is deployed in clusters. The cluster deployment model eliminates server single point failure, improves redundancy and increases service stability.

Alibaba Cloud provides the layer-4 (TCP protocol and UDP protocol) and layer-7 (HTTP protocol and HTTPS protocol) load balancing services.

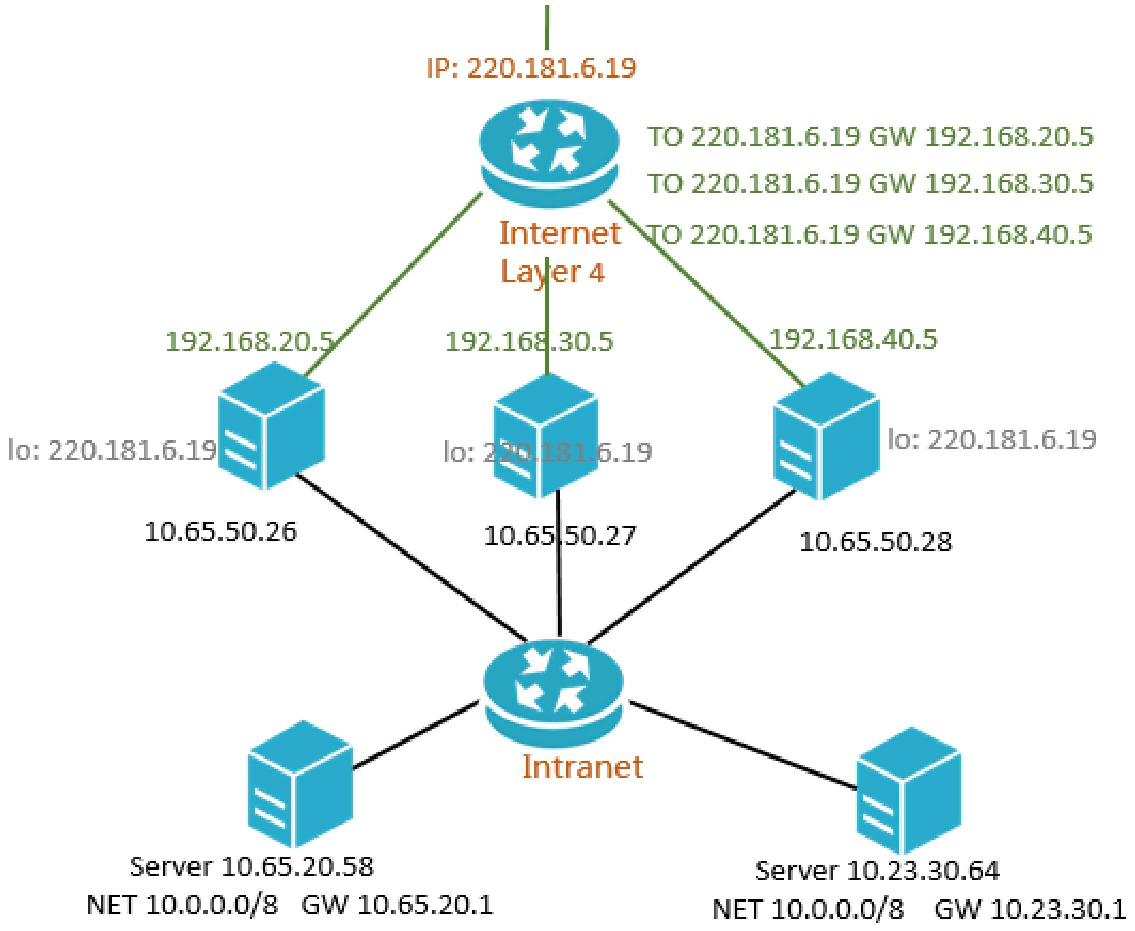
- Layer 4 uses the open source software Linux Virtual Server (LVS) with keepalived to achieve load balancing, and also makes some customization to it according to the cloud computing requirements.
- Layer 7 uses Tengine to achieve load balancing. Tengine is a Web server project launched by Taobao. Based on Nginx, it adds a wide range of advanced features dedicated for high-traffic websites.

Figure 10-2: SLB architecture



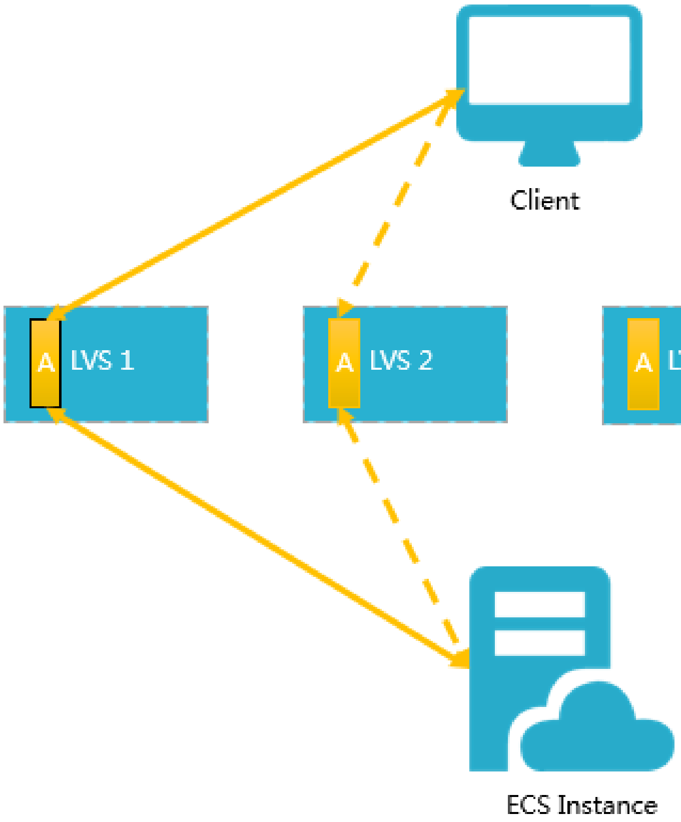
As shown in the following figure, the layer-4 load balancing in each region is actually run in a cluster of multiple LVS machines. The cluster deployment model strengthens the availability, stability, and scalability of the load balancing services in abnormal circumstances.

Figure 10-3: Layer-4 architecture



Additionally, the LVS machine in the LVS cluster uses multicast packets to synchronize sessions to other LVS machines. As shown in the following figure, session A established on LVS1 is synchronized to other LVS machines after three packets are transferred. In normal situations, the session request is sent to LVS1 as the solid line shows. If LVS1 is abnormal or being maintained , the session request will be sent to other machines working normally, as the dotted line shows. In this way, you can perform hot upgrades, machine failure maintenance, and cluster maintenance without affecting business applications.

Figure 10-4: Session persistence



10.4 Features

Layer-4 and layer-7 load balancing

Alibaba Cloud provides layer-4 (TCP and UDP) and layer-7 (HTTP and HTTPS) load balancing services. You can create different listeners to load balance for different applications. For example, create an HTTP listener to load balance HTTP applications.

Scheduling algorithms

Server Load Balancer supports the following scheduling algorithms:

- Round robin: Requests are distributed across the backend servers sequentially.
- Weighted least connections (WLC): In addition to weight set for each backend server, the number of connections to the client is also considered. The servers with a higher weight value will receive a larger percentage of live connections at any one time. If weights are the same, the system directs network connections to the server with the least established connections.

Health check

Server Load Balancer monitors the health of the added backend servers. When a backend server is declared as unhealthy, Server Load Balancer will stop forwarding requests to it and forward the request to other healthy backend servers.

Session persistence

Server Load Balancer supports session persistence. TCP listeners use IP addresses to establish sticky connections and HTTP/HTTPS listeners use cookies to establish sticky connections. With session persistence enabled, Server Load Balancer can forward requests from the same client to the same backend servers.

Access control

You can set a whitelist to control which IP addresses can access the load balancing service.

Certificate management

Server Load Balancer supports load balancing HTTPS applications and provides a certificate management function. You do not need to upload certificates to backend servers. Deciphering is performed on Server Load Balancer to reduce the CPU usage of backend servers.

Virtual server group

A virtual server group consists of a group of ECS instances. You can add ECS instances that run different applicants or provide different functions to different virtual server groups, and then you can create different forwarding tasks for different applicants to forward a specific request to a specified server group.

10.5 Scenarios

Server Load Balancer is applicable to the following scenarios:

Load balance high-traffic applications

If your application traffic is high, you can use Server Load Balancer to distribute the traffic to multiple ECS instances. Additionally, you can use session persistence feature to forward the session requests from a client to the same backend ECS instance to improve access efficiency.

Expand service capability for applications

You can extend the service capabilities by adding and removing backend ECS instances at any time based on the business demands. It is applicable to Web and App applications.

Eliminate the single point of failure (SPOF)

With health check, Server Load Balancer will automatically block unhealthy ECS instances and distribute requests to healthy ECS instances, eliminating any single point of failure.

10.6 Terms

Server Load Balancer

Server Load Balancer is a traffic distribution control service. It distributes incoming applicatio n traffic among multiple ECS instances according to the configured scheduling algorithm and listening rules.

Server Load Balancer instance

A Server Load Balancer instance is a running instance of the Server Load Balancer service. To use Server Load Balancer, you must first create a Server Load Balancer instance. The instance ID is a unique identifier for the Server Load Balancer instance.

Server Load Balancer IP

The IP address allocated to a Server Load Balancer instance. According to the instance type, the IP address is either a public IP or a private IP. You can resolve a domain name to the public IP address to provide external services.

Listener

A listener defines how the incoming requests are distributed. You must add at least one listener to a Server Load Balancer instance.

Backend server

The ECS instances that process the distributed requests.

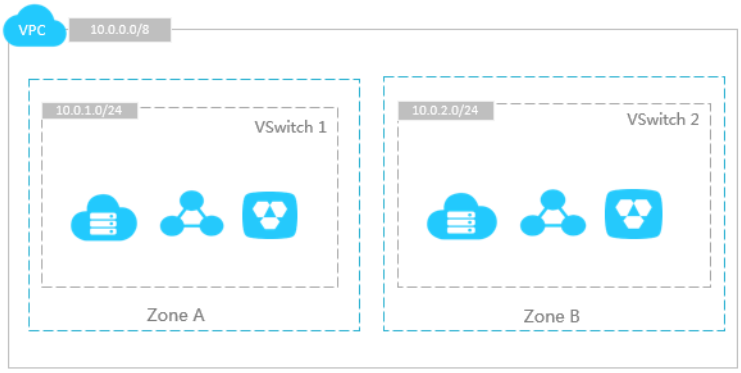
11 Virtual Private Cloud (VPC)

11.1 What is VPC

Virtual Private Cloud (VPC) is a private network established in Apsara Stack. VPCs are logically isolated from other virtual networks in Apsara Stack.

You have full control over your Alibaba Cloud VPC. For example, you can select its IP address range, further segment your VPC into subnets, as well as configure route tables and network gateways. Additionally, you can connect VPCs with a local network using a physical connection or VPN to form an on-demand customizable network environment. This allows you to smoothly migrate applications to the cloud with little effort.

Figure 11-1: Virtual Private Cloud



Each VPC consists of a private CIDR block, a VRouter and at least a VSwitch.

- CIDR block

  When creating a VPC or a VSwitch, you must specify the private IP address range in the form of Classless Inter-Domain Routing (CIDR) block. For more information, see *Classless Inter-Domain Routing*.

  You can use any of the following standard CIDR blocks and their subnets as the IP address range of the VPC.

  Note:

  To use a subnet of a standard CIDR block, you must use the `CreateVpc` API to create a VPC.

CIDR block	Number of available private IPs (system reserved ones not included)
192.168.0.0/16	65,532
172.16.0.0/12	1,048,572
10.0.0.0/8	16,777,212

- VRouter

  VRouter is the hub of a VPC. As an important component of a VPC, it connects VSwitches in a VPC and serves as the gateway connecting the VPC with other networks. After you successfully create a VPC, the system automatically creates a VRouter, which is associated with a route table.

- VSwitch

  VSwitch is a basic network device of a VPC and used to connect different cloud product instances. After creating a VPC, you can further segment your virtual private network to one or more subnets by creating VSwitches. The VSwitches within a VPC are interconnected. Therefore, you can deploy an application in VSwitches of different zones to improve the service availability.

11.2 Benefits

VPC features high security and flexible configuration, and supports multiple connection methods.

Secure

Each VPC has a unique tunnel ID, and each tunnel ID corresponds to only one VPC. VPCs are completed isolated from one another. Additionally, you can use security groups or whitelist to control access to cloud resources in the VPC.

Easy to use

You can quickly create and manage a VPC on the VPC console. After a VPC is created, the system automatically creates a VRouter and a route table for it.

Scalable

You can create multiple subnets in a VPC to deploy different services. Additionally, you can connect a VPC to a local data center or other VPCs to expand the network architecture.

11.3 Architecture

Based on tunneling technologies, VPC isolates virtual networks. Each VPC has a unique tunnel ID, and a tunnel ID corresponds to only one VPC.

Background information

With the continuous development of cloud computing, virtual network requirements are getting higher and higher, such as scalability, security, reliability, privacy, and higher requirements of connection performance. Therefore, a variety of network virtualization technologies is raised.

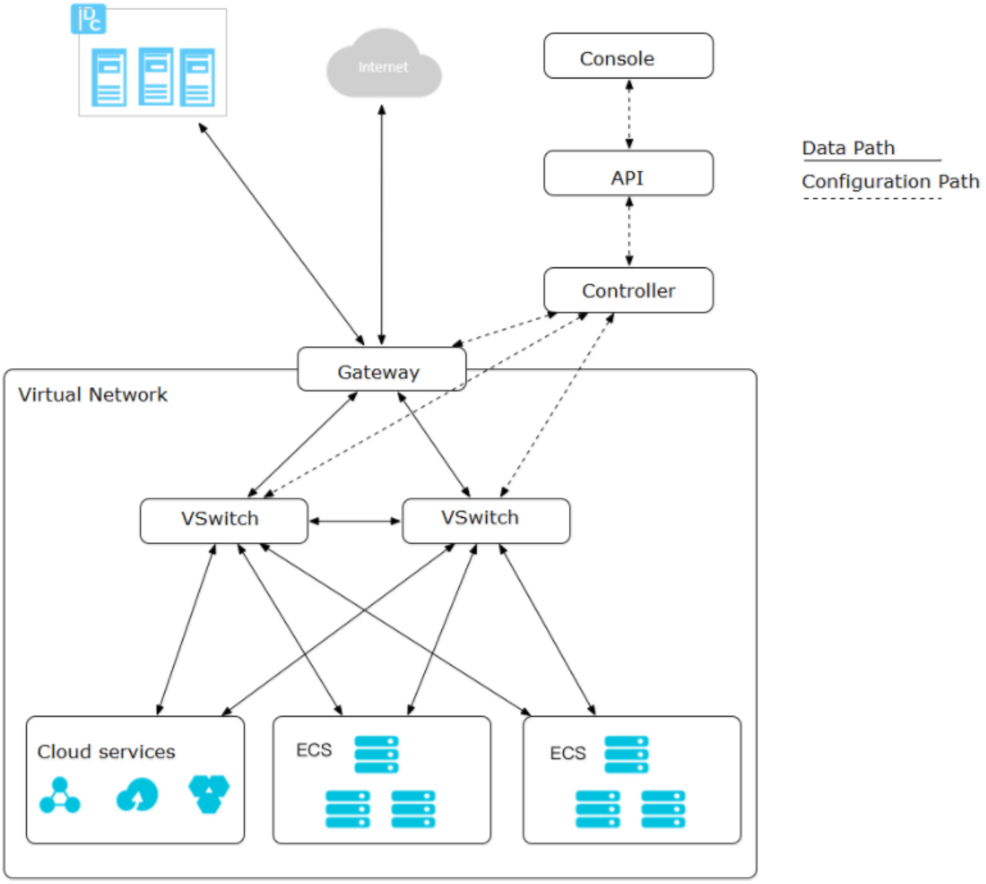
The earlier solutions combined the virtual machine's network with the physical network to form a flat network architecture, such as the large layer-2 network. With the increase of virtual network scalability, problems are getting more serious for the earlier solutions. These problems include ARP spoofing, broadcast storms, host scanning, and more. Various network isolation technologies emerged to resolve these problems by completely isolating the physical networks from the virtual networks. One technology isolates users with VLAN, but VLAN only supports up to 4096 nodes. It cannot support the huge amount of users in the cloud.

VPC basis

Based on tunneling technologies, VPCs isolate virtual networks. Each VPC has a unique tunnel ID, and a tunnel ID corresponds to only one VPC. A tunnel encapsulation carrying a unique tunnel ID is added to each data packet transmitted between the ECS instances within a VPC. Then, the data packet is transmitted over the physical network. Because the tunnel IDs are different for ECS instances in different VPCs and the IDs are located on two different routing planes, the ECS instances from different VPCs cannot communicate with each other and are isolated by nature. With the tunneling technologies and Software Defined Network (SDN) technology, Alibaba Cloud develops VPC in the basis of hardware gateways and self-developed switches.

Logical architecture

As shown in the following figure, the VPC architecture contains three main components: VSwitches, gateway, and controller. VSwitches and gateways form the key data path. Controller s use the self-developed protocol to forward the forwarding table to the gateway and VSwitches, completing the key configuration path. In the overall architecture, the configuration path and data path are separated from each other. VSwitches are distributed nodes, the gateway and controller are deployed in clusters, and all links have redundant disaster recovery. This improves the overall availability of the VPC.



11.4 Functions

VPC supports customizing the IP address range of the virtual network, controlling the traffic flow, and connecting with other networks.

Private network customization

VPC provides you with the capability to customize your private network. When creating a VPC and a VSwitch, you are allowed to specify the private IP address ranges for them. Additionally, you can further segment a VPC into one or more subnets by creating VSwitches. Then, deploy different services to different subnet to improve the service availability.

Traffic control

VPC provides you with the capability to control the traffic going through a VPC by adding custom route entries to the route table of the VPC.

The longest prefix match algorithm is used to route the network traffic when more than one route entries match the destination IP address. That is, the route entry with the longest subnet mask (the most specific route) is used.

VPC peer connection

VPC provides you with the router interface function to establish a private and secure connection between two VPCs.

Internet access

VPC provides you with the Elastic IP Address (EIP) and NAT Gateway services to enable the Internet access for the ECS instances in a VPC. Choose an Internet access method based on your business needs.

Local IDC connection

VPC provides you with the leased line function to connect a local data center with a VPC to build a hybrid cloud. Comparing with the Internet access, the leased line connection is more secure with low latency.

11.5 Scenarios

VPC applies to scenarios with high requirement on communication security and service availability.

Host applications

You can host an application that provides external services in a VPC and control Internet access by creating security group rules and whitelist. You can also control the access by isolating the application server from the database. For example, deploy the web server in a subnet that can access the Internet and deploy the database of the application in a subnet without Internet access.

Host applications requiring the access to the Internet

You can host an application that requires to access the Internet in a subnet of a VPC and route the traffic through NAT. By configuring SNAT rules, the instance in the subnet can access the Internet without exposing its private IP address and the private IP address can be changed to a public IP address any time to avoid external attacks.

Cross-zone disaster tolerance

You can create one or multiple subnets in a VPC by creating VSwitches. Different VSwitches in a VPC can communicate with one another through the intranet. You can deploy resources in VSwitches of different zones to achieve cross-zone disaster tolerance.

Business system isolation

Different VPCs are logically isolated from one another. If you must isolate multiple business systems, such as isolating the production environment from the test environment, you can create multiple VPCs.

11.6 Limits

VPC

Resource	Default limit
Maximum number of VRouters in a VPC	1
Maximum number of route tables in a VPC	1
Maximum number of VSwitches in a VPC	24
Maximum number of route entries in a route table	48

VRouter and VSwitch

Resource	Default limit
VRouter	• Each VPC can have only one VRouter • Each VRouter can have only one route table. • BGP and OSPF are not supported.
VSwitch	• Layer-two multicast and broadcast are not supported.

11.7 Concepts

Virtual Private Cloud (VPC)

Virtual Private Cloud (VPC) is a private network established in Apsara Stack. VPCs are logically isolated from other virtual networks in Apsara Stack.

VSwitch

A VSwitch is a basic network device of a VPC and used to connect different cloud product instances. When creating a cloud product instance in a VPC, you must specify the VSwitch that the instance is located.

VRouter

A VRouter is a hub in the VPC that connects all VSwitches in the VPC and serves as a gateway device that connects the VPC to other networks. VRouter routes the network traffic according to the configurations of route entries.

Route table

A route table is a list of route entries in a VRouter.

Route entry

Each entry in a route table is a route entry. A route entry specifies the next hop address for the network traffic destined to a CIDR block. It has two types of entries, system route entry and custom route entry.

12 Log Service (Log)

12.1 What is Log Service?

Log Service (or Log for short) is an all-in-one service for log-type data. It has been honed by countless big data scenarios at Alibaba Group. Without any development, you can quickly collect , consume, deliver, query, and analyze log data by using Log Service. It helps increase the O&M efficiency and build capabilities to process high-volume logs in this data technology (DT) era.

Log Service uses a Logtail agent or JS to collect events, binlogs, text logs, and logs in other formats in real time. It provides an interface for real-time consumption of the log data collected from the server, such as real-time retrieval and log analysis, and allows you to create data reports in diverse styles based on your analysis scenario and retrieval results.

12.2 Benefits

Managed security service

- Great accessibility allows you to set up and connect to Log Service within five minutes, and use Agents to collect data in any network environment.
- LogHub has all the functions of Kafka, and provides complete functional data, such as monitoring and alarms. It also supports auto scaling (by PB/day), saving costs of more than 50 % compared to self-deployed systems.
- LogSearch/Analytics provide query saving, dashboard, and alarm functions, saving costs of more than 80% compared to self-deployed systems.
- More than 30 access methods, seamless interworking with cloud products (OSS, E-MapReduce, MaxCompute, Table Store, MNS, and CDN) and open-source software (such as Storm and Spark).

Rich ecosystem

- LogHub supports over 30 collectors, including Logstash and Fluent, and can be easily connected using embedded devices, web pages, servers, and programs. It can also interconnect with consumption systems such as Spark Streaming, Storm, and Cloud Monitoring.
- LogShipper supports a variety of data formats (including TextFile, SequenceFile, and Parquet) and user-defined partitions. The data can be directly used by storage engines such as Presto, Hive, Spark, Hadoop, E-MapReduce, MaxCompute, and HybridDB.

- LogSearch/Analytics have complete syntaxes and are compatible with SQL-92. Supports interconnecting with Grafana by using JDBC protocol.

Real-time processing

- LogHub: Data can be used immediately after it is written. Logtail (the data collection agent) collects and transfers data to the server side within one second up to 99.9% of the time.
- LogSearch/Analytics: Data can be searched and analyzed immediately after it is written. If multiple search criteria are used, billions of data pieces can be searched within one second. When multiple aggregation conditions are used, hundreds of millions of data pieces can be analyzed within one second.

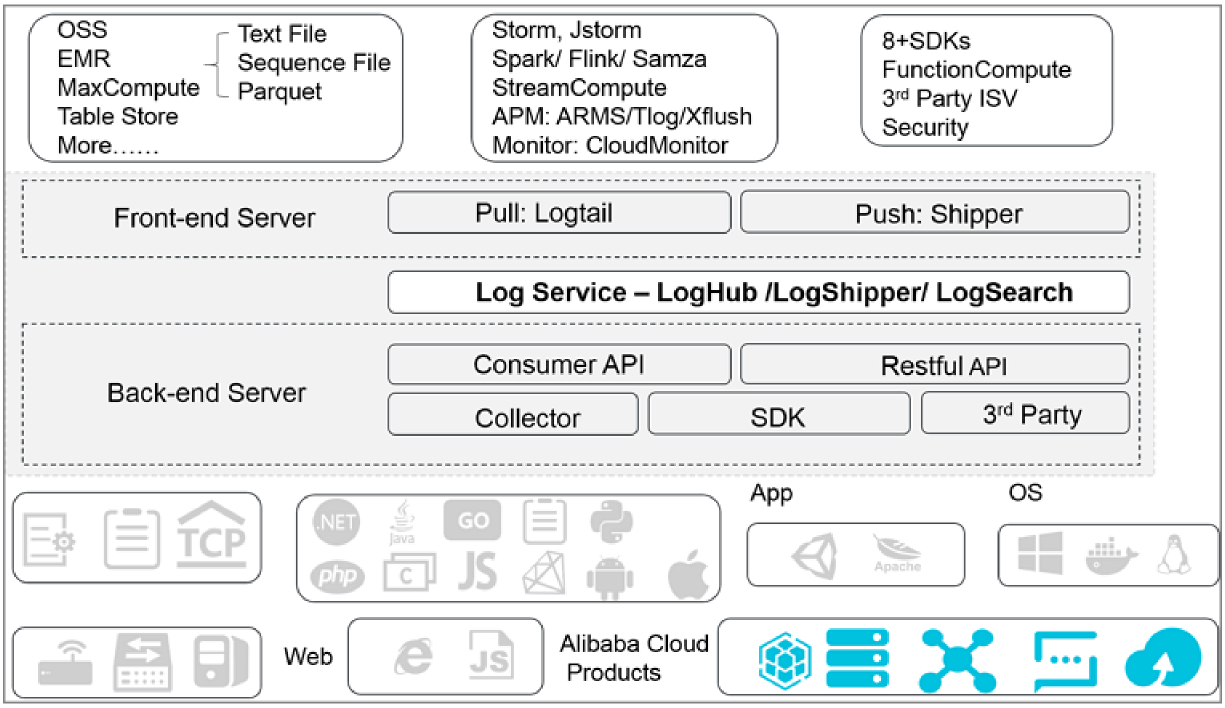
Complete API/SDK

- Log Service supports user-defined management and secondary development.
- All functions can be implemented using APIs/SDKs. SDKs for multiple languages are provided to facilitate service management.
- The query and analysis syntax is simple (compatible with SQL-92). The interfaces can be used to interconnect with the ecological softwares (supports Grafana interconnection solution).

12.3 Architecture

The Log Service architecture is shown in the following figure.

Figure 12-1: Log Service architecture



Logtail

- Non-invasive log collection based on log files
  - Only read files.
  - Non-invasive during reading process.
- Secure and reliable
  - Supports file rotation, so data are not lost.
  - Supports local caching.
  - Provides a network exception retry mechanism.
- Convenient management
  - Web client.
  - Visualize your configurations.
- Comprehensive self-protection
  - Real-time monitoring of process CPU and memory.
  - Restricts consumption of CPU/memory usage.

Frontend servers

Frontend machines are built using LVS+Nginx. Its features are as follows:

- HTTP and REST protocols
- Horizontal scaling
  - Support horizontal scaling when traffic increases.
  - Frontend machines can be quickly added to improve processing capabilities.
- High throughput, low latency
  - Pure asynchronous processing, a request exception will not affect other requests.
  - Lz4 compression is adopted to increase the processing capabilities of individual machines and reduce network bandwidth consumption.

Backend servers

The backend service is a distributed process deployed on multiple machines. It provides real-time Logstore data persistence, indexing, query, and shipping to MaxCompute. Features of the backend service are as follows:

- High data security
  - Each log you write is saved in triplicate.
  - Data are automatically recovered if damage to disks or machine downtime occurs.
- Stable service
  - Logstores automatically migrate in case of a process crash or machine downtime.
  - Automatic server load balancing ensures that traffic is distributed evenly among different machines.
  - Strict quota restrictions help prevent abnormal behavior of a single user from affecting other users.
- Horizontal scaling
  - Horizontal scaling is performed using shards as the basic unit.
  - You can dynamically add shards as needed to increase throughput.

12.4 Features

12.4.1 Core function

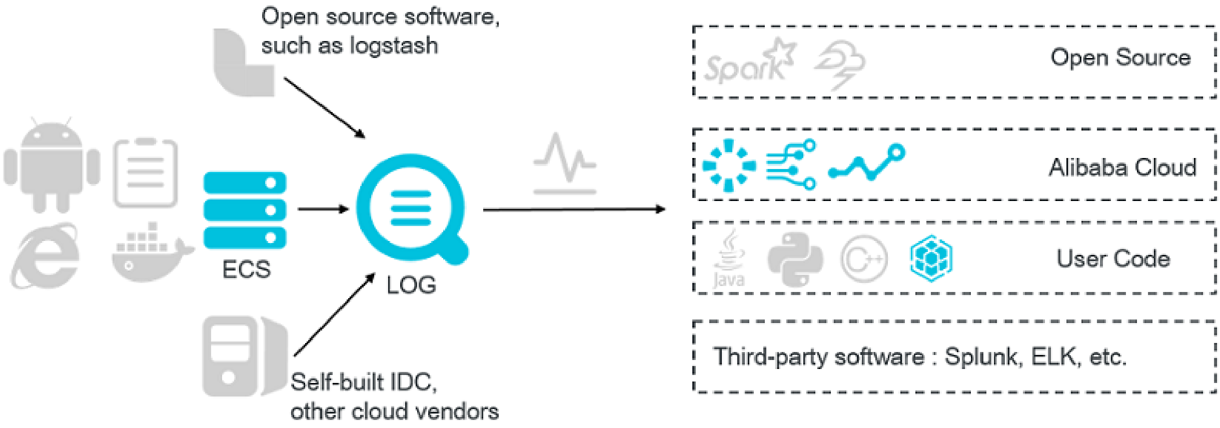
Real-time log collection and consumption (LogHub)

As a one-stop service for log data, Log Service (Log for short) experiences massive big data scenarios of Alibaba Group. Log Service allows you to quickly complete the collection, consumptio n, shipping, query, and analysis of log data without the need for development, which improves the Operation & Maintenance (O&M) efficiency and the operational efficiency, and builds the processing capabilities to handle massive logs in the DT (data technology) era.

Log Service has the following core functions.

- Use Elastic Compute Service (ECS), containers, mobile terminals, open-source softwares, and JS to access real-time log data (such as Metric, Event, BinLog, TextLog, and Click data).
- A real-time consumption interface is provided to interconnect with real-time computing and service.

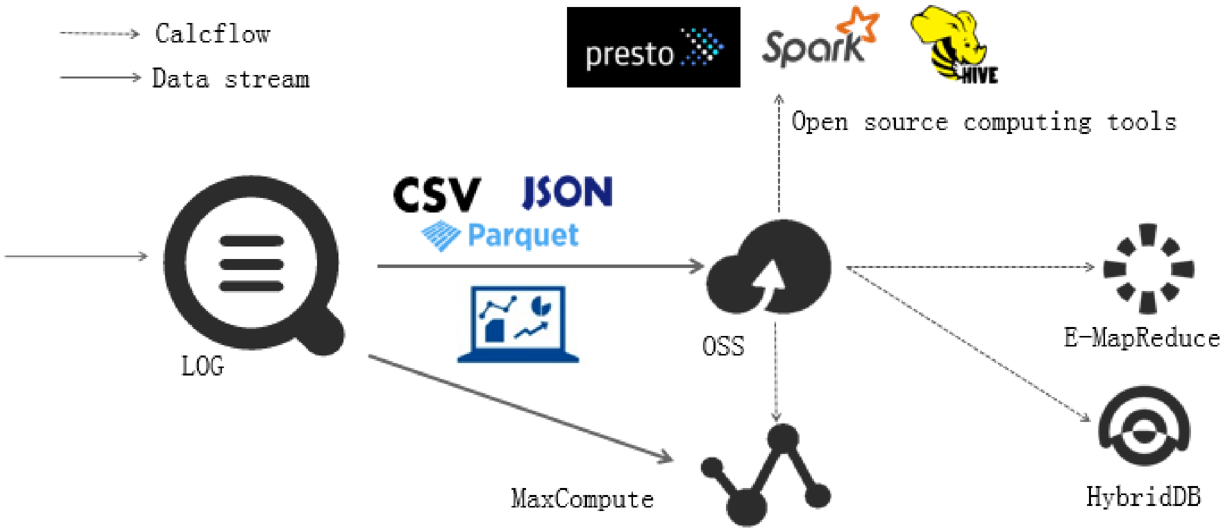
Figure 12-2: Real-time log collection and consumption



LogShipper

Stable and reliable log shipping ships LogHub data to storage services for storage and big data analysis. Supports various storage methods such as compression, user-defined partitions, row storage, and column storage.

Figure 12-3: LogShipper

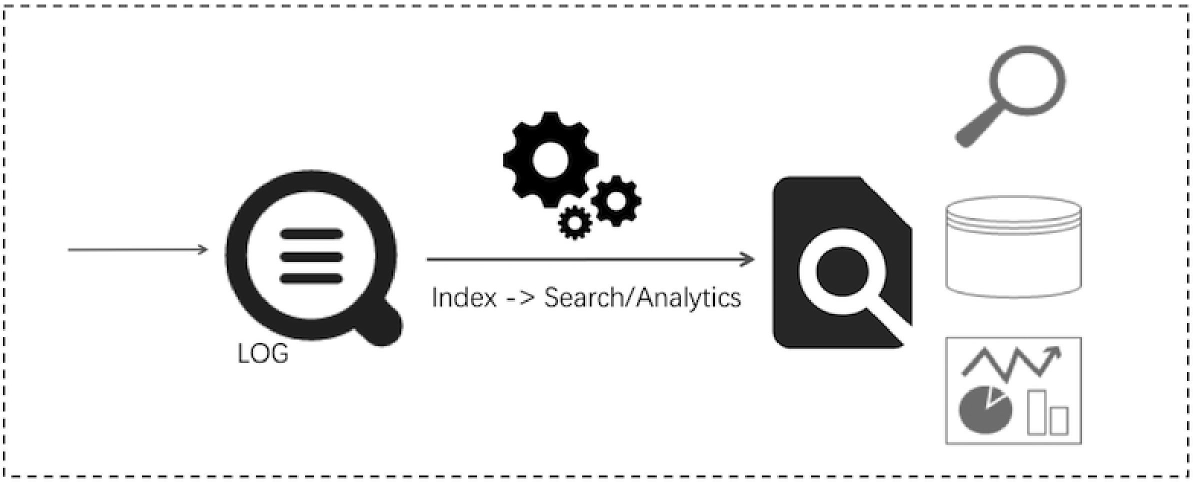


Query and real-time analysis (Search/Analytics)

Index, query, and analyze data in real time.

- Query: Keyword, fuzzy match, context, and range.
- Statistics: Rich query methods such as SQL aggregation.
- Visualization: Dashboard and report functions.
- Interconnection: Grafana and JDBC/SQL92.

Figure 12-4: Query and real-time analysis



12.4.2 Other functions

12.4.2.1 Log

Historically, the term "log" was associated with a thick notebook written by a ship captain or operator. Now with the advent of technology, logs are produced and consumed from all technological areas: servers, routers, sensors, GPS devices, orders, and various IoT devices generate and use logs.

What is a log?

Consider the example of a ship captain's log. In addition to a recorded timestamp, a log may contain all sorts of information, including a text record, an image, weather conditions, or sailing course. Presently, the concept of a "captain's log" has been expanded to include orders, payment records, user accesses, database operations, and many other fields that relate to today's technology.

The reason why logs are an enduring concept is that they are the simplest storage abstraction. A log is a series of records, that is, data arranged by chronological time order.

Each record has a unique log record number that fits into a definite sequence.

The log sequence is determined by time. From the preceding image, we can see that this log's time sequence runs from right to left. However, a log must always record the time an event happened.

Logs in Log Service

A log is an abstraction of system changes during the running process. The content is the time-ordered collection of some operations and operation results of specified objects. LogFile, Event, BinLog and Metric data are different carriers of logs. In LogFile, every log file is composed of one or more logs, and every log describes a single system event. A log is the minimum data unit processed in Log Service.

Log Service uses a semi-structured data model to define a log. This model is composed of four data fields: Topic, Time, Content and Source.

Furthermore, Log Service has different format requirements for different fields, as described in the following table.

Data Field	Description	Format
Topic	A custom field to mark a batch of logs. For example, access logs can be marked according to sites.	Any string up to 128 bytes in length, including null strings. By default, this field is a null string.
Time	This is a reserved field in the log and is used to indicate the generation time of the log. It is typically generated directly based on the time in the log.	It should be an Integer in standard UNIX time format. The unit is in seconds. This field indicates the number of seconds from 1970-1-1 00:00:00 UTC.
Content	This field is used to record the specific content of the log. Content is composed of one or more content items, and each content item is a Key-Value pair.	Key is a UTF-8 encoded string up to 128 bytes in length, and can contain letters, numbers, and underscores (_). It cannot start with a number. The following keywords cannot be used in the key: `__time__`, `__source__`, `__topic__`, `__partition_time__`, `_extract_others_`, and `__extract_others__`. The value can be any string up to 1024*1024 bytes.
Source	The source of the log, for example, the IP address of the machine generating the log.	Any string up to 128 bytes in length. By default, this field is null.

Various log formats are used in actual application scenarios. As an example, the following describes how to map an original Nginx access log to the Log Service log data model. Assume that the IP address of the user's Nginx server is `10.249.201.117`, and the following is the original log.

```
10.1.168.193 - - [01/Mar/2012:16:12:07 +0800] "GET /Send?AccessKeyId=
8225105404 HTTP/1.1" 200 5 "-" "Mozilla/5.0 (X11; Linux i686 on x86_64
; rv:10.0.2) Gecko/20100101 Firefox/10.0.2"
```

The original log is mapped to the Log Service log data model as follows.

Data Field	Content	Description
Topic	""	Use the default value (null string).
Time	1330589527	Precise generation time of the log, indicating the number of

Data Field	Content	Description
		seconds from 1970-1-1 00:00:00 UTC. This time is converted from the time stamp in the original log.
Content	Key-Value pair	Content of the log.
Source	"10.249.201.117"	Use the IP address of the server as the log source.

You can then decide how to extract the original content of the log and combine it into Key-Value pairs. The following table is an example.

Key	Value
ip	"10.1.168.193"
method	"GET"
status	"200"
length	"5"
ref_url	"-"
browser	"Mozilla/5.0 (X11; Linux i686 on x86_64; rv:10.0.2) Gecko/20100101 Firefox/10.0.2"

12.4.2.2 Project

A project is the Log Service's resource management unit. Projects isolate and control resources.

You can use a project to manage logs and related log sources of one application. A project manages Logstores of a user and the machine configurations of a log collection. It also serves as the portal for users to access the Log Service resources.

Projects can:

- Help you organize and manage different Logstores. You can use Log Service to collect and store different project, product, or environment logs in a centralized manner. You can classify different logs for management in different projects to facilitate subsequent log consumption, exporting, or indexing. In addition, projects also carry the log access permission management functions.

- Provide users' a portal to access Log Service resources. Log Service allocates a unique access portal to each created project. This access portal supports log writing, reading, and management through the network.

12.4.2.3 Logstore

Logstores are the units used in Log Service for log data collection, storage, and query. Each Logstore belongs to one project, and multiple Logstores can be created for a single project.

You can create multiple Logstores for one project according to your needs. Typically, an independent Logstore is created for each type of log in one application. For example, assume that you have the game application "big-game", and there are three types of logs on the server: operation_log, application_log, and access_log. You can first create a project named “big-game”, and then create three Logstores for the three types of logs under this project.

Whether writing or querying logs, you must specify a Logstore for the operation. If you want to ship the log data to MaxCompute for offline analysis, the data will be shipped in Logstore units for data synchronization (that is, the data in a single Logstore are shipped to a single MaxCompute table).

Logstores provide the following functions.

- Log collection, supports real-time logging
- Log storage, supports real-time consumption
- Index creation, supports real-time log query
- A data tunnel that ships logs to MaxCompute

12.4.2.4 Shard

Logstore read/write logs must be saved in a certain shard. Each Logstore is divided into several shards and each shard is composed of MD5 left-closed, right-open intervals. These intervals do not overlap and the interval ranges equates to the entire MD5 value range.

Range

When creating a Logstore, the entire MD5 range is automatically divided evenly based on the specified number of shards. Each shard has a certain range within the following value range: [00000000000000000000000000000000,ffffffffffffffffffffffffffffffff).

Each shard is composed of two keys, as follows:

- BeginKey: Indicates the start of the shard. This key is included in the shard range.
- EndKey: Indicates the end of the shard. This key is excluded from the shard range.

With the shard range, you can write logs by specifying the Hash Key, as well as split or merge shards. The corresponding shard must be specified when reading data from it. You can use load balancing mode or specified hash key mode when writing data. In load balancing mode, a data packet is written to an available shard at random. In specified Hash Key mode, data is written to the shard whose range includes the specified key.

In the following example, assume that the MD5 value range of the Logstore is [00,ff), and the Logstore contains 4 shards with the following ranges.

Shard No.	Range
Shard0	[00,40)
Shard1	[40,80)
Shard2	[80,C0)
Shard3	[C0,FF)

If you write a log and specify the MD5 key as 5F, the log data will be written into shard1 that contains the MD5 key 5F. If you specify the MD5 Key as 8C, the log data will be written into shard2 that contains the MD5 key 8C.

Shard read/write capacities

The service capacities of a shard are:

- Write: 5 MB/s, 2000 times/s
- Read: 10 MB/s, 100 times/s

You can calculate the number of shards needed based on the traffic. If you have set up several shards, you can also split or merge shards.

Note:

- When writing logs, if the API consistently reports a 403 or 500 error, refer to Logstore CloudMonitor metrics to view the traffic and status code and determine whether you need to increase the number of shards.
- For read/write operations that exceed a shard's service capacities, the system will attempt to provide services, but service quality cannot be guaranteed.

Shard status

- readwrite: capable of reading and writing data

- readonly: read-only data

When a shard is created, all the shards are in readwrite status. Split or merge operations change the shard status to readonly and generate a new shard in readwrite status. The shard status does not affect the performance of reading data. Shards in readwrite status maintain normal data writing performance, while shards in readonly status do not support writing data.

When splitting a shard, you must specify a ShardId in readwrite status and an MD5. The MD5 must be greater than the shard BeginKey and less than the shard EndKey. Split operations can split two other shards from one, that is, the number of shards is increased by 2 after the split. After the split, the status of the original shard specified to be split is changed from readwrite to readonly. Data can still be consumed, while new data cannot be written. The two newly generated shards are in readwrite status and arranged behind the original shard. The MD5 range of these two shards covers the range of the original shard.

When merging shards, you must specify a shard in readwrite status. Make sure the specified shard is not the last shard in readwrite status. The server automatically finds the adjacent shard at the right of the specified shard and merges these two shards. After the merge, the specified shard and the adjacent shard on the right are in readonly status. Data can still be consumed, while new data cannot be written. A new shard in readwrite status is generated and its MD5 range covers the total range of the original two shards.

12.4.2.5 Log topic

Logs in the same Logstore can be grouped by log topics. You can specify the topic when writing a log, and specify the log topic when querying logs. For example, users can use the user IDs as the log topics and write them into the logs. In this way, users can select to view only their own logs based on log topics. If there is no need to group specific logs in one Logstore, the same log topic can be used for all logs.

Note:

A null string is a valid log topic, and it is the default log topic for writing and querying logs. Therefore, if there is no need to use a log topic, the easiest method is to use a null string, when writing and querying logs.

12.5 Scenarios

Typical Log Service application scenarios include data collection, real-time computing, data warehousing and offline analysis, product operation and analysis, and O&M and management.

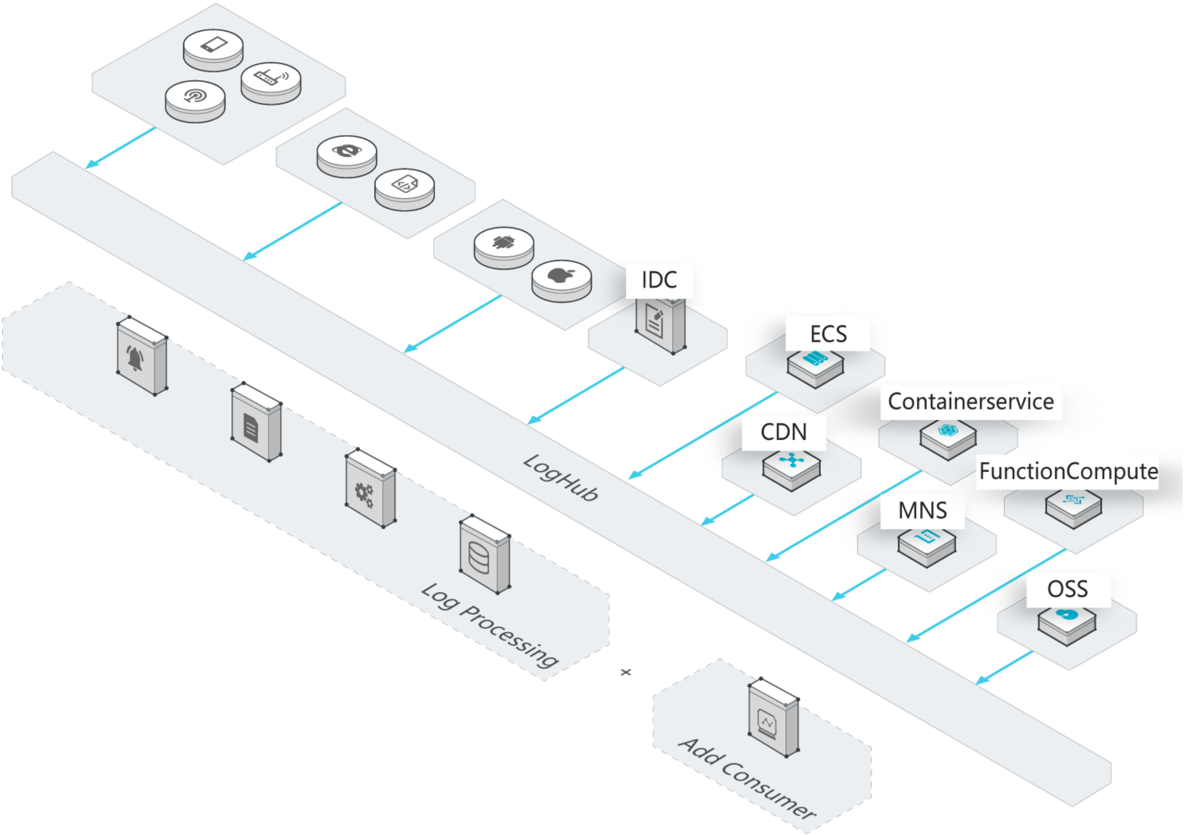
Data collection and consumption

The LogHub function of Log Service enables access to massive real-time log data (including Metric, Event, BinLog, TextLog, and Click data) at low costs.

Advantages:

- Easy-to-use: Over 30 real-time data collection methods are provided for you to quickly set up your platform. Log Service's powerful configuration and management capabilities can ease O&M workload with its globally distributed node network.
- Highly elasticity: Rapidly address traffic peaks and service growth.

Figure 12-5: Data collection and consumption

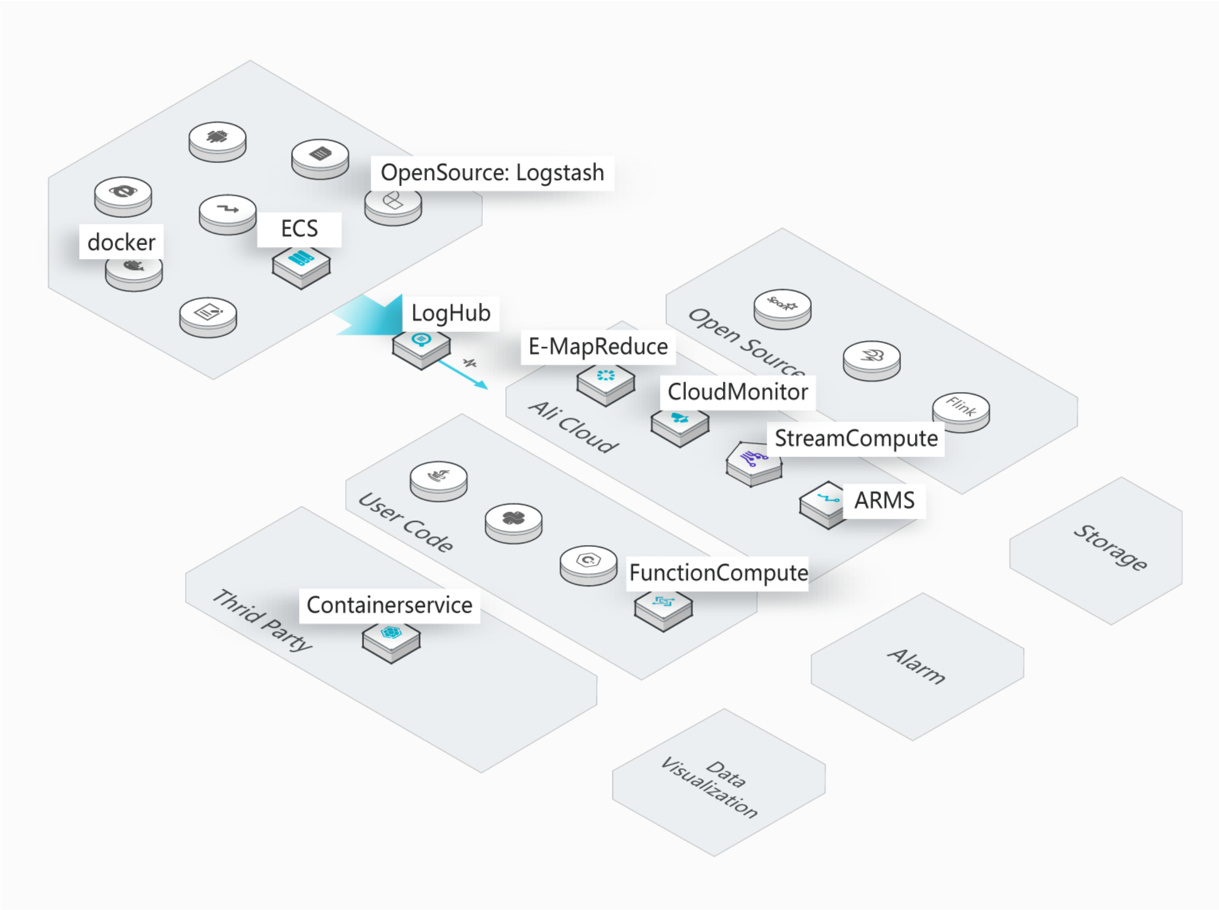


ETL/Stream Processing

The LogHub function can interwork with many real-time computing and services to provide complete progress monitoring and alerting functions, and support SDK/API-based custom consumption.

- Easy operation: Log Service provides various SDKs and programming frameworks, and can be seamlessly interconnected with various stream computing engines.
- Rich functions: Various monitoring data and alarm postponing are provided.
- Elastic: PB-level elasticity and zero latency.

Figure 12-6: ETL/Stream Processing

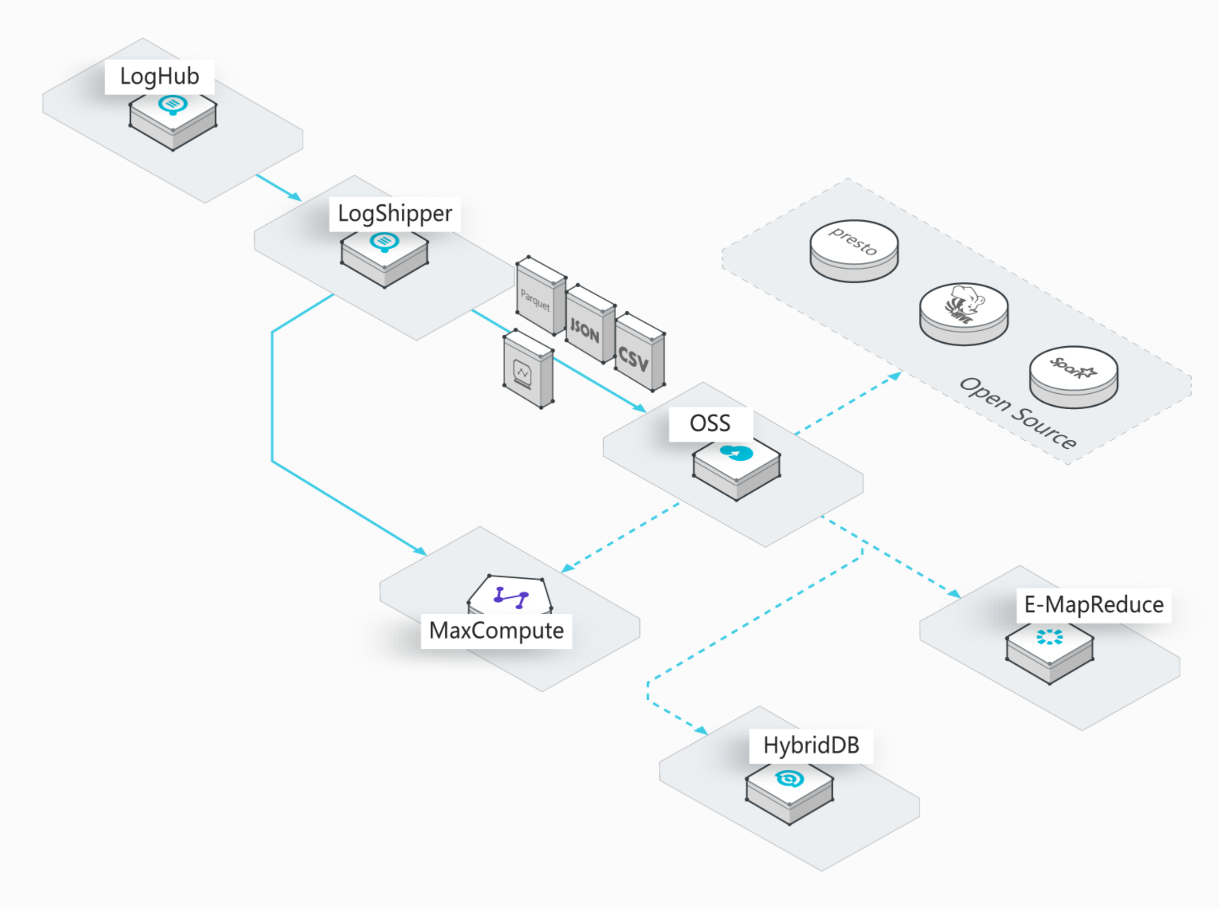


Data warehouse

The LogShipper function enables data in LogHub to be shipped to storage services and stored in a range of formats such as compression, custom partition, and row/column.

- Massive data: The amount of data is not limited.
- Rich storage formats: Supports various storage formats such as row storage, column storage, and TextFile.
- Flexible configuration: Supports configurations such as user-defined partitions.

Figure 12-7: Data warehouse

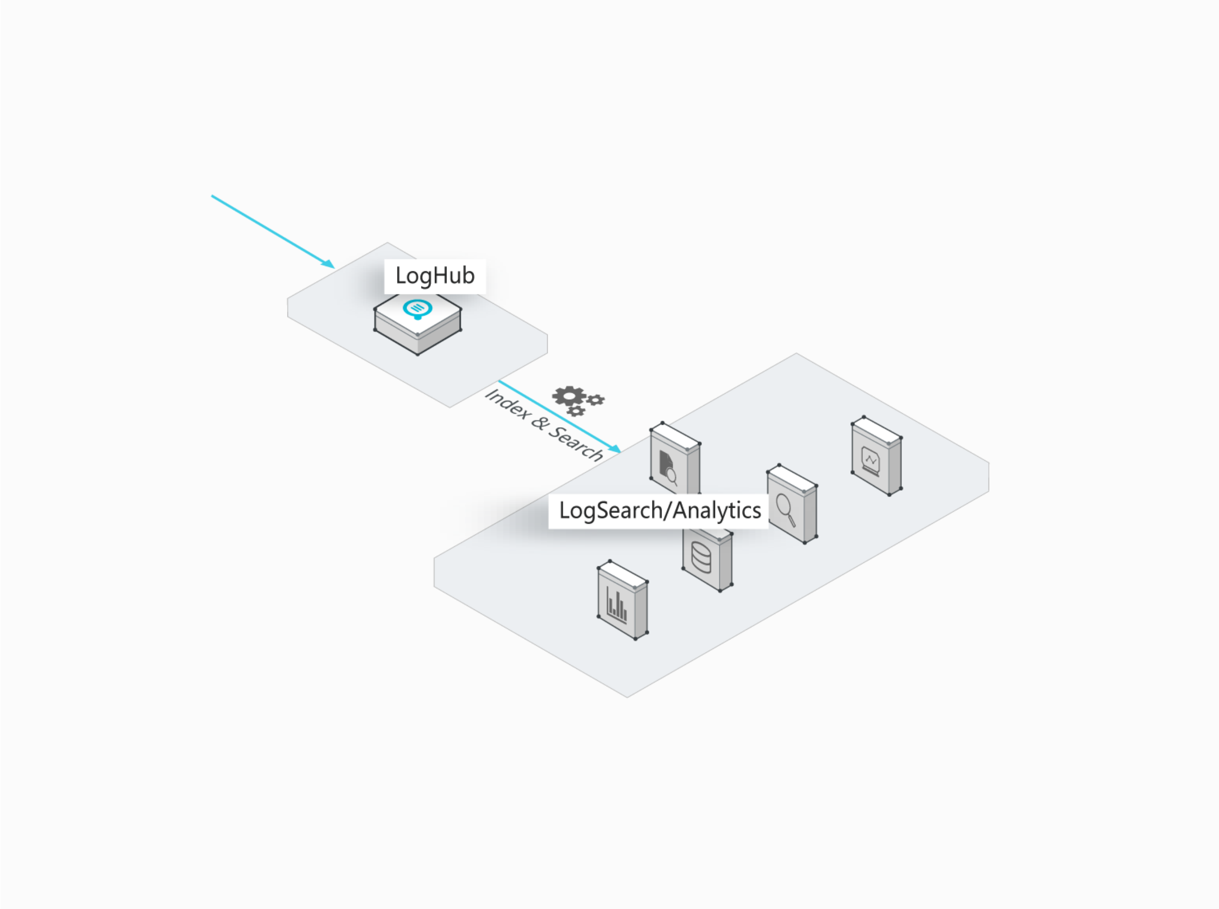


Log search and analytics

LogSearch/Analytics support real-time searching data in LogHub, and provide multiple inquiry methods, including keywords, fuzzy, context, scope, SQL aggregation.

- Strong real-time performance: Write after the query.
- Ultra low cost: Support PB/day indexing capacity, saving costs of up to 85% compared to self-built systems.
- Strong analytical skills: Support a variety of query means and SQL for aggregation analysis, and provide visualization and alarm function.

Figure 12-8: Log search and analytics



12.6 Limits

Resource limits

Item	Description	Note
Project	Up to 100 Projects can be created in each Department.	For more, open a ticket。
Logstore	Up to 100 Logstores can be created in each Project.	For more, open a ticket。
Shard	• Up to 10 Shards can be created in each Logstore. You can also split a shard increase the number of shards. • Up to 100 Shards can be created in each Project.	For more, open a ticket。

Item	Description	Note
Dashboard	• Up to 5 Dashboards can be created in each Project. • Up to 10 Charts can be added in each Dahboard.	For more, open a ticket。
Saved search	Up to 10 Saved search can be created in each Project.	For more, open a ticket。
Logtail configuration	Up to 100 Logtail configurat ions created in each Project.	For more, open a ticket。
ConsumerGroup	Up to 10 ConsumerGroups can be created in each Project.	For more, open a ticket。
Machine Group	Up to 100 Machine Groups can be created in each Project.	For more, open a ticket。
Log Retention Time	Collected logs can be kept for 1–365 days.	For more, open a ticket。

12.7 Glossary

Log

Log is an abstraction of system changes during the running process. The log content is a time-ordered collection of some operations and the corresponding operation results of specified objects . LogFile, Event, BinLog, and Metric data are different carriers of logs. In LogFile, every log file is composed of one or more logs, and every log describes a single system event. The log is the minimum data unit processed in Log Service.

Log group

A log group is a collection of logs and is the basic unit for writing and reading.

Log topic

Logs in a Logstore can be classified by log topics. You can specify the topic when writing and querying logs.

Project

The project is the resource management unit in Log Service and is used to isolate and control resources. You can manage all the logs and the related log sources of an application by using projects. Projects manage the information of all your Logstores and the log collection machine configuration, and serve as the portals where you can access the Log Service resources.

Logstore

The Logstore is a unit in Log Service for the collection, storage, and query of log data. Each Logstore belongs to a project, and each project can create multiple Logstores.

Shard

Each Logstore is divided into several shards and each shard is composed of MD5 left-closed and right-open intervals. Each interval range does not overlap with others and the total range of all the intervals is the entire MD5 value range.

13 Apsara Stack Security

13.1 What is Apsara Stack Security

Apsara Stack Security is a comprehensive Apsara Stack security solution that provides cloud security with network security, host security, application security, data security and security management dimensions.

In the cloud computing environment, the traditional border security protection that relies on the detection technology cannot guarantee the security of businesses on the cloud. Apsara Stack Security combines the powerful data analysis capabilities of cloud computing platform with professional security operation team, to provide a multi-level and integrated security protection service.

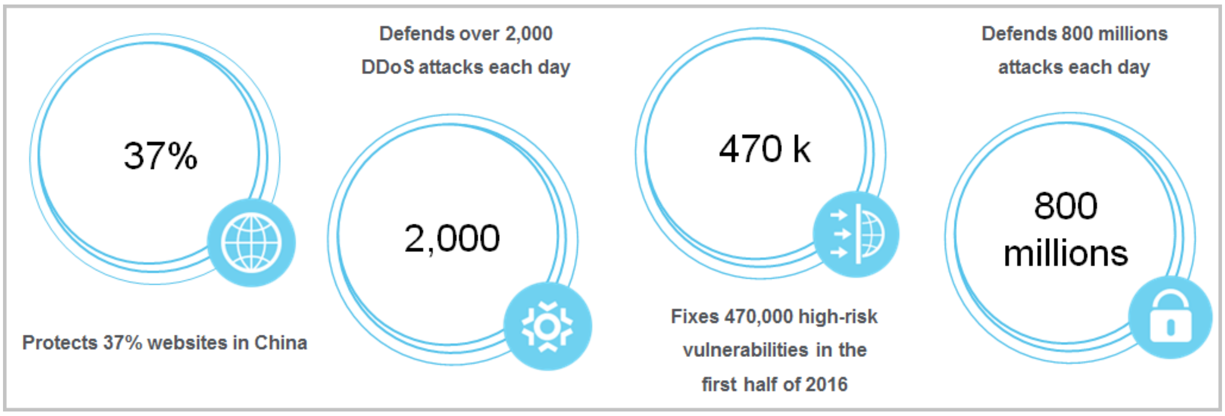
13.2 Benefits

13.2.1 Pioneer in cloud security

The Alibaba security team has been providing information security assistance to all internal business systems in the Alibaba Group since 2005, accumulating a wealth of security experience from that time. When Alibaba Cloud Security was first released in 2011, it began to offer comprehensive assurance for Alibaba Cloud security systems and has become a pioneer in cloud security.

Alibaba Cloud Security protects over 37% of Chinese websites. Each day, Alibaba Cloud Security defends against more than half of the large-traffic volume attacks in China and identifies and defends against 35,000 malicious IP addresses. Over the past year, Alibaba Cloud has helped users fix over 1.4 million vulnerabilities.

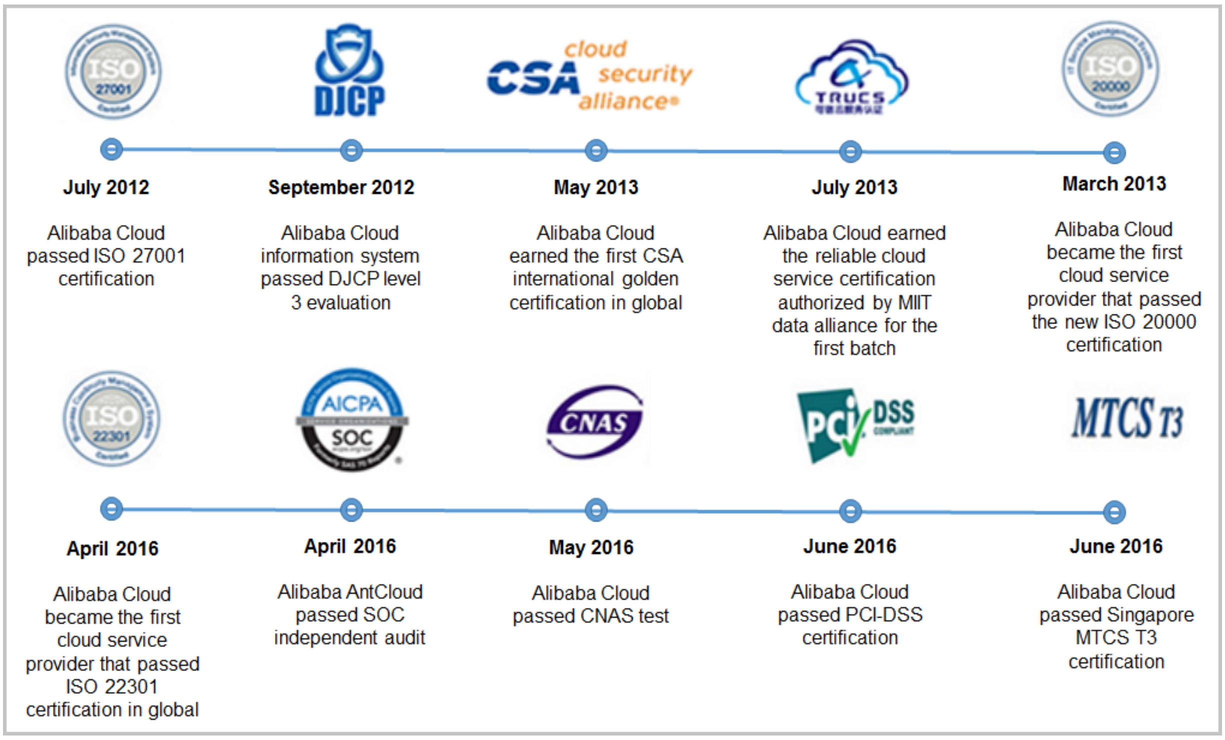
Figure 13-1: Alibaba Cloud Security processes massive Internet data



13.2.2 Authoritative certifications, secure and reliable

Alibaba Cloud has earned many international and domestic cloud security certifications because of the security features of the Alibaba Cloud platform and the attack prevention features of Alibaba Cloud Security.

Figure 13-2: Security certifications earned by Alibaba Cloud


July 2012
Alibaba Cloud passed ISO 27001 certification
September 2012
Alibaba Cloud information system passed DJCP level 3 evaluation
May 2013
Alibaba Cloud earned the first CSA international golden certification in global
July 2013
Alibaba Cloud earned the reliable cloud service certification authorized by MIIT data alliance for the first batch
March 2013
Alibaba Cloud became the first cloud service provider that passed the new ISO 20000 certification
April 2016
Alibaba Cloud became the first cloud service provider that passed ISO 22301 certification in global
April 2016
Alibaba AntCloud passed SOC independent audit
May 2016
Alibaba Cloud passed CNAS test
June 2016
Alibaba Cloud passed PCI-DSS certification
June 2016
Alibaba Cloud passed Singapore MTCS T3 certification

- First cloud service provider in the world to earn the CSA STAR Certification.
- First cloud security service provider in China to earn the ISO27001 international information security management system certification.

- First cloud computing system in China that passed the Ministry of Public Security's classified protection test (DJCP).
- First cloud classified protection pilot demonstration platform in China.
- Alibaba Cloud's e-government cloud platform was the first such platform to pass cloud service cyber-security review (enhanced level) by party and government departments.
- AntCloud passed the Level IV DJCP test with high marks, becoming the first Level IV cloud platform in China.

13.2.3 Complete systems, advanced technology

Apsara Stack Security is a product born from over ten years of protection experience. After a decade of experience in providing security services to Alibaba Group's internal business, Alibaba has accumulated a considerable number of security research achievements, massive security data, and extensive security operation and management approaches. On such a basis, Alibaba has built a professional cloud security team of experts. Apsara Stack Security brings together the rich experience of these experts to develop sophisticated systems that provide optimal security to cloud computing platforms. This product can effectively protect the security of the cloud platforms, cloud network environments, and cloud business systems of Apsara Stack users.

13.2.4 Comparison between Apsara Stack Security and traditional security products

Item	Traditional Security Products	Apsara Stack Security
Complete output of industry-leading security capabilities among Internet enterprises	Security vendors provide their own products with separate advantages, which cannot form an entire security protection system.	When defending against hacker attacks, Alibaba Cloud Security accumulates rich intelligence capabilities, discovers popular Internet attacks and 0-day attack measures in real time, and outputs complete security capabilities to users.
Analyzing risks in advance, and predicting eruption of risks	Without complete business monitoring capabilities, traditional security vendors cannot accurately analyze risks.	Apsara Stack Security could analyze and respond to major vulnerabilities and security events in time to prevent security risk explosion.

Item	Traditional Security Products	Apsara Stack Security
Implementation of security big data modeling analysis	The traditional single-feature detection method cannot discover any threats. Traditional security vendors analyze logs by data statistics and report display, which cannot achieve real security data modeling analysis.	Apsara Stack Security is the only product in the industry that discovers threats and globally displays the security situation. With the data model, Apsara Stack Security enables real situation awareness based on the historical data, network data, and host data.
Elastic security capability resizing, decoupled from hardware	Most security products are implemented by their custom hardware devices. Software-oriented security products are also implemented based on the virtual machines on the virtual platform.	• Apsara Stack Security uses the advanced cloud architecture design, where all function modules are based on the universal X86 hardware platform without dependency on the hardware. • In addition, based on the elastic resizing capabilities of the cloud, Apsara Stack Security's performance can be linearly expanded. When the performance is insufficient, more hardware devices can be smoothly installed without modifying the network structure.
Interacting between the network and host	Traditional security products achieve full coverage by device overlapping. However , no effective measures are provided for device interacting. Currently, traditional security vendors only manage platforms to collect and display logs and status of devices in a centralized manner. Detected vulnerabilities are output by report, and scan objects and	Apsara Stack Security provides complete Internet protection capabilities to ensure security of networks , applications, and hosts. Protection components interact with each other to form an overall protection system, so as to achieve the optimal effect when defending against known attacks.

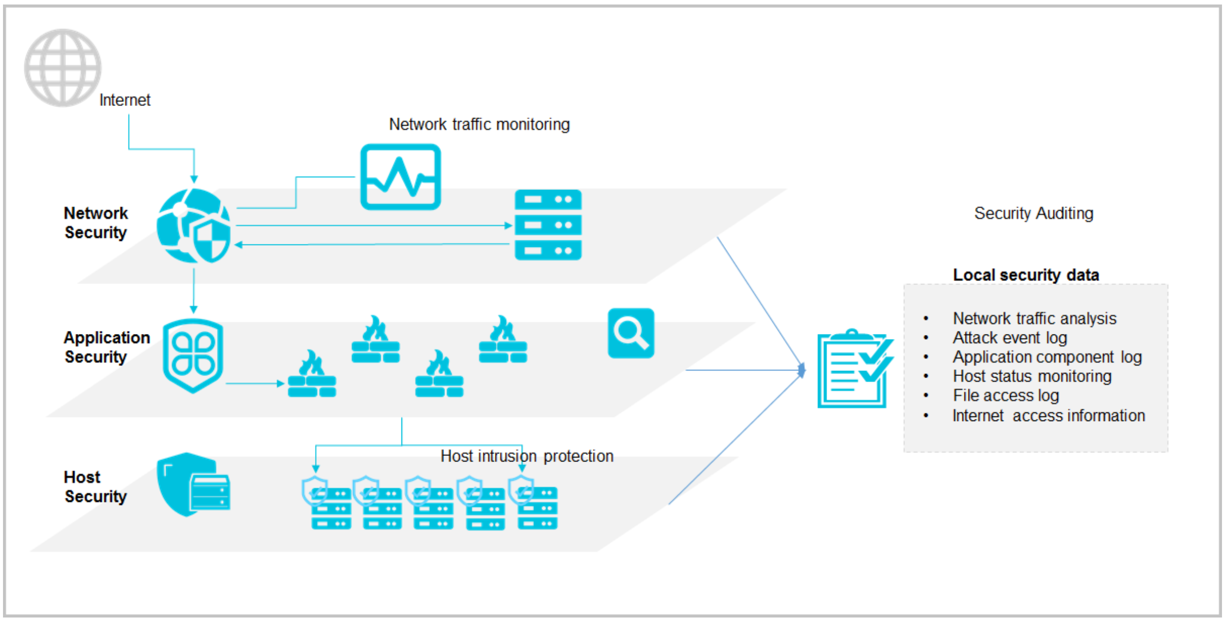
Item	Traditional Security Products	Apsara Stack Security
	ranges must be manually configured.	
interacting between the cloud and local device	Traditional security products are all deployed using offline hardware boxes, which only support feature database upgrade and do not interact with the cloud.	• Apsara Stack Security can interact with Alibaba Cloud Anti-DDoS Service Pro to achieve ultra-bandwidth defense. • Apsara Stack Security checks whether enterprise users' personal information exposed to the Internet leak out, for example, whether their enterprise emails, personnel accounts, and passwords leak out.
Compatible with all IDC environments, totally decoupled from the cloud platform	Most traditional security vendors provide security products using hardware boxes. As the SDN technology is widely used, the traditiona l mode cannot be fully compatible with the cloud environments.	Apsara Stack Security uses the architecture of network portal detection and server operating system interactin g. In terms of the security capabilities, Apsara Stack Security adopts the data analysis method to detect security threats. By using this architecture and method , Apsara Stack Security avoids the complex network structure in the IDC and is fully compatible with all IDC environments.

13.3 Architecture

Apsara Stack Security Basic Edition

Figure 13-3: Structure of Apsara Stack Security Basic Edition shows the structure of Apsara Stack Security Basic Edition in Apsara Stack Enterprise.

Figure 13-3: Structure of Apsara Stack Security Basic Edition



- **Network Traffic Monitoring**: This module is deployed on Apsara Stack network boundaries. It allows you to inspect and analyze each packet passing through the Apsara Stack network by traffic mirroring. The analysis results then are referenced by other Apsara Stack Security modules.
- **Host Intrusion Detection System (HIDS)**: This module is used to detect the integrity of the key folders in the host and send alerts for abnormal processes, ports, and network connections on the host.
- **Server Guard Basic Edition**: This module is deployed on an ECS instance, which detects and removes web Trojans, blocks brute force password cracking attacks, and sends alerts for abnormal logons.
- **Security Audit**: This module is used to collect database logs, host logs, console operation logs on the user and O&M sides, and network device logs in the Apsara Stack platform.

13.4 Features

Unlike traditional software and hardware security products, Apsara Stack Security adopts an in-depth defense and multi-point linkage cloud security architecture. Based completely on Alibaba Cloud's cloud computing environment R&D, Apsara Stack Security provides comprehensive and integrated cloud security protection capabilities to users, covering the network layer, application layer, host layer, and many other layers.

Functions of Apsara Stack Security Basic Edition

Table 13-1: Functions of Apsara Stack Security Basic Edition lists detailed functions provided by Apsara Stack Security Basic Edition.

Table 13-1: Functions of Apsara Stack Security Basic Edition

Module	Function	Function description
Network Traffic Monitoring	Traffic statistics	Collects inbound and outbound traffic of the interconnection switch (ISW) using a bypass in traffic mirroring mode and generates a traffic diagram.
	Abnormal traffic detection	Detects abnormal traffic that exceeds the threshold using the bypass in traffic mirroring mode.
	Web application attack protection	Conducts network-layer interception and bypass blocking for common web attacks based on embedded web application attack detection rules.
Host Intrusion Detection	Key folder integrity check	Checks integrity of files in specified folders of the system, detects tampering in time, and generates change alerts.
	Abnormal process alert	Detects startup of abnormal processes, and generates alerts.
	Abnormal port alert	Detects new port monitoring in time, and generates alerts.
	Abnormal network connection alert	Detects active connections with external public networks in time, and generates alerts.
Server Guard Basic Edition	Webshell detection and removal	Accurately detects and removes webshell scripts by means of rule matching, and allows you to manually isolate webshell scripts.
	Interception of brute-force password cracking	Detects and blocks brute force password cracking attacks initiated by hackers in real time.
	Remote logon alert	Analyzes and records users' frequently used logon locations to identify frequently used logon regions, and generates alerts for suspicious logon behaviors in non-frequently used logon regions.
Security audit	Raw log collection	Collects database logs, host logs, console operation logs on the user and O&M sides, and network device logs.

Module	Function	Function description
	Audit query	Allows you to query audit logs by audit type, audit object, operation type, operation risk level, alert, or creation time. In addition, full text retrieval of audit logs is supported.
	Policy setup	Allows you to configure audit rules using the following parameters: Initiator, Target, Command , Result, and Cause. In addition, the module can identify high-risk operations in raw logs, and generate alerts accordingly.

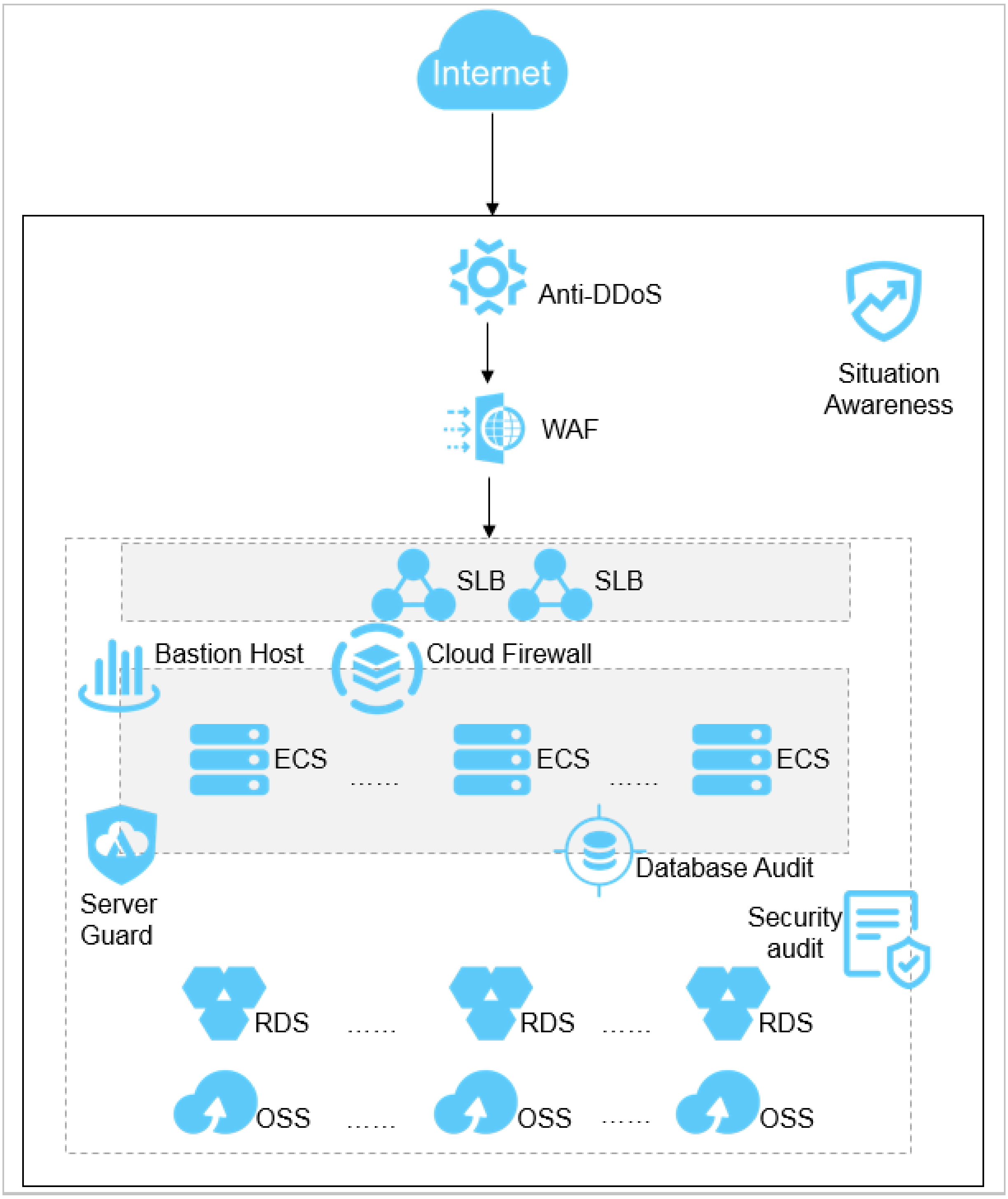
13.5 Scenarios

Internet-oriented

If your Apsara Stack platform is Internet-oriented and provides external services, we recommend that you reference the configuration shown in *Figure 13-4: Internet-oriented scenario*. You should protect Apsara Stack from multiple standpoints, such as network security, server security, and security management.

- Network security: DDoS Cleaning, WAF, and Cloud Firewall
- Server security: Server Guard
- Security management: Situation Awareness, Bastion Host, and Security Audit

Figure 13-4: Internet-oriented scenario

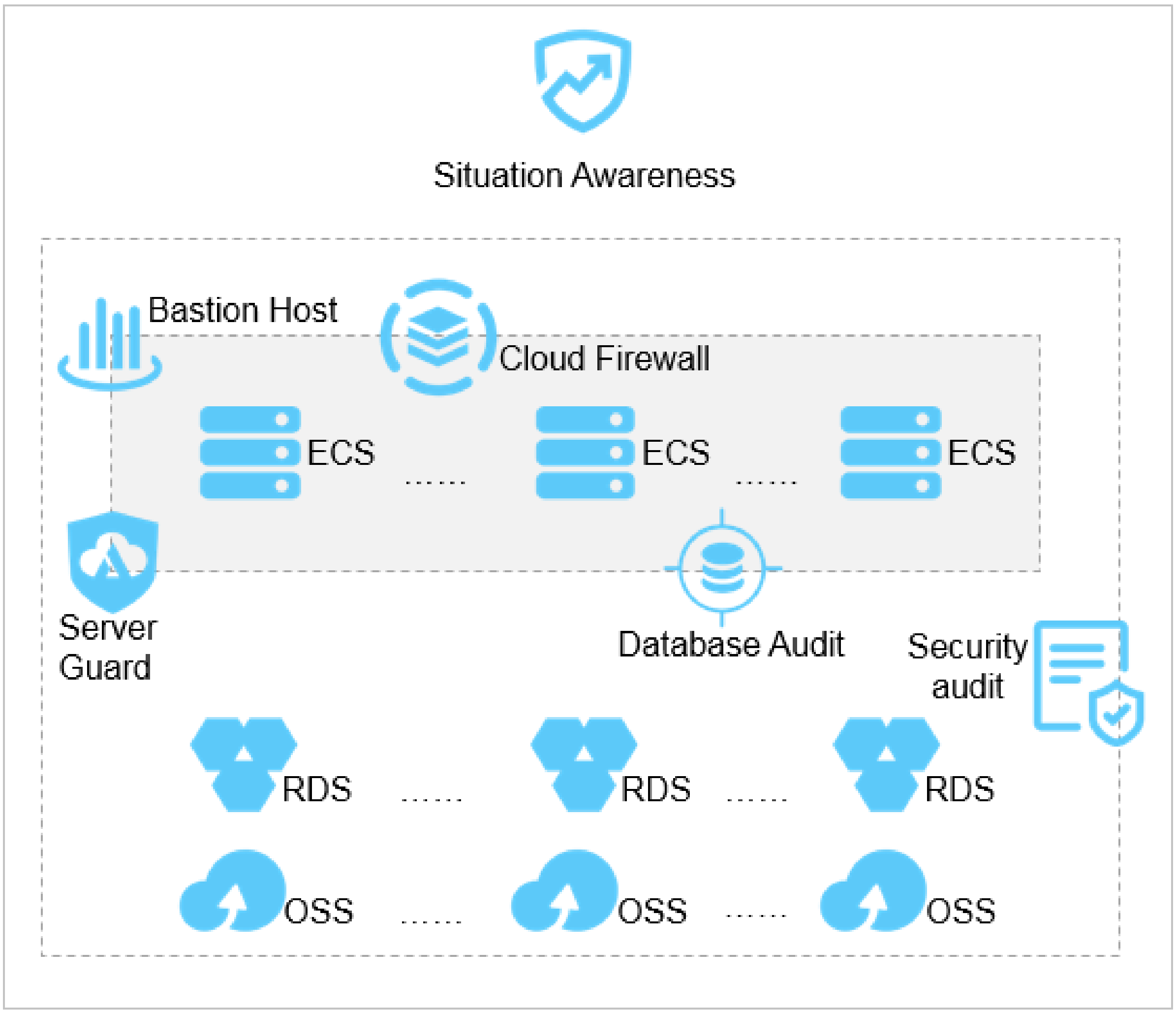


Intranet-oriented

If your Apsara Stack platform is Intranet-oriented and does not provide external services, we recommend that you reference the configuration shown in *Figure 13-5: Intranet-oriented scenario*. You should protect Apsara Stack from multiple standpoints, such as internal network security, server security, and security management.

- Internet network security: Cloud Firewall
- Server security: Server Guard
- Security management: Situation Awareness, Bastion Host, and Security Audit

Figure 13-5: Intranet-oriented scenario



13.6 Usage limitations

No

13.7 Basic concepts

DDoS attack

Distributed Denial of Service (DDoS) exploits client/server technology to combine multiple computers and form a platform to initiate an attack against one or more targets, which poses a threat that is orders of magnitude greater than that of a denial of service attack.

SQL injection

Web SQL injection is a security vulnerability that occurs at the database layer for apps. It is used to obtain website control permission illegally. Some apps may overlook checking on SQL instructions in input character strings. As a result, these instructions are falsely considered as normal SQL instructions and executed by the database. When this happens, the database is more prone to attacks, which may lead to data theft, or modification, deletion, or even insertion of malicious code and backdoors into websites.

Web Application Firewall

WAF is a cloud firewall service that protects core website data and safeguards the security and availability of your site.

Based on powerful Big Data cloud capabilities and underlying security, WAF provides protection against web-based attacks, including SQL injections, XSS, Malicious BOT, command execution vulnerabilities, and other common web attacks.

DDoS mitigation

DDoS mitigation is a set of techniques or tools for resisting or mitigating the impact of distributed denial-of-service (DDoS) attacks on networks attached to the Internet by protecting the target and relay networks. DDoS attacks are a constant threat to businesses and organizations by threatening service performance or to shut down a website entirely, even for a short time.

Brute-force attack

A brute-force attack consists of an attacker trying many passwords or passphrases with the hope of eventually guessing correctly. The attacker systematically checks all possible passwords and passphrases until the correct one is found. Alternatively, the attacker can attempt to guess the key which is typically created from the password using a key derivation function.

Webshell attack

A webshell attack is structured to write webpage-based Trojan viruses into websites to control corresponding servers.

14 Key Management Service (KMS)

14.1 What is KMS

Key Management Service (KMS) is a secure and easy-to-use key hosting service provided by Alibaba Cloud Apsara Stack. With KMS, you can easily create and manage your keys and use them to encrypt your data.

KMS is integrated with multiple Alibaba Cloud products and services to protect your cloud data.

KMS can solve the problems shown in *Table 14-1: Problems solved by KMS*.

Table 14-1: Problems solved by KMS

Roles	Problems and requirements	KMS solution
Application/Website developers	• My program needs keys or certificates for encryption or signature, and I want secure and independent key management. • I want to securely access keys wherever my application is deployed. I do not accept plaintext keys deployed everywhere. That is too risky.	Using envelop encryption technology, you can store the customer master key (CMK) in a KMS instance and deploy only the encrypted data keys . You need to call the KMS instance to decrypt data keys only when necessary.
Service developers	• I do not accept responsibility for the security of users' keys and data. • I want users to manage their own keys. I want to use specified keys to encrypt their data with their authorization. This way, I can focus on developing service features.	Based on envelop encryption technology and KMS APIs, service developers can use specified CMKs to encrypt and decrypt data keys, so **`plaintext is not directly stored on a storage device`**. This removes service developers' worries about how to manage users' keys.
Chief Security Officer (CSO)	• I expect our key management activities to meet compliance requirements.	KMS can connect to RAM for unified authorization management.

Roles	Problems and requirements	KMS solution
	• I need to ensure that keys are reasonably authorized and any use of keys is audited.	

14.2 Benefits

Low cost

In traditional key management solutions, purchasing secure key management equipment to construct a secure physical environment results in high hardware costs. Designing and executing secure key management specifications incurs high software costs.

By using KMS, you can manage your keys on the cloud platform, saving hardware and software costs.

Ease of use

KMS uses centralized and easy-to-use APIs and standard HTTPS protocol.

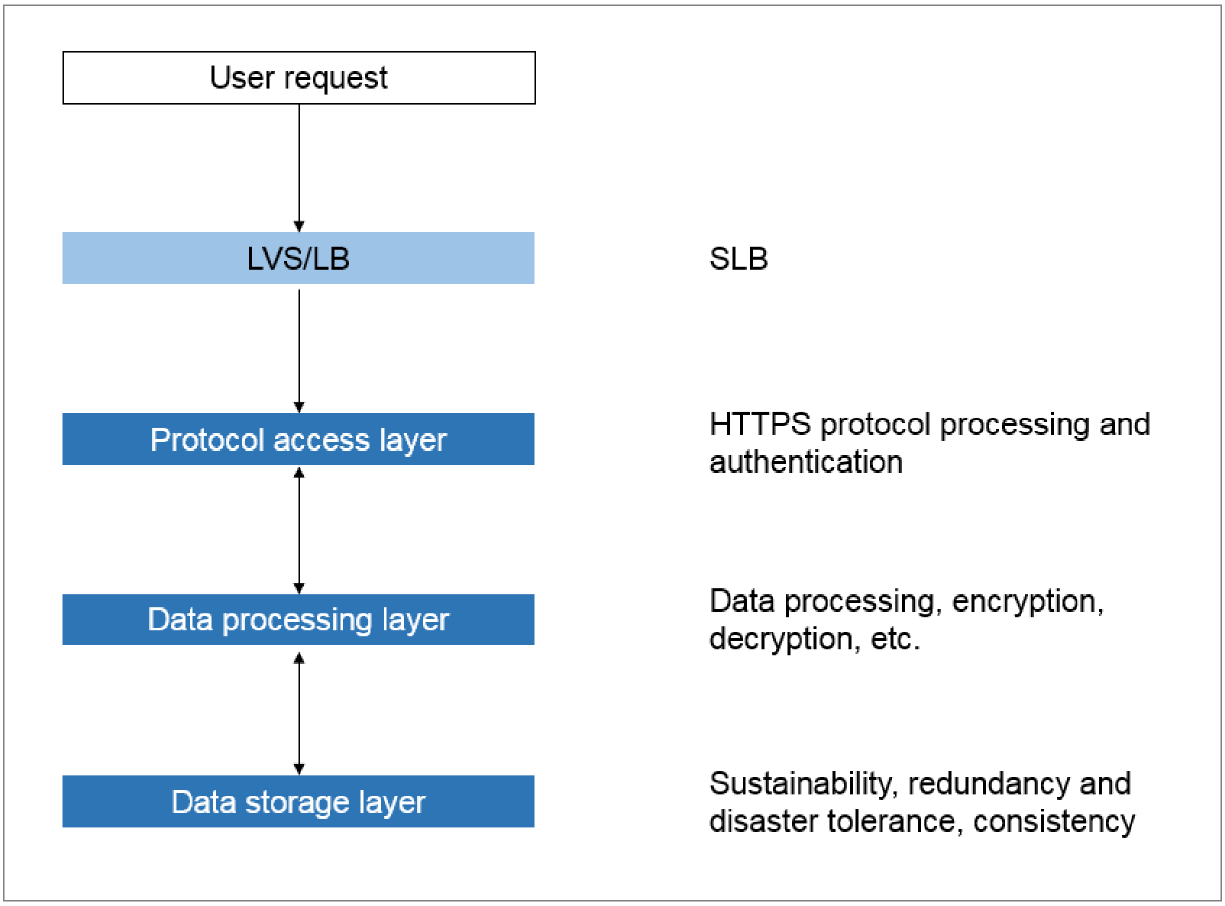
Reliability

KMS combines a distributed system and cryptographic hardware to enhance reliability.

14.3 Architecture

The KMS architecture is as follows.

Figure 14-1: KMS architecture


User request
LVS/LB
SLB
Protocol access layer
HTTPS protocol processing and authentication
Data processing layer
Data processing, encryption, decryption, etc.
Data storage layer
Sustainability, redundancy and disaster tolerance, consistency

- The responsibility of the protocol access layer is to receive HTTPS requests from users, and perform subsequent user authentication and permission identification. If a user passes the authentication, the protocol access layer forwards the user's request to the data processing layer to be processed. It sends the result to the user after it receives the processing result. If a user fails the authentication, the protocol access layer returns an error message to the user.
- The data processing layer is responsible for processing request data. Data processing work done by KMS is key-related operations, such as encryption and decryption. The protocol access layer uses the Remote Protocol Call (RPC) over Transport Layer Security (TLS) to communicate with the data processing layer. The data processing layer is deployed in a distributed manner, where each node is stateless and able to gracefully handle the requests from the protocol access layer.
- The data storage layer holds the core root key data of KMS and runs the Raft distributed consensus protocol to maintain data consistency. This layer leverages a TPM-based encryption solution to encrypt and store persistent data.

14.4 Features

The following table describes the features that KMS offers.

Feature	Description
Create Customer Master Keys (CMKs)	Used to create CMKs. You must create at least one CMK before you can use KMS. The CMK can be used directly to encrypt a small amount of data (less than 4 KB). However, it is typically used for the **`GenerateDataKey`** API to generate a data key.
Create data keys	Used to generate a data key with a specified CMK for KMS envelope encryption scenarios. You can use the data key to encrypt local data.
Encrypt data	Used to encrypt data with a specified CMK. Data no larger than 4 KB can be encrypted, such as Rivest-Shamir-Adleman (RSA) keys, database keys, or other sensitive customer data.
Decrypt data	Used to decrypt data encrypted by KMS.
View the key list	Used to obtain all CMK IDs in the current region under your account .
View key details	Used to obtain the details of a specified CMK. The details include the creation date and time, description, globally unique identifier, status, usage, pre-deletion period, creator, key material source, and key material expiration time.
Enable keys	Used to re-enable a disabled key.
Disable keys	Used to disable a key. The disabled key cannot be used to encrypt data. Data encrypted with a disabled key cannot be decrypted.
Key pre-deletion	Used to configure a pre-deletion period for a specified key. After the pre-deletion period expires, the key will be completely deleted.
Cancel key pre-deletion	Used to cancel the pre-deletion of a key, and re-enable the key.
API	Provides HTTPS API operations supported by KMS.
SDK	Provides SDKs for major languages.

14.5 Scenarios

Directly use KMS for encryption and decryption

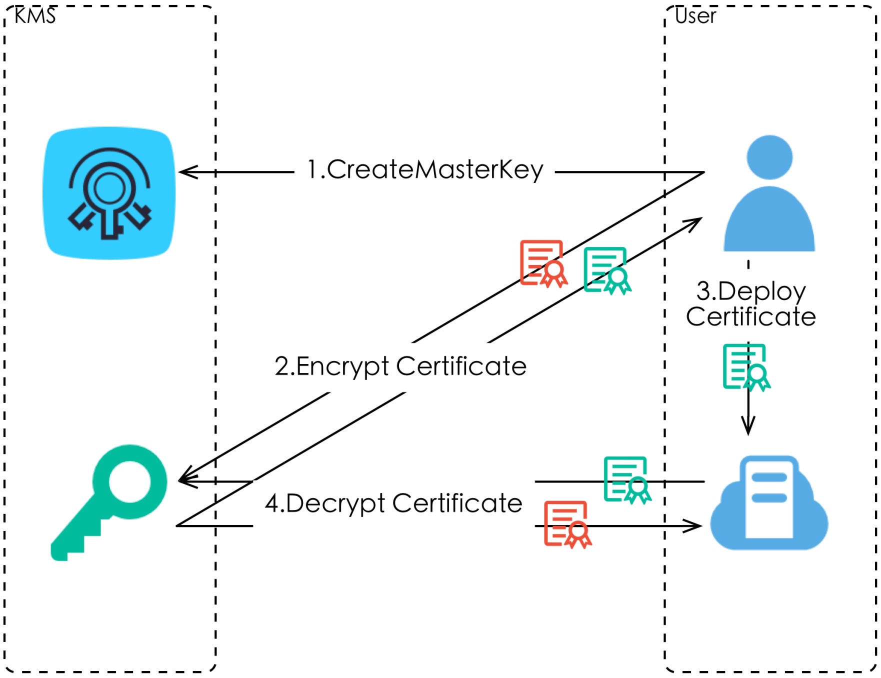
You can directly call KMS APIs and use the specified CMK to encrypt and decrypt data.

This scenario applies to the encryption and decryption of small amounts of data (less than 4 KB). Data is transmitted to the KMS server over secure channels, encrypted or decrypted on the server , and returned over secure channels.

Example: Protect the HTTPS certificate on the server.

Example: Protect the HTTPS certificate on the server, as shown in *Figure 14-2: Protect the HTTPS certificate on the server*.

Figure 14-2: Protect the HTTPS certificate on the server



Keys	Description	Keys	Description
🔑	Customer Master Key (CMK)	🔑	Ciphertext data key
📜	Plaintext certificate	📄	Plaintext file
📜	Ciphertext certificate	📄	Ciphertext file
🔑	Plaintext data key	-	-

The process is as follows:

1. First, create a CMK.
2. Call the KMS Encrypt interface to encrypt the plaintext certificate as a ciphertext certificate.
3. Deploy the ciphertext certificate on the server.
4. Call the KMS Decrypt interface to decrypt the ciphertext certificate as a plaintext certificate when the server gets started and needs the certificate.

Use envelop encryption to perform local encryption and decryption

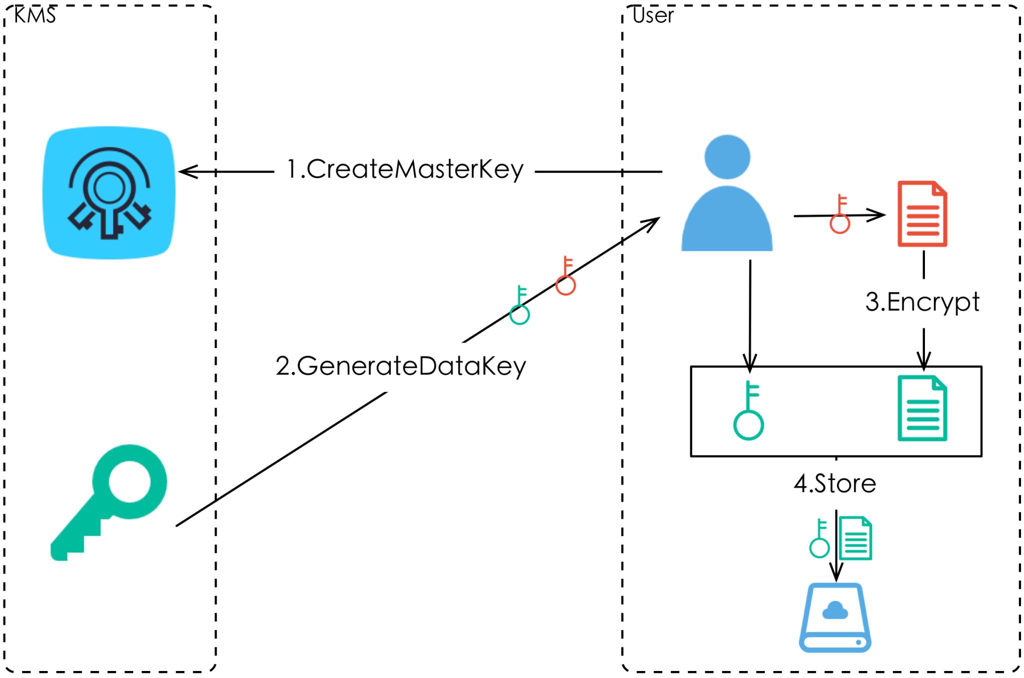
You can directly call KMS APIs, use the specified CMK to generate and decrypt the data key, and use the data key for local data encryption and decryption.

This scenario applies to encryption and decryption of large volumes of data , and does not require you to transmit large data volumes over the network, resulting in lower costs.

Example: Encrypt a local file.

Example: Encrypt a local file, as shown in *Figure 14-3: Encrypt a local file*.

Figure 14-3: Encrypt a local file



Keys	Description	Keys	Description
🔑	Customer Master Key (CMK)	⚷	Ciphertext data key
📜	Plaintext certificate	📄	Plaintext file
📜	Ciphertext certificate	📄	Ciphertext file
⚷	Plaintext data key	-	-

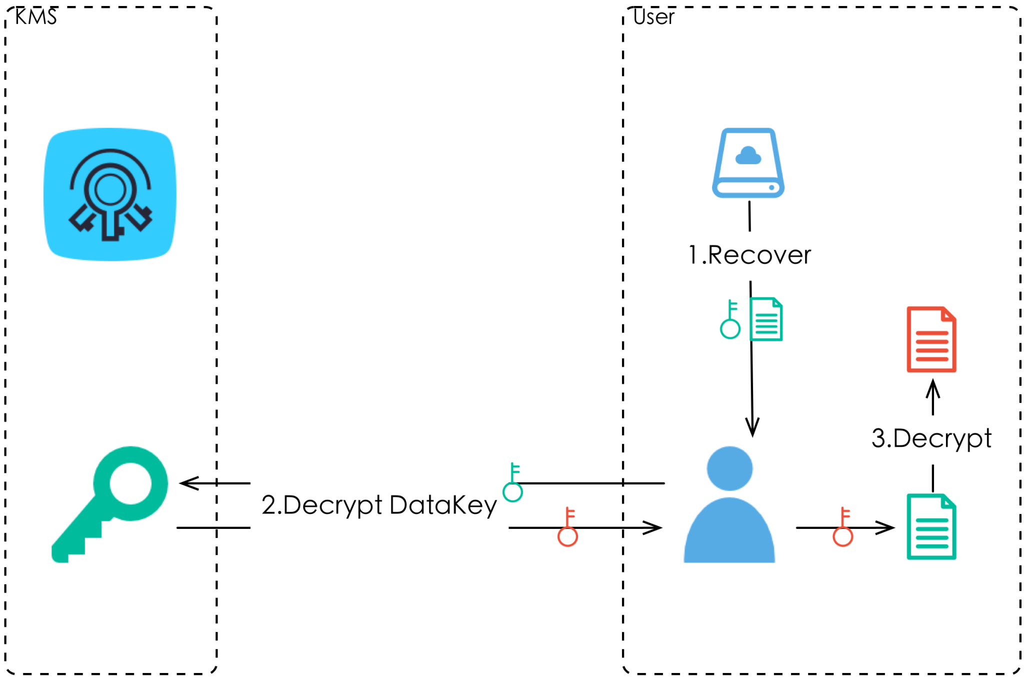
The encryption process is as follows:

1. First, create a CMK.
2. Call the KMS GenerateDataKey interface to generate data keys.

   You can obtain a plaintext data key and a ciphertext data key.
3. Use the plaintext data key to encrypt the file and generate a ciphertext file.
4. Save the ciphertext data key and the ciphertext file to a persistent storage device or service.

The decryption process is shown in *Figure 14-4: Decryption process*.

Figure 14-4: Decryption process



Keys	Description	Keys	Description
🔑	Customer Master Key (CMK)	⚷	Ciphertext data key
📜	Plaintext certificate	📄	Plaintext file
📜	Ciphertext certificate	📄	Ciphertext file
⚷	Plaintext data key	-	-

The decryption process is as follows:

1. Read the ciphertext data key and the ciphertext file from the persistent storage device or service.
2. Call the KMS Decrypt interface to decrypt the ciphertext data key to obtain the plaintext data key.
3. Use the plaintext data key to decrypt the file.

Note

1. You must authenticate the HTTPS certificate on the Alibaba Cloud server to prevent phishers from stealing your information.
2. We recommend that you grant different permissions to users using different keys.

14.6 Limits

CMK creation limit: A maximum of 200 CMKs can be created for each department.

14.7 Basic concepts

Envelope encryption

In envelope encryption, a symmetric key is generated **each time data is encrypted**. You can use a specific CMK to encrypt this symmetric key to an **"envelope"** for protection. The **enveloped key** is directly transferred in unsafe communication processes, such as data transmission and storage. You only need to remove the key from the envelope when you need the key.

Customer Master Key

A Customer Master Key (CMK) is the master key created by a user in Alibaba Cloud KMS, which is used to encrypt data keys and generate envelops. It can also be used to encrypt a small amount of data.

Enveloped Data Key/Data Key

Enveloped Data Key and Data Key are abbreviated as EDK and DK. A DK is the plaintext key for data encryption, and an EDK is the ciphertext key obtained after envelop encryption.

15 StreamCompute

15.1 What is StreamCompute

Alibaba Cloud StreamCompute is a general-purpose computing platform that provides streaming data computing services in real time.

Business Background

Currently, the demand for high timeliness and operability of information is increasing, which requires software systems to process more data in less time. The traditional big data processing model divides online transaction processing and offline analysis completely from time sequence, but it is obviously lagging behind customer's demand for real-time processing of massive data.

Alibaba Cloud StreamCompute is aimed at the stringent requirements for the timeliness of the data processing: the business value of the data decreases rapidly with the loss of time, so it must be calculated and processed as soon as possible. The traditional big data processing mode follows the day clearing model for data processing, that is, accumulating and processing data with hours or even days as the computing cycle. Obviously, this kind of processing mode can not meet the needs of real-time data calculation. In such fields as real-time big data analysis, risk control and alarm, real-time prediction, financial transactions and many other business scenes, batch processing is completely incompetent for the requirement. As a kind of real-time computing model for streaming data, Alibaba Cloud StreamCompute can effectively shorten the full link data stream latency, enable real-time computing logic, reduce computing costs, and fully meet business needs for real-time processing of big data..

What is streaming data?

Streaming data is a sequence of data that is computed on directly as the data are produced or received. Streaming data can be considered the opposite of static data (such as a file, or a record saved in a database). Application logic, analysis, and queries exist continuously and data flows through them constantly. Streaming data is generated from endless event streams, such as log files generated by a mobile device or web application, online shopping data, in-game player activity, social network information, financial transactions or geographic positioning services, and telemetry data from connected devices in a data center.

Key Features

Alibaba Cloud StreamCompute has three key features:

- Real-time and unbound data streams

  StreamCompute can compute directly on a real-time, streaming data source. Streaming data is subscribed to, and consumed by, StreamCompute in order of time. Due to the continuity of data , the stream is continuously integrated into the StreamCompute system over a length of time. For example, for a website's visit log, as long as the website does not close the log stream, it will continue to be generated and integrated into the StreamCompute system. Thereby, for a stream system, the streaming data is real-time and will not be terminated (unbounded).
- Continuous and efficient computing

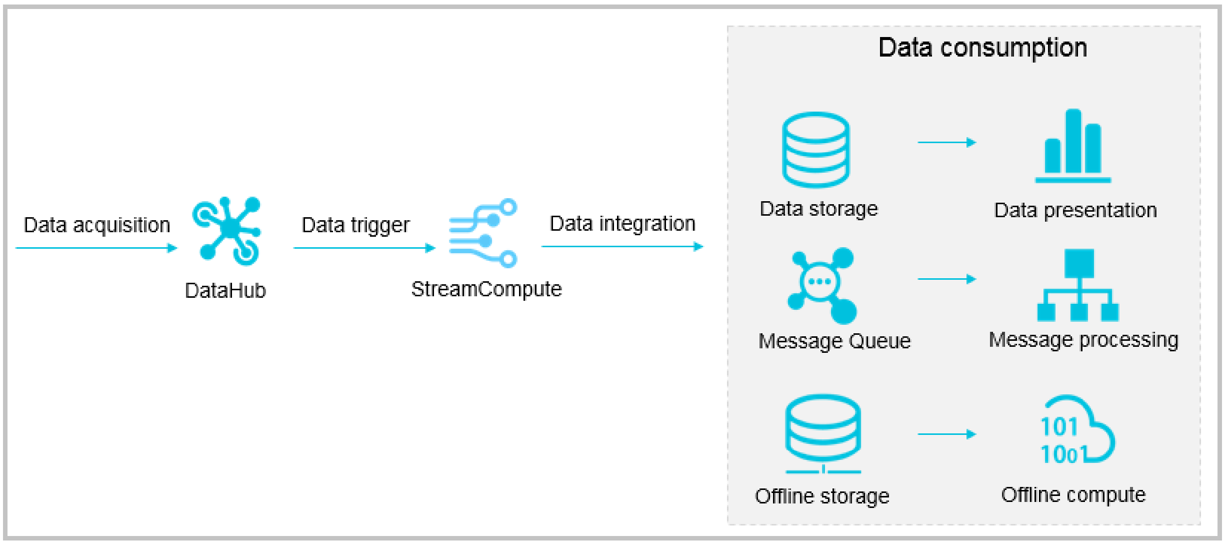
  StreamCompute operates as an event-triggered computing mode, whereby the trigger source is the previously mentioned unbounded stream data. Once new streaming data is entered into StreamCompute, StreamCompute immediately initiates and performs a computing task.
- Streaming and real-time data integration

  Streamdata triggers the computing result of a StreamCompute, and can be directly written into the destination data storage. For example, a computed report data is directly written into the RDS for report display. This means the computing result of the streaming data can be written to the destination data storage in a similar was as a streaming data source.

15.2 StreamCompute Full Link

Unlike existing offline/batch computing models, the StreamCompute full link focuses more on real-time data, including real-time data collection, computing, and integration. The real-time processing logic of the three types of data guarantees the low latency of StreamCompute on full link. The schematic diagram of full link StreamCompute is shown in *Figure 15-1: Full link StreamCompute*:

Figure 15-1: Full link StreamCompute



1. **Data collection**

   The user uses the streaming data collection tool to collect and transmit data to the big data Pub/Sub message system in real time. The system will provide continuous event sources for downstream StreamCompute to trigger the streaming computing tasks.

2. **Stream computing**

   The streaming data actuates StreamCompute as the trigger source. Therefore, a StreamComp ute task must use at least one streaming data as a data source. A batch of incoming data streams will directlytrigger a downstream StreamCompute for a streaming computing process, and return the results for the batch of streaming data.

3. **Data integration**

   StreamCompute directly writes the computed result data to the destination data source, which includes a variety of data sources including data storage systems, message delivery systems, including directly docking a business rules alert system to send alert information. Unlike batch computing (such as Alibaba Cloud MaxComputer or open source Hadoop), StreamCompute comes with a built-in data integration module, which can directly write the result data to the destination data source.

4. **Data consumption**

   Once StreamCompute delivers the result data to the destination data source, the subsequent data consumption is completely decoupled with StreamCompute in respect to system division . Users can use the data storage system to access data, use the message delivery system for information reception, or directly use the alarm system for alerts.

15.3 Differences between StreamCompute and batch computing

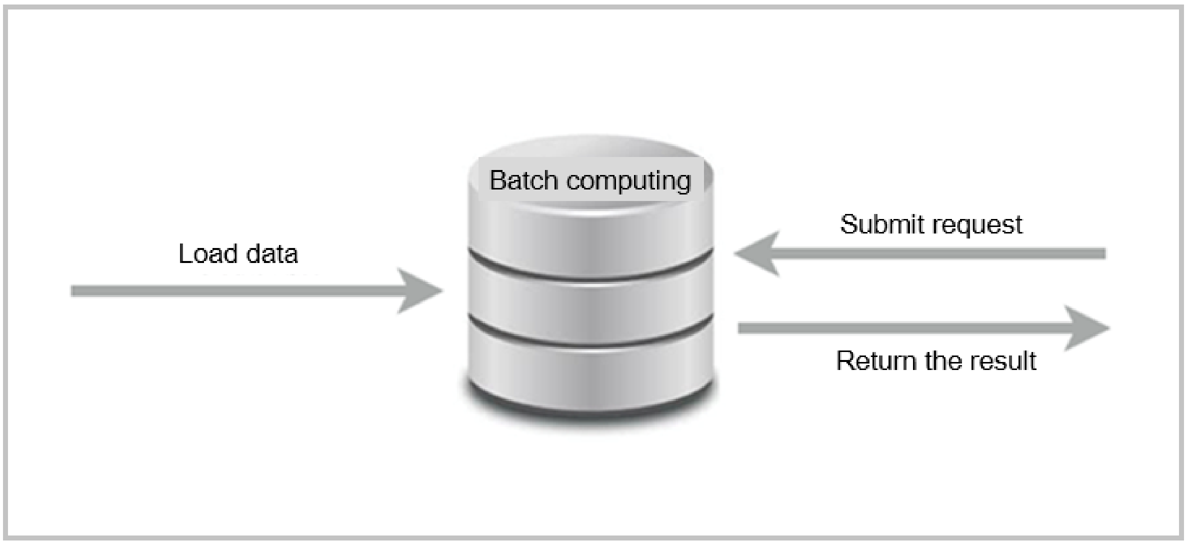
Compared to batch data computing, StreamCompute is still a relatively new computing concept. We are now introducing the difference between the two types of computing methods at the user/product level. It should be noted that the description here is not a rigorous scientific/theoretical explanation. For more detailed theoretical analysis, see the related section on Wikipedia*Stream Processing*.

15.3.1 Batch Computing

Presently, most of the traditional data computing and data analysis services are based on batch data processing models: Using the ETL or OLTP system to construct the data source, online data services (including Ad-Hoc query, DashBoard and other services) access the data source by constructing SQL and obtaining the analysis results. This data processing method has become more widely used with the evolution of relational databases in the industry. 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 streaming processing. The current processing models are facing the huge challenge of real-time processing. Traditional batch data processing is usually based on the following processing models:

1. The ETL or OLTP system is used to construct the original data source, which is provided to subsequent data services for data analysis and data computing. As in the following figure, the user loads the data, and the system performs index construction and a series of query optimizations for the loaded data based on the storage and computing conditions. Therefore , for batch computing, the data must be preloaded into the computing system, then the computing system can compute after the data load is complete.
2. The user/system initiates a computing task (such as a MaxCompute SQL task, or Hive SQL task) and requests to the data system. At this point, the computing system starts scheduling (initiating) the computing nodes to perform a large amount of data computing, which may take up to several minutes or even hours. Meanwhile, because the data accumulation is not real-time, data in the computing process must be historical data, and it cannot be guaranteed that the data is “fresh”. Users can adjust their own computing SQL at any time according to their needs, and even can enable real-time modification and real-time query when using AdHoc queries.
3. The data is returned to the users in the form of a result set when the computing task is completed, or the user can integrate the data into another system due to the large size of the

data stored in the computing system. If the data results are large, the overall data integration process will take long, maybe a few minutes or even hours.



Batch computing is a batch, high latency, actively initiated computing task. The order of operations for batch computing is:

1. Preload data.
2. Submit the computing task, which can be modified according to the business needs and submitted again.
3. Return the computing results.

15.3.2 Stream computing

Unlike the batch computing models, stream computing focuses more on the computing data stream and low latency. The data processing model of stream computing is as follows:

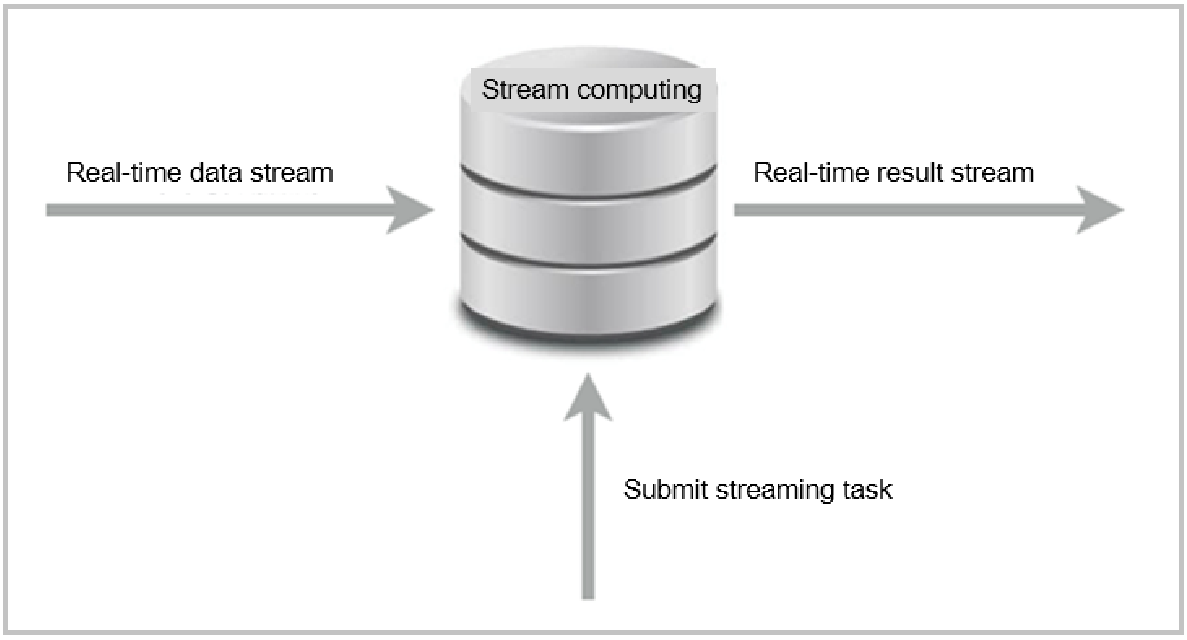
1. Real-time data integration tools are used to transfer real-time data changes to streaming data storage (that is, message queues, such as DataHub); thus the data transfer becomes real-time, and a large amount of data accumulated over a long period is divided into each time point for continuous small batch real-time transfer, so the latency of data integration can be reduced.

   At this point, the data will continue to be written into stream data storage, without the pre-loading process. Meanwhile, StreamCompute does not provide storage service for streaming data, the data is continuously streaming, and will be immediately discarded once the computing is complete.
2. There is a bigger difference in data computing between the streaming process model and batch process model, because the data integration becomes real-time from accumulation.

Unlike batch computing which initiates the computing task once the data integration is ready, a stream computing task is a resident computing service which will remain waiting for the event trigger once started, and once a small batch of data enters the streaming data storage, the StreamCompute immediately computes it and quickly gives the results. Meanwhile, Alibaba Cloud StreamCompute also uses the incremental computing model to incrementally compute a large amount of the data in batches, which further reduces the size of a single computation and effectively reduces the overall computing latency.

From a user perspective, for streaming tasks, the computing logic must be pre-defined and submitted to the stream computing system. During the entire operation, the task logic of the StreamCompute cannot be changed. The user can submit the task again after stopping the current task. The data that has been computed cannot be computed again.

3. Unlike the batch computing result data which needs to be transferred to the online system once the data computing result is complete, stream computing tasks can immediately write the data into the online/batch system after a small batch of data computing. Without having to wait for the overall data computing results, you can immediately deliver the results to the online system, and further enable real-time presentation of the real-time computing results.



StreamCompute is a continuous, low latency, event-triggered computing task. The order of operations for StreamCompute is:

1. Submit the StreamCompute task.
2. Wait for the streaming data to trigger the StreamCompute task.
3. Continuously output the computation results.

15.3.3 Model Comparison

The following table shows the differences between the computing model of StreamCompute and batch computing.

Table 15-1: Model comparison

Items	Batch computing	Stream computing
Data integration method	Preload data.	Load and compute data in real time.
Usage	The business logic can be modified and the data can be recomputed.	Once the business logic is modified, the previous data can not be recomputed (streaming data is perishable).
Data range	Query and process all or most of the data in the dataset.	Query and process data in the scrolling time window or only the recent data records.
Data volume	A large batch of data.	A single record or a small batch of data containing several records.
Performance	A latency of minutes or hours.	A latency of only seconds or milliseconds.
Analysis	Complex analysis.	Simple response functions, aggregation and scrolling items.

StreamCompute is an effective enhancement of batch computing, especially for processing timelines of event streams, and is an indispensable value-added service for big data computing.

15.4 Benefits

Compared with other stream computing products, Alibaba Cloud StreamCompute provides some competitive advantages. You can make full use of these advantages to solve the problem of big data analysis in real time.

Powerful real-time processing capability

StreamCompute does not require users to implement a large number of stream computing details from the beginning. Instead, StreamCompute integrates many full-link functions, which facilitates convenient full-link StreamCompute development for users, including:

- The powerful StreamCompute engine provides the standard StreamSQL, which supports automatic recovery from various types of failures, and ensures the accuracy of data processing during failures. It has multiple built-in functions such as string processing,time, and statistics, and accurately controls computing resources to guarantee the isolation of user operations.

- Key performance indicators, which are 3 to 4 times higher than that of the open-source Flink , and data computing latency is optimized to seconds or even milliseconds. A single task throughput can be in millions, and a single cluster can contain thousands of machines.
- Deep integration of various types of cloud data storage, including DataHub, Log Service, RDS , TableStore, AnalyticDB, IoTHub and other data storage systems, without additional data integration work. Alibaba Cloud StreamCompute can directly read/write the data of the above products.

Managed real-time computing services

Unlike open source or self-built stream processing services, StreamCompute is a fully managed stream computing engine that can run queries for streaming data without presetting or managing any infrastructure. With StreamCompute, you can enable streaming data service capability in one-click. StreamCompute seamlessly integrates services such as data development, data O&M, and monitoring alarms, which facilitate your management and transference of stream computing at low cost.

StreamCompute supports fully isolated management services with tenant, and provides the most effective isolation and comprehensive protection from the top work space environment to the bottom running machines, allowing users to use StreamCompute securely.

Good streaming development experience

Supports standard SQL (namely, BlinkSQL) and provides various built-in functions such as string processing, time, and statistics, replacing inefficient and complex Apache Storm development industry-wide. This enables users to complete real-time big data analysis and processing using simple BlinkSQL, greatly reducing the complexity and concerns of real-time big data processing.

Provides assistance kits for different stages including data development, data O&M, and monitoring alerts for full-link StreamCompute, so that it only takes a minimal amount of steps to completely release stream computing tasks.

Low cost operations

A large number of optimized SQL execution engines will produce more efficient, low-cost computing tasks than manually coded native Storm tasks. In both development costs and operating costs, StreamCompute offers great saving benefits compared to an open source streaming framework as it handles all code lines and the debugging, testing, tuning and releasing work of a task, allowing you to fully focus on your business goals.

15.5 Architecture

15.5.1 Business Architecture

StreamCompute is defined as a lightweight streaming compute engine that uses SQL expressions .

- **Data production**

  The source where data is produced. Data production generally takes place in server logs, database logs, sensors, and third-party systems. This streaming data enters the data integration module where it drives StreamCompute.
- **Data integration**

  This module integrates streaming data and acts as a hub for data publishing and subscription. The data can be collected from the DataHub service, the IoTHub service, and ECS's Log Service.
- **Data computing**

  StreamCompute subscribes to the streaming data provided by data integration to drive streaming computation of data.
- **Data storage**

  StreamCompute does not provide any storage resources. Instead, it writes the results of streaming processing and compute to other storage resources, including relational databases, NoSQL databases, and OLAP systems.
- **Data consumption**

  Different data storage resources allow you to consume data in various ways. For example, the storage for message queues can be used for alarms, while that for relational databases can be used for online business support.

15.5.2 Technical Architecture

StreamCompute is a real-time incremental computing platform. It provides StreamSQL-like syntax and uses the MapReduceMerge (MRM) model for incremental computing. StreamCompute provides an excellent failover mechanism to ensure data accuracy in the case of various exceptions.

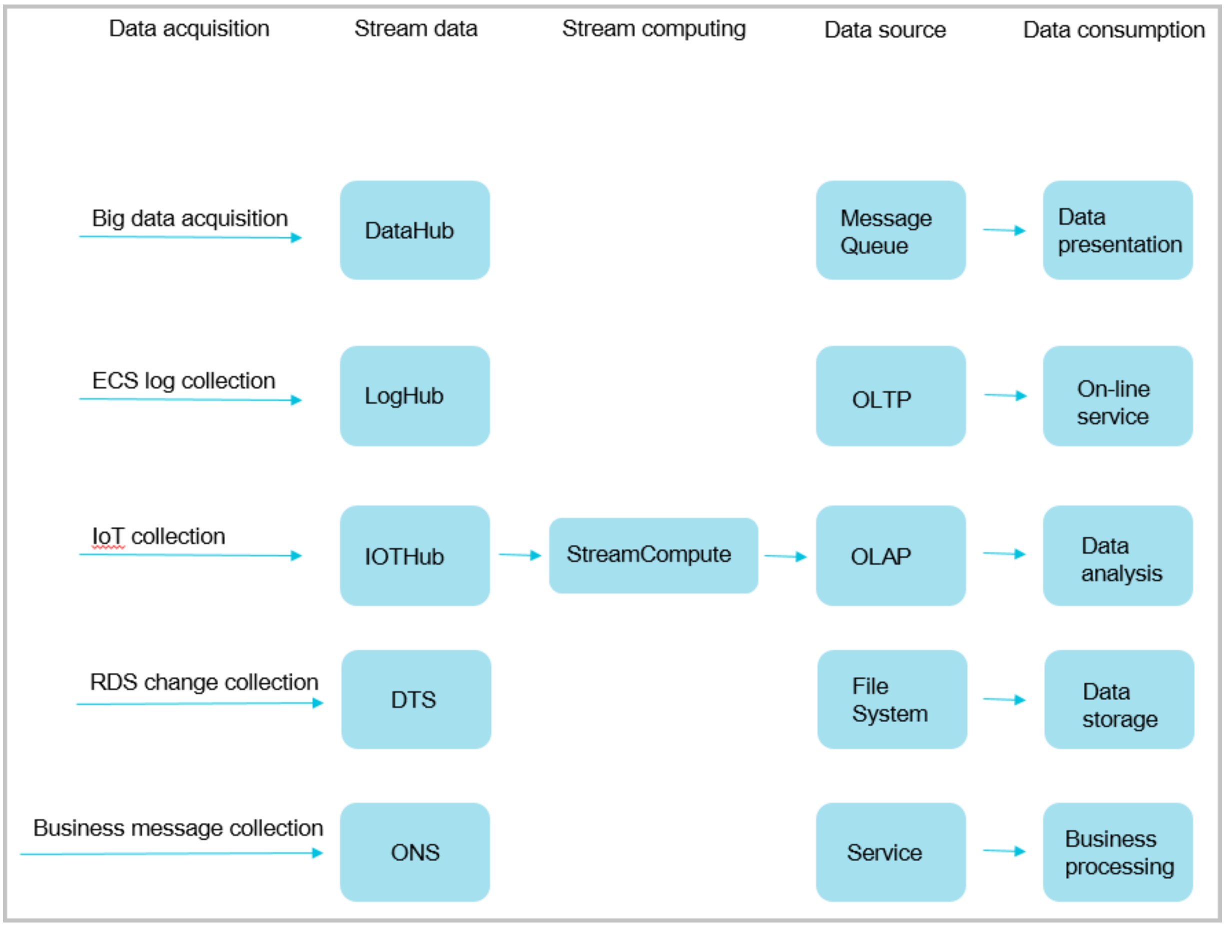
StreamCompute includes the following components:

- **Data application layer**: This provides a development platform for you to develop new business and submit jobs. A comprehensive monitoring and alarm system is provided to inform the business end of any job delay. You can also use Blink UI or another system to view the operating status of online jobs and performance bottlenecks, allowing you to quickly and effectively optimize your jobs.
- **Data development**: This layer parses Blink SQL statements, generates logical and physical execution plans, and ultimately converts execution plans into executable directed acyclic graphs (DAGs). This layer generates various directed graphs modeled by the DAGs obtained at the SQL layer. It is used to process specific business logic. Generally, a model contains three parts:
  - Map: This performs data filter, distribution (group), Join (MapJoin), and other operations.
  - Reduce: This performs aggregation within a single batch (StreamCompute packs stream data into batches, each of which contains multiple data entries).
  - Merge: This merges the computing results from the batch with the previous results (state) to get a new state. After processing N number (the value N is configurable) of batches, the checkpoint operation is performed to save the current state to the State system (such as HBase or Tair).
- **Blink Core**: It provides a variety of computing models, Table APIs and Blink SQL. The underlying layer supports DataStream API and DataSet API, and the lowest layer needs Blink Runtime which is responsible for resource scheduling to ensure jobs stable.
- **Distributed resource Scheduling**: The StreamCompute cluster is built on the Gallardo scheduling system, which ensures the effective operation and recovery of StreamCompute.
- **Physical layer**: It refers to the powerful cluster support provided by Alibaba Cloud.

15.5.3 Business process

The system architecture StreamCompute is shown in *Figure 15-2: System architecture*. It offers a simple understanding of how full-link streaming data is processed.

Figure 15-2: System architecture



1. Data collection

Real-time data collection occurs when a user uses the streaming data collection tool to collectand transmit data to the big data Pub/Sub message system in real time. The system will provide continuous event sources for downstream StreamCompute to trigger the stream computing tasks. The Alibaba Cloud big data ecosystem provides a number of streaming data Pub/Sub systems for different scenarios and fields, Alibaba Cloud StreamCompute integrates various Pub/Sub systems to facilitate integration of all types of streaming data storage systems . Due to some data storage systems can not match with Alibaba Cloud StreamCompute model , other types of data storage systems need to be used for transshipment. Some streaming data storage systems integrated with Alibaba Cloud StreamCompute are shown as follows:

- **DataHub**

  DataHub provides tools and UI to upload data (including logs, database BinLog, IoT data stream, etc.) from data sources to DataHub. It also integrates with some open-source commercial software. For more information, see DataHub documentation.

- **Log Service**

  Log Service is a one-stop service for logs, which has been developed by Alibaba group through a large number of big data scenarios. Log Service provides many functions such as collection, consumption, delivery, query and analysis for logs.
- **IoT Hub**

  IoT Hub is introduced by Alibaba Cloud for developers in the IoT field. The purpose is to help developers build a secure and powerful data channel, and facilitate the bidirectional communication between the terminal (such as sensors, executor, embedded devices or smart appliances, etc.) and the cloud.

  With IoT Hub rule engine, you can easily deliver IoT data to DataHub, and use Alibaba Cloud StreamCompute and MaxCompute to process data.
- **DTS**

  DTS supports data transmission between structured storage products with database as the core. It is a data transfer service that integrates data migration, data subscription and real-time data synchronization. With the data transmission function of DTS, RDS and other BinLog can be easily parsed and delivered to DataHub, and data processing can be made by using Alibaba Cloud StreamCompute and MaxCompute.
- **MQ**

  Alibaba Cloud MQ service is the core product of the enterprise level Internet architecture. Based on the high availability of distributed cluster technology, a complete set of message cloud services, including publish subscribe, message trajectory, resource statistics, timing (delay), monitoring and alarm, is built.

2. **Stream computing**

   The stream data actuates StreamCompute as the trigger source. Therefore, a StreamCompute task must use at least one piece of stream data as a data source. Furthermore, for scenarios with complex business, StreamCompute supports associating queries with static data storage. For example, for each piece of DataHub streaming data, StreamCompute will perform associating queries based on the primary key of the streaming data and RDS data. StreamCompute also supports associating operations for multiple data streams, and StreamSQL supports enterprise-level business complexity.
3. **Real-time data integration**

To minimize data processing latency while reducing data link complexity, StreamCompute directly writes the computing result data into the destination data source, to both minimize the total link data latency and ensure the latest data processing operations. StreamCompute seamlessly integrates OLTP (RDS product line, etc.), NoSQL (Table Store, etc.), OLAP (Analytic DB, etc.), MessageQueue (DataHub, ONS, etc.), and massive storage products (OSS , MaxCompute, etc.).

4. **Data consumption**

   When the result data of stream computing is written to the various data sources, user can use personalized applications to consume the resulting data. For example:

   - Use the data storage systemto access data.
   - Use the message delivery system for information reception.
   - Use the alarm system for alerts.

15.6 Features

Alibaba Cloud StreamCompute provides the following features:

- **Data Collection and Storage**

  All big data analysis systems are based on the premise that data needs to be collected into big data systems first. In order to maximize the user's streaming storage system, Alibaba Cloud StreamCompute integrates a variety of streaming storage, such as DataHub, LogService, IoTHub, DTS and MQ, allowing users to use the existing data stream storage without data collection and data integration.

  Users can enjoy more convenience of one-stop streaming computing development platform by registering data storage in advance. Alibaba Cloud StreamCompute provides management UI including data storage systems such as RDS, AnalyticDB and TableStore, so that you can use Alibaba cloud computing platform to manage your cloud data storage without crossing multiple product UIs.

- **Data Development**

  - Provides a fully hosted online development platform, which integrates a variety of SQL auxiliary functions, including BlinkSQL syntax checking, BlinkSQL intelligent hint and BlinkSQL syntax highlighting.

    - **BlinkSQL syntax checking**

  Users can automatically save configurations after modifying the IDE text. The save operation can trigger the SQL syntax checking function. If there is syntax error, the number of errors, the number of columns and the cause of the error will be prompted on the IDE interface.

  - **BlinkSQL intelligent hint**

    In the process of entering BlinkSQL, IDE provides hints including key words, built-in functions, table/field intelligent memory, and so on.

  - **BlinkSQL syntax highlighting**

    For the BlinkSQL keyword, Alibaba Cloud StreamCompute provide syntax highlighting with different colors to distinguish the different structures of BlinkSQL.

- **Version management for SQL**

  Data development covers key areas of daily development, including code assistance and code version. The data development module provides code version management function. Each submission can generate a code version, which is used to track changes and rollback later.

- Provides a set of convenient tools for data storage management, users can enjoy a variety of traversing data storage services, including data preview and DDL assisted generation by registering data storage.

  - **Data preview**

    The data development page provides data preview for all kinds of data storage types. It can effectively assist users to understand the upstream and downstream data characteri stics, identify key business logic, and quickly complete business development.

  - **DDL assisted generation**

    Most of the DDL generation work are mechanical translation work, that is, the data storage DDL statements that need to be mapped are manually translated into the StreamCompute DDL statements. StreamCompute provides an auxiliary generation of DDL function, which further reduces the complexity of manual writing of stream work by users, effectively reduces the error rate of manual writing of SQL, and ultimately provides the efficiency of StreamCompute business output.

- Supports using standard SQL for real-time data cleaning, statistical summary and data analysis, supports general aggregation functions, supports association queries of streaming data and static data .

  - Provides a set of debug environment. Users can customize upload data, simulate the operation and check the output results in the debug environment.
- **Data O&M**

  Provides the following operation and maintenance functions: job status, data curve, FailOver, CheckPoints, JobManager, TaskExecutor, blood relationship and attribute parameters.
- **Optimize performance**
  - **Optimize performance automatically**

    This function helps users solve the problem of performance such as insufficient throughput capacity and anti-pressure of full link.
  - **Optimize performance manually**
- **Monitoring and alarm**

  In order to help users monitor the statusof Jobs in real time, Alibaba Cloud StreamCompute is connected to the Cloud Monitor platform. Cloud Monitor services can be used to collect monitoring indicators of cloud resources or user-defined monitoring indicators, detect service availability, and set alert rules for indicators to enable you to fully understand the use of resources, the running status and health of the business on the cloud, and respond to the alarm in time to ensure the applications run smoothly. Alibaba Cloud StreamCompute supports the following four types of alarm:
  - Business delay
  - Read RPS
  - Write RPS
  - FailoverRate

15.7 Product positioning

StreamCompute provides standard StreamSQL-like semantics to help users simply and easily complete the processing of stream computing logic. And in scenarios where SQL code functions cannot meet your business needs, StreamCompute provides full-featured UDFs to enable data processing logic for customized business. In the field of streaming data analysis, users directly use StreamSQL+UDF to enable most of the streaming data analysis and processing logic. StreamCompute is suitable for streaming data analysis, statistics and processing. It is not recommended for fields using non-SQL solutions, such as complex iterative data processing and complex rules engine alarm.

StreamCompute is suitable for the following fields:

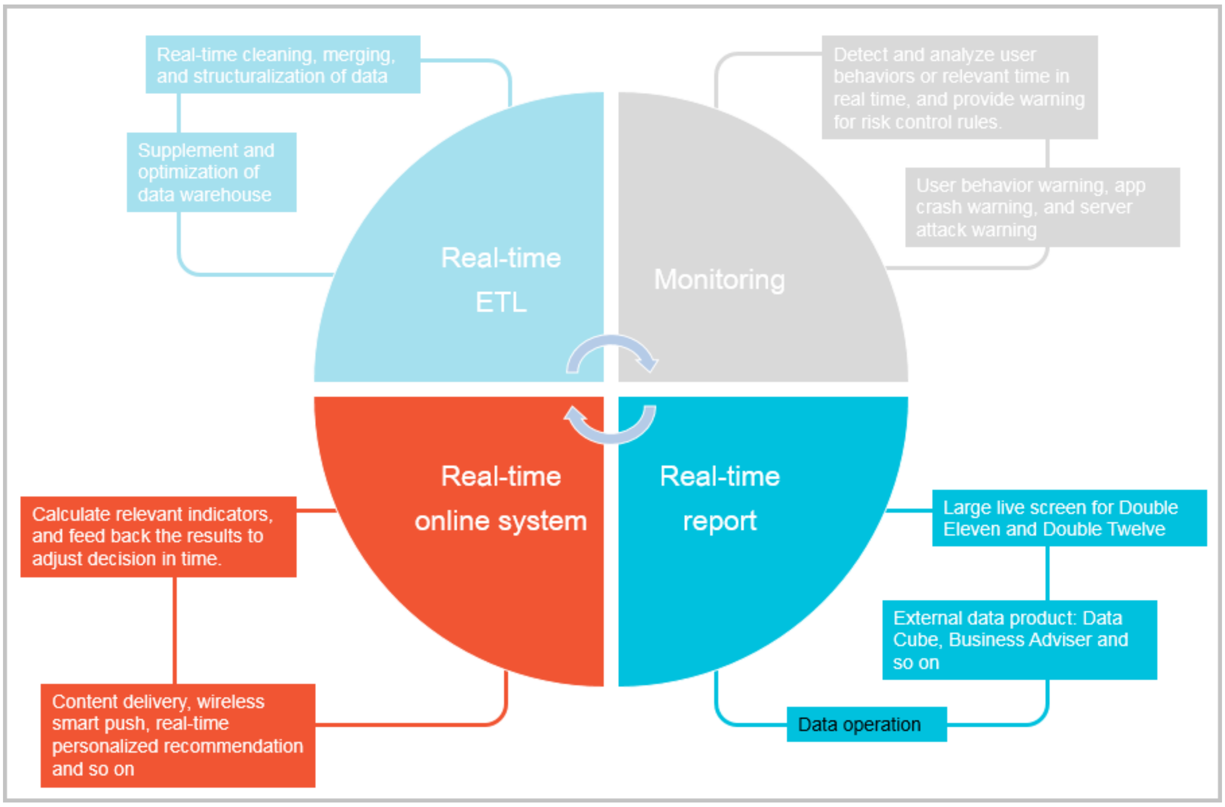
- PV and UV statistics of real-time website clicks
- Statistics of average traffic flow through a checkpoint over 5 minutes
- Statistics and presentation of dam pressure data
- Alerts for network payment involving financial theft rules

The external interface of StreamCompute is currently defined as StreamSQL/UDF, which provides a full set of development tools for streaming data analysis, statistics, and processing for customers such as database developers and data analysts who do not want to develop underlying code, but write StreamCompute SQL statements to complete streaming data analysis for their business.

15.8 Scenarios

StreamCompute uses StreamSQL mainly for streaming data analysis scenarios, as shown in *Figure 15-3: Use cases*.

Figure 15-3: Use cases


Real-time cleaning, merging, and structuralization of data
Supplement and optimization of data warehouse
Real-time ETL
Detect and analyze user behaviors or relevant time in real time, and provide warning for risk control rules.
User behavior warning, app crash warning, and server attack warning
Monitoring
Real-time online system
Calculate relevant indicators, and feed back the results to adjust decision in time.
Content delivery, wireless smart push, real-time personalized recommendation and so on
Real-time report
Large live screen for Double Eleven and Double Twelve
External data product: Data Cube, Business Adviser and so on
Data operation

- **Real-time ETL**

Integrates multiple existing data channels and the flexible SQL processing capacity of StreamCompute to clean, merge, and structure streaming data in real time. It is an effective addition and optimization of offline databases, and a computable channel for real-time data transfer.

- **Real-time report**

  Collects and processes streaming data in real time, and monitors and presents various indicators of business and customers, enabling real-time O&M.

- **Monitoring alert**

  Detects and analyzes systems and user behaviors in real time, and monitors and detects dangerous behaviors in real-time.

- **Online system**

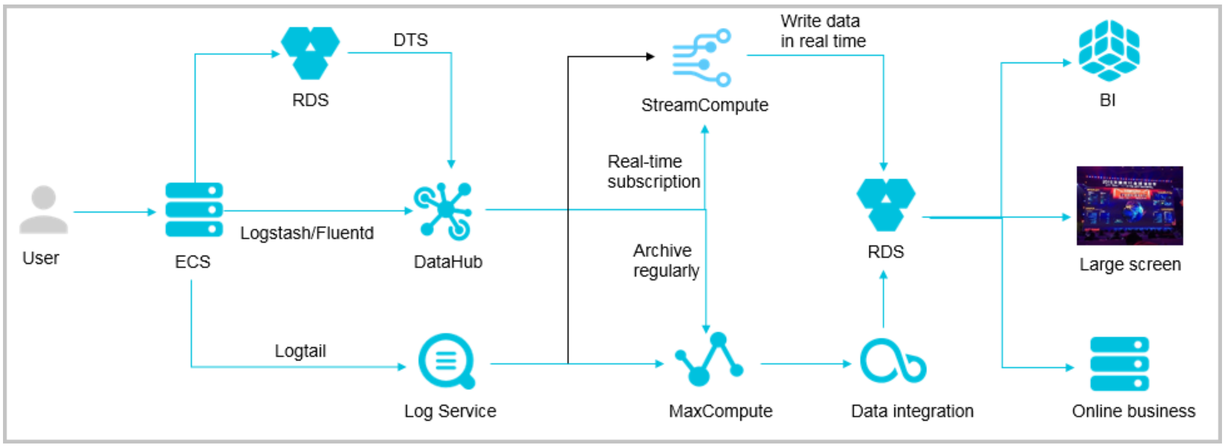
  Computes various data indicators in real time, and adjusts related policies of online systems using real-time results for such cases as content delivery and smart wireless push.

15.8.1 E-commerce Cases

StreamCompute meets demandsfor real-time processing of streaming data in such areas as:

- Real-timeanalysis of user behaviors, such as large screen display of user activity, transactions , and other statistics. Traditional offline analysis not only suffers from slow connection speed and high latency, but also experiences pressure on the online systems, which affects system stability.
- Real-time monitoring of users, services, and systems. For example, the full-site transaction time curve can help O&M or technical staff see site-wide transactions at the current point in time. If there are abnormal fluctuations (such as a sudden drop) in transactions, an alarm will immediately trigger to help facilitate rapid troubleshooting, reducing the impact of transaction fluctuations on the business.
- Real-time monitoring of promotional events, especially useful for when O&M staff need real-time access to a multitude of indicators to determine whether to change the O&M plan for the promotion.

StreamCompute, in combination with various computing and storage systems in Alibaba Cloud's ecosystem, can easily support various personalized streaming data analysis scenarios. Unlike other data analysis systems, Alibaba Cloud StreamCompute not only meets the business flexibility requirements, but also ensures a low threshold of business development using SQL.



15.8.2 IoT Application

Background

In the tide of economic globalization, the competition faced by industrial manufacturers is becoming increasingly fierce. Manufacturers of automobile, aviation, high-tech, food and beverage , textile and pharmaceutical industries need to innovate, but the replacement of existing infrastructure is a difficult task. Traditional systems and equipment have been used for decades, and maintenance costs are high. However, the replacement of systems and equipment may lead to a slowdown in production, and the quality is not guaranteed.

The security risk is more serious than before, and the demand for complex process automation is also more urgent. The manufacturing industry has made innovative preparations, but it strictly requires high reliability and high availability systems run safely and reliably in real time. Equipped with many active components, such as mechanical arm, assembly line, packaging machinery, and so on, remote applications must be able to seamlessly carry out equipment deployment, update, fault transfer and service life termination processing.

All new generation systems must be able to capture and analyze huge amounts of data generated by industrial facilities and respond to the result in time. In order to develop better , manufacturers need to optimize and upgrade. Alibaba Cloud StreamCompute combined with Alibaba Cloud IoT suite, can help users to analyze and diagnose industrial equipment running status, detect running faults, predict product scrap rate in real time. An example is shown as follows: an industrial equipment manufacturera use Alibaba Cloud StreamCompute to calculate and analyze large number of industrial sensor incoming data, which can realize real-time data cleaning and induction , real-time monitoring of the key indicators of the equipment, real-time data cleaning and writing to the online OLAP system.

Overview

The industrial customer has more than 1 thousand equipments, distributed in multiple factories of different cities, with 10 kinds of sensors on each equipment. These sensors' data are collected and uploaded to the Log Service per 5S, the format of each collection point is shown as follows .

s_id	s_value	s_ts
Sensor ID	The current value of sensor	Send time

At the same time, the sensors are distributed in different equipments and factories. Customer also records the distribution dimension table of sensors, equipments and factories in RDS, as shown below.

s_id	s_type	device_id	factory_id
Sensor ID	The monitor type of sensor	Equipment ID	Factory ID

The preceding information is stored in the RDS. The customer hopes that the data uploaded from the sensor can be associated with the above data and that the sensor data can be classified as a wide table per 1min according to the equipment, as shown below.

ts	device_id	factory_id	device_temp	device_pres
Time	Equipment ID	Factory ID	Equipment temperature	Equipment pressure

In order to simplify the subsequent calculations, it is assumed that only two types of monitoring sensors, that is, temperature and pressure, and the calculation logic is shown as follows.

1. Filter equipments with temperature greater than 80 and trigger alarm downstream. Users choose to use MQ as a message trigger source. StreamCompute filters and delivers equipments whose temperature are greater than 80 to the MQ, triggering the downstream user-defined alarm system.
2. Write the data to the online OLAP system, where the user selects HybridDB for MySQL. The user develops a set of BI system for multidimensional presentation of HybridDB for MySQL downstream.

Instructions

- How to extend to be a wide table?

Usually, the data of IoT is uploaded by a sensor with a dimension, which is not conducive to subsequent data processing and analysis. Alibaba Cloud StreamCompute can gather data according to a certain window, and filter data according to different dimensions of sensors, and extend it to a wide table.

- Why do you choose MQ as alarm source?

  In theory, Alibaba Cloud StreamCompute can write the results to any system, but in the alarm and notification scenes, we recommend that you deliver the result data to message storage such as MQ, so as to avoid the failure of the alarm system in the data delivery process leading to disclosure of alarm information, and make sure the accuracy of the alarm.

Code Instructions

The data uploaded by the sensor is entered into Log, and the data format is as follows.

```
{
    "sid": "t_xxsfdsad",
    "s_value": "85.5",
    "s_ts": "1515228763"
}
```

The Log source table is defined as s_sensor_data. The structure is shown below.

```
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 ='xxxxxxxxxxx',
      accessKey ='xxxxxxxxxxxxxxxxxxxxxxxxxxxx',
      project ='ali-cloud-streamtest',
      logStore ='stream-test',
);
```

The RDS dimension table associated with sensor and equipment is defined as d_sensor_device_data. The structure is shown below.

```
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='XXXXXX'
);
```

The alarm logic MQ table is defined as r_monitor_data. The structure is shown below.

```
CREATE TABLE r_monitor_data (
    ts    VARCHAR,
    device_id    BIGINT,
    factory_id    BIGINT,
    device_TEMP    DOUBLE,
    device_PRES DOUBLE
) WITH (
    TYPE='MQ'
);
```

The HybridDB table storing result data is defined as r_device_data, and the structure is as follows.

```
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'
);
```

Firstly, the sensor data is summarized by minute level to extend to a wide table. To make code structured to facilitate subsequent code maintenance, we use View here as shown below.

```
--Obtain the equipment and factory area corresponding to 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；
--Extend to a wide table.
CREATE VIEW v_device_data
AS
SELECT
    --Use the start time of the scroll window as the time of 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, --It is used to
calculate the average temperature of this 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 --It is used to
calculate the average pressure of this minute.
FROM
    v_sensor_device_data v
GROUP BY
    TUMBLE(v.ts, INTERVAL '1' MINUTE), v.device_id, v.factory_id;
```

The preceding is the core calculation logic. We will calculate the average value of temperature and pressure in this minute as the temperature value and pressure value of this minute. Since the Tumbling Window is used, it means that the data will be produced at the end of each minute. Next, filter and write data to MQ and HybridDB.

```
--Filter sensors whose temperature are greater than 80 and write them
to MQ, in order to trigger the downstream user-defined alarm system.
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 data to HybridDB for subsequent analysis.
INSERT INTO r_device_data
SELECT
    ts,
    device_id,
    factory_id,
    device_temp,
    device_pres
FROM
    v_device_data;
```

15.8.3 Tmall Double Eleven Screen

The annual Tmall Double Eleven Shopping Event is on course to become the largest online commercial promotion activity of its kind worldwide. Aside from consumers demonstrating a strong desire to purchase as much as possible, the most outstanding feature of this event has been the increase in overall turnover displayed on the Tmall Double Eleven Screen. This real-time big data presentation link is the result of months of hard work by Alibaba's top engineers. The screen displays eye-catching key indicators, such as the total link time from order placement to data collection, data computing, and data verification. Finally, the total link time compression for the Double Eleven Screen is 5s or less, with peak computing capacity being reached at 00:00.

This allows the screen to process hundreds of thousands of orders every second, while multiple backup link stream computing systems are used to prevent errors.

Alibab Cloud StreamCompute is the most powerful and essential supporting element of the Tmall Double Eleven Screen. Originally, the Tmall Double Eleven screen was developed using Storm, an open source compute solution. However, its development cycle was nearly one month in length . The Alibaba Data Department switched to using StreamCompute and StreamSQL. They reduced the entire development cycle to one week. The bottom layer of StreamCompute completely shields fault processing and optimizes the implementation. The final launched version using StreamCompute is faster and more efficient than the Storm version.

Tmall's use of the StreamCompute platform ensured a more stable and efficient processing link. For a description of the Double Eleven data flow, see the following section.

- **User online purchase rush**

  During Double Eleven, Tmall's purchasing system experienced a massive rush in online purchases by users. During peak hours, such as the 12am flash sales on Nov.11, the StreamCompute system handled hundreds of thousands of orders per second.
- **Real-time data collection**

  The data collection system collects the database change log and accesses the DataHub system. By using Alibaba's internal database change log for copying computation (called *DTS* in Alibaba Cloud), the online transactions database ensures that data is written into DataHub in a matter of seconds, even during peak transaction periods.
- **Real-time data computation**

  The StreamCompute system subscribes to DataHub stream data. It uses the data to constantly analyze and compute the total turnover of Tmall up to the current time. An Alibaba Cloud StreamCompute PC cluster includes thousands of PCs and supports a mega-scale throughput per second, which makes handling Tmall's hundreds of thousands of transactions per second a breeze. StreamCompute subscribes and computes in real time, and instantly writes the results data into the online RDS system.
- **Front-end visualization**

  The front-end visualization component team developed a variety of eye-catching customized features for the Tmall Double Eleven Screen, such as a ticker board and a display for global transaction hot spots. The front-end server regularly polls the RDS system and uses Web front-end technology to enhance the Double Eleven Screen.

The work of the StreamCompute team has evolved from the use of the Alibaba big data platform, and has made significant contributions to our Double Eleven shopping event over the past several years.

15.8.4 Wireless Data Analysis

Alibaba Cloud StreamCompute can help customers perform real-time wireless app data analysis . This includes the analysis of a mobile app's indexes, such as the app version distribution, crash detection and distribution, etc. Alibaba Cloud Mobile Analytics (MAN) is designed for mobile app data statistical analysis. It offers multi-dimensional user behavior analysis and supports independent log analysis. It also helps mobile developers realize big data technology-based refined operations, improves product quality and experience, and increases user stickiness. Alibaba Cloud MAN's bottom layer big data analytics fully relies on the use of Alibaba Cloud's big data products, including StreamCompute, MaxCompute, and more. For stream computing, MAN uses Alibaba Cloud StreamCompute as the bottom-layer big data analytics engine, and offers a full set of real-time mobile app analytics reporting services to customers with mobile data analytics needs.

MAN is already being utilized by several hundred users on Alibaba Cloud. When combined with Alibaba Cloud's big data platform (DTPlus), MAN allows users to access wireless analytics functions that are customized, work in real-time and support user-defined analysis logic. This significantly expands the overall functionality of MAN. The following section describes MAN's current full data flow.

- **Data collection**

  Developers use the SDK provided by “Alibaba Cloud Mobile Analytics” and build it into the installation package of their app. The SDK offers data collection components based on different mobile operating systems that collects both mobile data and user behavior data. This data is then analyzed using MAN's background analysis system.

- **Data Reporting**

  MAN's background analysis system offers a full set of SDK data reporting services that collect data reported by mobile phones using an SDK. The reporting system preliminarily de-noises this data and then sends it to DataHub.

  Note:

DataHub reports the mobile terminal data directly to the SDK. This allows for the reporting process of MAN's background analysis system to be omitted entirely (the de-noising process can be completed by StreamCompute), which further reduces MAN's hardware costs.

- **Stream computing**

  StreamCompute subscribes to the DataHub stream data. It then analyzes and uses the data to compute various app indexes. The results data for each time period is written into the online RDS/Table Store system instantly.

- **Data presentation**

  MAN provides a complete set of operating indexes that allow developers to quickly understand where their users come from, which pages they visited, how long they stayed on each page , and the status of users terminals and network conditions. They can also receive real-time feedback when their apps freeze or crash. The crash analysis function is accurate to a device level of granularity, which allows developers to view detailed crash analysis information for individual devices. The real-time data involved in this process is derived the results of the StreamCompute analysis.

15.9 Restrictions

None.

15.10 Glossary

- Project

  The most basic business organizational unit of Alibaba Cloud StreamCompute, as well as the basic unit used for managing clusters, jobs, resources, and staff. A user can create a project, or join another project as a sub-account. StreamCompute projects support the collaboration using Alibaba Cloud RAM user accounts.

- Job

  Similar to MaxCompute/Hadoop jobs, it is a complete set of stream data processing logic. It is the basic business unit of stream computing.

- Compute Unit (CU)

  For StreamCompute, the CU is the basic stream computing unit of a job. It is the minimum operational capacity used to run a StreamCompute job. It is the capacity to handle event streams with a defined CPU, memory, and IO. A StreamCompute job can run on one or more CUs.

Currently, StreamCompute defines it as: `1 CU = 1 Core CPU + 4G MEM`.

- StreamSQL

  StreamCompute differs from most open source stream data processing systems providing low-level programming APIs. It offers the higher-level business-facing StreamSQL (standard syntax extensions for the SQL syntax related to stream processing). This allows data developers to complete stream data computing and processing with standardized SQL alone. This makes StreamCompute a good choice for data analysts for quick and easy stream data processing.
- UDF

  Similar to Hive UDFs, StreamSQL also offers UDFs in addition to standardized stream data processing capacity. This allows users to customize the processing logic for businesses. Currently, StreamCompute only supports the Java UDF extension.
- Resource

  The Jar uploaded by users. Currently, StreamCompute only supports UDF (User Define Function) using Java.
- Data Collection

  In a broad sense, it is a process for collecting data from the data source and then transferring the data to a big data processing engine. StreamCompute generally follows this definition, but focuses more on collecting and transferring stream data to the data bus.
- Data Store

  StreamCompute is a light computing engine **that is built without a business data storage system**. Alibaba Cloud StreamCompute uses external data storage as both the data source and the destination. For StreamCompute, all external data storage are data stores. For example, the user RDS is used as the results table, which makes RDS a data store of StreamCompute.
- Data Develop

  The development process of stream computing (the process of writing StreamSQL). Alibaba Cloud StreamCompute offers a complete set of online IDEs for stream data processing that contain development and debugging functions.
- Data Operation

  The online operation and maintenance of stream computing jobs. StreamCompute offers a full set of management platforms where users can easily manage stream data operations.

16 E-MapReduce

16.1 What is EMR

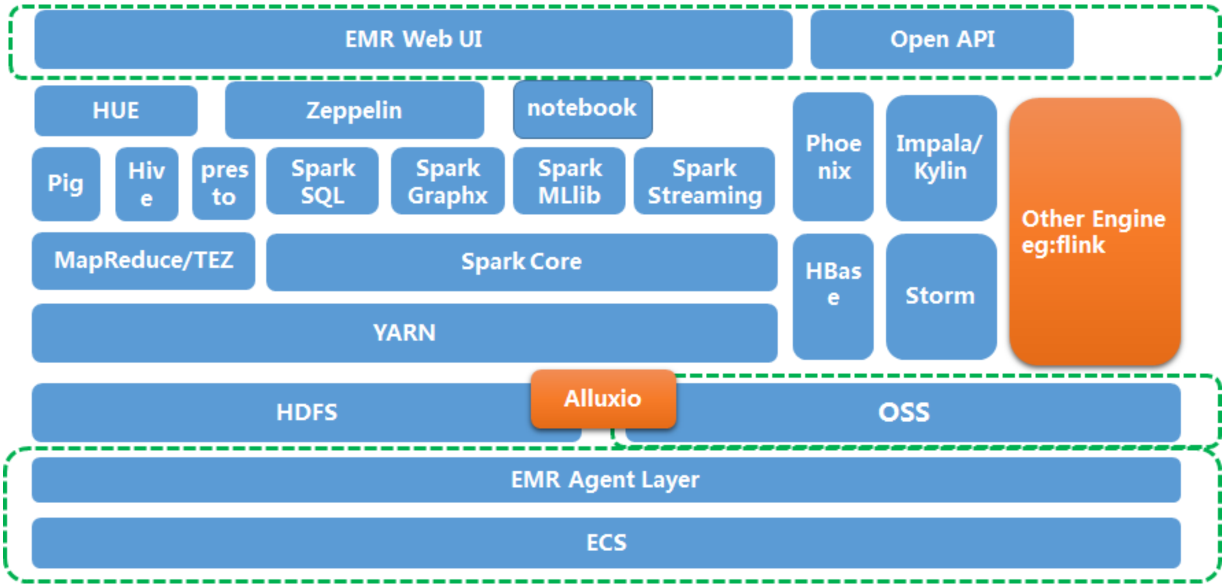
E-MapReduce (EMR) is a one-stop big data processing and analysis service that uses resources of the open-source big data ecosystem, including Hadoop, Spark, Kafka, and Storm, to provide users with the cluster, job, and data management functions.

EMR is built on Alibaba Cloud Elastic Compute Service (ECS) and is based on open-source Apache Hadoop and Apache Spark. It allows you to use other peripheral systems, such as Apache Hive, Apache Pig, and HBase, in the Hadoop and Spark ecosystems to analyze and process your own data. Moreover, you can use EMR to easily import and export the data to other cloud data storage and database systems of Alibaba Cloud, such as Alibaba Cloud Object Storage Service (OSS) and ApsaraDB for RDS.

16.2 Architecture

Figure 16-1: EMR architecture shows the architecture of EMR.

Figure 16-1: EMR architecture



EMR clusters are created based on the Hadoop ecosystem. EMR clusters can exchange data seamlessly with Alibaba Cloud services, such as Object Storage Service (OSS) and Relational Database Service (RDS). This enables you to share and transmit data between multiple systems to meet different business demands. E-MapReduce provides a series of OpenAPIs to facilitate you to operate clusters, jobs and execution plans.

For more introductions to the components in the Hadoop ecosystem, see *Glossary*.

16.3 Benefits

Compared with manually creating clusters, EMR offers an easier way for you to comprehensively manage the EMR clusters. Additionally, EMR offers the following benefits:

- Deep integration

  EMR is deeply integrated with other Alibaba Cloud services such as OSS, MNS, RDS, and MaxCompute. This enables these services to act as the input source or output destination of the Hadoop or Spartk compute engine in EMR.
- Security management

  EMR is integrated with Resource Access Management (RAM), using primary and sub-accounts to control service access.

16.4 Features

E-MapReduce provides the following functions:

- E-MapReduce supports a variety of job types including Spark, Hadoop, Hive, Pig, Sqoop, Spark SQL, Shell and so on, which can meet users' needs such as log analysis, data warehouse, business intelligence, machine learning and scientific simulation.

  You can define the command to execute and the policies for handling execution failures after choosing the job type according to your actual situation. You can also clone, modify, and delete jobs.
- Supports to create a flexible execution plan.

  An execution plan is a set of jobs that can be executed once or periodically through scheduling configuration. It can be executed on an existing E-MapReduce cluster and also can create a temporary cluster to execute jobs dynamically. Its biggest advantage is to use resources actually needed during execution to maximize resource savings. Its flexibility is shown as follows:

  - You can combine any jobs (including Hadoop/Spark/Hive/Pig) into execution plans.
  - There are two scheduling policies for execution plans, including manual execution and periodical execution.
- Provides notebook.

Notebook allows you to compile and run Spark, Spark SQL, and Hive SQL tasks directly on the E-MapReduce console. You can view the running results directly on Notebook. Notebook is ideal for processing debugging tasks that require shorter runtime and whose data results need to be viewed directly. For tasks with longer runtime and requiring regular execution, the job and execution plan function must be used.

- Supports alarm management.

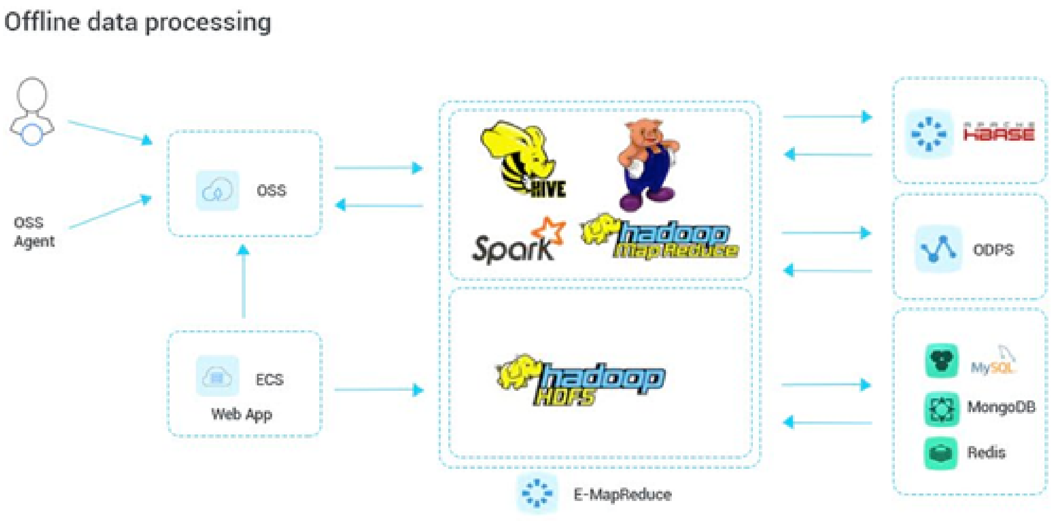
  E-MapReduce supports to associate execution plan with alarm contact group. After “Alarm notice” is enabled in the Execution Plan Details page, the contacts in the correlated alarm contact group can receive the short message notice once the plan is executed. The short message contains the execution plan name, plan execution (quantities of plans succeeded and failed), correspondinsg execution cluster name and specific execution time.

16.5 Scenarios

- **Offline data analysis**

  After synchronizing massive logs of games, Web applications, mobile apps, and other services from servers to EMR data nodes, you can quickly obtain data insights using tools such as Hue and mainstream computing frameworks such as Hive, Spark, and Presto. You can also load data distributed among ApsaraDB for RDS instances or other storage engines using tools, such as Sqoop, and synchronize the analyzed data to ApsaraDB for RDS instances, to provide data support for data visualization products.

Figure 16-2: Offline data analysis

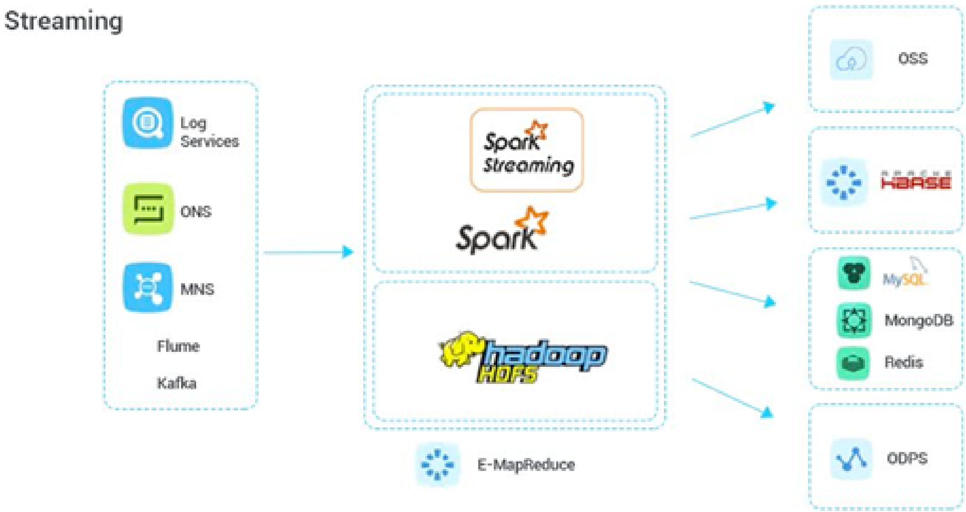


- **Streaming data analysis**

  With Spark Streaming and Storm, you can use and process real-time data of Alibaba Cloud Log Service (Log), Message Queue (MQ), Message Service (MNS), Apache Kafka, or other streaming data of data streams.

  You can also analyze the streaming data in a fault-tolerant way and write the corresponding results to Alibaba Cloud Object Storage Service (OSS) or Hadoop Distributed File System (HDFS).

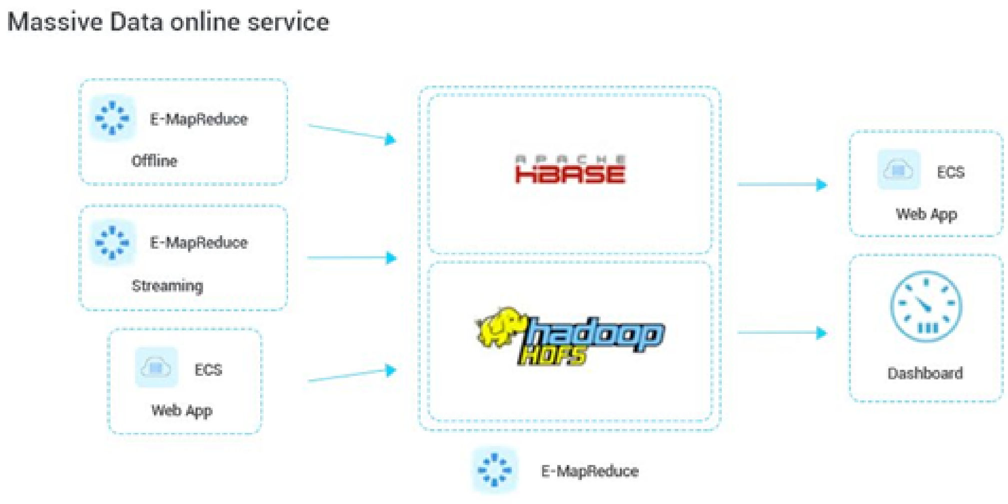
  Figure 16-3: Streaming data analysis



- **Online massive data analysis**

  EMR analyzes petabytes of structured, semi-structured, and unstructured data generated by your Web and mobile applications online, facilitating the Web applications or data visualization products to obtain the analysis results and display them in real time.

Figure 16-4: Online massive data analysis



16.6 Restrictions

None.

16.7 Glossary

- **Job**

  Similar to MaxCompute/Hadoop jobs, E-MapReduce job is the basic unit of big data processing and analysis services.

- **Hadoop**

  - YARN

    Schedules tasks and manages cluster resources.

  - HDFS

    Provides a distributed file storage system.

- **Hive**

  Hadoop-based offline data processing system that provides SQL-like query syntax for data analysis and processing, and stores data in tables with table management capabilities.

- **Spark**

  Memory-based new-generation distributed computing framework that supports offline and real-time computing, SQL syntax, and machine learning.

- **Hue**

  Visualized platform that manages open source components such as Hadoop, Hive, Oozie, and HBase.
- **Oozie**

  Job scheduling engine that supports complex DAG orchestration of jobs of various types.
- **Presto**

  Presto is a distributed SQL query engine that queries big data sets distributed among one or more data sources.
- **Zeppelin**

  Zeppelin is an interactive data query and analysis tool in the form of Web notes. You can query and analyze data and generate reports using Scala and SQL online.
- **ZooKeeper**

  Distributed open source application coordination service as an open source implementation of Google Chubby and an important component of Hadoop and HBase. As the software offering the consistency service for distributed applications, it provides features such as configuration maintenance, domain name services, distributed synchronization, and group services.
- **Sqoop**

  Data migration tool that supports migration of data between ApsaraDB for RDS and HDFS.
- **Kafka**

  Kafka is a high-throughput distributed message system featuring scalability, high reliability, and high performance. It is widely used for real-time computing, log processing, aggregation, and other scenarios.
- **HBase**

  As a distributed and column-oriented open source database, HBase is a subitem of Apache Hadoop. Different from common relational databases, HBase is applicable to unstructured data storage and works using the column oriented storage, instead of the row oriented storage.
- **Phoenix**

  Provides SQL-like syntax for analysis of HBase data.
- **MetaService**

  You can access resources of Alibaba Cloud without entering AccessKey based on this service in E-MapReduce.

- **Kerberos Authentication**

  It is a trusted third party authentication protocol designed for TCP/IP network, which is based on DES symmetric encryption algorithm.

17 Quick BI

17.1 What is Quick BI

Quick BI is a flexible and lightweight self-service platform, and it provides BI tools based on cloud computing.

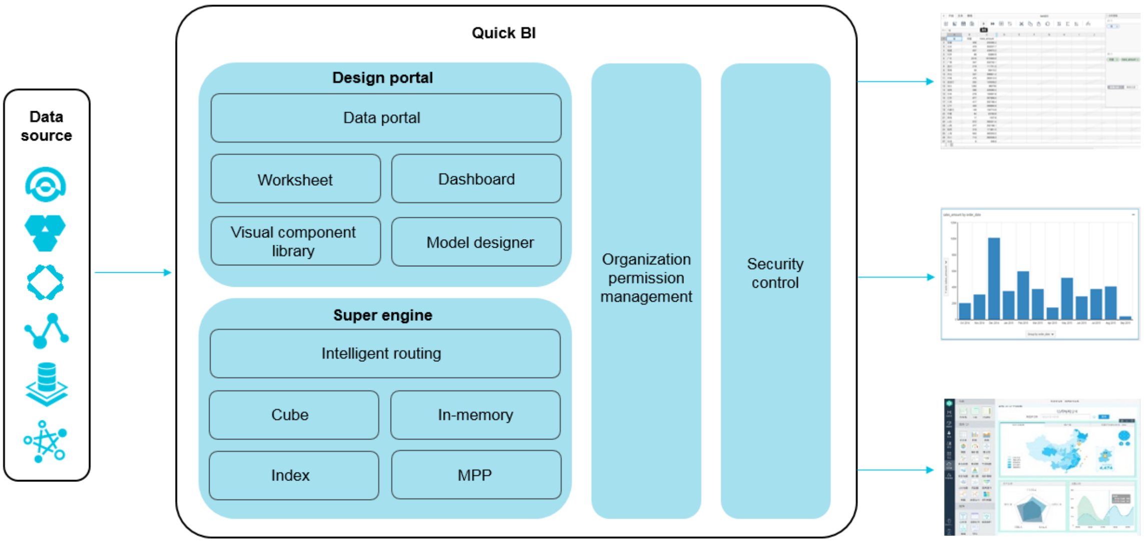
Quick BI can connect to multiple data sources, including cloud data sources such as MaxCompute (ODPS), HybirdDB for MySQL, AnalyticDB, and HybridDB (Greenplum), as well as your MySQL database on ECS, meanwhile, the data source from the VPC is also supported . Quick BI provides a real-time online analysis service tailored to massive data. With an intelligent data modeling tool, Quick BI reduces data acquisition cost by a large margin and makes it much easier to use . Besides, the drag-drop operation and the rich set of visual chart controls allow you to easily complete data perspective analysis, self-service data acquisition, business data profiling, report making, and data portal building.

In addition to a data viewer for business personnel, Quick BI also turns everyone into a data analyst to achieve data-based operation of enterprises.

17.2 Architecture

The architecture of Quick BI is shown in the following figure:

Figure 17-1: Quick BI architecture



Modules and features of Quick BI

- **Data connection module**

  Compatible with various cloud data sources, including but not limited to MaxCompute, RDS (MySQL, PostgreSQL, SQL Server), AnalyticDB, HybridDB (MySQL, PostgreSQL), used to encapsulate standard query APIs of meta data and other data from data sources.
- **Data preprocessing module**

  Provides lightweight ETC processing of data sources and supports custom SQL of MaxCompute. More data sources will be supported in the future.
- **Data modeling**

  Responsible for OLAP modeling of data sources, transforming data sources to multi-dimensional analysis model, supporting standard semantics such as dimensions (including date and geographic position), measurement, and star-type topology model, as well as computing field, and allowing you to process dimensions and measurements again by using SQL syntax of current data source.
- **Worksheet/Workbook**

  Provides operations related to online electronic spreadsheet (webexcel), including data analysis (such as row and column filtering, common/advanced filtering, classified aggregation , AutoSum, conditional formatting), data export, text processing, sheet processing, and other operations.
- **Dashboard**

  Assembles visual chart controls into a dashboard in a drag-drop manner, and supports 17 charts (such as line chart, pie chart, bar chart, funnel chart, tree chart, bubble map, color map , and indicator board), four basic controls (query conditions, TAB, IFRAME, and text box), and inter-chart data linkage.
- **Data portal**

  Assembles dashboards into a data portal in a drag-drop manner, and supports embedded link (dashboard), external link (third-party URL), and settings of template and menu bar.
- **QUERY engine**

  Queries data sources.
- **Organization permission management**

  Manages permissions based on <organization - workspace> architecture and user roles under workspace to achieve permission control and enable different users to view different tables.

- **Row-level permission management**

  Controls row-level permission of data and enables users of different roles to view different data from one report.
- **Share/Publish**

  Shares worksheets, dashboards, and data portals to other logged-in users, and publishes dashboards on the Internet for non-logged-in users to access.

17.3 Features

Quick BI offers the following functions:

Seamless integration with cloud-based database

Supports various Alibaba Cloud data sources, including but not limited to MaxCompute, HybirdDB for MySQL (MySQL, PostgreSQL, SQL Server), AnalyticDB, and HybridDB (MySQL, PostgreSQL).

Chart

Diverse options for data visualization. The built-in 17 types of visual charts (such as bar chart, line chart, pie chart, radar chart, and scatter chart) can meet data presentation demands of different scenarios. Besides, it can automatically recognize data features and recommend an appropriate visualization solution.

Analysis

Multi-dimensional data analysis. The web page based environment supports Microsoft Excel-like drag-and-drop operations, data import with one click, and real-time analysis. This allows you to analyze data from different perspectives without having to build a new model.

Quick building of data portal

Drag-and-drop operations, powerful data modeling, and rich visual charts help you build a data portal in a short time.

Real-time

Supports online analysis of massive data without preprocessing, thus greatly improving the analysis efficiency.

Secure management of data permissions

Provides organizational member management, and supports row-level data permissions to enable users of different roles to view different reports as well as to view different data from a same report table.

17.4 Benefits

The benefits of Quick BI can be summarized as follows:

High compatibility

Supports multiple data sources such as HybirdDB for MySQL, MaxCompute, and AnalyticDB.

Fast response

Responds in seconds for hundreds of millions of data.

Powerful capabilities

The built-in complete spreadsheet tools allow you to easily create complex Chinese statements.

Ease of use

Rich data visualization, automatic identification of data features, and automatic intelligence function can help you to generate the most appropriate chart.

17.5 Limitations

None.

17.6 Scenarios

The following scenarios showcase how Quick BI can be integrated to achieve powerful data analytics and drive forward business goals.

17.6.1 Fast data-driven decision making

Business goals:

- Easier method to read data

  Currently, businesses must rely on IT professionals to design SQL statements for multi-dimensional analytics.
- Easier report making processes

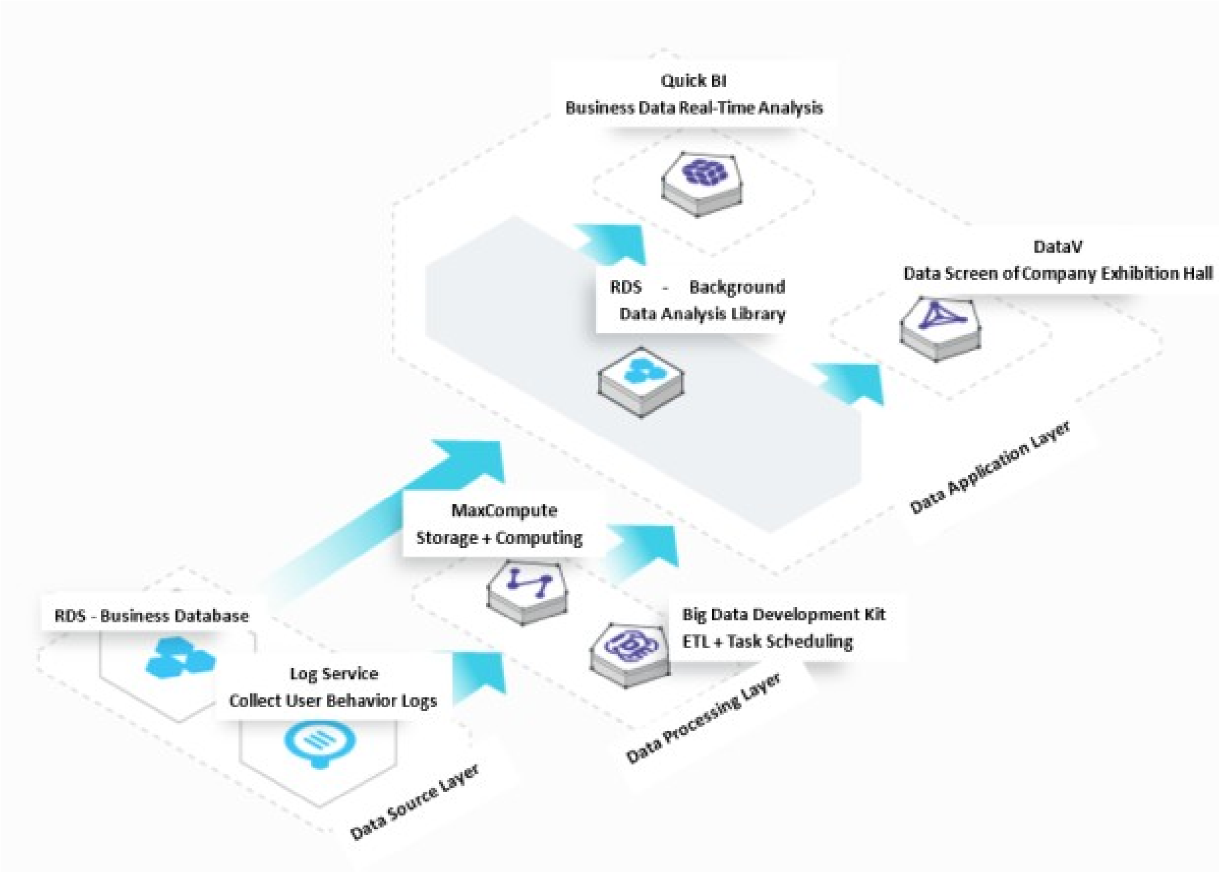
Delivering updates and new code to the backend analytics system is time-consuming and tedious.

- Reduction of associated costs

  Many reporting libraries usually require a tremendous upfront cost.

Quick BI integration

Figure 17-2: RDS + Quick BI



17.6.2 Integrate with existing systems

Business goals:

- Easy adoption

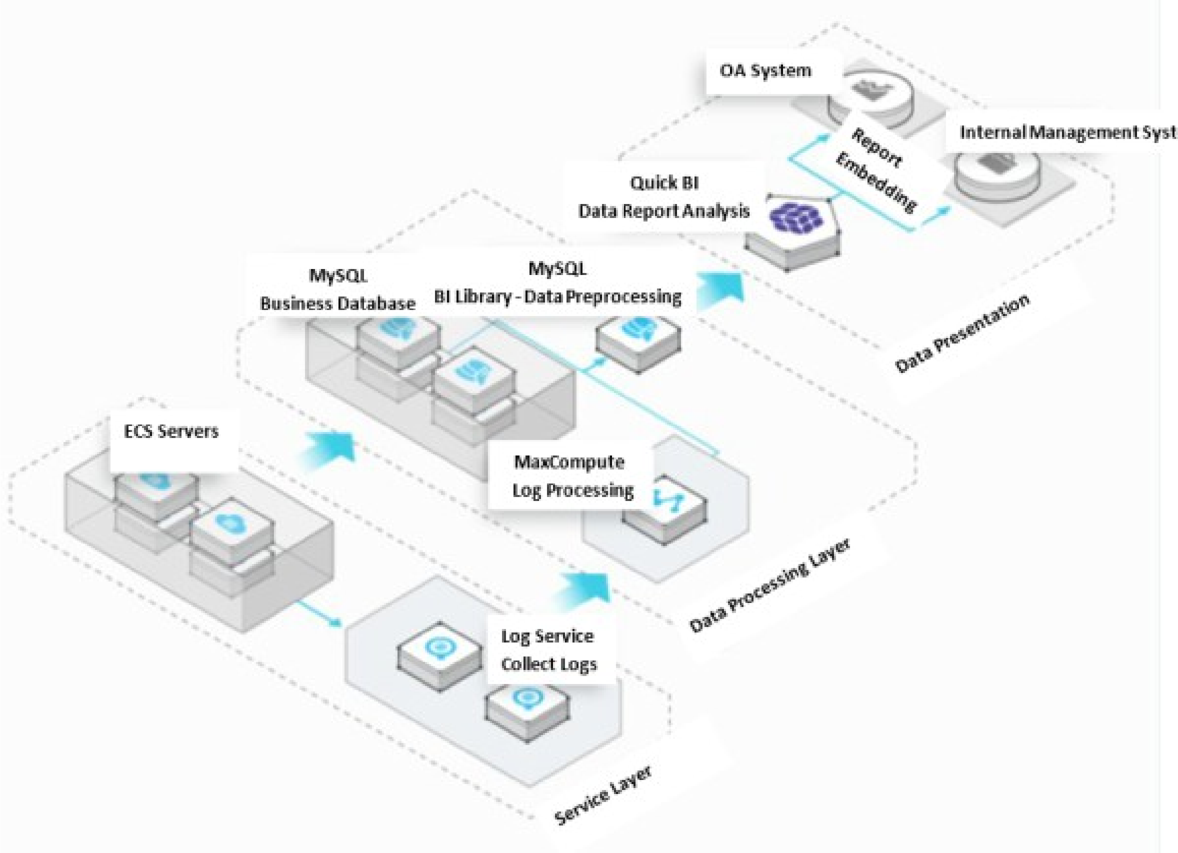
  It can be difficult to find a data analytics platform that is both easy to adopt and suitable for users with varying degrees of expertise.
- Faster data consumption

  Smooth integration with existing systems allows for rapid analytics and data consumption.
- Unified interface

An all-in-one interface enables users to easily access data and eliminate the need to use multiple systems.

Quick BI integration

Figure 17-3: RDS + Quick BI



17.6.3 Control transactional data

Business goals:

- Row-level access control

  Implement fine-grained access control in order for users to view and analyze data directly related to their specific tasks and goals.

- Rapid response to change

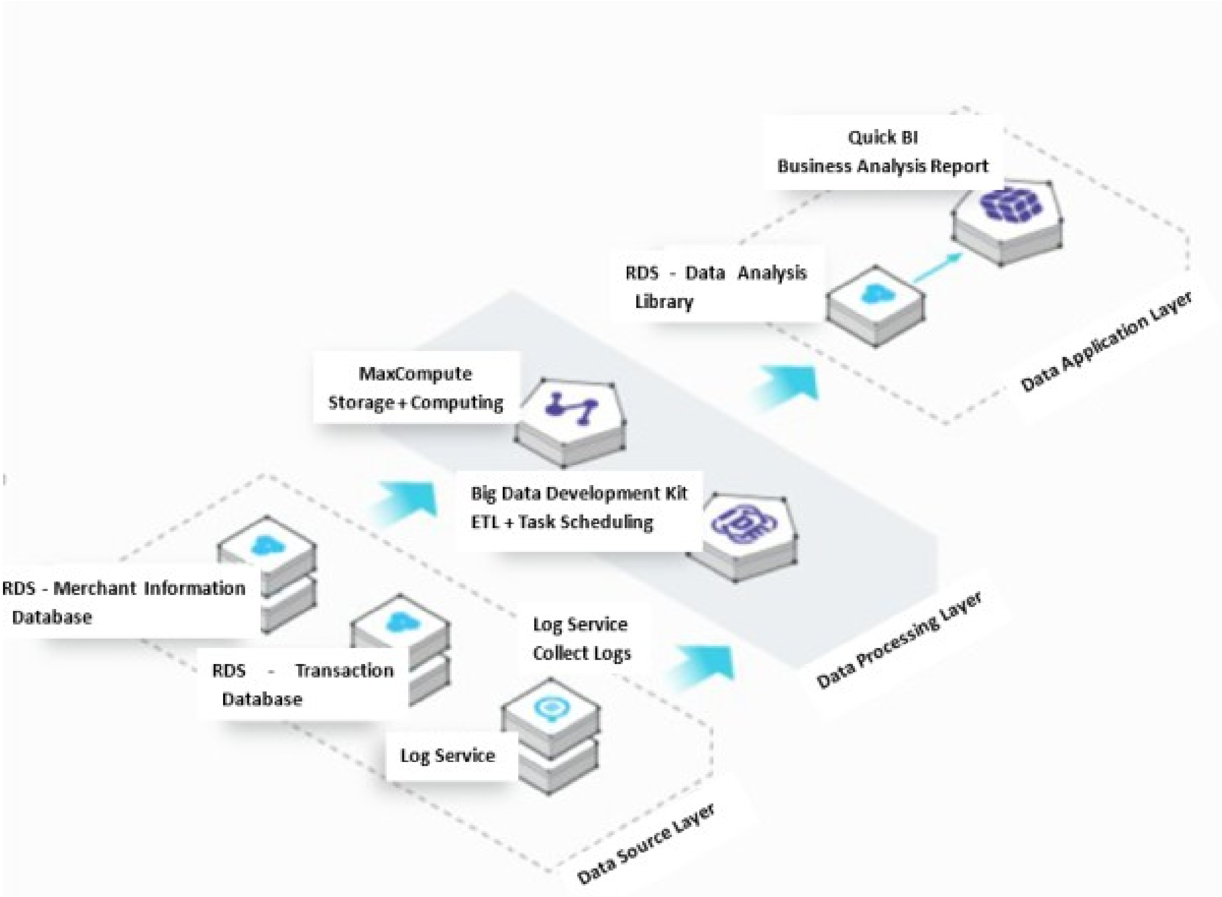
  When business growth adjusts, the analytics platform must respond to these changes quickly.

- Consistent computing performance across multiple data sources

  A strong cloud infrastructure and data analytics platform underlayer allows you to connect to different data sources without compromising on performance.

Quick BI integration

Figure 17-4: Log + RDS + Quick BI + MaxCompute

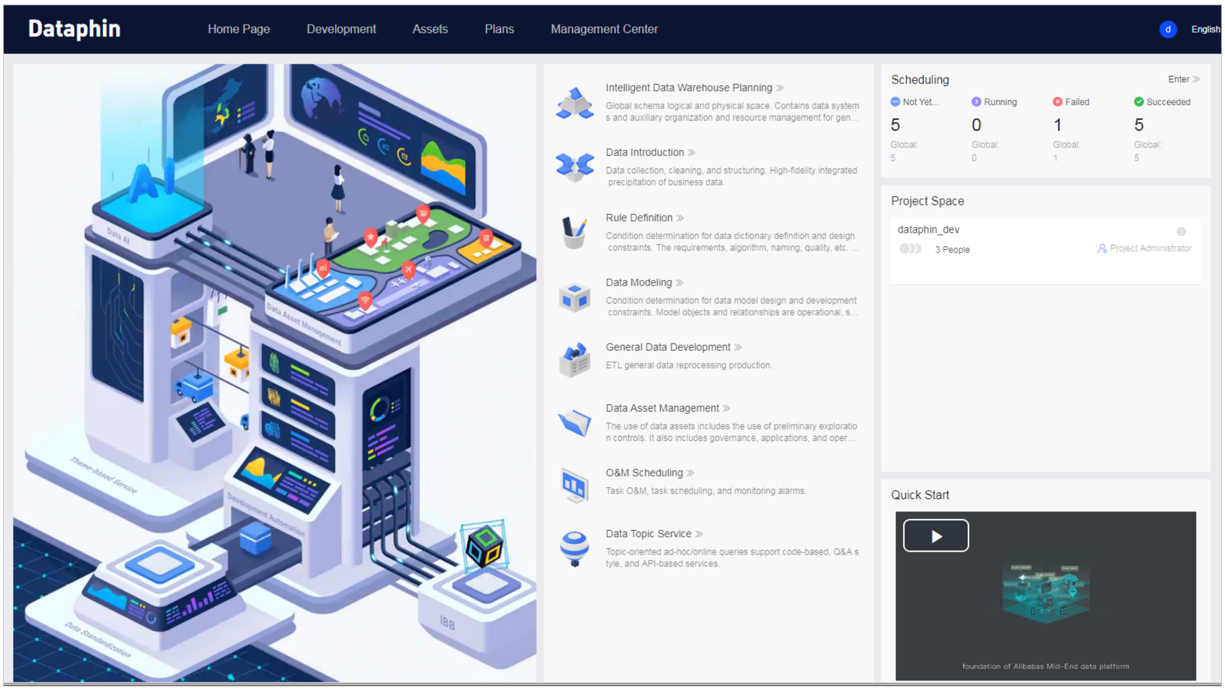


18 Dataphin

18.1 What is Dataphin

Dataphin is an intelligent data construction and management engine that has been designed for utilization in multiple industries. Dataphin applies the OneData, OneID, and OneService data construction technology that has been tested by business in Alibaba Group for 10 years. Dataphin provides an end-to-end intelligent data construction and management service, which includes data importing, data standardization, data modeling, data development, data distilling, asset management, and other data services. These features can be used to help the government and enterprises build an intelligent data system that includes standardization, integration, assets, services, and closed-loop optimization.

Dataphin aims to support different compute and storage environments. By using Dataphin, you can quickly import data, construct standardized data, and build data models. The service also allows you to create a tag system using customer and product data to gain business knowledge , create data assets, and resolve business issues. Dataphin also provides multiple types of data services including data table search and intelligent voice search.



Intelligent Data Warehouse Planning »
Global schema logical and physical space. Contains data systems and auxiliary organization and resource management for gen...

Data Introduction »
Data collection, cleaning, and structuring. High-fidelity integrated precipitation of business data.

Rule Definition »
Condition determination for data dictionary definition and design constraints. The requirements, algorithm, naming, quality, etc. ...

Data Modeling »
Condition determination for data model design and development constraints. Model objects and relationships are operational, s...

General Data Development »
ETL general data reprocessing production.

Data Asset Management »
The use of data assets includes the use of preliminary exploration controls. It also includes governance, applications, and oper...

O&M Scheduling »
Task O&M, task scheduling, and monitoring alarms.

Data Topic Service »
Topic-oriented ad-hoc/online queries support code-based, Q&A style, and API-based services.

Scheduling

Not Yet...	Running	Failed	Succeeded
5	0	1	5
Global: 5	Global: 0	Global: 1	Global: 5

Project Space

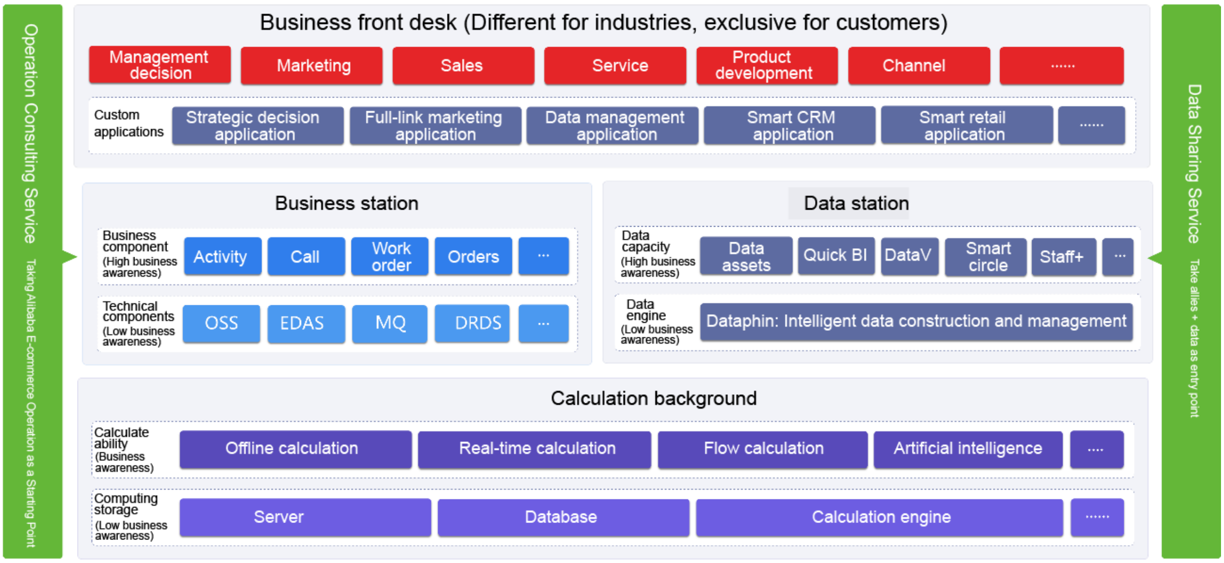
dataphin_dev
3 People
Project Administrator

18.2 Benefits

- **Data standardization:** Defines the standards based on the dimensions, dimension attributes, business workflows, and metrics and fields in dimensional modeling. This ensures the quality of the data and prevents the ambiguity of metrics.
- **High-performance and automatic coding:** Defines logical components for common data computing by using functions. Dataphin also supports statistics metric customization, which enables you to create data models and construct production data through automatic coding.
- **Intelligent computing optimization:** Supports logical modeling for business data overview. The system can automatically create physical models and encode data based on the released logical models, which reduces your dependence on professional data developers.
- **One-stop development:** Integrates data importing, modeling, development, O&M, data search, and probing, to implement centralized and high-performance development.
- **Deep data distillation:** Simple object definition and compute parameter settings allow you to quickly verify IDs, create correlations with these IDs, and build a tag system that uses personas based on IDs and tags. This simplifies the establishment of the data management platform for marketing.
- **Systematic data directories:** Supports standardized modeling, as well as high-performance and automatic metadata distillation. You can use a standardization model to create standardized and user-readable business data directories and data asset maps for data search, analysis, and transmission.
- **Semantic intelligent data search:** Metadata-based data profiling for simple, fast, and intelligent data and table search.
- **Data asset visualization:** Systematically creates data asset charts and data maps for your business system, gains business knowledge, and refines data in key components of your business.
- **Ease of use and reliability:** Metric definitions are services. Quick search and access are available to logical tables created based on business topics, which saves 80% of the encoding time.
- **Low costs:** Dataphin is metadata-based and AI algorithm-driven. The intelligent and automatic production feature can help you separate the physical plane from the logical plane. In addition to the full link analysis, tracking, and the optimization feature for data assets, you are also able to create the optimal strategy for computation and storage resource allocation, so that you can greatly reduce the costs.

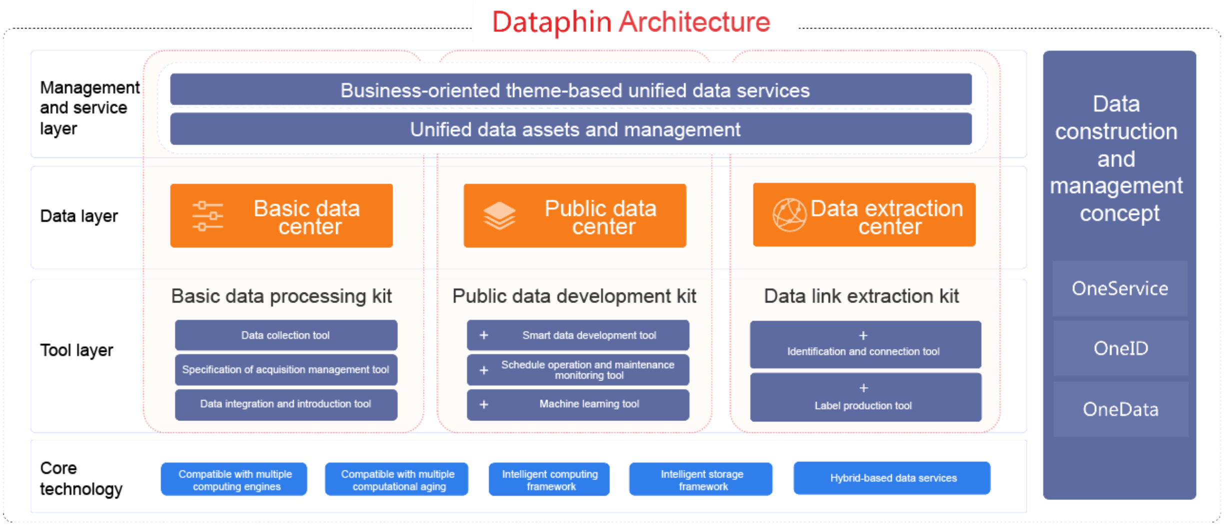
18.3 System architecture

The following figure shows where Dataphin has been deployed in the business system:


Operation Consulting Service
Taking Alibaba E-commerce Operation as a Starting Point
Business front desk (Different for industries, exclusive for customers)
Management decision
Marketing
Sales
Service
Product development
Channel
......
Custom applications
Strategic decision application
Full-link marketing application
Data management application
Smart CRM application
Smart retail application
......
Business station
Business component (High business awareness)
Activity
Call
Work order
Orders
...
Technical components (Low business awareness)
OSS
EDAS
MQ
DRDS
...
Data station
Data capacity (High business awareness)
Data assets
Quick BI
DataV
Smart circle
Staff+
...
Data engine (Low business awareness)
Dataphin: Intelligent data construction and management
Calculation background
Calculate ability (Business awareness)
Offline calculation
Real-time calculation
Flow calculation
Artificial intelligence
....
Computing storage (Low business awareness)
Server
Database
Calculation engine
......
Data Sharing Service
Take allies + data as entry point

For business, data services are the most fundamental feature in digital transformation (DT). Dataphin is the underlying foundation of platform as a service (PaaS), and can quickly process raw data in your business system, output standardized data, and make the data services easy-to-use. Dataphin offers reliable and stable support for a variety of data products and applications. Industries adopting Dataphin can deliver high-quality and high-performance data monitoring, digital marketing, data-driven business services, and data products. Based on these services, Dataphin interconnects infrastructure as a service (IaaS) and software as a service (SaaS). This allows Dataphin to work with different types of hardware (such as servers, databases, and compute engines), and serve multiple products and applications (such as Quick BI report analysis and enterprise strategy making).

The following figure shows the architecture of Dataphin:


Dataphin Architecture
Management and service layer
Business-oriented theme-based unified data services
Unified data assets and management
Data layer
Basic data center
Public data center
Data extraction center
Tool layer
Basic data processing kit
Data collection tool
Specification of acquisition management tool
Data integration and introduction tool
Public data development kit
Smart data development tool
Schedule operation and maintenance monitoring tool
Machine learning tool
Data link extraction kit
Identification and connection tool
Label production tool
Core technology
Compatible with multiple computing engines
Compatible with multiple computational aging
Intelligent computing framework
Intelligent storage framework
Hybrid-based data services
Data construction and management concept
OneService
OneID
OneData

Dataphin is based on the OneData, OneID, and OneService data construction and management concept. The architecture of Dataphin includes four layers:

- Core technology: A technical framework that tolerates the differences in underlying compute, storage, and software systems. This ensures that data development can be compatible with multiple compute engines, ensures data validity, and provides support for automatic encoding, intelligent storage and computing, and hybrid storage.
- Tool layer: This layer provides data construction and management tools for developers, including data standardization, data integration, basic data importing, public data standardiz ation, intelligent modeling development, scheduling, O&M, machine learning, ID verification and association in data distillation, and tag creation.
- Data layer: This layer utilizes the core technology and data handling tools to output three types of structured data: basic data, public data, and distilled data. The basic data center provides highly accurate business data. The public data center provides topic-oriented and calculated data. The distilled data center provides deep levels of business entity processing.
- Management and service layer: This layer provides an overview of data assets and data services, allowing both the developers and sales to obtain high-quality and unified data assets . From the business perspective, the existing data is packaged and processed into topical data services to ensure that business data can be centrally searched and accessed.

18.4 Functions

18.4.1 Overview

- **Platform**: This module allows you to learn about the entire product system, the global settings, and the features. It also enables you to implement system management and control to ensure that all the other modules are running correctly.
- **Comprehensive design**: This module allows you to gain a comprehensive overview of your business and build a business data bus. You can partition data center name spaces, define topic domains and relevant metrics, create management units (projects), and create data sources.
- **Data import**: Projects and physical data sources created based on the comprehensive design module. You can build a basic data center in the procedures of data distillation, data synchronization and integration, and data scrubbing.
- **Standardization**: A basic data center built based on the data import module and the service bus defined in the comprehensive design module. This module provides tags and metrics as components and definitions to meet business data requirements and to guarantee standardized data construction.
- **Modeling development**: Data elements created based on standardization. This module enables you to visually design, build, and submit data models. The system then automatically generates code and dispatches tasks for full management over the public data center.
- **Encoding development**: This module provides a common encoding interface for you to flexibly develop data encoding and submit tasks.
- **Resource and function management**: This module allows you to manage resource packages (such as JAR files and archives) to meet data processing requirements. The module supports native system function search and utilization, as well as custom functions.
- **Data distillation**: The key concept for data distillation is target objects. Based on the basic data center and public data center, this module provides optional parameters for you to implement full management. Full management includes verifying and associating object IDs, distilling object behavior and tags, implementing data exploring and deep mining, generating code, and scheduling tasks.
- **Task scheduling and maintenance**: This module supports the scheduling and management of encoding tasks for modeling development, encoding development, and data distillation. The module allows you to deploy data construction tasks, view task running status and

dependencies, and manage and maintain tasks to ensure that all tasks are executed at the specified time.

- **Metadata center:** This module collects, analyzes, and manages metadata in the basic data center, public data center, and distilled data center.
- **Asset analysis:** Based on the metadata center, this module supports metadata deep analysis and data asset management, and allows you to gain an overview of asset distribution and detailed information about metadata. In this way, you can quickly and deeply learn about data assets.
- **Security management:** This module supports quality and security management, including metric definition, analysis result display, workflow management, monitoring and alarms, full link tracking from data sources to applications. With these features, you are able to locate asset optimization problems and provide solutions.
- **Ad-Hoc query:** This module supports asset data search by using custom SQL queries. You can use the search and analysis engine to quickly search data in physical tables and topic-oriented logical tables. Topic-oriented logical tables are also known as data models or logical models.

18.4.2 Platform management

As a basic function of Dataphin, this module ensures that all users carry out data R&D in a controlled, orderly, and smooth manner. This module provides global functions such as account management, computation management, quick start guide, home page, and data acquisition, to ensure that the super administrator manages the entire platform and that other users quickly access target modules.

18.4.2.1 Account management

To secure use of system functions, this module controls user accounts and identifies and configures the scope of eligible users for the product on the basis of the enterprise's existing account system. The user with the maximum permissions can manage other users' accounts and permissions.

18.4.2.2 Computation management

As a platform product at the PaaS layer, Dataphin allows you to set and manage computing types and underlying hardware configurations and management from the perspective of stable and unified system computing. It is compatible with computing engines at the IaaS layer for data construction in various environments. Dataphin supports mainstream ecological calculation engines, MaxCompute, and Hadoop, as well as automatic collection and analysis of their metadata. For details about metadata collection, deployment, and initialization, see *Metadata center*.

18.4.2.3 Home page guide

The home page is an entry for data construction and management and a portal for unified product guide and integrated workspace. It shows the entire process of data production, management, and services, helping you systematically understand products and quickly access functional modules.

18.4.2.4 Internationalization - language support

Dataphin identifies your system language and displays a corresponding language, Chinese or English, which facilitates users from different countries and regions. The default language is Chinese.

18.4.3 Global design

The top-level global design of data architecture is a foundational step in data construction. It ensures that data management is controllable, the data system defined and designed during data R&D, extraction, management meets mid- and long-term business requirements, and the data acquired by the business teams is consistent with services, topic-oriented, and easy to use.

This module includes the business bus divided based on business characteristics (maintenance and permission control of business module and data domain definitions, and globally statistical period setting and management for public definitions), project space divided based on independent data management and development collaboration requirements (basic project information and computation resource configuration management and membership management), and data sources defined based on project computation resources and business data requirements (data source configuration management).

18.4.3.1 Business bus

The business bus defines logical namespaces, subject categories, and terminology based on business characteristics to standardize data definitions in top management design and construction control.

18.4.3.2 Project space

A project space is a physical namespace defined for resource isolation, user member grouping, and data construction constraint configuration based on the requirements of data R&D and management teams for independent management of data R&D projects and efficient management of data resource quality.

18.4.3.3 Physical data source

Dataphin allows you to register and log out of a database by performing data source operations such as creation and modification. The supported data source types include MaxCompute, MySQL, SQL Server, and Postgre SQL. On the one hand, the data source serves as the source or target of data synchronization transmission; on the other hand, special data source types (such as MaxCompute) can be used for project calculation and storage after you set the calculation engine type.

18.4.4 Data introduction

This module selects the required business data for storage based on the basic data layer design in the global architecture of enterprise data, and formulates data synchronization, cleaning, and structuring polices based on requirements for storage and data timeliness and quality.

As an initial stage in data platform construction, the data synchronization suite is developed based on Alibaba's years of practice in the synchronization and exchange of business data, log data, and other types of data. The suite introduces raw business data efficiently, and collects statistics of transmission metadata through pipes. In terms of data transmission volume and content, it supports simple rule check and flexible management of custom fault tolerance mechanisms, realizing high-quality data synchronization.

18.4.4.1 Data source configuration

This module supports data source access and management. The data source list allows you to manage accessed data sources conveniently and add data sources of various types. Currently , the data synchronization center supports data sources including MaxCompute, MySQL, SQL Server, PostgreSQL, and Hive.

18.4.4.2 Data synchronization

This module allows you to select source data and target data, configure incremental or full-synchronization parameters, identify the mappings between source data fields and target data fields, configure transmission traffic and the number of concurrent transmissions, and generate and dispatch task nodes.

18.4.5 Standard definition

In traditional data R&D, specific and important data construction and R&D such as data modeling and index definition depend on R&D personnel's professional capabilities in many cases. With no uniform naming rules, R&D standards and design are transferred based on individual and changing documents, which are likely to cause a series of problems such as index name conflicts or duplicate calculation.

Dataphin, based on the OneData methodology, standardizes the definition of important data elements such as dimensions, business processes, and indexes. This ensures the uniqueness of calibers, algorithms, and names, and eliminates index ambiguities from the initial stage of data design. In addition, Dataphin helps you create indices in batches based on forms, which facilitates data R&D, quickly enables business personnel with basic data analysis capabilities, and greatly increases the R&D efficiency.

The standard definition mainly involves five modules: dimension, business process, atomic index, business limitation, and derivative indexes. Based on the data domain, Dataphin further implements data construction bus design, common definition reuse, standard data element precipitation - data warehouse topics (including objects, relationships, and transactions), and index construction elements (minimum calculation logical units and decorations).

18.4.5.1 Dimension

- The dimension is unique in the business module and belongs to a unique data domain, which normalizes and standardizes naming and topic classification.
- Dataphin supports the definition of the relationship between main dimensions and sub-dimensions to unify dimensional objects and normalize dimension features.
- Dataphin supports various types of dimensions, including enumerated dimension, virtual dimension, normal (level) dimension, and normal dimensions.
- Dataphin allows you to view and manage the existing dimension list in the business module and project, and quickly view and edit individual dimensions.

18.4.5.1.1 View and manage a dimension list

Dataphin allows you to view the dimension list in a selected project, including the dimension name, creator, and publish status. You can search for a dimension in the dimension list and edit, deprecate, or delete it.

18.4.5.1.2 View and manage dimensions

Dataphin allows you to view dimensions in a list, and create and edit dimensions in a standard manner. As a key business concept, dimensions contain the following information:

- Basic information: Belonging data domain, Chinese name, English name, and description. The English name is prefixed with dim_ by default to ensure uniqueness.
- Logical information: Defines the range of dimension objects to logicalize object characteristics , and ensure true and unique dimensions. For each dimension type, fill in different content to meet the construction requirements for various dimension objects.
- Quick view of dimensions: You can quickly view basic dimension information without affecting existing operations, and quickly reach related operations.

18.4.5.2 Business process

The business process refers to a collection of smallest-unit behaviors or transactions, such as creating an order and browsing the web page. The behavior details in the business process, such as paying an order and browsing a web page, is recorded in a fact table, which will focus on a particular business process in most cases.

Like dimensions, the business process is a top-level design concept of the OneData methodology and clearly defines the data construction architecture together with dimensions. Dataphin supports the standard definition of business processes, which allows you to view the organization's overall business, and classify and manage fact tables easily by using business processes.

To ensure that the fact model is constructed in a unified and standard manner, the business process is unique within the business module and uniquely belongs to a data domain, which standardizes and normalizes naming and topic classification.

Dataphin allows you to view and manage the business process list in the business module and project, and quickly view and edit individual business processes.

18.4.5.2.1 View and manage a business process list

Dataphin allows you to view the business process list in a selected project, including the business process name, creator, and publish status. You can search a specific business process in the list and edit and delete it.

18.4.5.2.2 View and manage a business process

Dataphin allows you to view business processes in a list and create and edit a business process in a standard manner. As a key business concept, the business process contains the belonging data domain, Chinese name, English name, and description.

18.4.5.3 Atomic index

The atomic index is an abstraction of the index statistics caliber and specific algorithms. To eliminate definition and R&D inconsistency, Dataphin innovatively puts forward the concept of " Design is development". When an index is defined, the statistical caliber (that is, calculation logic) is defined. Secondary or repeated development of ETL is not required, which increases the R &D efficiency and ensures the consistency of statistical results. According to the complexity of calculation logic, Dataphin classifies atomic indexes into two types: raw atomic indexes, such as payment amount, and derivative atomic indexes, which are constructed based on combined atomic indexes. For example, the customer price is the payment amount divided by the number of buyers.

To ensure that all statistical indexes are constructed in a unified and standard manner, the atomic index is unique in the business module and uniquely belongs to a source logic table. The calculation logic also uses the fields of the source logic table model as the definition criterion. Each atomic index tracks its own type by using all the logical tables related to the source logic table model, so it may belong to multiple data domains, which achieves normalized naming and logics as well as standardized and systematic topic classification.

18.4.5.3.1 View and manage an atomic index list

Dataphin allows you to view the atomic index list in a selected project, including the atomic index name, creator, and publish status. You can search for a specific atomic index in the list and edit, deprecate, or delete it.

18.4.5.3.2 View and manage an atomic index

- Atomic index

  To ensure standard production of atomic indexes, Dataphin allows you to define an atomic index only in the logical table and its model. You can select a source table, select a field in the snowflake or star model, and define the calculation logic as an atomic index based on the field.
- Derivative atomic index

Derivative atomic indexes are calculated from other atomic indexes. For example, you can obtain the "Order-payment conversion rate" atomic index after you define atomic indexes " Number of payment buyers" and "Number of order buyers" in advance and then calculate " Number of payment buyers"/"Number of order buyers".

18.4.5.4 Business limitation

The atomic index is the standardized definition of the calculation logic, and the business limitation is the standardized definition of condition restrictions. Similar to the atomic index, to ensure that all the statistical indexes are constructed in a unified and standard manner, the business limitation is unique in the business module and uniquely belongs to a source logic table. The calculation logic uses the fields of the source logic table model as the definition criterion. Each business limitation tracks its own type by using all the logical tables related to the source logic table model, so it may belong to multiple data domains, which achieves normalized naming and logics as well as standardized and systematic topic classification.

18.4.5.4.1 View and manage a business limitation list

Dataphin allows you to view the business limitation list in a selected project, including the name , creator, and publish status of each business limitation. You can search for a specific business limitation on the list and edit, deprecate, or delete the business limitation.

18.4.5.4.2 View and manage business limitations

To standardize creation of business limitations, Dataphin allows you to define a business limitation only based on a logical table and its model. You can select a source table, select a field in the snowflake or star schema, and define the computing logic based on this field as a business limitation.

18.4.5.5 Derived indicators

Derived indicators are commonly used statistical indicators. To create statistical indicators in a standard, regular, and unambiguous way, the OneData methodology is abstracted into the following parts:

- Atomic indicator: Statistical criterion, namely, the computing logic.
- Business limitation: The business scope for statistics collection, used to find the records complying with business rules.
- Statistical period: A time period during which statistics are collected, for example, the latest 1 or 30 days.

- Statistics granularity: A statistical object or perspective that defines the extent of data summarization. It can be considered as a grouping condition for aggregation computing (group by clauses in SQL statements). Granularity is a combination of dimensions. For example, if a seller's turnover in a province is used as a statistical indicator, the statistics granularity is the combination of the seller and region dimensions.

The combination of preceding definitions enables quick batch creation of derived indicators without repetitions, and ensures clear and non-repetitive definitions and computing logic. A derived indicator is of the same level as a field and unique within a statistics granularity, ensuring unique and definite statistical data definition for an object combination.

18.4.5.5.1 View and manage derived indicator lists

Dataphin allows you to view the derived indicator list in a selected project, including the name , creator, and publish status of each derived indicator. You can search for a specific derived indicator on the list and edit, deprecate, or delete the derived indicator.

18.4.5.5.2 View and manage derived indicators

To standardize creation of derived indicators, the statistical scope and objects must be determined based on the statistical computing logic. Therefore, you must pick an atomic indicator, and select a combination of the statistics granularity, statistical period, and business limitation related to the atomic indicator. Then, you can create new derived indicators in batches following a standard process with one click.

1. Determine the statistics granularity

   To apply a statistics granularity, select the corresponding dimension combination. Select all the correlated dimensions of the logical table model where the atomic indicator resides in the box to guarantee meaningful and implementable statistical computing based on the selected statistics granularity.

2. Determine the statistical period

   Dataphin supports conventional statistical periods, and allows you to add custom statistical periods in Global Design > Business Bus to meet your own statistical computing requirements.

3. Determine business limitations

   A business limitation is a constraint or filtering condition defined in a logical table. To meet a group or a type of business data requirements, you need to define indicators for the same statistical scope and computing logic for different statistical periods, such as the latest 1 day, 7 days, or 30 days. Therefore, Dataphin allows you to define multiple statistics granularities, statistical periods, and business limitations, which can be flexibly combined for batch indicator production. This ensures standard indicator creation and improves the development efficiency.

18.4.6 Modeling development

Dataphin provides systematic modeling and development functions to implement the data warehouse theory semi-automatically with tools. It can create business dimensions and processes in a top down order, refine development of dimension tables, fact tables, summary tables, and the application layer, and accumulate data assets following unified standards, to facilitate data application in the business hierarchy and optimize computing and storage.

18.4.6.1 Dimension logical table

A dimension logical table shows a detailed dimension logical model and contains details about dimensions. Dataphin allows you to view and manage the list of dimension logical tables, and to view and edit a specific dimension logical table on a GUI.

18.4.6.1.1 View and manage the list of dimension logical tables

Dataphin allows you to view the list of dimension logical tables in a selected project, including the name, creator, creation time, and publish status of each table. You can search for a dimension logical table on the list and edit, deprecate, or delete the table.

You can view details about a dimension logical table model, including the defined primary key , associated dimensions, and attributes of the primary table, summary of the star schema and snowflake schema covering the associated dimension table, and information about the parent and child dimension table models (if an inheritance relation has been defined). You can also publish a logic after unlocking and editing it, zoom in or out the canvas, and switch to an online version . Dataphin also provides a GUI for you to edit a specific dimension logical table model by setting the dimension attributes, associated dimensions, sub-dimensions, and physical parameters of the logical table.

18.4.6.2 Fact logical table

Dataphin allows you to describe the data warehouse model of a specific procedure (for example, place an order or pay for a commodity) or state metric (such as the account balance and inventory) attribute using a fact logical table. Fact logical tables are created in an optimized semi-snowflake schema, which allows attributes other than metrics and associated dimensions to be degraded into fact attributes in a fact table and be categorized. This reduces complexity of the model design and makes it more user friendly.

View and manage the list of fact logical tables

Dataphin allows you to view the list of fact logical tables in a selected project, including the name, creator, and publish status of each table. You can search for a fact logical table on the list and edit , deprecate, or delete the table.

View and edit a fact logical table

Dataphin provides a list of fact logical tables and allows you to view details about a specific fact logical table model on a GUI, including the associated dimensions, metrics, and fact attributes of the primary table, and the dimension logical table associated with the primary table. You can also publish a logic after unlocking and editing it, zoom in or out the canvas, and switch to an online version. Dataphin also provides a GUI for you to edit the model of a specific fact logical table by defining the basic information, primary key, and fields, and setting physical parameters of the logical table.

18.4.6.3 Summary logical table

The summary logical table model is an important data warehouse model. It contains two types of elements: 1. Various statistical values (derived indicators such as the turnover in the latest seven days) used to describe a statistics granularity (a combination of N dimensions, N >= 0, for example, province + product line). 2. Attributes (such as the province name, product line name, and product line level) of the statistics granularity dimensions (such as the province and product line).

View and manage the list of summary logical tables

Dataphin allows you to view the list of summary logical tables in a selected project, including the name and creation time of each table. You can search for a summary logical table on the list and edit, deprecate, or delete the table.

View and edit a summary logical table

A summary logical table can be created in two modes: automatically aggregating the derived indicators defined following the standard process, or mounting fields of physical tables developed with compatible code.

18.4.6.4 Code automation

After the required dimension logical table, fact logical table, and summary logical table are submitted and published, Dataphin automatically designs the physical model, compiles code, and generates scheduling tasks (one logical table usually involves multiple tasks) to produce required data. You can view the execution logic in *Scheduling O&M*.

18.4.7 Code development

Code development is an important data development method parallel with model development during data processing and development in Dataphin. Dataphin allows you to edit scripts using the code compiling mode of your computing engine and submit the scripts to the scheduling system for task generation. You can also trace back versions of nodes to complete development of common data. Different types of scripts (including SQL, Shell, and MapReduce scripts) have different requirements for code compiling and configuration (such as code syntax and scheduling configuration). After a script is submitted and published successfully, Dataphin creates corresponding tasks to run and produce data. The tasks are called nodes in a scheduling O&M directed acyclic graph (DAG). Core functions of code development include code file management (adding, deleting, modifying, and viewing code files), code editing, task scheduling configuration and publishing, and node version management. For details about data synchronization task DataX, see *Data import*.

18.4.7.1 Code editor

The code editor provides an online code editing interface to complete data development tasks. It supports SQL, MR, Spark, and Shell programming.

18.4.7.2 Task scheduling configuration and publishing

Scheduling configuration

Dataphin supports scheduling configuration for manual and periodic tasks. You can publish the tasks after completing scheduling configuration. The system automatically checks integrity of the scheduling configuration and allows the tasks to be published only after verifying that the scheduling configuration is complete. All the published tasks are displayed on the periodic task list on the Scheduling O&M page.

Submission and publishing

Members of a project can submit and publish tasks if they have corresponding permissions. Only the scheduling configuration with complete parameter settings, valid dependence, and no circular dependence can be submitted and published to create scheduling tasks, guaranteeing stable and orderly data production on schedule.

18.4.7.3 Code management

Dataphin supports various code operations to facilitate code file management and use. You can add, delete, update, rename, and sort code files in different folders.

File management

Dataphin allows you to edit, delete, deprecate, and rename a specific file, and check the publish state, creator, and creation time of a code file. These file management capabilities enable convenient creation, clear display, and systematic management of files in the entire big data code development process.

Folder management

When many code files are available, sort them in different folders to save and display these files orderly. You can create, rename, and delete folders, and move historical and new code files to specified folders to facilitate management. Dataphin supports hierarchical folder management.

18.4.7.4 Collaborative programming

Node version management

Dataphin supports trace back of task node versions. You can view the version number, submitter, submission time, and remarks of each task node version and check detailed code of each version to find their differences. Dataphin supports multiple node types, including MaxCompute_SQL, ODPS MR, and Shell.

Collaborative development

To allow for more efficient development through concurrent code editing by multiple developers , Dataphin provides a script locking mechanism, which prevents conflicts during collaborative development. This mechanism ensures that a line of code can be edited by only one user at a time. A user can get a lock or steal the lock of another user to obtain the code editing permission . The user whose lock is stolen can also obtain the lock theft information and take appropriate actions.

18.4.8 Resource and function management

Resource and function management is an important auxiliary function for code development. Data developers can upload their local resources and configure task nodes to call these resources

, so as to meet special data processing requirements. They can also complete common data processing using the built-in functions of the script programming language system supported by the computing engine. Particularly, if a data logic (such as data conversion in compliance with a business logic) needs to be processed at a high frequency but this cannot be achieved with built-in functions of the system, developers can define custom functions based on the resources they uploaded.

18.4.8.1 Resource management

Dataphin allows data developers of a project to add, modify, or perform other operations on resources in the project. You can use the create and edit functions to name and upload the resource files, and then copy them to or reference them into the code. You can also manually delete unnecessary resource files.

Create and upload resource files

The following types of local resource files can be uploaded. New file types can be quickly added in three days by using the standard interfaces. Each resource name is unique within a project. The file name and resource package cannot be changed after a resource file is successfully submitted . Only one resource file can be uploaded each time, and the type of the uploaded file must be the same as the selected file type.

Reference resources

You can click Copy and Reference to copy and paste a selected resource to a specific position in the code editing box, and write a statement to call this resource.

Update resources

You can update descriptions of managed resources and delete existing resources to release storage space.

18.4.8.2 Function management

The function management module allows you to find, use, and manage functions. Functions are classified into two types: default built-in functions of the system and custom functions defined based on uploaded resources such as JAR and Python packages. Custom functions can be extended by referencing standard functions.

Create a custom function

Each custom function must have a unique name within a project and cannot be renamed after being registered.

Referencing functions

You can click Copy and Reference to copy and paste a built-in or custom function name to a specific position in the code editing box, and write a statement in the format of the sample command to process this function.

Update functions

You can update custom functions by editing their information (except their names) and delete unnecessary custom functions.

18.4.9 Scheduling O&M

The scheduling O&M sub-product is used for routine maintenance and control in the late stage of data research and development. It provides a list of all data processing tasks (periodic and manual tasks), task dependence DAGs, list of instances with running tasks (periodic, manual, and data population tasks), and dependence and state DAGs of running instances. You can use this sub-product to set the task execution sequence, split processes, achieve optimal distribution of machine resources, and discover abnormal tasks, ensuring that all the tasks can be stably and reliably executed on schedule. It also reports exceptions during task execution to ensure that exceptions can be handled in time. The scheduling O&M sub-product consists of two function modules: task list and task O&M.

18.4.9.1 Task list

The task list module provides lists of periodic and manual tasks and task dependence DAGs in different projects.

18.4.9.2 Periodic tasks

For periodic tasks, you can view the task list, search for specific tasks, and view dependence of a single task. You can switch between different projects to view and search for tasks of specific projects, or perform a fuzzy match by task node name or node ID. Dataphin supports secondary filtering of period tasks by task nodes of a specified user and task nodes published today, helping to narrow down the scope of tasks or locate tasks accurately for task O&M.

18.4.9.3 Manual tasks

For manual tasks, you can view task lists, search for specific tasks, and view details about a single task. You can switch between different projects to view and search for tasks of specific projects, or perform a fuzzy match by task node name or node ID. Dataphin supports secondary filtering of period tasks by task nodes of a specified user and task nodes published today, helping to narrow down the scope of tasks or locate tasks accurately for task O&M.

18.4.9.4 Instance O&M

The instance O&M module provides lists of periodic task instances, manual task instances, and data population instances in different projects, and details about task instance operation.

18.4.9.5 Periodic instances

For periodic instances, you can view instance lists, search for specific instances, and view details about a single instance. You can view the running states of all common instances, and information about a specific task, including its unique node ID, node name, owner, task start time, end time, and duration. In addition, you can switch between different projects to view and search for task instances of specific projects, or perform a fuzzy match by task node name or node ID. Dataphin supports secondary filtering of periodic instances by my instances, abnormal instances, unfinished nodes and task execution time, helping to narrow down the scope of instances or locate instances accurately for instance O&M.

18.4.9.6 Manual instances

For periodic instances, you can view instance lists, search for specific instances, and view details about a single instance. You can view the running states of all manual instances, and information about a specific task, including its unique node ID, node name, owner, task start time, end time, and duration. In addition, you can switch between different projects to view and search for task instances of specific projects, or perform a fuzzy match by task node name or node ID. Dataphin supports secondary filtering of manual instances by my instances and instances published today, helping to narrow down the scope of instances or locate instances accurately for instance O&M.

18.4.9.7 Supplementary data instance

In a list of supplementary data instances, you can view data population task names, data population time zones and states, information about task nodes with supplementary data (including the node IDs, names, and owners), and data population duration. Search and filtering of data population instances help you find a specific instance quickly and easily.

18.4.10 Metadata center

Dataphin provides powerful metadata management capability. It can collect and extract metadata of various computing and storage engines such as MaxCompute, Hadoop, Hive, MySQL, PostgreSQL, and Oracle, and supports real-time tracing of metadata in these engines. In addition , Dataphin can abstract metadata of different storage engines to build a unified metadata model , and also allows for quick extension of various metadata. The metadata center of Dataphin provides a wide variety of metadata, follows unified standards, and ensures stable operation, delivering comprehensive metadata for data maps and data governance.

The metadata center is the core foundation of data asset management. You must determine the following items when developing the metadata center:

1. Metadata collection standard: A unified data collection standard must be used to ensure consistency of models, data tables, and Data lineage dependence records, improving availability of metadata in retrieval and services.
2. Metadata timeliness and quality: The metadata output time and quality must be guaranteed to improve the speed of data application in asset management and accuracy of data retrieval by developers.
3. Metadata model system: A unified public metadata model must be built to ensure compatibility with various data and deliver one-stop service of data maps.

18.4.11 Asset analysis

The asset analysis sub-product is designed for data discovery and collection at the late stage of data development. It classifies and manages data like assets, discovers exceptions, and takes optimization measures based on the design and application principles of OneData and data asset methodology, leveraging technologies of metadata collection, extraction, analysis, management , and processing. These functions minimize the cost of data, maximize the value of data, and enable customers to boost their business with the value.

Data asset management is implemented with a series of technologies. The real-time event and subscription service enables real-time updates of tables, tasks, and other metadata. The rule engine ensures efficient and accurate judgment of data governance rules and creation of health scoring models. Dynamic log analysis supports execution of numerous production tasks every day and analysis of machine O&M logs. Graph computing supports analysis and establishment of data lineage. Onelog all-link tracing interlinks end-to-end metadata during data production , service, and consuming. The plugin metadata access and processing architecture ensures compatibility between multiple computing and storage engines. Data asset management is a set of data collection, analysis, governance, application, and operation methodologies and products developed by Alibaba based on extensive experience of massive data management, serving the whole lifecycle of data, including data creation, management, application, and destruction.

Data asset analysis involves two keywords: population and fusion. Population is a process of checking all data and establishing a data asset chart based on factors in the OneData system, including the dimensions, business processes, and correlations. Population describes data assets using a modeling language. Fusion is a process of analyzing the cost and value of data assets during production. This process describes the functions of different data sets in the asset chart based on the connection and contribution models.

The data map module of Dataphin uses the metadata profiling technology and search engine to enable efficient retrieval of an enterprise's data assets based on the data asset category established after analysis of these data assets and users' habits in use of the data.

18.4.11.1 Asset panorama

Data assets of an enterprise established based on the OneData system can be displayed in a structural chart, in which material components in different shapes represent business entities, whereas lines of different styles represent business relations between entities. This chart shows a clear panorama of data in the same business section.

18.4.11.2 Asset map

An asset map summarizes the relation between dimensions and business processes in the data domain of a business section to show the composition of an enterprise's data, corresponding to the asset panorama of the enterprise. In addition, the asset map provides an entry to efficient, fast , and accurate data search and exploration for users based on their self-initiated behaviors such as search, visits, and save as favorites.

18.4.12 Security management

As big data application extends, data security becomes an important issue. The Cybersecurity Law of the People's Republic of China took effect on June 1, 2017 to encourage development of cyber data protection and exploitation technologies. The General Data Protection Regulation issued by the EU will take effect on May 25, 2018. All these legislations aim to strengthen protection of personal information. Dataphin focuses on intelligent data creation and management and places great importance on data security management. It provides comprehensive data security protection throughout the entire lifecycle from data production to destruction, including data access control and isolation, data security classification, personal information law compliance management, data masking, and data use security auditing.

Data access control and isolation should be given a top priority in data security management. Dataphin provides well-developed data access permission application, approval, and lifecycle management functions, supports multi-tenant data access isolation and per-field permission control, and offers an ACL-based data access authorization model.

Dataphin establishes a comprehensive data security guarantee system covering the entire lifecycle of data. This system provides technical and management measures to protect data based on data access behaviors, data content, and data environment. During big data development and management, Dataphin works with Alibaba Cloud data security management system to offer an "available but invisible" environment for secure big data exchange in addition to comprehensive security guarantee capabilities, including per-field access permission control, strict permission application approval process control, and all-round data use behavior tracing and auditing. All these guarantee data security during storage, transfer, and use of big data.

Dataphin offers a hierarchical permission control system and a full range of management processes covering the application, approval, assignment, return, and authentication of data access permissions.

18.4.12.1 Types of permission

Dataphin controls data access permissions based on user roles and resources, enabling users to use products and access data in a secure and controlled way.

Role permission

To manage operations of users on the platform in a centralized manner, Dataphin provides an account management mechanism, which controls user access on the platform based on roles of the super administrator and system members. In addition, Dataphin provides a project management mechanism to control users' permission to obtain and handle data resources on a per-project basis. These mechanisms are called role management, which assigns the permission to obtain and handle a group of data resources to a batch of users in the system under proper control.

Resource permission

Dataphin offers a data access control mechanism for centralized management of users' operations on data resources in projects. This mechanism controls access to storage resources between different projects when project spaces are managed independently and project members and resources are logically isolated. In this way, data can be shared among different project spaces without migration.

18.4.12.2 Permission management

Permission application

After data developers find a required data table on the Data Map page and view detailed metadata of this table, they must apply for the permission to use this table.

In a permission application process, Dataphin can display information about the source data table, including the table type and the home business section. Field metadata in the table is also displayed. Dataphin supports permission application following the least permission principle. It allows users to apply for the field access permission, select different validity periods for the permission (start and end dates or a period of 30 days, 90 days, 180 days, or 1 year), and enter the business scenario and purpose of the permission they apply for. Then, the approver can determine whether to assign the permission accordingly.

Application record management

Dataphin allows you to view your application records and the current state of the applications on the permission list. You can click Details to view application information and click Cancel to cancel an application. After your application is approved, you can view the accessible data list and specific fields.

Permission application approval

After a permission application is submitted, the system randomly assigns the approval ticket to an administrator of the project to which the data table belongs. The administrator can view information about the application in My Approval and determine whether to accept or reject the application.

Permission return

Users must return their permissions before shifting to another position or leaving Alibaba, to ensure that related data and production tasks can be handed over to appropriate personnel. Click Return on the My Permission page to return the permission to the project administrator. Dataphin can then reclaim the permission.

18.4.13 Ad hoc queries

The powerful OneService engine of Dataphin provides high-performance support for temporary data query and exploration. It supports both traditional simple queries and theme queries, and shows excellence in code simplicity and query processing speed.

18.4.13.1 Syntax characteristics

1. Dataphin supports offline queries on all modeled logical tables. The intelligent engine selects the optimal physical table based on factors such as the output time and query performance.
2. The associated query function based on the snowflake schema makes SQL more simple and intelligent.
3. Dataphin supports queries on physical tables, logical tables, and combination of physical tables and logical tables.
4. Dataphin supports syntax of multiple computing engines, such as MaxCompute SQL and Hive SQL.
5. Dataphin provides intelligent prompt, pre-compiling, and formatting functions for SQL.
6. Dataphin can manage permissions and authenticate users for access to fields in a logical or physical table.

18.4.13.2 Query execution

You can enter any query statements in a query script. The script editor provides intelligent prompts based on the input content, locates the required data table or field quickly, and verifies validity of the script syntax.

18.5 Scenarios

- The modeling development module can help you visually create SQL expressions to build a model. The system then automatically allocates the relevant tasks and production data. All metrics and indexes are unambiguous.
- The asset analysis module enables you to build and manage data assets and gain an overview of your business data.
- Data extraction supports customizing business master data centered on entity objects and DMP construction in three steps by customizing configuration parameters to realize ID-based identification, automatic standardization of labeling standards, and elimination of data islands.

- The Ad-Hoc query module can provide topic-oriented logical table search. The module greatly simplifies SQL queries, improves the performance of data search, and ensures that the data output to applications are normalized, standardized, and unambiguous.

18.6 Limits

None.

18.7 Concept

Business unit

Define the name and business space of the data warehouse. If the business in the scenario is retail-oriented and there is less isolation between systems, you need only create a new business unit.

Public definitions

Public definitions are used to define conceptual objects to ensure uniform global concepts and universal references.

Project management

Physical space division facilitates the isolation of physical resources and the grouping of developers. After the physical space configuration name, modeling and development can be performed.

Physical data source

Dataphin allows you to register and log out of a database by performing data source operations such as creation and modification. On the one hand, the data source serves as the source or target of data synchronization transmission. On the other hand, special data source types (such as MaxCompute) can be used for project calculation and storage after you set the calculation engine type.

Dimension

The dimension is unique in the business module and belongs to a unique data domain, which normalizes and standardizes naming and topic classification.

Business process

The business process refers to a collection of smallest-unit behaviors or transactions, such as creating an order and browsing the web page. The behavior details in the business process, such as paying an order and browsing a web page, is recorded in a fact table, which will focus on a particular business process in most cases.

Dimension logical table

A dimension logical table shows a detailed dimension logical model and contains details about dimensions. Dataphin allows you to view and manage the list of dimension logical tables, and to view and edit a specific dimension logical table on a GUI.

Fact logical table

Dataphin allows you to describe the data warehouse model of a specific procedure (for example, place an order or pay for a commodity) or state metric (such as the account balance and inventory) attribute using a fact logical table. Fact logical tables are created in an optimized semi-snowflake schema, which allows attributes other than metrics and associated dimensions to be degraded into fact attributes in a fact table and be categorized. This reduces complexity of the model design and makes it more user friendly.

Atomic index

The atomic index is an abstraction of the index statistics caliber and specific algorithms. To eliminate definition and R&D inconsistency, Dataphin innovatively puts forward the concept of " Design is development". When an index is defined, the statistical caliber (that is, calculation logic) is defined. Secondary or repeated development of ETL is not required, which increases the R &D efficiency and ensures the consistency of statistical results. According to the complexity of calculation logic, Dataphin classifies atomic indexes into two types: raw atomic indexes, such as payment amount, and derivative atomic indexes, which are constructed based on combined atomic indexes. For example, the customer price is the payment amount divided by the number of buyers.

Business limitation

The atomic index is the standardized definition of the calculation logic, and the business limitation is the standardized definition of condition restrictions. Similar to the atomic index, to ensure that all the statistical indexes are constructed in a unified and standard manner, the business limitation is unique in the business module and uniquely belongs to a source logic table. The calculation logic uses the fields of the source logic table model as the definition criterion. Each business limitation tracks its own type by using all the logical tables related to the source logic table model, so it may belong to multiple data domains, which achieves normalized naming and logics as well as standardized and systematic topic classification.

Derived indicators

Derived indicators are commonly used statistical indicators. To create statistical indicators in a standard, regular, and unambiguous way, the OneData methodology is abstracted into the following parts:

- Atomic indicator: Statistical criterion, namely, the computing logic.
- Business limitation: The business scope for statistics collection, used to find the records complying with business rules.
- Statistical period: A time period during which statistics are collected, for example, the latest 1 or 30 days.
- Statistics granularity: A statistical object or perspective that defines the extent of data summarization. It can be considered as a grouping condition for aggregation computing (group by clauses in SQL statements). Granularity is a combination of dimensions. For example, if a seller's turnover in a province is used as a statistical indicator, the statistics granularity is the combination of the seller and region dimensions.