ALIBABA CLOUD

Alibaba Cloud

Apsara Stack Enterprise

Product Introduction

Product Version: 2006, Internal: V3.12.0

Document Version: 20210915

Alibaba Cloud

Legal disclaimer

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

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.

2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company or individual in any form or by any means without the prior written consent of Alibaba Cloud.

3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should 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 this document based on the "status quo", "being defective", and "existing functions" of its products and services. 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 take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, 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 directly contact Alibaba Cloud for any errors of this document.

Document conventions

Style	Description	Example
Danger	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	Danger: Resetting will result in the loss of user configuration data.
Warning	A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	Warning: Restarting will cause business interruption. About 10 minutes are required to restart an instance.
Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	Notice: If the weight is set to 0, the server no longer receives new requests.
Note	A note indicates supplemental instructions, best practices, tips, and other content.	Note: You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click **Settings**> **Network**> **Set network type**.
Bold	Bold formatting is used for buttons , menus, page names, and other UI elements.	Click **OK**.
`Courier font`	Courier font is used for commands	Run the `cd /d C:/window` command to enter the Windows system folder.
Italic	Italic formatting is used for parameters and variables.	`bae log list --instanceid` *Instance_ID*
[] or [a|b]	This format is used for an optional value, where only one item can be selected.	`ipconfig [-all|-t]`
{} or {a|b}	This format is used for a required value, where only one item can be selected.	`switch {active|stand}`

Table of Contents

1.Introduction to Alibaba Cloud Apsara Stack

1.1. What is Apsara Stack?

Apsara Stack is an open, unified, and trusted cloud platform tailored for enterprise customers. It is developed based on the same distributed architecture as Alibaba Cloud public cloud. Enterprises customers can deploy public cloud products and services in on-premises environments, expand capabilities to the public cloud with ease, and enjoy hybrid cloud services anytime and anywhere.

Private clouds

A private cloud is a cloud computing system that is deployed on premises for government and enterprise customers by cloud computing service providers. Cloud infrastructure, software, and hardware resources are deployed in the private cloud behind a firewall to allow internal users of enterprises to share resources within their data centers. The private cloud can be managed by the enterprise itself or by a third party, and located within or outside the enterprise. Private clouds provide better privacy and exclusivity than public clouds.

Private clouds are divided into the following types based on enterprise scale or business requirements:

- Multi-tenant comprehensive private clouds for industries and large groups: an end-to-end cloud system created in a top-down manner. The system is designed to drive hyper-scale digital applications and meet IT requirements such as the continuous integration and development of DevOps applications and the operation support of production environments.
- Single-tenant basic private clouds for small and medium-sized enterprises and scenarios: a cloud system that can perform local computing tasks and host technical systems such as large-scale Software as a Service (SaaS) applications, industrial clouds, and large group clouds.

Apsara Stack

More and more enterprises migrate their IT infrastructure to the cloud, and they must consider construction requirements such as security compliance, reuse of existing data centers, and the benefits of a collocated data center. Some enterprises may prefer to use their own data centers but want to deliver a service experience that relies on large-scale cloud computing.

Apsara Stack is an extension of Alibaba Cloud public cloud that brings public cloud technology to private clouds. Apsara Stack delivers complete and customizable Alibaba Cloud software solutions and allows enterprises to experience the same hyper-scale cloud computing and big data products provided by Alibaba Cloud public cloud within their own data centers. Apsara Stack also provides enterprises with a consistent hybrid cloud experience where you can obtain IT resources and ensure business continuity.

Apsara Stack supports local deployment and can be independently run and managed outside Alibaba Cloud.

Benefits

Apsara Stack helps government and enterprise customers digitally transform their businesses and services based on a variety of products, services, and the digitalization practices of Alibaba Group in combination with mature solutions and developed from experience. Apsara Stack provides the following benefits:

- Elastic

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

- Agile

  Uses Internet and microservice integration to accelerate innovation.

- Digital

  Allows data to flow vertically between businesses and forms a mid-end to handle large amounts of data.

- Smart

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

Platform features

Apsara Stack is an enterprise-level cloud platform. It has the following features:

- Software-defined platform: masks underlying hardware differences, enables resources to scale up or out, and does not affect the performance of upper-layer applications.
- Production-level reliability and security compliance: ensures the continuity and security of enterprise data.
- Centralized access management: isolates permissions of different roles to facilitate subsequent O&M management.

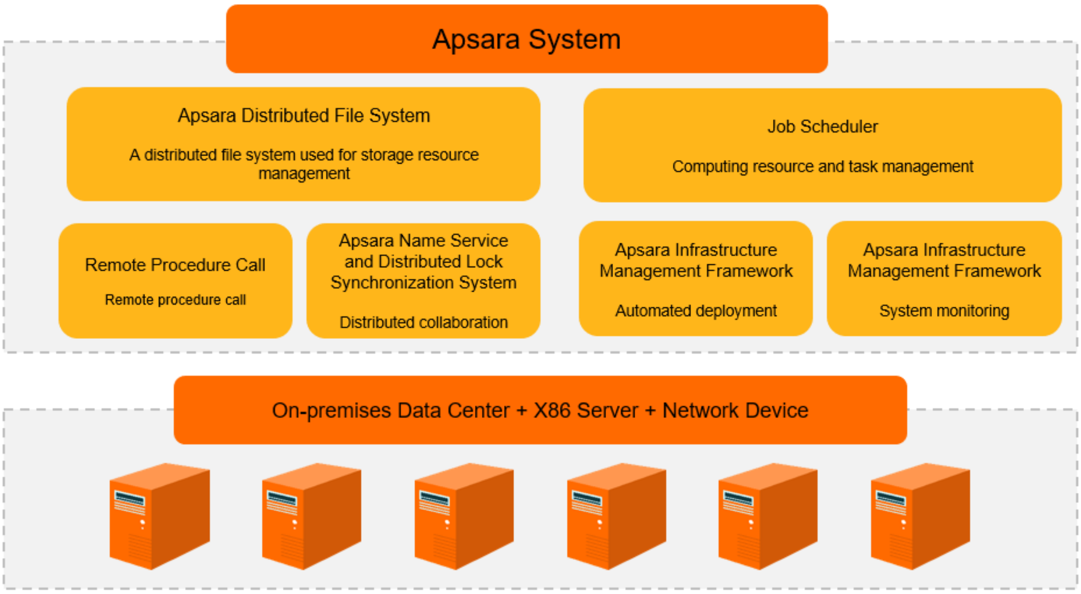
1.2. The reasons to choose Apsara Stack

1.2.1. Hyper-scale distributed cloud operating system

Both Apsara Stack and Alibaba Cloud public cloud are based on the Apsara system. The Apsara system provides underlying services such as storage, computing, and scheduling services for upper-layer services.

The Apsara system is a hyper-scale universal operating system developed by Alibaba Cloud for use both inside and outside China. It connects millions of servers around the world to act as a supercomputer. This provides powerful, universal, and accessible computing capabilities as online public services.

Apsara system kernel architecture


Apsara System
Apsara Distributed File System
A distributed file system used for storage resource management
Job Scheduler
Computing resource and task management
Remote Procedure Call
Remote procedure call
Apsara Name Service and Distributed Lock Synchronization System
Distributed collaboration
Apsara Infrastructure Management Framework
Automated deployment
Apsara Infrastructure Management Framework
System monitoring
On-premises Data Center + X86 Server + Network Device

The Apsara system kernel consists of the following modules:

- Underlying services for distributed systems

  This module provides coordination, remote procedure call, security management, and resource management services needed in a distributed environment. These services provide support for upper-layer modules such as the distributed file system and task scheduling module.

- Distributed file system

  This module provides a reliable and scalable service to store large amounts of data. The distributed file system aggregates the storage capabilities of each node in a cluster and automatically protects against hardware and software faults to provide uninterrupted access to data. This module also supports incremental scaling and automatic data balancing. An API similar to Portable Operating System Interface of UNIX (POSIX) is provided to access user space files. Additionally, the module supports random read/write and append write operations.

- Task scheduling

  This module schedules tasks in the cluster system and supports both online services that rely on a quick response speed and offline tasks that require high data processing throughput. The module can automatically detect faults and hot spots in the system. The module ensures stable and reliable service operations by using methods such as error retry and concurrent backup for long-tail operations.

- Cluster monitoring and deployment

  This module monitors the status of clusters as well as the running status and performance metrics of upper-layer application services, and generates alerts and records of exception events. In addition, this module provides O&M personnel with deployment and configuration management of the entire Apsara system and upper-layer applications. The module supports both the online elastic scaling of clusters and the online upgrade of application services.

1.2.2. Apsara Infrastructure Management Framework

Apsara Infrastructure Management Framework provides cloud services with underlying support capabilities such as unified deployment, verification, authorization, and control.

Apsara Infrastructure Management Framework includes modules such as deployment framework, resource library, metadatabase, authentication and authorization, API Gateway, Log Service, and control service.

- The deployment framework provides unified deployment of access platforms and manages service dependencies.
- The resource library stores the executable files of all cloud services and their dependent components.
- The authentication and authorization module provides access control capabilities for cloud services and supports multi-tenant isolation.
- API Gateway provides a centralized API management platform for cloud services.
- Log Service provides log storage, retrieval, and access for cloud services.
- The control service module monitors the basic health status of each cloud service and supports the Apsara Stack O&M system.

1.2.3. Highly reliable disaster recovery solutions

Apsara Stack provides zone-disaster recovery solutions.

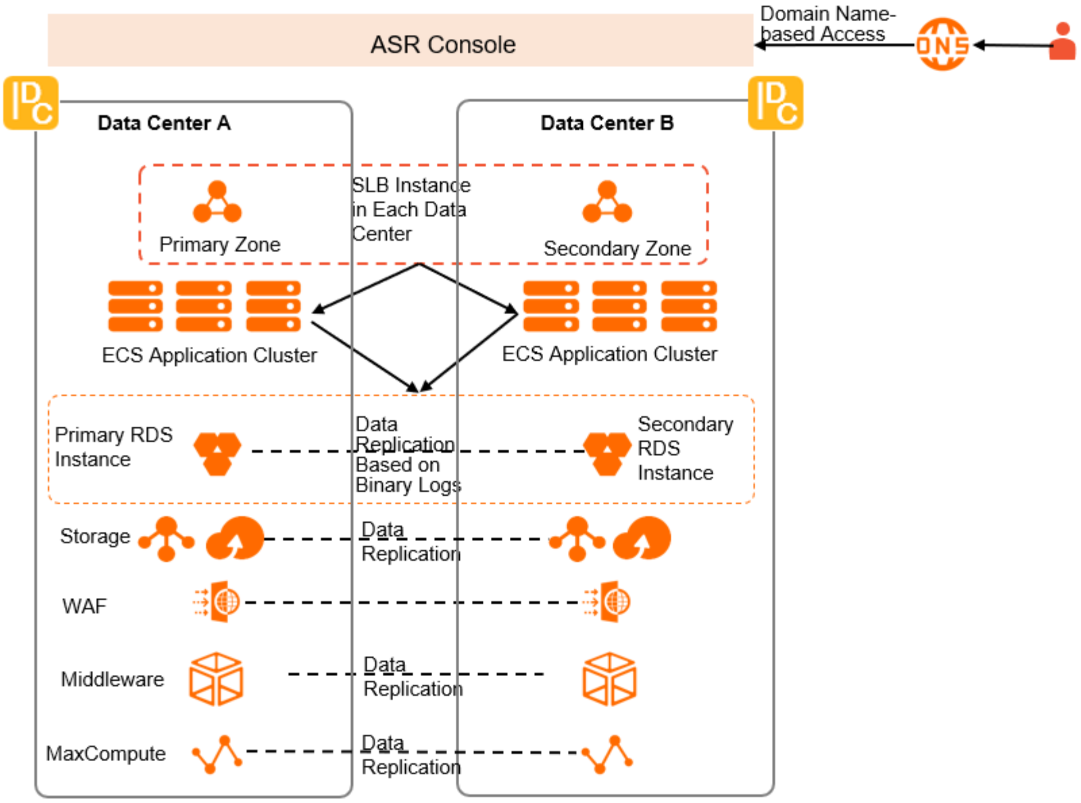
A disaster recovery system includes two or more systems that provide the same functions in distant locations. These systems mutually monitor health status and switch functions. If one system stops due to an unexpected incident (such as a fire, flood, earthquake, or vandalism), services can be switched over to the system in one location to guarantee normal system functions.

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

Zone-disaster recovery

Zone-disaster recovery refers to two independent, mutually backed up IDCs within the same zone. When an exception occurs in the active data center, the standby data center is switched online by using Apsara Stack Resilience (ASR).

Architecture of double data centers

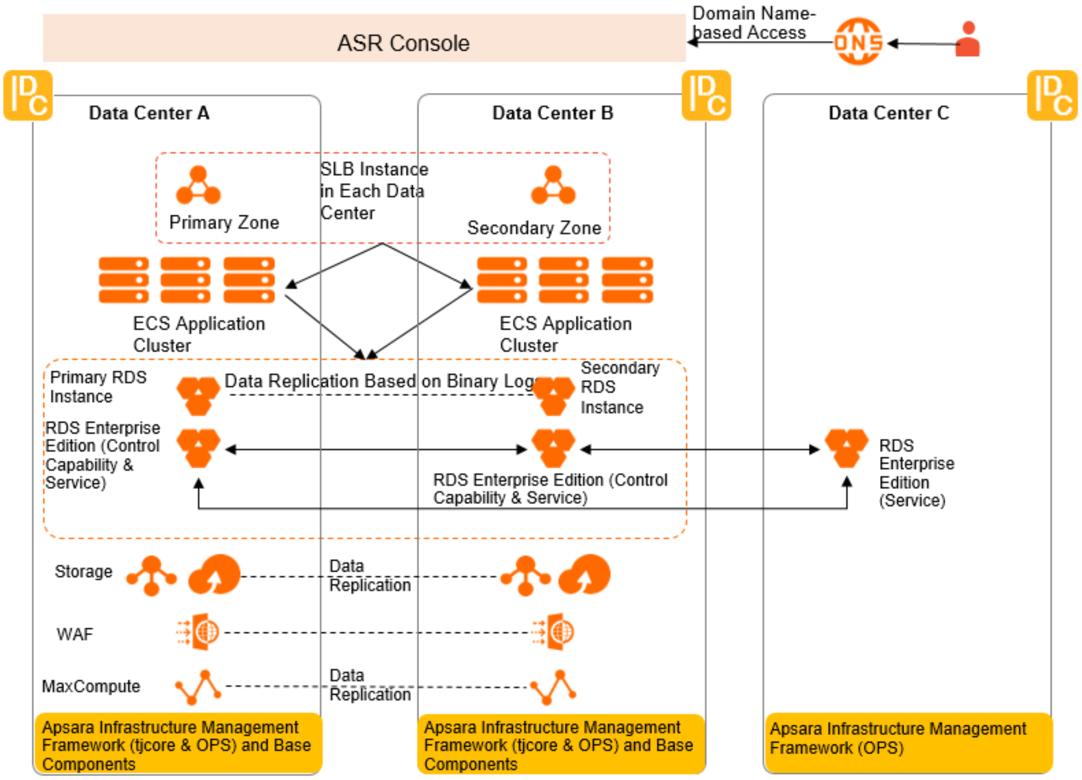


Note Data center A is the active data center. Data center B is the standby data center.

You can use domain names to connect to cloud services in the two data centers. The domain names of cloud services do not change when services are switched to the standby data center. You do not need to remodel your applications. This simplifies application development, makes cloud services easy to use, and allows you to focus on business development.

Based on the architecture of double data centers, you can add a third data center and deploy distributed databases to ensure zero data loss in the finance industry and zero recovery point objective (RPO). In the architecture of triple data centers, the disaster recovery mechanism for services in active-standby or active-active mode remains unchanged. Failovers are implemented based on the policies used in the architecture of double data centers.

Architecture of triple data centers

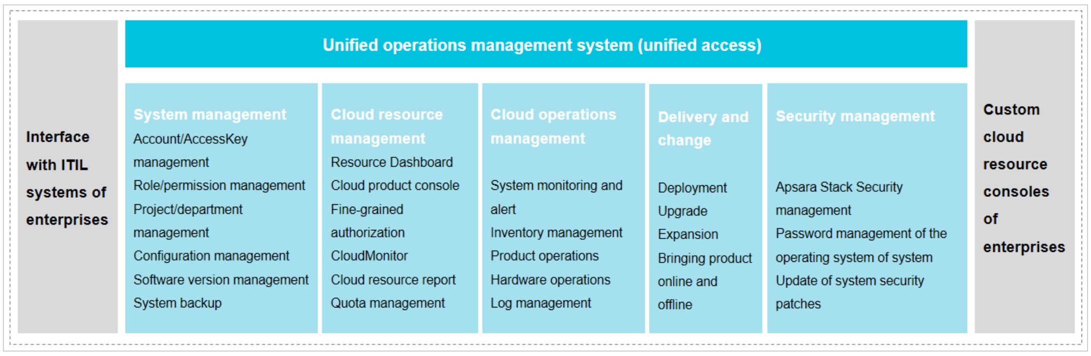

ASR Console
Domain Name-based Access
DNS
Data Center A
Data Center B
Data Center C
SLB Instance in Each Data Center
Primary Zone
Secondary Zone
ECS Application Cluster
ECS Application Cluster
Primary RDS Instance
Data Replication Based on Binary Log
Secondary RDS Instance
RDS Enterprise Edition (Control Capability & Service)
RDS Enterprise Edition (Control Capability & Service)
RDS Enterprise Edition (Service)
Storage
Data Replication
WAF
MaxCompute
Data Replication
Apsara Infrastructure Management Framework (tjcore & OPS) and Base Components
Apsara Infrastructure Management Framework (tjcore & OPS) and Base Components
Apsara Infrastructure Management Framework (OPS)

1.2.4. Centralized operation management and automated O&M capability

Apsara Stack provides a centralized management portal. You can configure different management permissions for different roles.

OpenAPI enables you to manage O&M tasks and customize your cloud resource console. Apsara Stack can be synchronized and integrated with the existing Information Technology Infrastructure Library (ITIL) systems of enterprises.

Centralized O&M management



Interface with ITIL systems of enterprises

Unified operations management system (unified access)

System management	Cloud resource management	Cloud operations management	Delivery and change	Security management
Account/AccessKey management Role/permission management Project/department management Configuration management Software version management System backup	Resource Dashboard Cloud product console Fine-grained authorization CloudMonitor Cloud resource report Quota management	System monitoring and alert Inventory management Product operations Hardware operations Log management	Deployment Upgrade Expansion Bringing product online and offline	Apsara Stack Security management Password management of the operating system of system Update of system security patches

Custom cloud resource consoles of enterprises

1.2.5. OpenAPI

Apsara Stack provides a wide range of SDKs and RESTful APIs on the OpenAPI platform.

OpenAPI provides flexible access to a variety of Apsara Stack services. You can also use OpenAPI to obtain the basic control information of Apsara Stack and integrate Apsara Stack with your centralized control system.

1.3. Product architecture

1.3.1. Types of private cloud architectures

This topic describes the following types of private cloud architectures: cloud-native architecture and integrated cloud architecture.

- Cloud-native architecture

  The cloud-native architecture is derived from Internet-based open architecture. Based on a distributed system framework, the cloud-native architecture was originally used for big data and web applications and later used to provide a range of basic services.

- Integrated cloud architecture

  The integrated cloud architecture focuses on virtualized computing services. Integrated cloud architecture is a breakthrough from traditional computing architecture developed by OpenStack and has become the most popular choice for private cloud architecture.

Alibaba Cloud adopts the cloud-native architecture and is built based on self-developed operating systems, distributed technologies and products of Alibaba Cloud. Apsara Stack uses a single architecture for a variety of deployment environments to support all cloud products. The architecture provides a full set of enterprise-class services, accessibility, disaster recovery capabilities, and self-developed and controllable capabilities.

1.3.2. System architecture

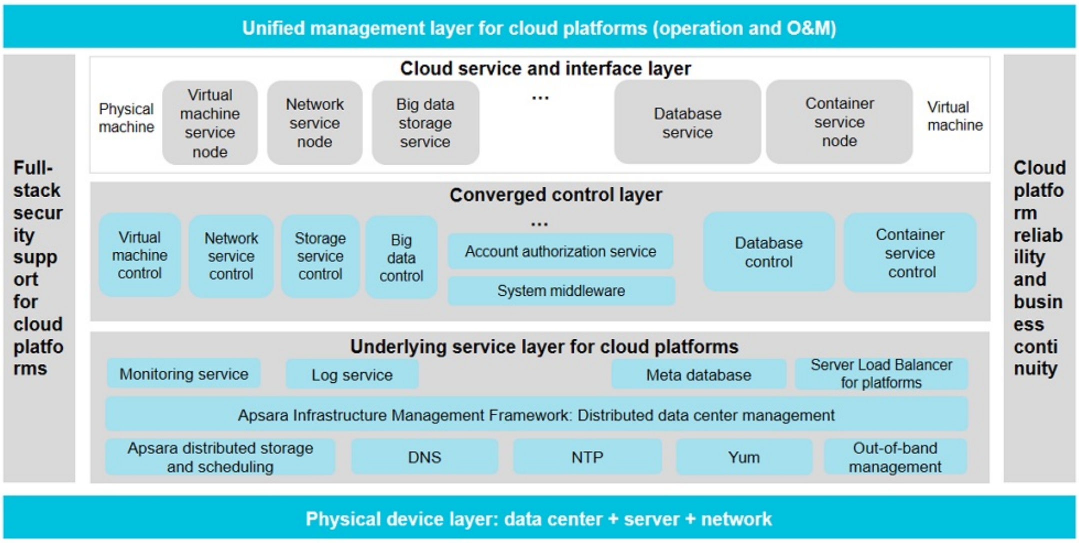
This topic describes the system architecture, logical architecture, and functions of each module.

As shown in System architecture of Apsara Stack, the system architecture of Apsara Stack consists of the following layers:

- Physical device layer: includes hardware devices for cloud computing, such as physical data centers, servers, and network.
- Underlying service layer: provides services for upper-layer applications based on the underlying physical environment.
- Converged control layer: provides unified scheduling for upper-layer applications or services by using the converged control architecture.
- Cloud service and interface layer: provides centralized management and O&M for virtual machines and physical machines through converged service node management, and uses the OpenAPI platform to provide centralized API management and support custom development.
- Centralized management layer for cloud platforms: provides centralized O&M management.

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

System architecture of Apsara Stack

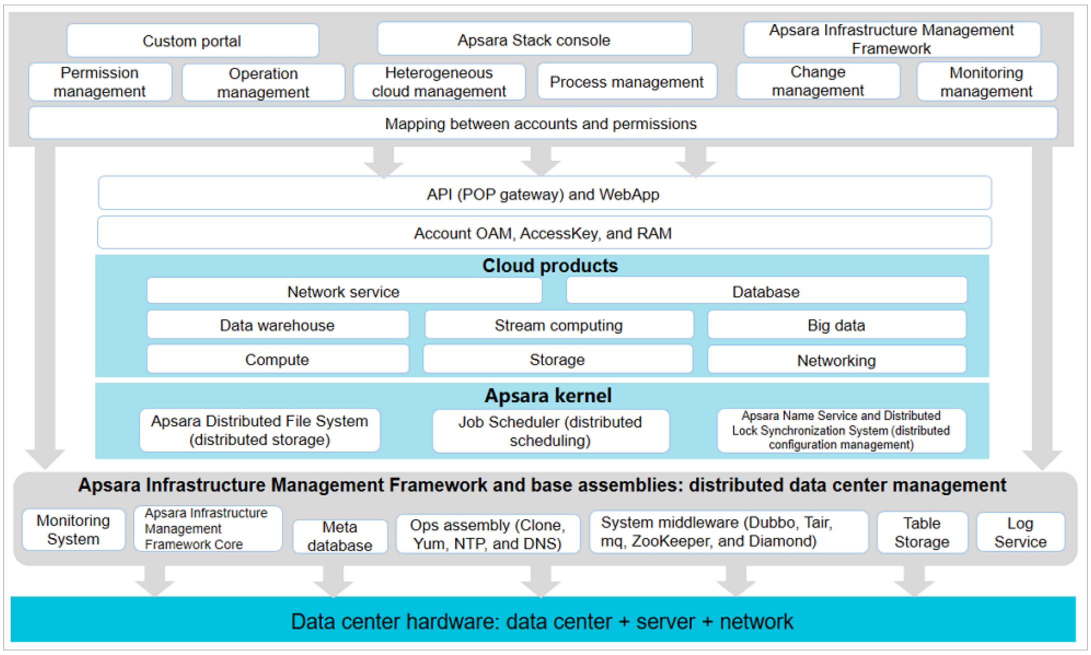

Unified management layer for cloud platforms (operation and O&M)
Full-stack security support for cloud platforms
Cloud service and interface layer
Physical machine
Virtual machine service node
Network service node
Big data storage service
...
Database service
Container service node
Virtual machine
Converged control layer
Virtual machine control
Network service control
Storage service control
Big data control
...
Account authorization service
System middleware
Database control
Container service control
Underlying service layer for cloud platforms
Monitoring service
Log service
Meta database
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
Cloud platform reliability and business continuity
Physical device layer: data center + server + network

Logical architecture

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

- The hardware infrastructure consists of data centers, servers, and network devices.
- A variety of cloud services are provided based on the Apsara kernel (distributed engine).
- All cloud services are required to comply with a unified API framework, security system, and O&M and management system (accounts, authorization, monitoring, and logs).
- A consistent user experience is guaranteed across all services.

Logical architecture of Apsara Stack


Custom portal
Apsara Stack 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
Database
Data warehouse
Stream computing
Big data
Compute
Storage
Networking
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
Monitoring System
Apsara Infrastructure Management Framework Core
Meta database
Ops assembly (Clone, Yum, NTP, and DNS)
System middleware (Dubbo, Tair, mq, ZooKeeper, and Diamond)
Table Storage
Log Service
Data center hardware: data center + server + network

1.3.3. Network architecture

1.3.3.1. Network architecture overview

Apsara Stack adopts the flat Clos network architecture. The network architecture isolates the service plane from the out-of-band management plane, and supports linear extension and load sharing of switches.

The network architecture contains two logical sections: business service section and integrated access section, as shown in Logical sections.

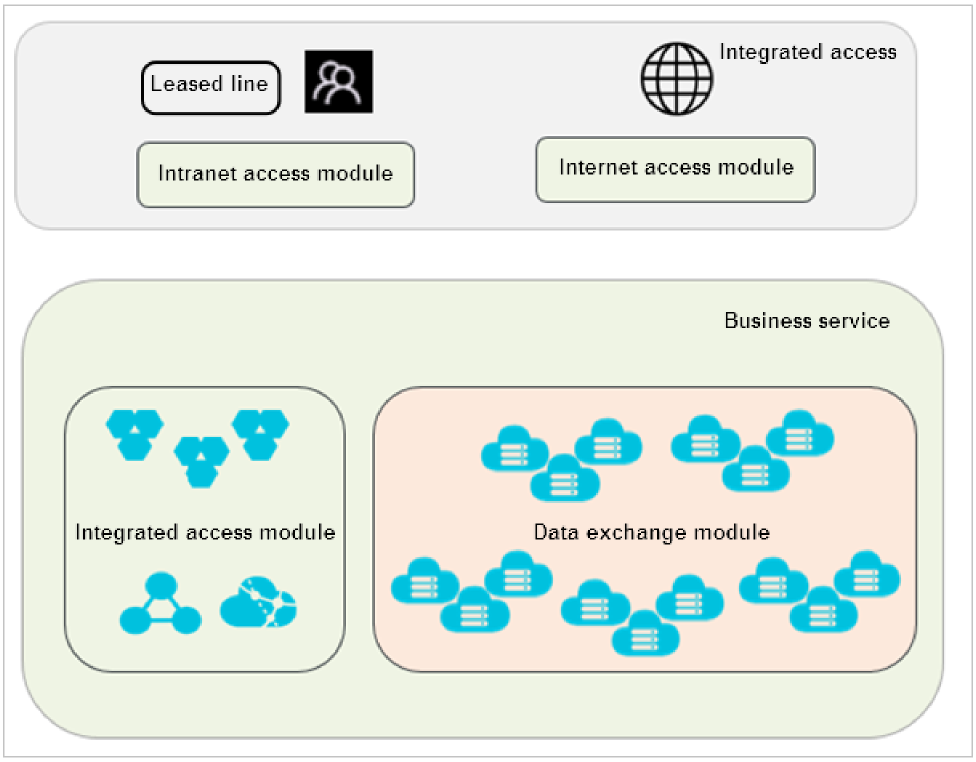
- Business service section

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

- Integrated access section

  The integrated access section is an extension of the business service section, and provides a channel for user management, and the access to Apsara Stack networks by using Internet and user networks. This section can be tailored based on actual deployment requirements.

Logical sections



The following table describes roles and functions of switches in each module.

Role	Module	Description
Inter-connection Switch (ISW)	Internet access module	ISW is an egress switch and provides access to Internet service providers (ISPs) or user backbone networks.

Role	Module	Description
Customer Access Switch (CSW)	Internal network access module	CSW facilitates access to user internal backbone networks, including access to Virtual Private Cloud (VPC) by using leased lines. It performs route distribution and interaction between the internal network and Internet.
Distribution Switch (DSW)	Data exchange module	DSW is a core switch to connect each Access Switch (ASW).
ASW	Data exchange module	ASW provides access to Elastic Compute Service (ECS) and is uplinked with the core switch DSW.
LVS Switch (LSW)	Integrated access module	LSW provides access to cloud services, such as VPC and Server Load Balancer (SLB).

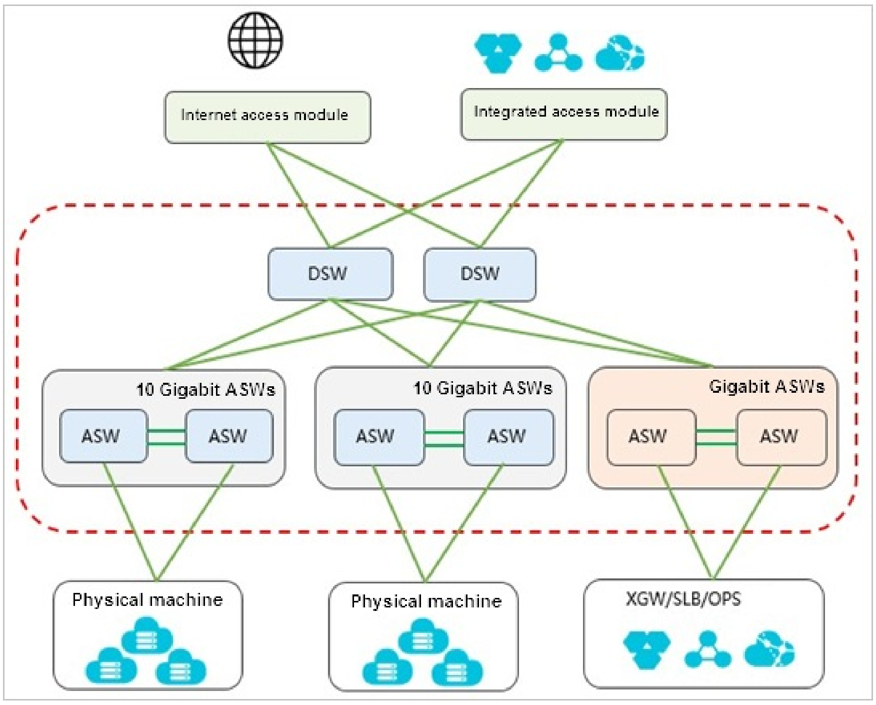
1.3.3.2. Business service section

This topic describes the data exchange module and the integrated service module that are contained in the business service section.

- Data exchange module

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

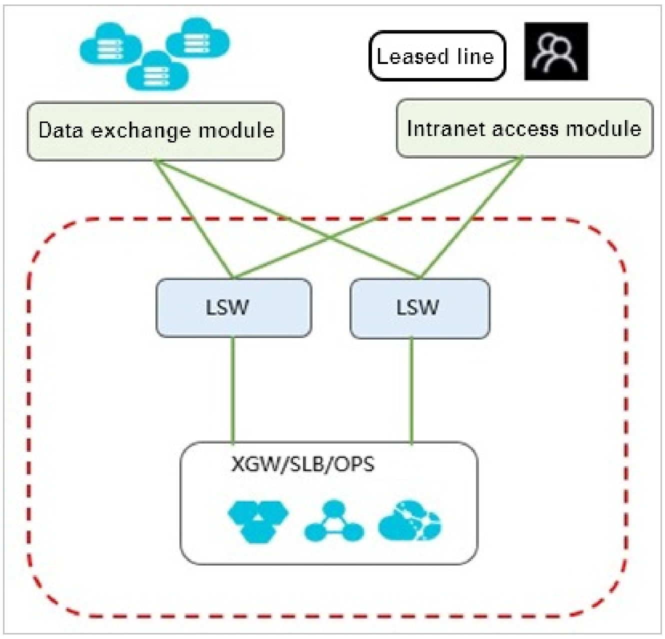
  Data exchange module



- Integrated service module

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

  Integrated service module



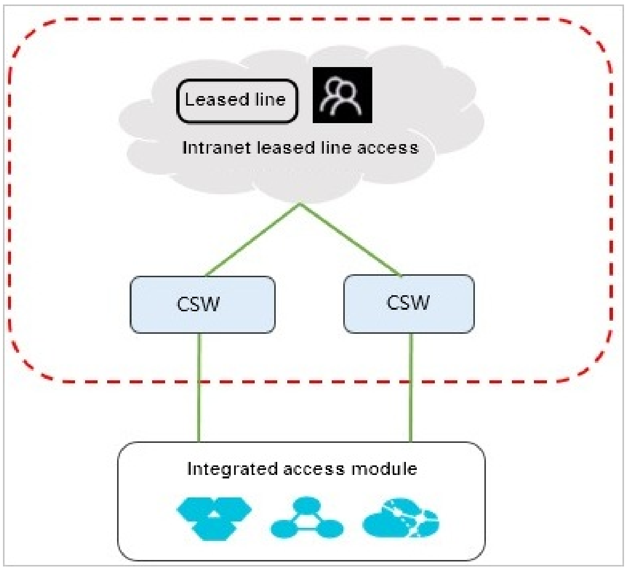
1.3.3.3. Integrated access section

This topic describes the internal network access module and Internet access module that are contained in the integrated access section.

- Internal network access module

  Two CSWs provide internal users with access to VPCs and general cloud services. CSWs establish mappings between internal users and VPCs to route these users to different VPCs. Different groups are isolated from each other on CSWs. For access to general cloud services, CSWs are connected to the integrated service module by using EBGP and allow direct access to all resources in the business service section.

  Internal network access module



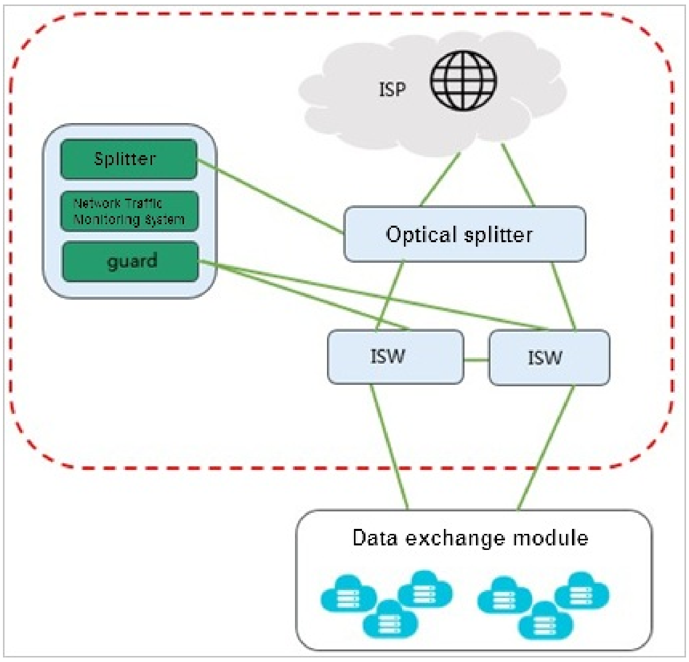
- Internet access module

  The Internet access module consists of two ISWs. This module facilitates the access to Internet service providers (ISPs) or user public backbone networks and performs route distribution and interaction between the internal network and Internet. The two ISWs use IBGP to back up routes between each other. ISWs can use static routing or EBGP to uplink with ISPs or user public backbone networks based on actual conditions. The link bandwidth is determined by the Alibaba Cloud network size of users and the bandwidth of their public backbone networks. We recommend that ISWs use BGP to connect with multiple ISPs to improve reliability. Each ISP has two 10 GE lines. The Internet access module uses EBGP to exchange routes with the data exchange module. The Internet access module releases relevant Internet routes to the data exchange module and receives internal cloud service routes that are sent by the data exchange module to implement the interaction between the internal network and Internet.

  The Internet access module is connected to the Alibaba Cloud security protection system in one-arm mode. The traffic that is transmitted from the Internet to cloud networks is diverted to Network Traffic Monitoring System by using an optical splitter. When Network Traffic Monitoring System detects malicious traffic, it releases the corresponding routes by using Apsara Stack Security to divert the malicious traffic to Apsara Stack Security for scrubbing. The scrubbed traffic is injected back into the Internet access module.

  Internet access module

Internet access module

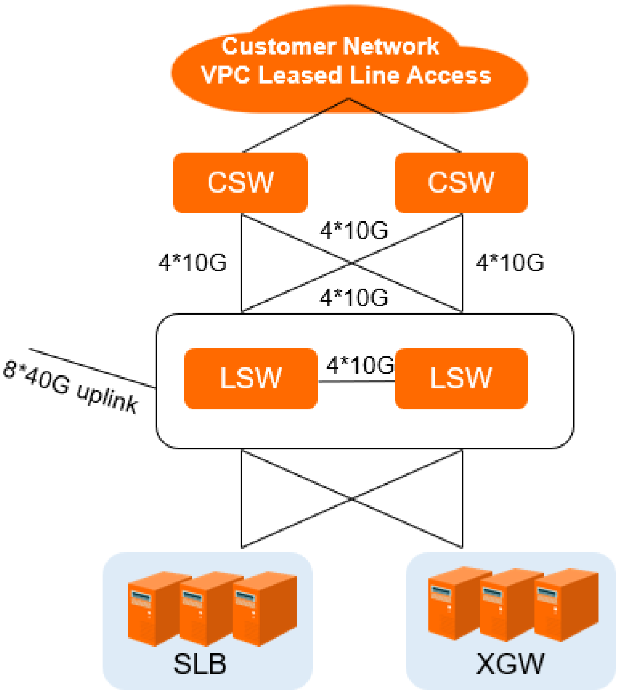


1.3.3.4. VPC leased line access

The leased line access solution of VPC allows users to control their own virtual networks, such as selecting their own CIDR blocks and configuring route tables and gateways. Users can also connect their VPC 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 or SLB) is connected to two LSWs. These servers exchange routing information by using OSPF. Two LSWs exchange routing information between each other by using IBGP. LSWs exchange routing information with CSWs by using EBGP.

VPC leased line access



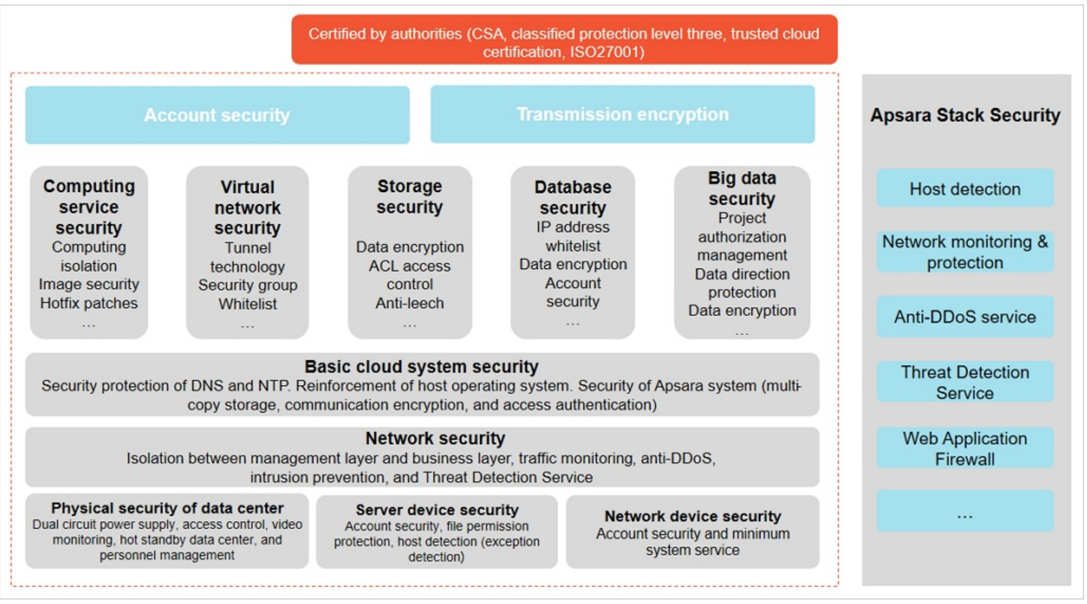
1.3.4. Security architecture

Apsara Stack provides comprehensive security capabilities from underlying communication protocols to upper-layer applications to guarantee the security of your access and data.

Access to all consoles in Apsara Stack is allowed only with HTTPS certificates. Apsara Stack provides a comprehensive role authorization mechanism to guarantee a secure and controllable access to resources in multi-tenant mode. It supports a variety of security roles, such as a security administrator, system administrator, and security auditor.

In addition, Apsara Stack V3.0 and later introduce Alibaba Cloud Security services to provide you with a multi-level and integrated cloud security solution.

Hierarchical security architecture of Apsara Stack


Certified by authorities (CSA, classified protection level three, 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
...
Basic cloud system security
Security protection of DNS and NTP. Reinforcement of host operating system. Security of Apsara system (multi-copy storage, communication encryption, and access authentication)
Network security
Isolation between management layer and business layer, traffic monitoring, anti-DDoS, intrusion prevention, and Threat Detection Service
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
Apsara Stack Security
Host detection
Network monitoring & protection
Anti-DDoS service
Threat Detection Service
Web Application Firewall
...

1.3.5. Base modules

The Apsara Stack base consists of three module types, all of which provide support for the deployment and O&M of the cloud platform.

Base modules

Module		Function description
	YUM	The installation package. Software repositories are deployed in the initial installation stage to install the operating system and deploy application packages such as the Apsara system and ECS, and dependent modules of Apsara Stack on hosts.
	Clone	The virtual machine cloning service.

Module		Function description
OPS modules	NTP	The clock source service. NTP is deployed on hosts of Apsara Stack to synchronize time from the standard NTP clock source to other hosts.
	DNS	The domain name resolution service. DNS provides forward and reverse resolution of domain names for the internal Apsara Stack environment. It runs a bind instance on each of the two OPS machines and uses keepalived to provide high availability services. When one machine fails, the other machine automatically takes over its work.
Base middleware	Dubbo	The distributed remote procedure call (RPC) service.
	Tair	The cache service.
	Message Queue (MQ)	The message queuing service.
	ZooKeeper	The distributed collaboration service.
	Diamond	The configuration management service.
	SchedulerX	The scheduled task service.
Basic modules of the base	Apsara Infrastructure Management Framework	The data center management system.
	Monitoring System	The data center monitoring system.
	OTS-inner	The table storage service.
	SLS-inner	The log service of cloud platform.
	Metadatabase	The metadatabases.
	POP	The Apsara Stack OpenAPI platform.

Module		Function description
	OAM	The account system.
	RAM	The authentication and authorization system.
	WebApps	The service that provides support for the Apsara Stack Operations (ASO) console.

1.4. Product panorama

Apsara Stack offers a wide range of services to meet the diverse needs of different users.

Infrastructure

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

Storage services

Apsara Stack provides various storage services for different storage objects. The main services include Object Storage Service (OSS), Apsara File Storage NAS, and Log Service.

Middleware and applications

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

Domain and website

Apsara Stack provides secure, fast, stable, and reliable domain name services. The main services include Alibaba Cloud DNS.

Database services

Apsara Stack provides a variety of database engines that can communicate with each other. The main services include ApsaraDB for RDS, KVStore for Redis, ApsaraDB for MongoDB, Data Transmission Service (DTS), Data Management (DMS), Tablestore, ApsaraDB for OceanBase, AnalyticDB for MySQL, AnalyticDB for PostgreSQL, and PolarDB-X.

Big data

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

Artificial intelligence

Apsara Stack provides a machine learning algorithm platform based on the distributed computing engine developed by Alibaba Cloud. The main services include Machine Learning (PAI).

Security

Apsara Stack provides comprehensive protection from underlying communication protocols to upper-layer applications to guarantee the security of your access and data. The main services include Apsara Stack Security.

1.5. Scenarios

Apsara Stack provides flexible and scalable industrial solutions for customers of different scales and sectors.

Apsara Stack can create customized solutions based on the unique business traits of different sectors such as industry, agriculture, transportation, government, finance, and education to provide users with end-to-end products and services. This topic describes the following 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. Urban data has a greater value as it has more time and larger space to flow. Cloud computing becomes urban infrastructure. Data becomes a new means of production and strategic resources. AI technology becomes the nerve center of a smart city. All of these together form the City Data Brain.

City Brain has the following values and features:

- A breakthrough of urban governance mode. City Brain uses urban data as resources to improve government management capabilities, resolve prominent issues of urban governance, and implement an intelligent, intensive, and humane form of governance.
- A breakthrough of urban service mode. City Brain provides more accurate and convenient services for enterprises and individuals, makes 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.

Alibaba Finance Cloud

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

Alibaba Finance Cloud has the following values and features:

- Independent resource clusters

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

1.6. Compliance security solution

1.6.1. Overview

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

1.6.2. Interpretation on key points

Network and communication security

Interpretation on clauses

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

Coping strategies

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

Device and computing security

Interpretation on clauses

- Avoid account sharing, record, and audit operations actions, which is an elementary security requirement.

- Secure system layer with necessary security measures and prevent servers from intrusions.

Coping strategies

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

Application and data security

Interpretation on clauses

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

Coping strategies

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

Security management policies

Interpretation on clauses

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

Coping strategies

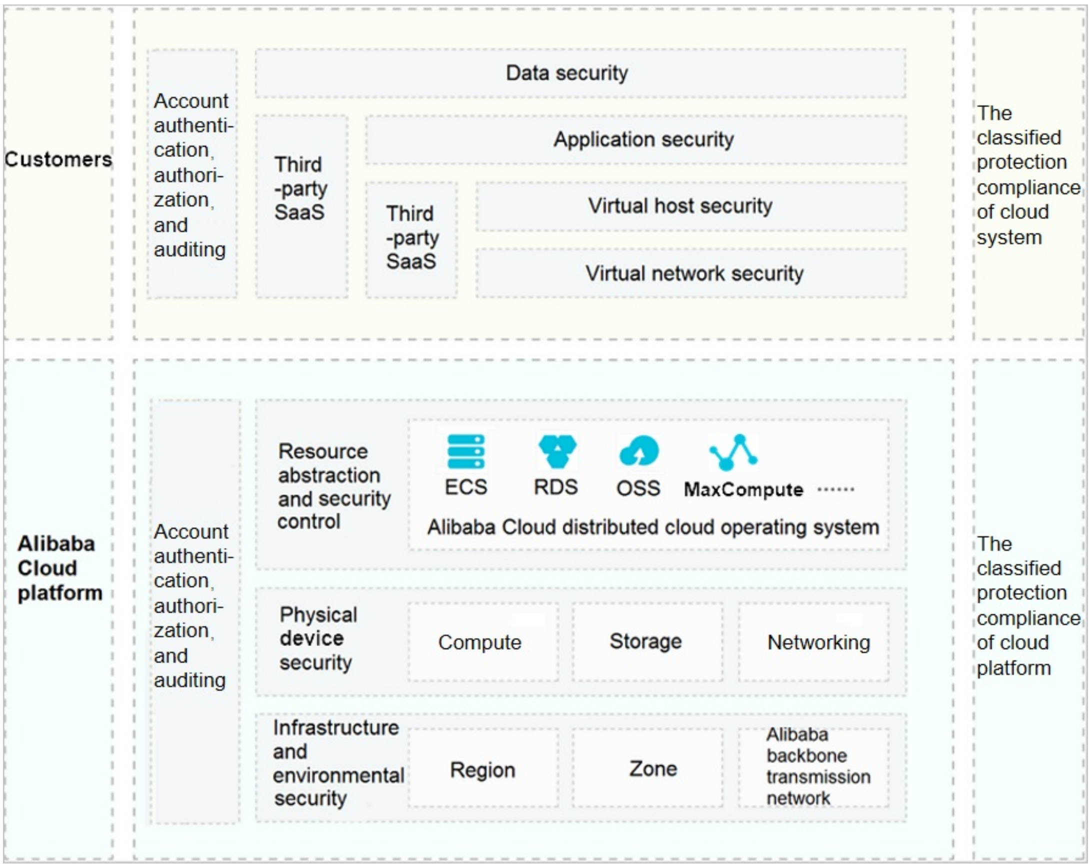
- The customer management staff can arrange, prepare, and fulfill the security policy, regulation, and management personnel according to the actual condition of enterprise and form specialized documents.
- For the technical means required in the process of vulnerability management, we recommended that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

1.6.3. Cloud-based classified protection compliance

Shared compliance responsibilities

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

Shared compliance responsibilities


Customers
Account authenti-cation, authori-zation, and auditing
Data security
Third -party SaaS
Application security
Third -party SaaS
Virtual host security
Virtual network security
The classified protection compliance of cloud system
Alibaba Cloud platform
Account authenti-cation, authori-zation, and auditing
Resource abstraction and security control
ECS
RDS
OSS
MaxCompute ......
Alibaba Cloud distributed cloud operating system
Physical device security
Compute
Storage
Networking
Infrastructure and environmental security
Region
Zone
Alibaba backbone transmission network
The classified protection compliance of cloud platform

Alibaba Cloud provides the following contents:

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

More details about shared responsibilities are as follows:

- Alibaba Cloud is the unique cloud service provider in China that participates in and passes the pilot demonstration of cloud computing classified protection standards. Public Cloud and E-Government Cloud pass the filing and assessment of the third level of classified protection. Finance Cloud passes the filing and assessment of the fourth level of classified protection.
- According to the regulatory authority, you can use the assessment conclusions of physical security, partial network security, and security management for the classified protection assessment of the tenant systems on Alibaba Cloud, and Alibaba Cloud can provide supporting details.

- With the complete security technology, management architecture, and protection system of Apsara Stack Security, Alibaba Cloud platform makes it easy for tenants to pass the classified protection assessment.

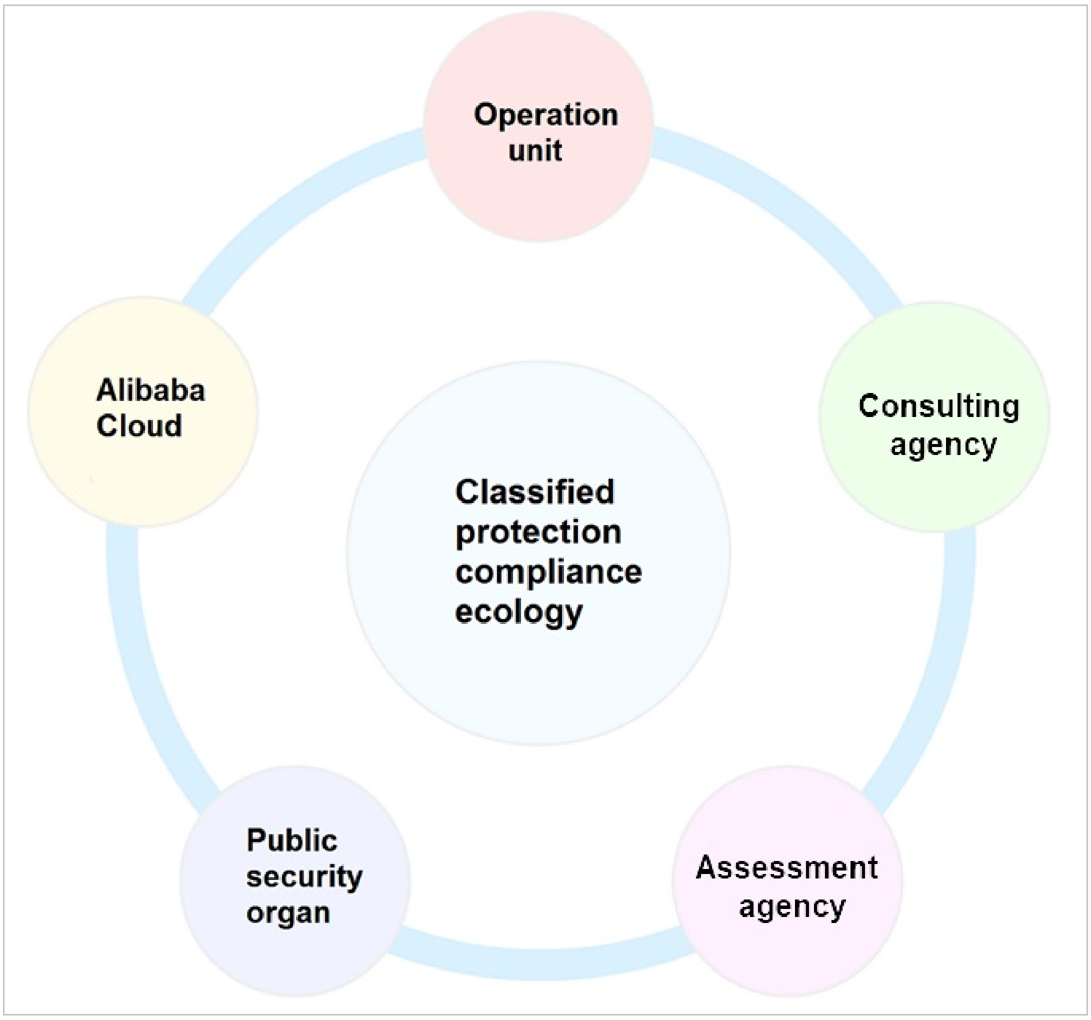
Classified protection compliance ecology

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

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

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

Classified protection compliance ecology

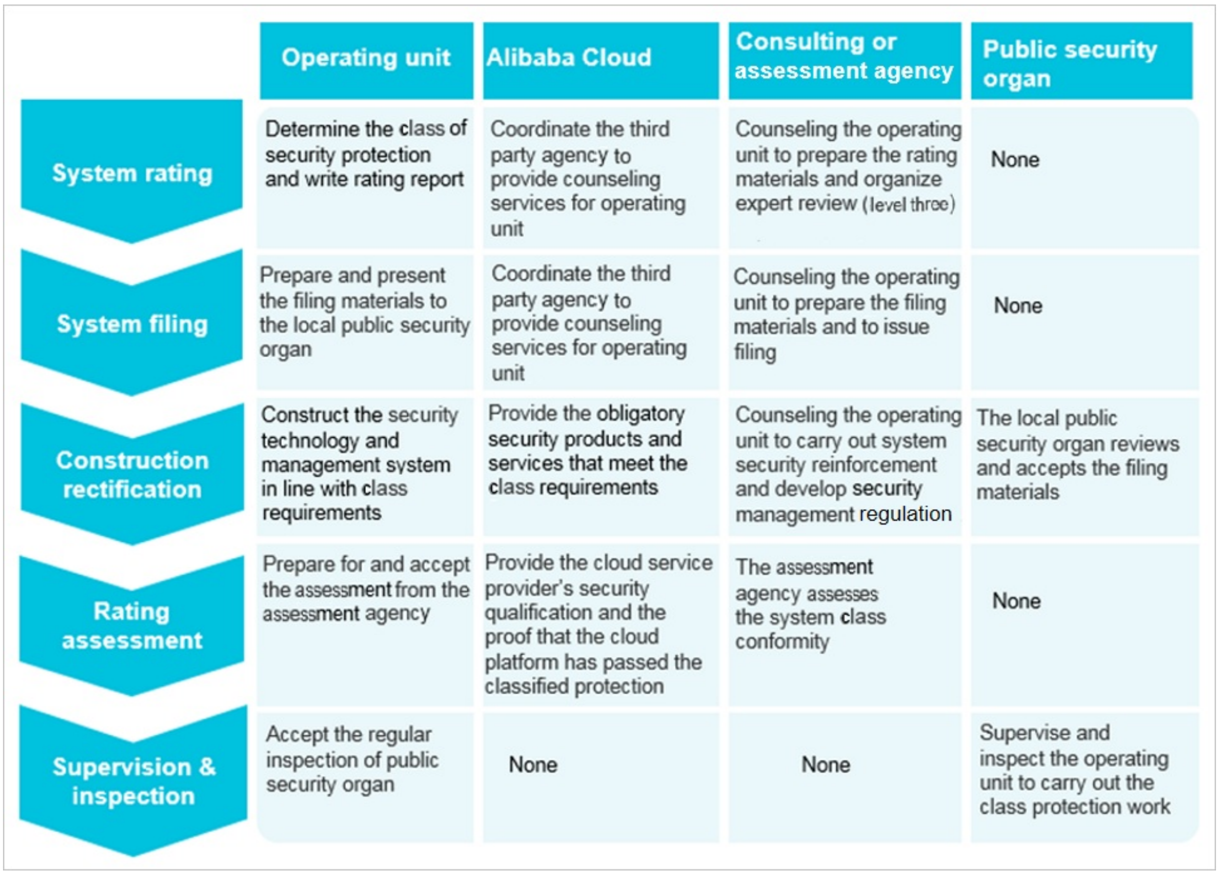


Work division of classified protection:

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

- Public security organ: reviews filing and supervises services

1.6.4. Classified protection implementation process

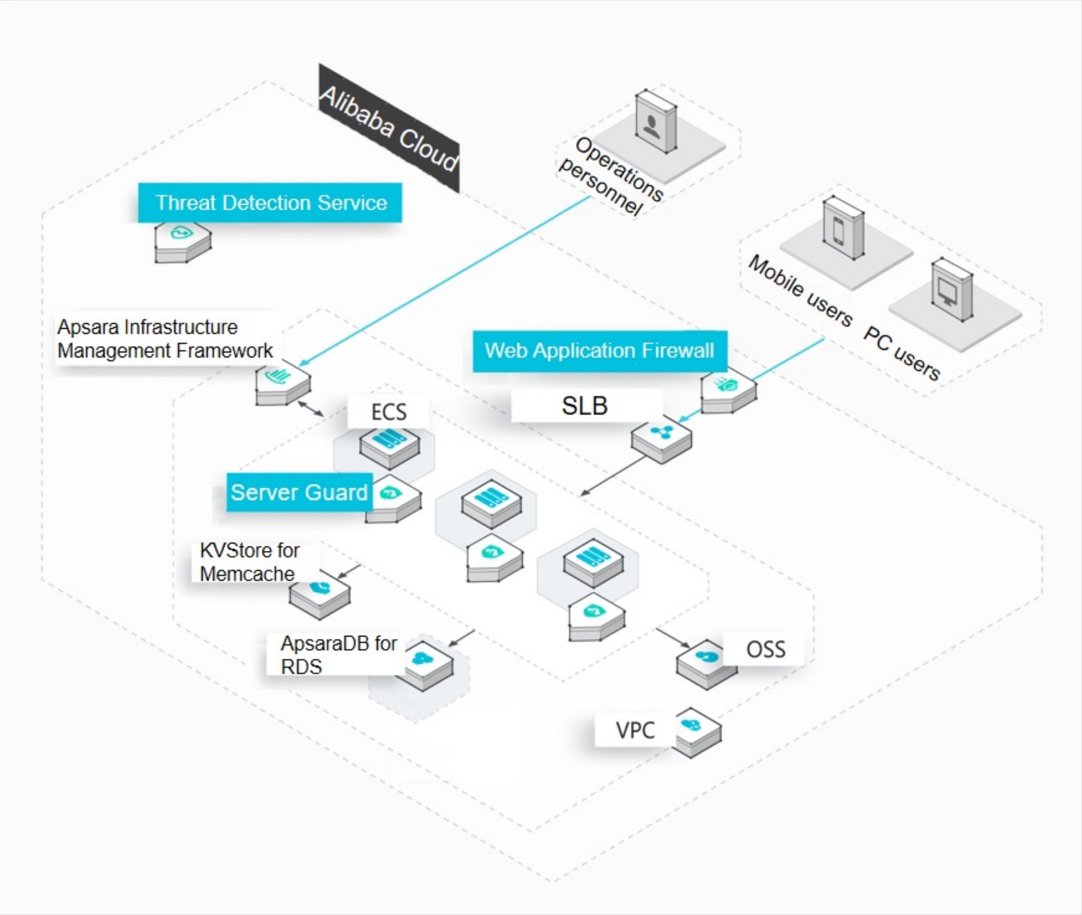


	Operating unit	Alibaba Cloud	Consulting or assessment agency	Public security organ
System rating	Determine the class of security protection and write rating report	Coordinate the third party agency to provide counseling services for operating unit	Counseling the operating unit to prepare the rating materials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

1.6.5. Security compliance architecture

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

Security compliance architecture



Basic requirements of classified protection are as follows:

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

1.6.6. Solution benefits

One-stop assessment service of classified protection

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

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

A complete security protection system

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

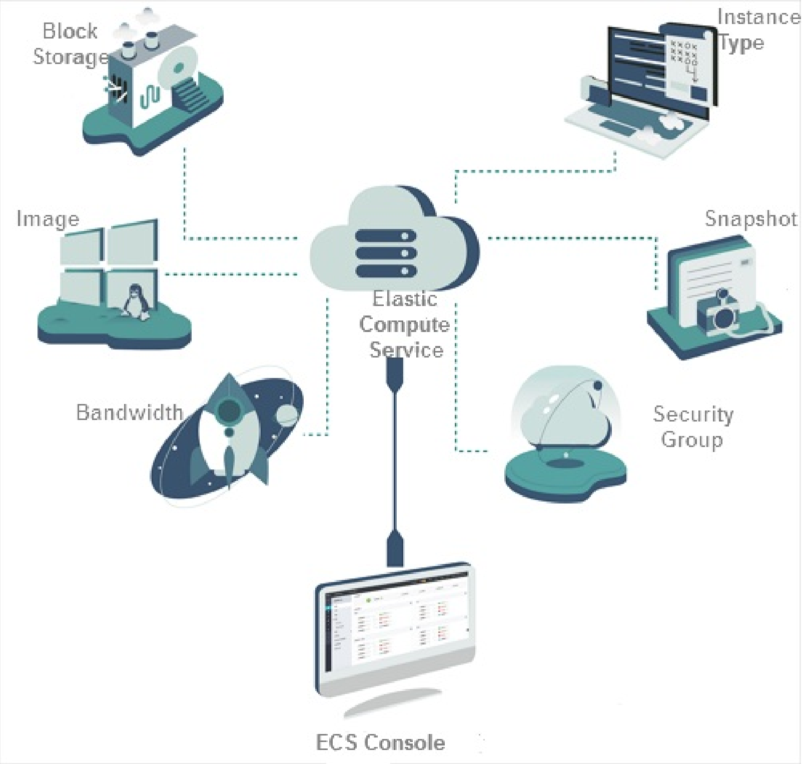
2.Elastic Compute Service (ECS)

2.1. What is ECS?

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

An ECS instance is a virtual computing environment that contains the most basic components of computers such as the CPU, memory, and storage. Users perform operations on ECS instances. Instances are core components of ECS, and operations can be performed on instances through the ECS console. Other resources, such as block storage, images, and snapshots, can only be used after they are integrated with ECS instances. For more information, see ECS components.

ECS components



2.2. Benefits

Compared with the services provided by other server vendors and common Internet data centers (IDCs), ECS provides high availability, security, and elasticity.

High availability

ECS employs stringent IDC standards, server access standards, and O&M standards to ensure data reliability and high availability at the infrastructure and instance levels.

If you want to increase availability, you can create active/standby or active/active ECS instances in multiple zones. You can build fault tolerant systems across multiple regions and zones to implement a financial-grade solution that spans three data centers across two regions. Apsara Stack provides mature solutions for fault tolerant services such as disaster recovery.

Apsara Stack provides the following support services:

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

Security

Security and stability are two of the primary concerns for any cloud service user. Alibaba Cloud has received a host of international information security certifications that demand strict confidentiality of user data and user privacy protection, including ISO 27001 and Multi-Tier Cloud Security (MTCS).

- **With a simple configuration to connect your business environment to global IDCs,** Apsara Stack Virtual Private Cloud (VPC) can increase the flexibility, scalability, and stability of your business.
- **You can connect your own IDC to Apsara Stack VPC** by using leased lines to build a hybrid cloud architecture. You can use a variety of hybrid cloud architectures to provide network services and robust networking. A superior business ecosystem is made possible with the ecosystem of Apsara Stack.
- **VPCs are more stable and secure.**
  - **Stability:** After you construct a VPC, you can update your network architecture and obtain new functions each day to constantly evolve your infrastructure and ensure the smooth operation of your business. VPCs allow you to divide, configure, and manage your network.
  - **Security:** VPCs provide traffic isolation and attack isolation to protect your services against cyber attacks. You can establish the first line of defense against malicious attacks and traffic by building your business in a VPC.

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

Elasticity

Elasticity is a key benefit of cloud computing.

- **Elastic computing**
  - **Vertical scaling**

    Vertical scaling is the process where the configurations of an ECS instance are modified. It may be difficult to change the configurations of a single server in a traditional IDC. However, you can change the configurations of your ECS instances based on the volume of your business.

- ○ Horizontal scaling

  Horizontal scaling allows the scaling of resources for applications. A traditional IDC may not be able to immediately provide sufficient resources for online gaming or live video streaming applications during peak hours. The elasticity of cloud computing makes it possible to provide the resources required during peak hours. When the load returns to normal levels, you can release unnecessary resources to reduce operation costs.

  The combination of ECS vertical and horizontal elasticity and Auto Scaling enables you to scale resources up and down by specified quantities as scheduled or against business loads.

- Elastic storage

  Apsara Stack provides elastic storage. In a traditional IDC, you must add servers to increase the storage space. However, the number of servers that you can add is limited. Apsara Stack provides large-capacity storage. You can purchase storage products such as cloud disks at any time.

- Elastic network

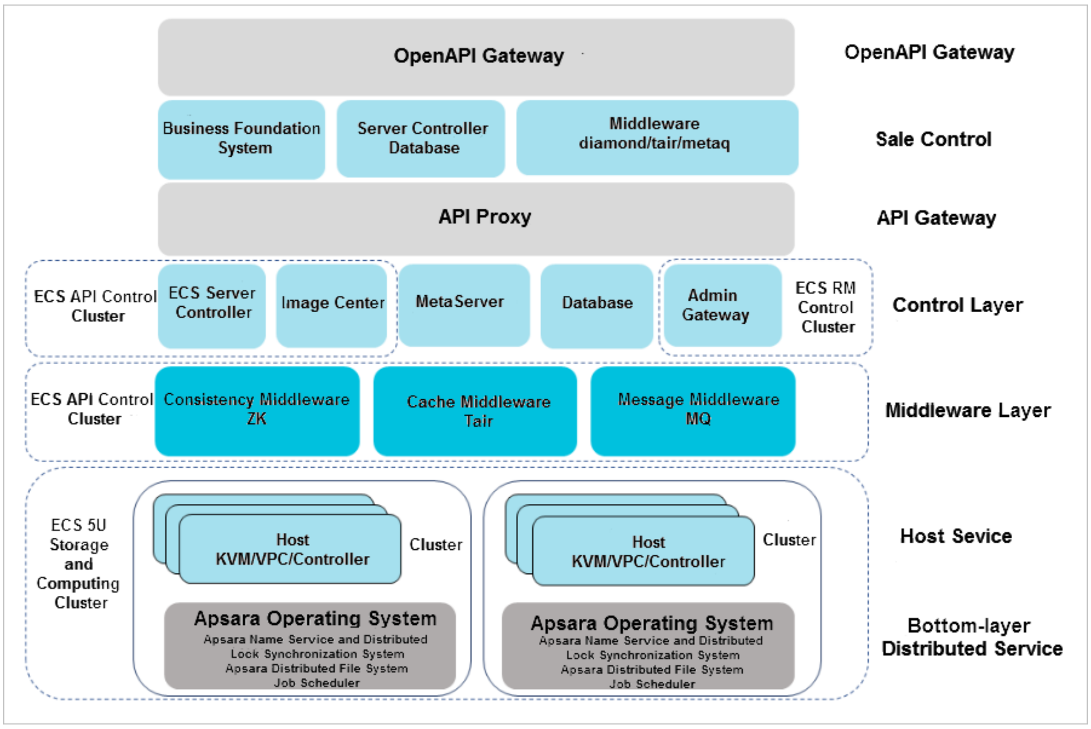
  Apsara Stack provides network elasticity. When you purchase Apsara Stack VPCs, you can configure the VPCs in the same way as your IDCs. In addition, VPCs have the following benefits: interconnection between data centers, separate secure domains in data centers, and flexible network configurations and planning within a VPC.

In conclusion, Apsara Stack provides elastic computing, storage, networking as well as business architecture planning. By using Apsara Stack, you can build your business portfolio in any way.

2.3. Architecture

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

ECS architecture



Architecture description

Component	Description
Apsara Name Service and Distributed Lock Synchronization System	A basic module that provides services related to distributed consensus in Apsara Stack. As a key distributed coordination system of Apsara Stack, this module provides three types of basic services: distributed lock services, subscription and notification services, and lightweight metadata storage services.
Apsara Distributed File System	A distributed storage system developed by Alibaba Cloud. As of 2017, hundreds of clusters and hundreds of thousands of storage nodes using Apsara Distributed File System have been deployed in the production environments. Apsara Distributed File System manages tens of exabytes of disk space.
Job Scheduler	A distributed resource scheduler that manages and allocates resources in the distributed systems.
Server Controller	The ECS scheduling system that schedules storage, network, and computing resources in a unified manner and produces virtual machines (ECS instances) that are deliverable to users.
Scheduling process	API > Business layer > Control system > Host service.
OpenAPI Gateway	Provides services such as authentication and request forwarding.
Business Foundation System	Creates and releases instances and snapshot policies, processes sales requests, and provides APIs to users.
API Proxy	Forwards requests to services that are deployed in the region specified by region_id.
Server Controller database	Stores control data and status data.
Tair	Provides cache services for Server Controller.
Zookeeper	Provides the distributed lock service for Server Controller.
MQ	Provides message queuing services of virtual machine statuses.
Image Center	Provides image management services such as import and copy.
MetaSever	Provides metadata management services for ECS instances.
Host service	Provides services such as KVM for computing virtualization, VPC for network virtualization, and control through interaction with Libivrt.

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

2.4. Features

2.4.1. Instances

2.4.1.1. Overview

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

2.4.1.2. Instance families

An ECS instance is the smallest unit that can provide compute capabilities and services for your business. The compute capabilities of instances vary by instance type. ECS instances are categorized into different instance families based on the business scenarios for which they are suited.

> **Note** The instance families and instance types described in this topic are for reference only. The physical server where an instance is hosted determines the instance type.

Instance family	Feature	Scenario
n4, shared general purpose instance family	• Offers a CPU-to-memory ratio of 1:2. • Equipped with 2.5 GHz Intel® Xeon® E5-2680 v3 (Broadwell) processors. • Paired with the latest DDR4 memory. • I/O optimized by default.	• Small and medium-sized web servers • Batch processing • Advertising services • Distributed analysis
mn4, shared balanced instance family	• Offers a CPU-to-memory ratio of 1:4. • Equipped with 2.5 GHz Intel® Xeon® E5-2680 v3 (Broadwell), Intel® Xeon® E5-2680 v4 (Haswell), Intel® Xeon® E5-2682 v4 (Broadwell), or Intel® Xeon® E5-2650 v2 (Haswell) processors. • Paired with the latest DDR4 memory. • I/O optimized by default.	• Small and medium-sized web servers • Batch processing • Advertising services • Distributed analysis • Hadoop clusters

Instance family	Feature	Scenario
xn4, shared compact instance family	- Offers a CPU-to-memory ratio of 1:1. - Equipped with 2.5 GHz Intel® Xeon® E5-2680 v4 (Haswell) or Intel® Xeon® E5-2682 v4 (Broadwell) processors. - Paired with the latest DDR4 memory. - I/O optimized by default.	- Minisite web applications - Small databases - Development and testing environments - Code storage servers
e4, shared memory optimized instance family	- Offers a CPU-to-memory ratio of 1:8. - Equipped with 2.5 GHz Intel® Xeon® E5-2680 v4 (Broadwell), Intel® Xeon® E5-2680 v3 (Broadwell), Intel® Xeon® E5-2650 v2 (Haswell), or Intel® Xeon® E5-2682 v4 (Broadwell) processors. - I/O optimized by default.	Applications that involve large numbers of memory operations, queries, and computations, such as Cache, Redis, search applications, and in-memory databases
n4v2, shared general purpose instance family	- Supports IPv6. - Offers a CPU-to-memory ratio of 1:2. - Equipped with 2.5 GHz Intel® Xeon® E5-2680 v3 (Broadwell) processors. - Paired with the latest DDR4 memory. - I/O optimized by default.	- Small and medium-sized web servers - Batch processing - Advertising services - Distributed analysis
mn4v2, shared balanced instance family	- Supports IPv6. - Offers a CPU-to-memory ratio of 1:4. - Equipped with 2.5 GHz Intel® Xeon® E5-2680 v3 (Broadwell), Intel® Xeon® E5-2680 v4 (Haswell), Intel® Xeon® E5-2682 v4 (Broadwell), or Intel® Xeon® E5-2650 v2 (Haswell) processors. - Paired with the latest DDR4 memory. - I/O optimized by default.	- Small and medium-sized web servers - Batch processing - Advertising services - Distributed analysis - Hadoop clusters

Instance family	Feature	Scenario
xn4v2, shared compact instance family	• Supports IPv6. • Offers a CPU-to-memory ratio of 1:1. • Equipped with 2.5 GHz Intel® Xeon® E5-2680 v4 (Haswell) or Intel® Xeon® E5-2682 v4 (Broadwell) processors. • Paired with the latest DDR4 memory. • I/O optimized by default.	• Minisite web applications • Small databases • Development and testing environments • Code storage servers
e4v2, shared memory optimized instance family	• Supports IPv6. • Offers a CPU-to-memory ratio of 1:8. • Equipped with 2.5 GHz Intel® Xeon® E5-2680 v4 (Broadwell), Intel® Xeon® E5-2680 v3 (Broadwell), Intel® Xeon® E5-2650 v2 (Haswell), or Intel® Xeon® E5-2682 v4 (Broadwell) processors. • I/O optimized by default.	Applications that involve large numbers of memory operations, queries, and computations, such as Cache, Redis, search applications, and in-memory databases
t5, burstable instance family	• Equipped with 2.5 GHz Intel® Xeon® processors. • Paired with the DDR4 memory. • Offers multiple CPU-to-memory ratios. • Provides baseline CPU performance and is burstable, but limited by accumulated CPU credits. • Offers a balance of compute, memory, and network resources. • Supports VPCs only.	• Web application servers • Lightweight applications and microservices • Development and testing environments
sn1ne, compute optimized instance family with enhanced network performance	• Offers a CPU-to-memory ratio of 1:2. • Provides an ultra-high packet forwarding rate. • Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors for consistent computing performance. • Provides high network performance based on large computing capacity.	• Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are received and transmitted • Web frontend servers • Frontend servers of massively multiplayer online (MMO) games • Data analysis, batch processing, and video encoding • High-performance scientific and engineering applications

Instance family	Feature	Scenario
g6, general purpose instance family	• Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture. • I/O optimized. • Supports enhanced SSDs, standard SSDs, and ultra disks. • Provides high storage I/O performance based on large computing capacity. • Offers a CPU-to-memory ratio of 1:4. • Allows you to enable or disable Hyper-Threading. • Provides an ultra-high packet forwarding rate. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors that deliver a maximum turbo frequency of 3.2 GHz for consistent computing performance. • Provides high network performance based on large computing capacity.	• Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are received and transmitted • Enterprise-level applications of various types and sizes • Websites and application servers • Game servers • Small and medium-sized database systems, caches, and search clusters • Data analysis and computing • Compute clusters and memory intensive data processing
g5, general purpose instance family	• I/O optimized. • Supports enhanced SSDs, standard SSDs, and ultra disks. • Offers a CPU-to-memory ratio of 1:4. • Provides an ultra-high packet forwarding rate. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or 8269CY (Cascade Lake) processors for consistent computing performance. • Provides high network performance based on large computing capacity.	• Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are received and transmitted • Enterprise-level applications of various types and sizes • Small and medium-sized database systems, caches, and search clusters • Data analysis and computing • Compute clusters and memory intensive data processing

Instance family	Feature	Scenario
sn2ne, general purpose instance family with enhanced network performance	• Offers a CPU-to-memory ratio of 1:4. • Provides an ultra-high packet forwarding rate. • Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors for consistent computing performance. • Provides high network performance based on large computing capacity.	• Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are received and transmitted • Enterprise-level applications of various types and sizes • Small and medium-sized database systems, caches, and search clusters • Data analysis and computing • Compute clusters and memory intensive data processing
r6, memory optimized instance family	• Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture. • I/O optimized. • Supports enhanced SSDs, standard SSDs, and ultra disks. • Provides high storage I/O performance based on large computing capacity. • Provides an ultra-high packet forwarding rate. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors that deliver a maximum turbo frequency of 3.2 GHz for consistent computing performance. • Offers a CPU-to-memory ratio of 1:8. • Allows you to enable or disable Hyper-Threading. • Provides high network performance based on large computing capacity.	• Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are received and transmitted • High-performance databases and in-memory databases • Data analysis, data mining, and distributed memory caching • Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance family	Feature	Scenario
r5, memory optimized instance family	- I/O optimized. - Supports enhanced SSDs, standard SSDs, and ultra disks. - Offers a CPU-to-memory ratio of 1:8. - Provides an ultra-high packet forwarding rate. - Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or Intel® Xeon® Platinum 8269CY (Cascade Lake) processors for consistent computing performance. - Provides high network performance based on large computing capacity.	- Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are received and transmitted - High-performance databases and in-memory databases - Data analysis, data mining, and distributed memory caching - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications
se1ne, memory optimized instance family with enhanced network performance	- Offers a CPU-to-memory ratio of 1:8. - Provides an ultra-high packet forwarding rate. - Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors for consistent computing performance. - Provides high network performance based on large computing capacity.	- Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are received and transmitted - High-performance databases and in-memory databases - Data analysis, data mining, and distributed memory caching - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications
se1, memory optimized instance family	- Provides consistent computing performance. - Offers a CPU-to-memory ratio of 1:8. - Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors. - Paired with the latest DDR4 memory. - Provides high network performance based on large computing capacity. - I/O optimized by default.	- Cache/Redis - Search applications - In-memory databases - Databases with high I/O requirements, such as Oracle and MongoDB - Hadoop clusters - Computing scenarios where consistent performance is required, such as large-volume data processing

Instance family	Feature	Scenario
ebmg5s, general purpose ECS Bare Metal Instance family with enhanced network performance	• Supports enhanced SSDs, standard SSDs, and ultra disks. • Offers a CPU-to-memory ratio of 1:4. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors that have 96 vCPUs and a maximum turbo frequency of 2.7 GHz. • Provides high network performance with a packet forwarding rate of 4,500 Kpps. • Supports VPCs only. • Provides dedicated hardware resources and physical isolation.	• Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware • Third-party virtualization (including but not limited to Xen and KVM), and AnyStack (including but not limited to OpenStack and ZStack) • Containers (including but not limited to Docker, Clear Containers, and Pouch) • Enterprise-level applications such as large and medium-sized databases • Video encoding
ebmg5, general purpose ECS Bare Metal Instance family	• Offers a CPU-to-memory ratio of 1:4. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors that have 96 vCPUs and a maximum turbo frequency of 2.9 GHz. • Provides high network performance with a packet forwarding rate of 4,500 Kpps. • Supports standard SSDs and ultra disks.	• Deployment of Apsara Stack services such as OpenStack and ZStack • Deployment of services such as Docker containers • Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are received and transmitted • Enterprise-level applications of various types and sizes • Large and medium-sized databases, caches, and search clusters • Data analysis and computing • Compute clusters and memory intensive data processing
i2, instance family with local SSDs	• Attached with high-performance local NVMe SSDs that deliver high IOPS, high I/O throughput, and low latency. • Offers a CPU-to-memory ratio of 1:8, which is designed for high-performance databases. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors. • Provides high network performance based on large computing capacity.	• Online transaction processing (OLTP) and high-performance relational databases • NoSQL databases such as Cassandra and MongoDB • Search scenarios that use solutions such as Elasticsearch

Instance family	Feature	Scenario
d1, big data instance family	• Attached with high-capacity local SATA HDDs that deliver high throughput and bandwidth of up to 17 Gbit/s between instances. • Offers a CPU-to-memory ratio of 1:4, which is designed for big data scenarios. • Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors. • Provides high network performance based on large computing capacity.	• Hadoop MapReduce, HDFS, Hive, and HBase • Spark in-memory computing and MLlib • Enterprises in Internet, finance, and other industries that need to compute, store, and analyze large volumes of data • Elasticsearch and logging
d2, big data instance family	• Attached with high-capacity local SATA HDDs that deliver high throughput and bandwidth of up to 10 Gbit/s between instances. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors. • Provides high network performance based on large computing capacity.	• Hadoop MapReduce, HDFS, Hive, and HBase • Spark in-memory computing and MLlib • Enterprises in Internet, finance, and other industries that need to compute, store, and analyze large volumes of data • Elasticsearch and logging

Instance family	Feature	Scenario
sccg5ib, general purpose Super Computing Cluster (SCC) instance family	• Offers a CPU-to-memory ratio of 1:8. • Provides an ultra-high packet forwarding rate. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors for consistent computing performance. • Supports 100 Gbit/s InfiniBand networks that deliver ultra high bandwidth and ultra low latency.	• Data analysis and computing • Artificial intelligence computing • Manufacturing simulation • High performance computing clusters • Genetic analysis • Pharmaceutical analysis
scch5ib, SCC instance family with high clock speed	• Offers a CPU-to-memory ratio of 1:6. • Provides an ultra-high packet forwarding rate. • Equipped with 3.1 GHz Intel® Xeon® Gold 6149 (Skylake) processors for consistent computing performance. • Supports 100 Gbit/s InfiniBand networks that deliver ultra high bandwidth and ultra low latency.	• Data analysis and computing • Artificial intelligence computing • Manufacturing simulation • High performance computing clusters • Genetic analysis • Pharmaceutical analysis

Instance family	Feature	Scenario
c6, compute optimized instance family	• Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture. • I/O optimized. • Supports enhanced SSDs, standard SSDs, and ultra disks. • Provides high storage I/O performance based on large computing capacity. • Allows you to enable or disable Hyper-Threading. • Offers a CPU-to-memory ratio of 1:2. • Provides an ultra-high packet forwarding rate. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors that deliver a maximum turbo frequency of 3.2 GHz for consistent computing performance. • Provides high network performance based on large computing capacity.	• Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are received and transmitted • Web frontend servers • Frontend servers of MMO games • Data analysis, batch processing, and video encoding • High-performance scientific and engineering applications
sn1, compute optimized instance family	• Offers a CPU-to-memory ratio of 1:2. • Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or E5-2680 v3 (Haswell) processors for consistent computing performance. • Provides high network performance based on large computing capacity.	• Web frontend servers • Frontend servers of MMO games • Data analysis, batch processing, and video encoding • High-performance scientific and engineering applications
sn2, general purpose instance family	• Offers a CPU-to-memory ratio of 1:4. • Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or E5-2680 v3 (Haswell) processors for consistent computing performance. • Provides high network performance based on large computing capacity.	• Enterprise-level applications of various types and sizes • Small and medium-sized database systems, caches, and search clusters • Data analysis and computing • Compute clusters and memory intensive data processing

Instance family	Feature	Scenario
f1, compute optimized FPGA-accelerated instance family	• Equipped with Intel® ARRIA® 10 GX 1150 FPGAs. • Offers a CPU-to-memory ratio of 1:7.5. • Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors. • Provides high network performance based on large computing capacity.	• Deep learning and inference • Genomics research • Financial analysis • Image transcoding • Computational workloads such as real-time video processing and security management
f3, compute optimized FPGA-accelerated instance family	• Uses Xilinx Virtex UltraScale+ VU9P. • Offers a CPU-to-memory ratio of 1:4. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors. • Provides high network performance based on large computing capacity.	• Deep learning and inference • Genetic computation • Video encoding and decoding • Chip prototype verification • Database acceleration

Instance family	Feature	Scenario
gn5, compute optimized GPU-accelerated instance family	• Uses NVIDIA P100 GPUs. • Offers multiple CPU-to-memory ratios. • Attached with high-performance NVMe SSDs. • Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors. • Provides high network performance based on large computing capacity.	• Deep learning • Scientific computing applications such as computational fluid dynamics, computational finance, genomics, and environmental analysis • Server-side GPU computational workloads such as high performance computing, rendering, and multi-media coding and decoding
gn4, compute optimized GPU-accelerated instance family	• Uses NVIDIA M40 GPUs. • Offers multiple CPU-to-memory ratios. • Equipped with 2.5 GHz Intel® Xeon® E5-2680 v4 (Haswell) processors. • Provides high network performance based on large computing capacity.	• Deep learning • Scientific computing applications such as computational fluid dynamics, computational finance, genomics, and environmental analysis • Server-side GPU computational workloads such as high performance computing, rendering, and multi-media coding and decoding

Instance family	Feature	Scenario
ga1, compute optimized GPU-accelerated instance family	• Uses AMD S7150 GPUs. • Offers a CPU-to-memory ratio of 1:2.5. • Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors. • High-performance NVMe SSDs. • Provides high network performance based on large computing capacity.	• Rendering and multi-media encoding and decoding • Machine learning, high performance computing, and high performance databases • Server-side workloads that require powerful concurrent floating-point computing capacity
gn5i, compute optimized instance family with GPU capabilities	• Uses NVIDIA P4 GPUs. • Offers a CPU-to-memory ratio of 1:4. • Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or E5-2680 v4 (Haswell) processors. • Provides high network performance based on large computing capacity.	• Deep learning and inference • Server-side GPU computational workloads such as multimedia encoding and decoding
gn5e, compute optimized GPU-accelerated instance family	• I/O optimized. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors for consistent computing performance. • Uses NVIDIA P4 GPUs. • Provides high network performance based on large computing capacity.	• Deep learning and inference • Video and image processing, such as noise reduction, encoding, and decoding

Instance family	Feature	Scenario
gn6i, compute optimized GPU-accelerated instance family	• I/O optimized. • Offers a CPU-to-memory ratio of 1:4. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors. • Supports enhanced SSDs, ultra disks, and enhanced SSDs that deliver millions of IOPS. • Provides better performance with the new-generation X-Dragon compute architecture. • Uses NVIDIA T4 GPU computing accelerators. ◦ Powered by the new NVIDIA Turing architecture. ◦ Equipped with up to 320 Turing Tensor Cores per GPU. ◦ Equipped with 2,560 CUDA Cores per GPU. ◦ Mixed-precision Tensor Cores support 65 FP16 TFLOPS, 130 INT8 TOPS, and 260 INT4 TOPS. ◦ Equipped with 16 GB GPU memory (320 GB/s bandwidth) per GPU. ◦ Provides high network performance based on large computing capacity.	• AI (deep learning and machine learning) inference for computer vision, speech recognition, speech synthesis, natural language processing (NLP), machine translation, and recommendation systems • Real-time rendering for cloud gaming • Real-time rendering for AR and VR applications • Graphics workstations or overloaded graphics computing • GPU-accelerated databases • High performance computing

Instance family	Feature	Scenario
gn6v, compute optimized GPU-accelerated instance family	• I/O optimized. • Supports enhanced SSDs, standard SSDs, and ultra disks. • Uses NVIDIA V100 GPUs. • Offers a CPU-to-memory ratio of 1:4. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors. • Uses NVIDIA V100 GPU computing accelerators (SXM2-based). ○ Powered by the new NVIDIA Volta architecture. ○ Equipped with 16 GB HBM2 GPU memory (900 GB/s bandwidth) per GPU. ○ Equipped with 5,120 CUDA cores per GPU. ○ Equipped with 640 Tensor cores per GPU. ○ Supports up to six NVLink connections and a total bandwidth of 300 GB/s (25 GB/s per connection). • Provides high network performance based on large computing capacity.	• Deep learning applications such as training and inference applications of AI algorithms used in image classification, autonomous vehicles, and speech recognition • Scientific computing applications such as computational fluid dynamics, computational finance, molecular dynamics, and environmental analysis

Instance family	Feature	Scenario
sccgn6p, compute optimized GPU-accelerated SCC instance family	• I/O optimized. • Offers a CPU-to-memory ratio of 1:8. • Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors for consistent computing performance. • Provides all features of ECS Bare Metal Instance. • Storage: ○ Supports standard SSDs, ultra disks, and enhanced SSDs that deliver millions of IOPS. ○ Supports high performance Cloud Parallel File System (CPFS). ○ Supports Apsara File Storage NAS. • Networking: ○ Supports VPCs equipped with two 25 Gbit/s ports. ○ Supports 100 Gbit/s InfiniBand networks for low-latency RDMA communication. • Uses NVIDIA V100 GPU computing accelerators (SXM2-based). ○ Powered by the new NVIDIA Volta architecture. ○ Equipped with 32 GB HBM2 GPU memory. ○ CUDA Cores 5120 ○ Tensor Cores 640 ○ Offers a GPU memory bandwidth of up to 900 GB/s. ○ Supports up to six NVLink connections and a total bandwidth of 300 GB/s (25 GB/s per connection).	• Ultra-large-scale training for machine learning on a distributed GPU cluster • Large-scale high performance scientific computing and simulations • Large-scale data analysis, batch processing, and video encoding

Instance family	Feature	Scenario

Instance family	Feature	Scenario
Custom instance type ecs.anyshare	• Allows you to create, modify, and delete custom instance types in Apsara Stack Operation (ASO). • Allows you to create instances of custom instance types. • Custom instance types are shared instance types.	Scenarios where existing instance types cannot meet the requirements and instance types need to be customized

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

Instance family	Feature	Scenario
n1, shared computing optimized instance family	• Offers a CPU-to-memory ratio of 1:2. • Equipped with 2.5 GHz Intel® Xeon® E5-2680 v3 (Haswell) processors. • Provides high network performance based on large computing capacity. • I/O optimized. • Supports standard SSDs and ultra disks.	• Small and medium-sized web servers • Batch processing • Distributed analysis • Advertising services
n2, shared general purpose instance family	• Offers a CPU-to-memory ratio of 1:4. • Equipped with 2.5 GHz Intel® Xeon® E5-2680 v3 (Haswell) processors. • Provides high network performance based on large computing capacity. • I/O optimized. • Supports standard SSDs and ultra disks.	• Medium-sized web servers • Batch processing • Distributed analysis • Advertising services • Hadoop clusters
e3, shared memory optimized instance family	• Offers a CPU-to-memory ratio of 1:8. • Equipped with 2.5 GHz Intel® Xeon® E5-2680 v3 (Haswell) processors. • Provides high network performance based on large computing capacity. • I/O optimized. • Supports standard SSDs and ultra disks.	• Cache/Redis • Search applications • In-memory databases • Databases with high I/O requirements, such as Oracle and MongoDB • Hadoop clusters • Large-volume data processing

Instance family	Feature	Scenario
c1, generation I instance family	• Equipped with Intel® Xeon® E5-2420 processors that deliver a minimum operating frequency of 1.9 GHz. • Paired with the latest DDR3 memory. • I/O optimized and non-I/O optimized instances are available. • I/O optimized instances support both standard SSDs and ultra disks. • Non-I/O optimized instances support only basic disks.	Generation 1 instance types are legacy shared instance types. They are still categorized based on the number of cores (1, 2, 4, 8, and 16 cores) and are insensitive to instance families.
c2, generation I instance family	• Equipped with Intel® Xeon® E5-2420 processors that deliver a minimum operating frequency of 1.9 GHz. • Paired with the latest DDR3 memory. • I/O optimized and non-I/O optimized instances are available. • I/O optimized instances support both standard SSDs and ultra disks. • Non-I/O optimized instances support only basic disks.	Generation 1 instance types are legacy shared instance types. They are still categorized based on the number of cores (1, 2, 4, 8, and 16 cores) and are not sensitive to instance families.
m1, generation I instance family	• Equipped with Intel® Xeon® E5-2420 processors that deliver a minimum operating frequency of 1.9 GHz. • Paired with the latest DDR3 memory. • I/O optimized and non-I/O optimized instances are available. • I/O optimized instances support both standard SSDs and ultra disks. • Non-I/O optimized instances support only basic disks.	Generation 1 instance types are legacy shared instance types. They are still categorized based on the number of cores (1, 2, 4, 8, and 16 cores) and are insensitive to instance families.

Instance family	Feature	Scenario
m2, generation I instance family	- Equipped with Intel® Xeon® E5-2420 processors that deliver a minimum operating frequency of 1.9 GHz. - Paired with the latest DDR3 memory. - I/O optimized and non-I/O optimized instances are available. - I/O optimized instances support both standard SSDs and ultra disks. - Non-I/O optimized instances support only basic disks.	Generation 1 instance types are legacy shared instance types. They are still categorized based on the number of cores (1, 2, 4, 8, and 16 cores) and are insensitive to instance families.
s1, generation I instance family	- Equipped with Intel® Xeon® E5-2420 processors that deliver a minimum operating frequency of 1.9 GHz. - Paired with the latest DDR3 memory. - Non-I/O optimized. - Supports basic disks only.	Generation 1 instance types are legacy shared instance types. They are still categorized based on the number of cores (1, 2, 4, 8, and 16 cores) and are insensitive to instance families.
s2, generation I instance family	- Equipped with Intel® Xeon® E5-2420 processors that deliver a minimum operating frequency of 1.9 GHz. - Paired with the latest DDR3 memory. - I/O optimized and non-I/O optimized instances are available. - I/O optimized instances support both standard SSDs and ultra disks. - Non-I/O optimized instances support only basic disks.	Generation 1 instance types are legacy shared instance types. They are still categorized based on the number of cores (1, 2, 4, 8, and 16 cores) and are insensitive to instance families.
s3, generation I instance family	- Equipped with Intel® Xeon® E5-2420 processors that deliver a minimum operating frequency of 1.9 GHz. - Paired with the latest DDR3 memory. - I/O optimized and non-I/O optimized instances are available. - I/O optimized instances support both standard SSDs and ultra disks. - Non-I/O optimized instances support only basic disks.	Generation 1 instance types are legacy shared instance types. They are still categorized based on the number of cores (1, 2, 4, 8, and 16 cores) and are insensitive to instance families.

Instance family	Feature	Scenario
t1, generation I instance family	• Equipped with Intel® Xeon® E5-2420 processors that deliver a minimum operating frequency of 1.9 GHz. • Paired with the latest DDR3 memory. • Non-I/O optimized. • Supports basic disks only.	Generation 1 instance types are legacy shared instance types. They are still categorized based on the number of cores (1, 2, 4, 8, and 16 cores) and are insensitive to instance families.

2.4.1.3. Instance types

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

An ECS instance type defines the basic properties of ECS instances, such as CPU, clock speed, and memory. In addition to the instance type, you must also configure the Block Storage devices, image, and network type when you create an ECS instance. The following table describes instance families and lists the instance types of each instance family.

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

Instance family **mn4**	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.mn4.xlarge	4	16.0	N/A	0.8	150	1	2
	ecs.mn4.2xlarge	8	32.0	N/A	1.2	300	1	3
	ecs.mn4.4xlarge	16	64.0	N/A	2.5	400	1	8
	ecs.mn4.8xlarge	32	128.0	N/A	5.0	500	2	8
xn4	ecs.xn4.small	1	1.0	N/A	0.5	50	1	1
e4	ecs.e4.small	1	8.0	N/A	0.5	50	1	1
	ecs.e4.large	2	16.0	N/A	0.5	100	1	1
	ecs.e4.xlarge	4	32.0	N/A	0.8	150	1	2
	ecs.e4.2xlarge	8	64.0	N/A	1.2	300	1	3
	ecs.e4.4xlarge	16	128.0	N/A	2.5	400	1	8
sn1ne	ecs.sn1ne.large	2	4.0	N/A	1.0	300	2	2
	ecs.sn1ne.xlarge	4	8.0	N/A	1.5	500	2	3
	ecs.sn1ne.2xlarge	8	16.0	N/A	2.0	1,000	4	4
	ecs.sn1ne.3xlarge	12	24.0	N/A	2.5	1,300	4	6
	ecs.sn1ne.4xlarge	16	32.0	N/A	3.0	1,600	4	8
	ecs.sn1ne.6xlarge	24	48.0	N/A	4.5	2,000	6	4

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.sn1ne.8xlarge	32	64.0	N/A	6.0	2,500	8	8
g6	ecs.g6.large	2	8.0	N/A	1.0	300	2	2
	ecs.g6.xlarge	4	16.0	N/A	1.5	500	4	3
	ecs.g6.2xlarge	8	32.0	N/A	2.5	800	8	4
	ecs.g6.3xlarge	12	48.0	N/A	4.0	900	8	6
	ecs.g6.4xlarge	16	64.0	N/A	5.0	1,000	8	8
	ecs.g6.6xlarge	24	96.0	N/A	7.5	1,500	12	8
	ecs.g6.8xlarge	32	128.0	N/A	10.0	2,000	16	8
g5	ecs.g5.large	2	8.0	N/A	1.0	300	2	2
	ecs.g5.xlarge	4	16.0	N/A	1.5	500	2	3
	ecs.g5.2xlarge	8	32.0	N/A	2.5	800	2	4
	ecs.g5.3xlarge	12	48.0	N/A	4.0	900	4	6
	ecs.g5.4xlarge	16	64.0	N/A	5.0	1,000	4	8
	ecs.g5.6xlarge	24	96.0	N/A	7.5	1,500	6	8
	ecs.g5.8xlarge	32	128.0	N/A	10.0	2,000	8	8
	ecs.g5.16xlarge	64	256.0	N/A	20.0	4,000	16	8

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
sn2ne	ecs.sn2ne.large	2	8.0	N/A	1.0	300	2	2
	ecs.sn2ne.xlarge	4	16.0	N/A	1.5	500	2	3
	ecs.sn2ne.2xlarge	8	32.0	N/A	2.0	1,000	4	4
	ecs.sn2ne.3xlarge	12	48.0	N/A	2.5	1,300	4	6
	ecs.sn2ne.4xlarge	16	64.0	N/A	3.0	1,600	4	8
	ecs.sn2ne.6xlarge	24	96.0	N/A	4.5	2,000	6	4
	ecs.sn2ne.8xlarge	32	128.0	N/A	6.0	2,500	8	8
	ecs.sn2ne.14xlarge	56	224.0	N/A	10.0	4,500	14	8
r6	ecs.r6.large	2	16.0	N/A	1.0	300	2	2
	ecs.r6.xlarge	4	32.0	N/A	1.5	500	4	3
	ecs.r6.2xlarge	8	64.0	N/A	2.5	800	8	4
	ecs.r6.3xlarge	12	96.0	N/A	4.0	900	8	6
	ecs.r6.4xlarge	16	128.0	N/A	5.0	1,000	8	8
	ecs.r6.6xlarge	24	192.0	N/A	7.5	1,500	12	8
	ecs.r6.8xlarge	32	256.0	N/A	10.0	2,000	16	8
	ecs.r5.large	2	16.0	N/A	1.0	300	2	2

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
r5	ecs.r5.xlarge	4	32.0	N/A	1.5	500	2	3
	ecs.r5.2xlarge	8	64.0	N/A	2.5	800	2	4
	ecs.r5.3xlarge	12	96.0	N/A	4.0	900	4	6
	ecs.r5.4xlarge	16	128.0	N/A	5.0	1,000	4	8
	ecs.r5.6xlarge	24	192.0	N/A	7.5	1,500	6	8
	ecs.r5.8xlarge	32	256.0	N/A	10.0	2,000	8	8
	ecs.r5.16xlarge	64	512.0	N/A	20.0	4,000	16	8
se1ne	ecs.se1ne.large	2	16.0	N/A	1.0	300	2	2
	ecs.se1ne.xlarge	4	32.0	N/A	1.5	500	2	3
	ecs.se1ne.2xlarge	8	64.0	N/A	2.0	1,000	4	4
	ecs.se1ne.3xlarge	12	96.0	N/A	2.5	1,300	4	6
	ecs.se1ne.4xlarge	16	128.0	N/A	3.0	1,600	4	8
	ecs.se1ne.6xlarge	24	192.0	N/A	4.5	2,000	6	4
	ecs.se1ne.8xlarge	32	256.0	N/A	6.0	2,500	8	8
	ecs.se1ne.14xlarge	56	480.0	N/A	10.0	4,500	14	8
	ecs.se1.large	2	16.0	N/A	0.5	100	1	2

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
se1	ecs.se1.xlarge	4	32.0	N/A	0.8	200	1	3
	ecs.se1.2xlarge	8	64.0	N/A	1.5	400	1	4
	ecs.se1.4xlarge	16	128.0	N/A	3.0	500	2	8
	ecs.se1.8xlarge	32	256.0	N/A	6.0	800	3	8
	ecs.se1.14xlarge	56	480.0	N/A	10.0	1,200	4	8
ebmg5s	ecs.ebmg5s.24xlarge	96	384.0	N/A	30.0	4,500	8	32
ebmg5	ecs.ebmg5.24xlarge	96	384.0	N/A	10.0	4,000	8	32
i2	ecs.i2.xlarge	4	32.0	1 × 894	1.0	500	2	3
	ecs.i2.2xlarge	8	64.0	1 × 1,788	2.0	1,000	2	4
	ecs.i2.4xlarge	16	128.0	2 × 1,788	3.0	1,500	4	8
	ecs.i2.8xlarge	32	256.0	4 × 1,788	6.0	2,000	8	8
	ecs.i2.16xlarge	64	512.0	8 × 1,788	10.0	4,000	16	8
	ecs.d1.2xlarge	8	32.0	4 × 5,500	3.0	300	1	4
	ecs.d1.3xlarge	12	48.0	6 × 5,500	4.0	400	1	6
	ecs.d1.4xlarge	16	64.0	8 × 5,500	6.0	600	2	8
	ecs.d1.6xlarge	24	96.0	12 × 5,500	8.0	800	2	8

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
d1	ecs.d1-c8d3.8xlarge	32	128.0	12 × 5,500	10.0	1,000	4	8
	ecs.d1.8xlarge	32	128.0	16 × 5,500	10.0	1,000	4	8
	ecs.d1-c14d3.14xlarge	56	160.0	12 × 5,500	17.0	1,800	6	8
	ecs.d1.14xlarge	56	224.0	28 × 5,500	17.0	1,800	6	8
d2	ecs.d2-zyy-d0.4xlarge	16	64.0	N/A	3.0	300	2	8
	ecs.d2-zyy-d0.6xlarge	24	96.0	N/A	4.0	400	2	8
	ecs.d2-zyy.4xlarge	16	64.0	6 × 7,500	3.0	300	4	8
	ecs.d2-zyy.6xlarge	24	96.0	12 × 7,500	4.0	400	4	8
	ecs.d2-zyy.7xlarge	56	224.0	12 × 7,300	17.0	1,800	6	8
	ecs.d2-zyy.8xlarge	32	160.0	12 × 7,300	17.0	1,800	6	8
	ecs.d2-zyy.22xlarge	88	352.0	12 × 7,300	17.0	1,800	6	8
	ecs.d2-zyy-m40.8xlarge	32	128.0	12 × 7,500	6.0	600	4	8

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.d2-gab.4xlarge	16	64.0	6 × 1,150	3.0	300	4	8
	ecs.d2-gab.8xlarge	32	128.0	12 × 1,150	6.0	600	4	8
sccg5ib	ecs.sccg5ib.24xlarge	96	384.0	N/A	10.0	4,500	8	32
scch5ib	ecs.scch5ib.16xlarge	64	192.0	N/A	10.0	4,500	8	32
c6	ecs.c6.large	2	4.0	N/A	1.0	300	2	2
	ecs.c6.xlarge	4	8.0	N/A	1.5	500	4	3
	ecs.c6.2xlarge	8	16.0	N/A	2.5	800	8	4
	ecs.c6.3xlarge	12	24.0	N/A	4.0	900	8	6
	ecs.c6.4xlarge	16	32.0	N/A	5.0	1,000	8	8
	ecs.c6.6xlarge	24	48.0	N/A	7.5	1,500	12	8
	ecs.c6.8xlarge	32	64.0	N/A	10.0	2,000	16	8
sn1	ecs.sn1.medium	2	4.0	N/A	0.5	100	1	2
	ecs.sn1.large	4	8.0	N/A	0.8	200	1	3
	ecs.sn1.xlarge	8	16.0	N/A	1.5	400	1	4
	ecs.sn1.3xlarge	16	32.0	N/A	3.0	500	2	8

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.sn1.7xlarge	32	64.0	N/A	6.0	800	3	8
sn2	ecs.sn2.medium	2	8.0	N/A	0.5	100	1	2
	ecs.sn2.large	4	16.0	N/A	0.8	200	1	3
	ecs.sn2.xlarge	8	32.0	N/A	1.5	400	1	4
	ecs.sn2.3xlarge	16	64.0	N/A	3.0	500	2	8
	ecs.sn2.7xlarge	32	128.0	N/A	6.0	800	3	8
	ecs.sn2.14xlarge	56	224.0	N/A	10.0	1,200	4	8

The following table describes FPGA-accelerated instance families.

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	FPGA
f1	ecs.f1-c8f1.2xlarge	8	60.0	N/A	3.0	400	4	2	Intel ARRIA 10 GX 1150
	ecs.f1-c8f1.4xlarge	16	120.0	N/A	5.0	1,000	4	2	2 × Intel ARRIA 10 GX 1150
	ecs.f1-c28f1.7xlarge	28	112.0	N/A	5.0	2,000	8	2	Intel ARRIA 10 GX 1150

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	FPGA
	ecs.f1-c28f1.14xlarge	56	224.0	N/A	10.0	2,000	14	2	2 × Intel ARRIA 10 GX 1150
f3	ecs.f3-c16f1.4xlarge	16	64.0	N/A	5.0	1,000	4	8	1 × Xilinx VU9P
	ecs.f3-c16f1.8xlarge	32	128.0	N/A	10.0	2,000	8	8	2 × Xilinx VU9P
	ecs.f3-c16f1.16xlarge	64	256.0	N/A	20.0	2,000	16	8	4 × Xilinx VU9P

The following table describes GPU-accelerated instance families.

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPU
	ecs.gn5-c4g1.xlarge	4	30.0	440	3.0	300	1	3	1 × NVIDIA P100
	ecs.gn5-c8g1.2xlarge	8	60.0	440	3.0	400	1	4	1 × NVIDIA P100
	ecs.gn5-c4g1.2xlarge	8	60.0	880	5.0	1,000	2	4	2 × NVIDIA P100
	ecs.gn5-c8g1.4xlarge	16	120.0	880	5.0	1,000	4	8	2 × NVIDIA P100

gn5 Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPU
	ecs.gn5-c28g1.7xlarge	28	112.0	440	5.0	1,000	8	8	1 × NVIDIA P100
	ecs.gn5-c8g1.8xlarge	32	240.0	1760	10.0	2,000	8	8	4 × NVIDIA P100
	ecs.gn5-c28g1.14xlarge	56	224.0	880	10.0	2,000	14	8	2 × NVIDIA P100
	ecs.gn5-c8g1.14xlarge	54	480.0	3520	25.0	4,000	14	8	8 × NVIDIA P100
gn4	ecs.gn4-c4g1.xlarge	4	30.0	N/A	3.0	300	1	3	1 × NVIDIA M40
	ecs.gn4-c8g1.2xlarge	8	30.0	N/A	3.0	400	1	4	1 × NVIDIA M40
	ecs.gn4.8xlarge	32	48.0	N/A	6.0	800	3	8	1 × NVIDIA M40
	ecs.gn4-c4g1.2xlarge	8	60.0	N/A	5.0	500	1	4	2 × NVIDIA M40
	ecs.gn4-c8g1.4xlarge	16	60.0	N/A	5.0	500	1	8	2 × NVIDIA M40
	ecs.gn4.14xlarge	56	96.0	N/A	10.0	1,200	4	8	2 × NVIDIA M40

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPU
ga1	ecs.ga1.xlarge	4	10.0	1 × 87	1.0	200	1	3	0.25 × AMD S7150
	ecs.ga1.2xlarge	8	20.0	1 × 175	1.5	300	1	4	0.5 × AMD S7150
	ecs.ga1.4xlarge	16	40.0	1 × 350	3.0	500	2	8	1 × AMD S7150
	ecs.ga1.8xlarge	32	80.0	1 × 700	6.0	800	3	8	2 × AMD S7150
	ecs.ga1.14xlarge	56	160.0	1 × 1,400	10.0	1,200	4	8	4 × AMD S7150
gn5i	ecs.gn5i-c2g1.large	2	8.0	N/A	1.0	100	2	2	1 × NVIDIA P4
	ecs.gn5i-c4g1.xlarge	4	16.0	N/A	1.5	200	2	3	1 × NVIDIA P4
	ecs.gn5i-c8g1.2xlarge	8	32.0	N/A	2.0	400	4	4	1 × NVIDIA P4
	ecs.gn5i-c16g1.4xlarge	16	64.0	N/A	3.0	800	4	8	1 × NVIDIA P4
	ecs.gn5i-c16g1.8xlarge	32	128.0	N/A	6.0	1,200	8	8	2 × NVIDIA P4

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPU
	ecs.gn5i-c24g1.12xlarge	48	192.0	N/A	10.0	2,000	8	8	2 × NVIDIA P4
	ecs.gn5i-c28g1.14xlarge	56	224.0	N/A	10.0	2,000	14	8	2 × NVIDIA P4
gn5e	ecs.gn5e-c11g1.3xlarge	10	58.0	N/A	2.0	150	1	6	1 × NVIDIA P4
	ecs.gn5e-c11g1.5xlarge	22	116.0	N/A	4.0	300	1	8	2 × NVIDIA P4
	ecs.gn5e-c11g1.11xlarge	44	232.0	N/A	6.0	600	2	8	4 × NVIDIA P4
	ecs.gn5e-c11g1.22xlarge	88	464.0	N/A	10.0	1,200	4	8	8 × NVIDIA P4
	ecs.gn6i-c10g1.2xlarge	10	42.0	N/A	5.0	800	2	4	1 × T4
	ecs.gn6i-c10g1.5xlarge	20	84.0	N/A	8.0	1,000	4	6	2 × T4
	ecs.gn6i-c10g1.10xlarge	40	168.0	N/A	15.0	2,000	8	8	4 × T4

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPU
gn6i	ecs.gn6i-c10g1.20xlarge	80	336.0	N/A	30.0	4,000	16	8	8 × T4
	ecs.gn6i-c14g1.3xlarge	14	56.0	N/A	5.0	1,000	4	6	1 × T4
	ecs.gn6i-c14g1.7xlarge	28	112.0	N/A	10.0	2,000	8	8	2 × T4
	ecs.gn6i-c14g1.14xlarge	56	224.0	N/A	20.0	4,000	12	8	4 × T4
	ecs.gn6i-c20g1.5xlarge	20	80.0	N/A	10.0	1,500	4	6	1 × T4
	ecs.gn6i-c20g1.10xlarge	40	160.0	N/A	20.0	3,000	8	8	2 × T4
gn6v	ecs.gn6v-c8g1.2xlarge	8	32.0	N/A	2.5	800	4	4	1 × NVIDIA V100
	ecs.gn6v-c8g1.8xlarge	32	128.0	N/A	10.0	2,000	8	8	4 × NVIDIA V100
	ecs.gn6v-c8g1.16xlarge	64	256.0	N/A	20.0	2,500	16	8	8 × NVIDIA V100

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	GPU
sccgn6p	ecs.sccgn6p.24xlarge	96	768.0	N/A	30.0	4,500	8	32	8 × NVIDIA V100

The following table describes shared instance families that support IPv6.

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)
n4v2	ecs.n4v2.small	1	2.0	N/A	0.5	50	Yes	1	1
	ecs.n4v2.large	2	4.0	N/A	0.5	100	Yes	1	1
	ecs.n4v2.xlarge	4	8.0	N/A	0.8	150	Yes	1	2
	ecs.n4v2.2xlarge	8	16.0	N/A	1.2	300	Yes	1	2
	ecs.n4v2.4xlarge	16	32.0	N/A	2.5	400	Yes	1	2
	ecs.n4v2.8xlarge	32	64.0	N/A	5.0	500	Yes	1	2
mn4v2	ecs.mn4v2.small	1	4.0	N/A	0.5	50	Yes	1	1
	ecs.mn4v2.large	2	8.0	N/A	0.5	100	Yes	1	1
	ecs.mn4v2.xlarge	4	16.0	N/A	0.8	150	Yes	1	2
	ecs.mn4v2.2xlarge	8	32.0	N/A	1.2	300	Yes	1	2
	ecs.mn4v2.4xlarge	16	64.0	N/A	2.5	400	Yes	1	2

Instance family	Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)
	ecs.mn4v2.8xlarge	32	128.0	N/A	5.0	500	Yes	2	8
xn4v2	ecs.xn4v2.small	1	1.0	N/A	0.5	50	Yes	1	1
e4v2	ecs.e4v2.small	1	8.0	N/A	0.5	50	Yes	1	1
	ecs.e4v2.large	2	16.0	N/A	0.5	100	Yes	1	1
	ecs.e4v2.xlarge	4	32.0	N/A	0.8	150	Yes	1	2
	ecs.e4v2.2xlarge	8	64.0	N/A	1.2	300	Yes	1	3
	ecs.e4v2.4xlarge	16	128.0	N/A	2.5	400	Yes	1	8

The following table describes burstable instance families.

Instance family	Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.t5-lc2m1.nano	1	0.5	20%	12	288	N/A	0.1	40	1	1
	ecs.t5-lc1m1.small	1	1.0	20%	12	288	N/A	0.2	60	1	1

Instance family	Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
t5	ecs.t5-lc1m2.small	1	2.0	20%	12	288	N/A	0.2	60	1	1
	ecs.t5-lc1m2.large	2	4.0	20%	24	576	N/A	0.4	100	1	1
	ecs.t5-lc1m4.large	2	8.0	20%	24	576	N/A	0.4	100	1	1
	ecs.t5-c1m1.large	2	2.0	25%	30	720	N/A	0.5	100	1	1
	ecs.t5-c1m2.large	2	4.0	25%	30	720	N/A	0.5	100	1	1
	ecs.t5-c1m4.large	2	8.0	25%	30	720	N/A	0.5	100	1	1
	ecs.t5-c1m1.xlarge	4	4.0	25%	60	1,440	N/A	0.8	200	1	2

Instance family	Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
	ecs.t5-c1m2.xlarge	4	8.0	25%	60	1,440	N/A	0.8	200	1	2
	ecs.t5-c1m4.xlarge	4	16.0	25%	60	1,440	N/A	0.8	200	1	2
	ecs.t5-c1m1.2xlarge	8	8.0	25%	120	2,880	N/A	1.2	400	1	2
	ecs.t5-c1m2.2xlarge	8	16.0	25%	120	2,880	N/A	1.2	400	1	2
	ecs.t5-c1m4.2xlarge	8	32.0	25%	120	2,880	N/A	1.2	400	1	2
	ecs.t5-c1m1.4xlarge	16	16.0	25%	240	5,760	N/A	1.2	600	1	2
	ecs.t5-c1m2.4xlarge	16	32.0	25%	240	5,760	N/A	1.2	600	1	2

The following table describes the ecs.anyshare custom instance type.

Instance family	vCPUs (x)	Memory (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	Disk bandwidth (Gbit/s)
Custom instance type ecs.anyshare	0 < x <=2	1 to 16	0.5	100	1	2	0.5
	2 < x <=4	2 to 32	0.8	100 + (x - 2)/0.2	1	2	0.8
	4 < x <=8	4 to 64	0.8 + (x - 5)/4	200 + (x - 4)/0.2	1	3	0.8 + (x - 5)/4
	8 < x <=12	8 to 96	1.5 + (x - 8)/8	400 + (x - 8)/0.8	2	3	1.5 + (x - 8)/4
	12 < x <=16	12 to 128	2 + (x - 12)/4	450 + (x - 12)/0.8	3	4	1
	16 < x <=24	16 to 196	3 + (x - 16)/8	500 + (x - 16)/0.8	3	5	1 + (x - 16)/8
	24 < x <=32	24 to 256	4 + (x - 24)/8	600 + (x - 24)/0.4	4	6	2 + (x - 24)/8
	x > 32	32 to 352	min(5 + (x - 32)/8, 10)	min(800 + (x-32)/0.8, 1200)	4	8	min(3 + (x - 32)/8, 8)

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

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

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

Instance family	Instance type	vCPUs	Memory (GiB)
t1	ecs.t1.small	1	1.0

2.4.1.4. UserData

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

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

Methods

- UserData-Scripts: are applicable to users who need to initialize instances by executing the shell scripts. The UserData-Scripts begin with `#! /bin/sh` . A review of user data shows that most users input UserData by running UserData-Scripts. UserData-Scripts are also suitable for complicated deployment scenarios.
- Cloud-Config: is a special script supported by cloud-init. It packs frequently-used personalized configurations into YAML files, which enable you to complete the frequently-used configurations more conveniently. The script starts with `# Cloud-config` in the first line and is followed by an array containing ssh_authorized_keys, hostname, write_files, and manage_etc_hosts.

Scenarios

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

2.4.1.5. Instance lifecycle

The lifecycle of an ECS instance begins when the instance is created and ends when the instance is released. This topic describes the instance states in the ECS console, state attributes, and corresponding instance states in API responses.

The following table describes the states that an ECS instances may go through during its lifecycle.

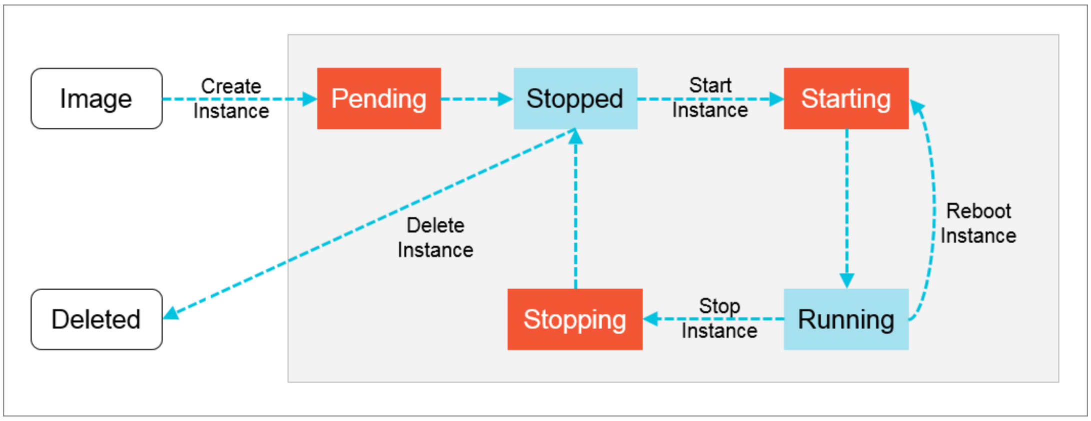
Instance states

State	State attribute	Description	State in an API response
Instance being created	Intermediate	The instance is being created and waiting to be started. If an instance remains in the Instance being created state for an extended period of time, an exception has occurred.	Pending

State	State attribute	Description	State in an API response
Starting	Intermediate	When you start or restart an instance by using the ECS console or calling an API operation, the instance enters this state before it enters the Running state. If an instance remains in the Starting state for an extended period of time, an exception has occurred.	Starting
Running	Stable	While an instance is in the Running state, the instance can function normally and can accommodate your business needs.	Running
Stopping	Intermediate	When you stop an instance by using the ECS console or calling an API operation, the instance enters this state before it enters the Stopped state. If an instance remains in the Stopping state for an extended period of time, an exception has occurred.	Stopping
Stopped	Stable	An instance enters this state when it is stopped. An instance in the Stopped state cannot provide external services.	Stopped
Reinitializing	Intermediate	When you re-initialize the system disk or a data disk of an instance by using the ECS console or calling an API operation, the instance enters this state before it enters the Running state. If an instance remains in the Reinitializing state for an extended period of time, an exception has occurred.	Stopped
Changing system disk	Intermediate	When you replace the system disk of an instance by using the ECS console or calling an API operation, the instance enters the Changing system disk state before it enters the Running state. If an instance remains in the Changing system disk state for an extended period of time, an exception has occurred.	Stopped

Instance states describes the relationships between instance states in the ECS console and instance states in API responses. The following figure shows the transitions between instance states in API responses.

Transitions between instance states in API responses



2.4.1.6. EBM Instances

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

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

Benefits

EBM Instances provides the following benefits through technological innovation:

- **Exclusive computing resources**

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

- **Encrypted computing**

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

- **Any Stack on Alibaba Cloud**

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

- **Heterogeneous instruction set processor support**

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

Configuration features

The following table lists the configuration features of EBM Instances.

Features

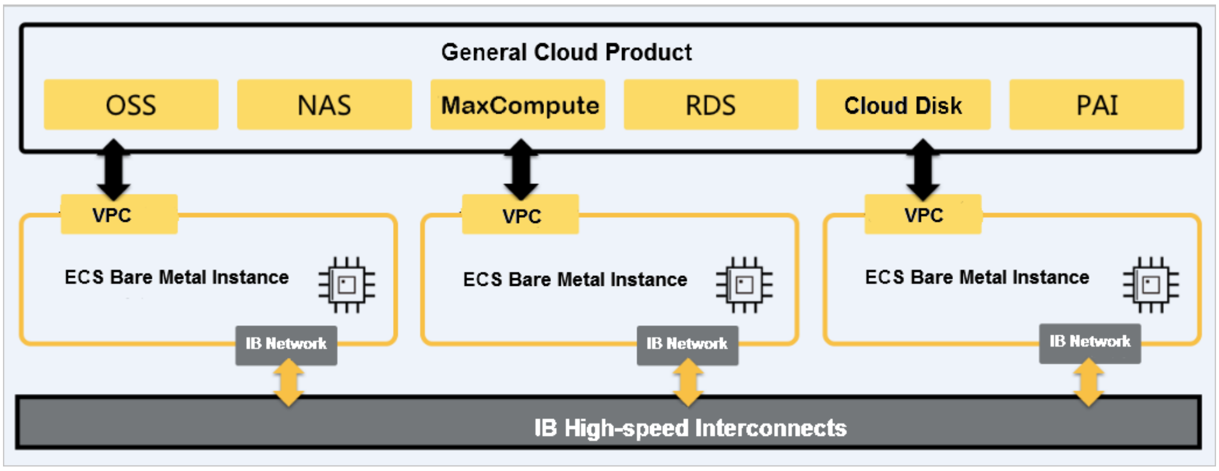
Item	Description
CPU configuration	Only supports the ebmg5 general-purpose EBM Instance type family.
Memory configuration	Supports expansion from 32 GiB to 384 GiB as needed. The ratio of CPU to memory is 1:2 or 1:4 to provide better computing performance.
Storage configuration	Supports startup from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

2.4.1.7. Super Computing Clusters

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

SCC architecture

The following figure shows the SCC architecture.



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

- SCCs have all the benefits of ECS Bare Metal Instances. The underlying architecture allows you to use exclusive cloud servers or physical servers to create a secure and controllable underlying environment where you can configure security groups and VPCs for your SCC instances to implement traffic control.

- SCCs adopt InfiniBand, a conversion cable technology that supports multiple concurrent connections. InfiniBand is the next-generation I/O standard for compute server platforms and features high scalability, high bandwidth, and low latency. InfiniBand is ideal for establishing communication between servers such as replication servers and distributed servers, between servers and storage devices such as SAN and direct-attached storage, and between servers and networks such as LANs, WANs, and the Internet. The InfiniBand architecture is commonly used in high-performance computing and provides higher bandwidth, lower latency, and more reliable connections than the Ethernet architecture.

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

Scenarios

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

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

Scenario	Description
Industry verticals	The application of HPC in traditional industries is mature and continues to evolve. Apsara Stack SCCs are suited for all the industries where traditional supercomputing is applied. The industries include the following: • Petrochemical: seismic data processing and reservoir simulation • Finance: financial derivative analysis, actuarial analysis, asset and liability management, and investment risk analysis • Industrial manufacturing: collision analysis, failure analysis, and thermodynamic analysis • Life science: drug discovery, protein folding, DNA sequencing, and medical imaging • Media and entertainment: video post-production and animation rendering • Government and higher education: weather forecast, high-energy physics, and geophysics

2.4.2. Block storage

2.4.2.1. Overview

This topic describes the different types of Block Storage devices, including elastic Block Storage services based on a distributed storage architecture and local storage services based on the local hard disks of physical servers that hosts ECS instances.

Definitions of elastic Block Storage and local storage:

- Elastic Block Storage provides ECS instances with block-level storage that features low latency, high performance, durability, and high reliability. A triplicate distributed mechanism is used to ensure data reliability. Elastic Block Storage devices can be created, released, and resized at any time.
- Local storage, also known as local disks, are temporary disks attached to physical machines that host ECS instances. Local storage is designed for business scenarios that require high storage I/O performance. It provides block-level data access capabilities for ECS instances, and features low latency, high random IOPS, and high throughput.

Differences among Apsara Stack storage services

Apsara Stack provides the following data storage services: Block Storage, Object Storage Service (OSS), and Apsara File Storage NAS. The following table describes the differences between these services.

Comparison between data storage services

Data storage service	Feature	Scenario

Data storage service	Feature	Scenario
Block Storage	A high-performance and low-latency block-level storage device provided by Alibaba Cloud for ECS instances. It supports random read and write operations. You can format a Block Storage device and create file systems on it in the same way as you do with a physical disk.	Block Storage can meet the data storage requirements of most business scenarios.
OSS	A huge storage space designed to store for unstructured data on the Internet, such as images, audios, and videos. You can access data stored in OSS anytime and anywhere by calling API operations.	OSS is applicable to business scenarios such as website construction, separation of dynamic and static resources, and acceleration of domain name access through CDN.
Apsara File Storage NAS	A storage space designed for storing large volumes of unstructured data that can be accessed based on standard file access protocols, such as the Network File System (NFS) protocol for Linux and the Common Internet File System (CIFS) protocol for Windows. You can set permissions to allow different clients to access the same file at the same time.	Apsara File Storage NAS is applicable to business scenarios such as file sharing across departments in an enterprise, non-linear editing in radio and television industries, high-performance computing, and Docker containers.

2.4.2.2. Elastic block storage

2.4.2.2.1. Overview

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

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

2.4.2.2.2. Cloud disks

Cloud disks are block-level storage devices provided by Apsara Stack for ECS instances. Cloud disks can be classified based on performance or purpose.

Performance-based classification

Cloud disks are divided by performance into ultra disks and standard SSDs.

- Ultra disks are ideal for medium I/O load scenarios and provide a random IOPS performance of up to 3,000 for ECS instances.
- Standard SSDs are ideal for I/O-intensive scenarios and provide stable and high random IOPS performance.

The following table provides a comparison of the performance of standard SSDs and ultra disks.

Item	Standard SSD	Ultra disk
Maximum capacity of a single disk	32,768 GiB	32,768 GiB
Maximum IOPS	25,000	5,000
Maximum throughput	300 MB/s	140 MB/s
Formula for calculating the IOPS per disk	min{1800 + 30 × Capacity, 25000}	min{1800 + 8 × Capacity, 5000}
Formula for calculating the throughput per disk	min{120 + 0.5 × Capacity, 300} MB/s	min{100 + 0.15 × Capacity, 140} MB/s
Data durability	99.9999999%	99.9999999%
Single-channel random write access latency	0.5 ms to 2 ms	1 ms to 3 ms
API operation	cloud_ssd	cloud_efficiency
Scenario	Small and medium-sized development and test applications that require high data durability	● Development and test applications ● System disks

Purpose-based classification

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

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

2.4.2.2.3. Shared Block Storage

Shared Block Storage is a block-level data storage service that supports concurrent read and write operations on multiple ECS instances and offers high performance and high reliability.

A single Shared Block Storage device can be attached to a maximum of four ECS instances. Shared Block Storage devices can be used only as data disks and must be created separately. Shared access is allowed. You can configure Shared Block Storage devices to be released when the associated ECS instances are released.

Shared Block Storage devices can be divided into the following types based on performance:

- **SSD Shared Block Storage device**: uses SSDs as the storage medium to provide stable and high performance storage that offers enhanced random I/O and data reliability.
- **Ultra Shared Block Storage device**: uses a hybrid SSD and HDD storage medium.

An ECS instance can have up to 16 data disks including cloud disks and Shared Block Storage devices. In other words, the sum of cloud disks and Shared Block Storage devices used as data disks cannot exceed 16.

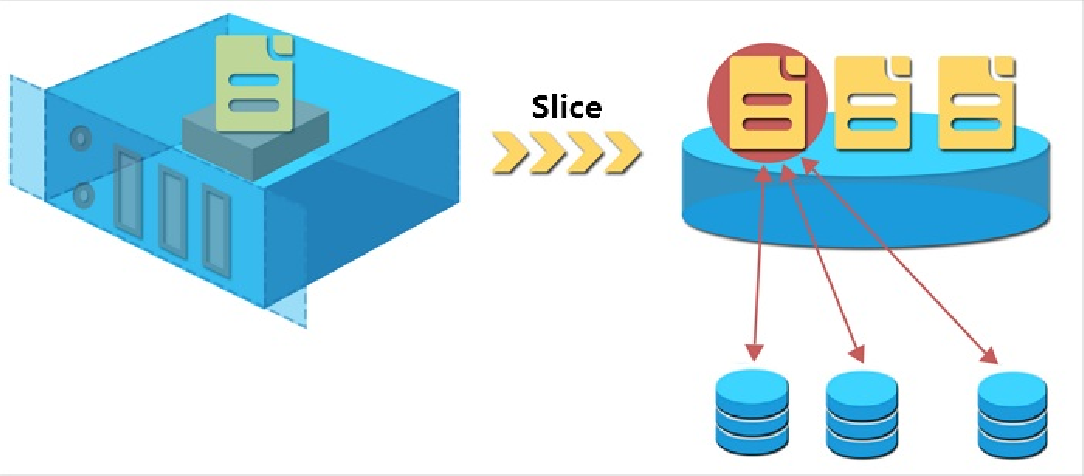
2.4.2.2.4. Triplicate storage

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

Chunks

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

Triplicate backup



How triplicate technology works

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

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

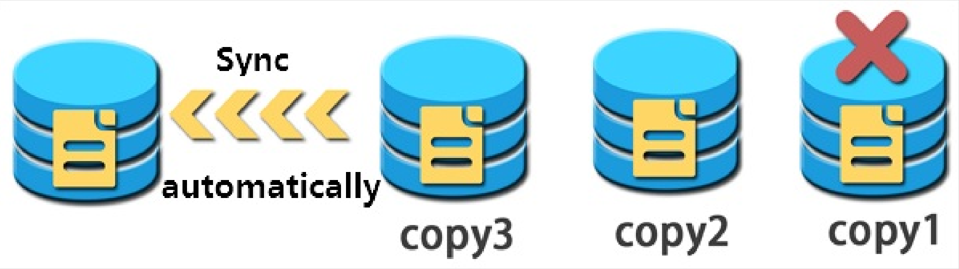
returned.

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

Data protection mechanism

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

Automatic replication



All user-level operations for data on cloud disks are synchronized across the three chunk replicas at the underlying layer. Operations that are synchronized include adding, modifying, and deleting data. This mode ensures the reliability and consistency of user data.

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

2.4.2.2.5. Erasure coding

Erasure coding (EC) can improve storage reliability. Compared to triplicate storage, EC can provide higher data reliability at lower data redundancy levels.

What is EC?

EC involves the following concepts:

- Data fragments (m): Data is divided into m data fragments.
- Parity fragments (n): n parity fragments are computed from the m data fragments.

The m data fragments and n parity fragments compose an erasure coding group. The data fragments and parity fragments are located on different servers. When n or less than n segments are lost, the lost segments can be restored based on the erasure coding algorithm. Both m and n are configurable. The typical configuration for Apsara Stack is 8 + 3, with the number of servers being no less than 14.

Comparison between EC and triplicate storage

Compared to triplicate storage, EC is a better solution in terms of storage usage and data reliability.

Item	EC	Triplicate storage
Storage usage	`m/(m + n)` : When m is 8 and n is 3, the storage usage is calculated based on the following formula: 8/(8 + 3) = 72.7%.	1/3 = 33.3%
Reliability	Allows up to n fragments to be lost. Failures on up to n servers are allowed in the worst case. For example, when m is 8 and n is 3, failures on up to three servers are allowed.	Allows up to two replicas to be lost. Failures on up to two servers are allowed in the worst case.

2.4.2.3. ECS disk encryption

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

With ECS disk encryption, you do not need to create or maintain your own key management infrastructure, change existing applications and maintenance procedures, or perform additional encryption operations. Disk encryption does not have any negative impact on your business processes. The following types of data can be encrypted:

- Data on cloud disks.
- Data transmitted between cloud disk and instances. Data is not encrypted again within the instance operating system.
- All snapshots created from encrypted cloud disks. Such snapshots are encrypted snapshots.

Data transmitted from ECS instances to cloud disks is encrypted on the hosts where the ECS instances are deployed.

Note ECS disk encryption supports Chinese cryptographic algorithms.

All available cloud disks (basic disks, ultra disks, and SSD disks) and Shared Block Storage devices (ultra Shared Block Storage devices and SSD Shared Block Storage devices) in Apsara Stack ECS can be encrypted.

2.4.2.4. Local storage

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

Local storage provides block-level data access for instances with high random IOPS, high throughput, and low latency. The reliability of data stored in local disks depends on the reliability of the physical server to which the disks are attached. This is a single point of failure risk which may cause data loss. We recommend that you implement data redundancy at the application layer to ensure the availability of the data.

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

Local disk types

Apsara Stack provides two types of local disks:

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

2.4.3. Images

An image is a template for running environments within ECS instances. An image includes an operating system and pre-installed software.

An image works as a copy that stores data from one or more disks. An ECS image may store data from a system disk or from both system and data disks. You can use an image to create an ECS instance or replace the system disk of an ECS instance.

Image types

ECS provides a variety of image types for you to access image resources.

Image description

Type	Description
Public image	Public images provided by Apsara Stack support Windows and most popular versions of Linux operating systems, including: • Windows • CentOS • CoreOS • Debian • Gentoo • FreeBSD • OpenSUSE • SUSE Linux • Ubuntu
Custom image	Custom images created based on your existing physical servers, virtual machines, or cloud hosts. This image type is flexible enough to meet all of your specific business needs.

Obtain an image

You can use one of the following methods to obtain images:

- Create a custom image based on an existing ECS instance.
- Choose an image shared by another Apsara Stack tenant account.
- Import an offline image file to an ECS cluster to generate a custom image.
- Copy a custom image to another region to implement unified deployment of environments and applications across regions.

Image formats

ECS supports images in the VHD, RAW, and qcow2 formats. Images in other formats must be converted to the supported formats before they can be run in ECS. For more information, see Convert image formats in *ECS User Guide*.

2.4.4. Snapshots

2.4.4.1. Overview

A snapshot is a copy of data on a cloud disk at the point in time that the snapshot is created.

You can use snapshots in scenarios such as environment replication and disaster recovery:

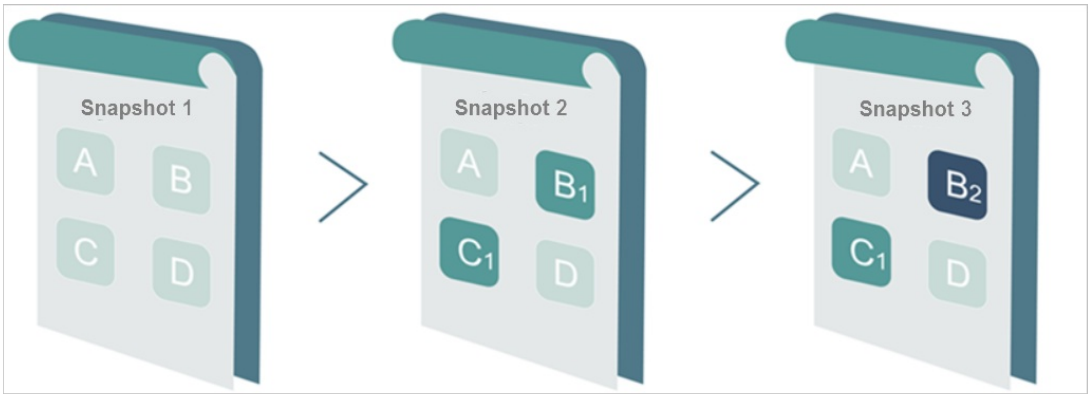
- You may want to use the data of one disk as the basis to write or store data to a different disk. To achieve this, you can create a snapshot for a cloud disk and then create another cloud disk from the snapshot. The new disk contains the basic data of the original disk.
- While cloud disks are a secure way to store data, their data may be subject to errors caused by application errors or malicious read and write operations and requires additional safeguard mechanisms. You can create snapshots at regular intervals to restore data to a previous point in time in case of data errors.

2.4.4.2. Mechanisms

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

Snapshots are created incrementally such that only data changes between two snapshots are copied instead of all of the data, as shown in Snapshots.

Snapshots



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

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

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

2.4.4.3. Specifications of ECS Snapshot 2.0

Built on the features of the original snapshot service, the ECS Snapshot 2.0 data backup service provides a higher snapshot quota and a more flexible automatic snapshot policy. This service has less impact on business I/O.

Comparison of snapshot specifications

Item	Traditional snapshot specification	Snapshot 2.0 specification	Benefit	Example
Snapshot quota	Maximum allowable number of snapshots: Number of disks × 6 + 6.	Each disk can have up to 64 snapshots.	Longer protection cycle and smaller protection granularity.	• A snapshot is created for the data disks of non-core business at 00:00 every day. Snapshots taken within the last two months are retained. • A snapshot is created for the data disks of core business every four hours. Snapshots taken within the last ten days are retained.

Item	Traditional snapshot specification	Snapshot 2.0 specification	Benefit	Example
Automatic snapshot policy	By default, the task is scheduled to be triggered once a day and cannot be modified manually.	You can customize the time of day and days of the week that snapshots are scheduled to be created and the retention period of snapshots. The disk quantity and related details associated with an automatic snapshot policy can be queried.	More flexible protection policy.	• You can schedule snapshots to be created on the hour several times in a single day. • You can specify the days of the week for which to create snapshots. • You can specify the snapshot retention period or choose to retain a snapshot permanently. When the number of automatic snapshots reaches the upper limit, the oldest automatic snapshot will be automatically deleted.
Implementation	Copy-on-write (COW)	Redirect-on-write (ROW)	Mitigates the impact of snapshot tasks on business I/O performance.	Snapshots can be taken at any time without interruptions to your business.

2.4.4.4. Technical comparison

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

Comparison of technical advantages

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

2.4.5. Deployment sets

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

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

- **Improve business reliability**

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

- **Improve network performance**

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

Deployment granularities and policies

- **Deployment granularities**
  - Host: indicates physical-server-level scheduling.
  - Rack: indicates rack-level scheduling.
  - Switch: indicates Switch-level scheduling.
- **Deployment policies**
  - LooseAggregation
  - StrictlAggregation
  - LooseDispersion
  - StrictDispersion

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

Granularities and policies lists the deployment policies and business scenarios corresponding to each deployment granularity.

Granularities and policies

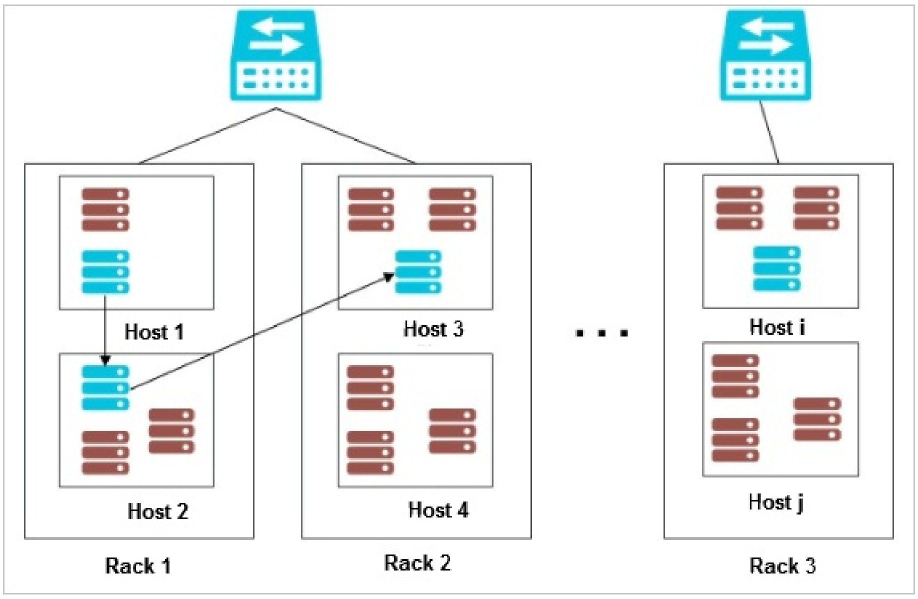
Deployment granularity	Deployment policy	Business scenario
Host	StrictDispersion	General purposes
	LooseDispersion	
Rack	StrictDispersion	Big data and databases
	LooseDispersion	Game customers
Switch	StrictDispersion	VPN
	LooseDispersion	Game customers
	StrictAggregation	Big data and databases
	LooseAggregation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

Typical examples

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

Typical example

Typical example



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

2.4.6. Network and security

2.4.6.1. IP addresses of ECS instances of VPC type

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

IP address types

ECS instances have the following IP address types:

- **Private IP addresses**

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

- **Elastic IP (EIP)**

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

Scenarios

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

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

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

2.4.6.2. Elastic network interfaces

This topic describes the concepts, scenarios, types, attributes, and limits of elastic network interfaces (ENIs).

What is ENI?

An ENI is a virtual network interface controller (NIC) that can be bound to a VPC-type ECS instance. You can use ENIs to deploy high availability clusters and perform low-cost failovers and fine-grained network management.

Scenarios

ENIs can be used in the following scenarios:

- Deployment of high availability clusters

  An ENI can meet the demand for multiple NICs on a single instance within a high availability architecture.

- Low-cost failover

  You can unbind an ENI from a failed ECS instance and bind it to another normal instance to quickly redirect traffic destined for the failed instance to the normal instance and immediately restore the service.

- Fine-grained network management

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

ENI types

ENIs are classified into two types:

- Primary ENI

  A primary ENI is created by default when an instance is created in a VPC. The lifecycle of the primary ENI is the same as that of the instance and you cannot unbind the primary ENI from the instance.

- Secondary ENI

  You can create a secondary ENI and bind it to or unbind it from an instance. The maximum number of ENIs that can be bound to a single instance varies with the instance type. For more information, see Instance families.

ENI attributes

The following table describes the attributes of an ENI.

Attribute description

Attribute	Quantity
Primary private IP address	1
MAC address	1

Attribute	Quantity
Security group	1 to 5
Description	1
ENI name	1

Limits

ENIs have the following limits:

- A single account can own up to 100 ENIs within a single region.
- The ECS instance must be in the same zone of the same VPC as the ENI, but does not have to use the same VSwitch.
- For instance types that support ENIs and the number of ENIs supported by each instance type, see Instance types.
- Binding multiple ENIs does not increase the instance bandwidth.

> ? **Note** The instance bandwidth varies with the instance type.

2.4.6.3. Internal network

If you need to transmit data between two ECS instances within the same region, we recommend that you transmit data over the internal network. ECS instances can connect to ApsaraDB for RDS, SLB, and OSS over the internal network.

In the internal network, each non-I/O optimized instance has a shared bandwidth of 1 GiB and each I/O optimized instance has a shared bandwidth of 10 GiB. The internal network is a shared network. Therefore, the bandwidth may fluctuate.

> ? **Note** Most mainstream instances are I/O optimized instances, and the actual bandwidth is related to the physical hardware.

ECS instances can communicate with RDS instances, SLB instances, and OSS buckets within the same region over the internal network.

The following rules apply to VPC-type ECS instances in the internal network:

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

2.4.6.4. Security group rules

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

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

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

2.5. Scenarios

ECS instances can be used either independently as simple web servers or with other Apsara Stack services such as OSS and CDN to provide advanced multimedia solutions. The following sections describe the typical application scenarios of ECS instances:

Official websites for enterprises and simple web applications

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

Multimedia and high-traffic applications or websites

When you use ECS instances together with OSS, you can store static images, videos, and downloaded packages in OSS to reduce storage costs. You can also use ECS in combination with CDN or SLB to shorten user response time, reduce bandwidth fees, and improve availability.

Applications or websites that have large traffic fluctuations

Some applications and websites may encounter large fluctuations in traffic within a short period of time. ECS provides elastic processing capabilities. The number of ECS instances automatically increases or decreases in response to changes in traffic to meet resource requirements and preserve cost efficiency. ECS can be used in combination with SLB to implement a high availability architecture.

Databases

Databases with high I/O requirements are supported. High-configuration I/O optimized ECS instances can be used together with standard SSDs to support high I/O concurrency and higher data reliability. Alternatively, multiple low-configuration I/O optimized ECS instances can be used in combination with SLB to implement a high availability architecture.

2.6. Limits

This topic describes the limits of ECS.

ECS has the following limits:

- ECS instances with 4 GiB or higher memory must use a 64-bit operating system. A 32-bit operating system can address a maximum of 4 GiB of memory.
- A 32-bit Windows operating system can use a maximum of four cores in its CPU.

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

The following table lists additional limits of ECS.

Other limits

Type	Description
Instance type	For more information, see Instance families and Instance types.
Block Storage	● Limits on specifications ○ Number of system disks per instance: 1. ○ Number of data disks per instance: 16. ○ Maximum number of instances to which a single Shared Block Storage device can be attached: 4. ○ System disk capacity: 40 GiB to 500 GiB. ○ Capacity of a single basic disk: 5 GiB to 2,000 GiB. ○ Capacity of a single SSD disk: 20 GiB to 32,768 GiB. ○ Capacity of a single ultra disk: 20 GiB to 32,768 GiB. ○ Total capacity of a single ultra Block Storage device: 32,768 GiB. ● Service limits ○ Only data disks can be encrypted. System disks cannot be encrypted. ○ Unencrypted disks cannot be directly converted into encrypted disks. ○ Encrypted disks cannot be directly converted into unencrypted disks. ○ Unencrypted snapshots cannot be directly converted into encrypted snapshots. ○ Encrypted snapshots cannot be directly converted into unencrypted snapshots. ○ Images with encrypted snapshots cannot be shared. ○ Images with encrypted snapshots cannot be exported.
Snapshot quota	Number of disks × 64.
Image	● Maximum number of users to whom a single image can be shared: 50. ● Instance types with 4 GiB or higher memory do not support 32-bit images.

Type	Description
Security group	- A single security group can contain a maximum of 1,000 instances. If more than 1,000 instances need access to each other over the internal network, you can distribute them to different security groups and authorize mutual access among the security groups. - Each instance can belong to a maximum of five security groups. - Each user can have a maximum of 100 security groups. - Each security group can have a maximum of 100 security group rules. - Modifications to security groups do not affect your services. - Security groups are stateful. If a security group permits outbound traffic over a connection, it also permits inbound traffic over this connection.
Elastic network interface	The number of elastic network instances that can be bound to different instance type families. For more information, see Instance types.
User data	ECS user data supports VPC-type I/O optimized instances. You can use the user data when you create such instances. Because user data depends on the cloud-init service, cloud-init must be installed in the image.

2.7. Terms

This topic describes the basic terms in ECS to help you better understand ECS.

ECS

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

instance

An independent resource entity that contains basic resource elements.

security group

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

image

A template for running environments in ECS instances. An image includes an operating system and pre-installed software. Images can be divided into public images and custom images. You can use an image to create an ECS instance or replace the system disk of an ECS instance.

snapshot

Data backup of a disk at a certain point in time. Snapshots consist of automatic snapshots and user-created snapshots.

cloud disk

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

Block Storage

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

throughput

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

performance test

A world-leading SaaS performance test platform with powerful distributed stress test capabilities. It can simulate real business scenarios with a large number of users to find all application performance problems.

virtual private cloud (VPC)

A virtual private cloud built and customized based on Apsara Stack. Full logical isolation is implemented between VPCs. Users can create and manage cloud service instances, such as ECS instances, SLB instances, and RDS instances in their own VPCs.

internal endpoint

A service connection address for clients that use private IP addresses as their source.

GPU-accelerated instance

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

3.Container Service for Kubernetes

3.1. What is Container Service?

Container Service provides high-performance, enterprise-class management for scalable Kubernetes-based containerized applications throughout the application lifecycle.

Container Service simplifies the creation and scaling of container management clusters. It integrates Apsara Stack virtualization, storage, network, and security capabilities, providing the optimal environment to run Kubernetes-based containerized applications in the cloud. Alibaba Cloud is a Kubernetes certified service provider, with Container Service being among the first services to pass the Certified Kubernetes Conformance Program. Container Service provides professional container support and services.

3.2. Benefits

Overview

Easy to use

- You can easily create Kubernetes clusters in the Container Service console.
- You can easily upgrade Kubernetes clusters in the Container Service console.

  When you use custom Kubernetes clusters, you may need to handle clusters of different versions. Currently, each time you upgrade the clusters, you need to make major adjustments and high operation and maintenance costs are incurred. Container Service allows you to perform rolling upgrades based on images and supports full metadata backups. You can easily roll back clusters to previous versions.
- Allows you to easily scale Kubernetes clusters in the Container Service console.

  Kubernetes clusters enable you to quickly scale up or down applications to handle traffic fluctuations in a timely manner.

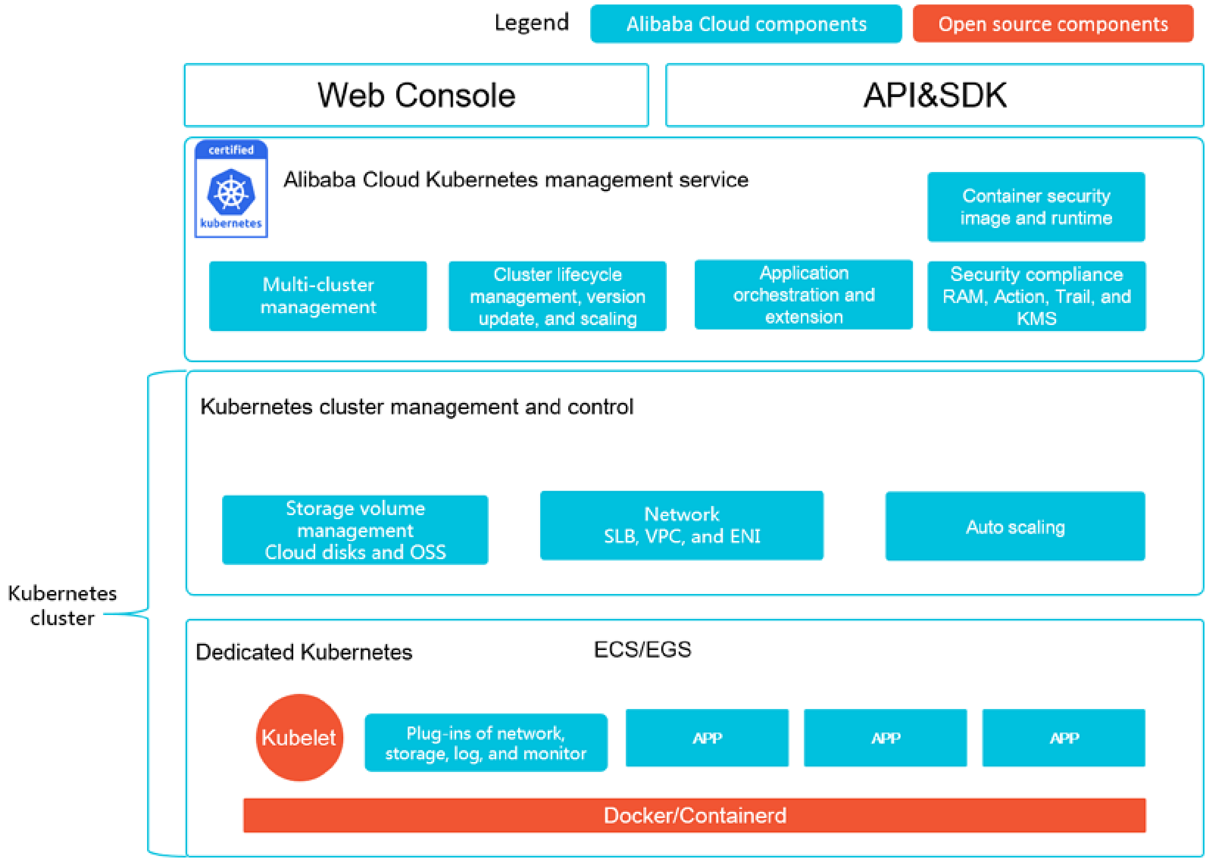
Features

Feature	Description
Network	Supports continuous network integration to optimize network performance.

Feature	Description
Load balancing	Allows you to create public and internal SLB instances. If you use an Ingress to control access to your Kubernetes cluster, frequent service releases may negatively affect the performance of the Ingress and increase the error rate. Container Service allows you to create SLB instances, which provide high availability load balancing and can automatically modify network configurations to suit your business needs. This solution is adopted by a large number of users and has been proven to be a more stable and reliable alternative to Ingresses.
Storage	Supports Apsara Stack cloud disks, Network Attached Storage (NAS), and Block Storage, and provides FlexVolume drivers. Supports seamless integration with cloud storage services for custom Kubernetes clusters that cannot use cloud storage resources.
O&M	• Supports integration with Apsara Stack Log Service. • Supports automatic scaling.
Image repository	• Provides high availability and high concurrency. • Supports accelerated image retrieval. • Supports peer-to-peer image distribution. Custom image repositories may stop responding when millions of clients attempt to pull images at the same time. Container Service provides an image repository system that offers enhanced reliability and reduces O&M and upgrade costs.
Stability	• Dedicated support teams guarantee the stability of containers. • All Linux and Kubernetes versions must pass rigorous testing before they are available to the public. Container Service supports Docker CE and provides a Docker community to help you communicate with other Docker enthusiasts and solve problems. Best practices are provided to help you address issues, such as network interruptions, kernel incompatibilities, or Docker crashes.

Feature	Description
Technical support	- Allows you to quickly upgrade Kubernetes clusters to the latest version. - Provides professional technical support services to help you solve the issues that may occur when you use containers.

3.3. Architecture


Legend
Alibaba Cloud components
Open source components
Web Console
API&SDK
certified
kubernetes
Alibaba Cloud Kubernetes management service
Container security image and runtime
Multi-cluster management
Cluster lifecycle management, version update, and scaling
Application orchestration and extension
Security compliance RAM, Action, Trail, and KMS
Kubernetes cluster management and control
Storage volume management Cloud disks and OSS
Network SLB, VPC, and ENI
Auto scaling
Kubernetes cluster
Dedicated Kubernetes
ECS/EGS
Kubelet
Plug-ins of network, storage, log, and monitor
APP
APP
APP
Docker/Containerd

Container Service is adapted and enhanced on the basis of native Kubernetes. This service simplifies cluster creation and scaling and integrates Apsara Stack virtualization, storage, network, and security capabilities, providing the optimal environment to run Kubernetes-based containerized applications in the cloud.

Feature	Description
Dedicated Kubernetes mode	Integrated with Apsara Stack virtualization technologies, the service allows you to create dedicated Kubernetes clusters. ECS, Elastic GPU Service (EGS), and ECS Bare Metal instances can all be used as cluster nodes. Instances support a wide range of plug-ins and can be flexible configured to different specifications.

Feature	Description
Alibaba Cloud Kubernetes cluster management and control service	The service provides powerful network, storage, cluster management, scaling, and application extension features.
Alibaba Cloud Kubernetes management service	The service supports secure images and is highly integrated with Apsara Stack Resource Access Management (RAM), Key Management Service (KMS), and logging and monitoring services to provide a secure and compliant Kubernetes solution.
Convenient and efficient use	Container Service for Kubernetes provides services through the Web console, APIs, and SDKs.

3.4. Features

Features

Cluster management

- With the Container Service console, you can easily create a classic dedicated Kubernetes cluster supporting GPU servers within 10 minutes.
- Provides container-optimized OS images as well as Kubernetes and Docker versions that have undergone **stability testing and security enhancement**.
- Supports multi-cluster management, cluster upgrades, and cluster scaling.

Provides end-to-end container lifecycle management

- **Network**

  Provides high performance VPC and elastic network interface (ENI) plug-ins optimized for Apsara Stack, boasting 20% increased performance compared with regular network solutions.

  Supports container access and throttling policies.

- **Storage**

  Container Service is integrated with Apsara Stack disks and OSS, and provides the standard FlexVolume drive.

  Supports real-time creation and migration of volumes.

- **Logs**

  Provides high-performance log collection integrated with Apsara Stack Log Service.

  Supports the integration with third-party open-source logging solutions.

- **Monitoring**

  Supports both container-level and VM-level monitoring. Integration with third-party open-source monitoring solutions is supported.

- **Permissions**

  Supports cluster-level Resource Access Management (RAM).

  Supports application-level permission configuration management.

- Application management

  Supports phased release and blue-green release.

  Supports application monitoring and scaling.

High-availability scheduling policies that allow you to easily handle upstream and downstream delivery processes

- Supports service-level affinity policies and scale-out.
- Provides high availability and disaster recovery across zones.
- Provides cluster and application management APIs to easily implement continuous integration and private system deployment.

3.5. Scenarios

DevOps continuous delivery

Optimized continuous delivery pipeline

Container Service works with Jenkins to automate the DevOps pipeline, from code submission to application deployments. The service ensures that code is only submitted for deployment after passing automated testing, and provides a better alternative to traditional delivery models that involve complex deployments and slow iterations.

Benefits

- DevOps pipeline automation

  Automates the DevOps pipeline, from code updates to code builds, image builds, and application deployments.

- Consistent environment

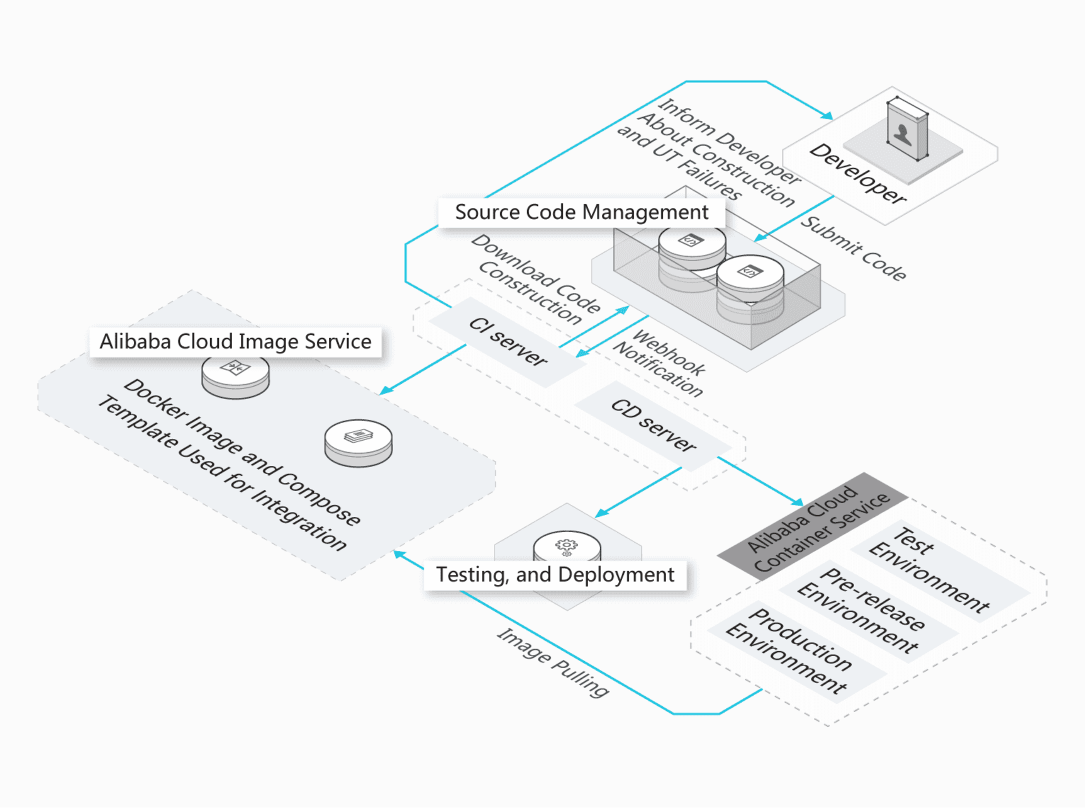
  Allows you to deliver code and runtime environments based on the same architecture.

- Continuous feedback

  Provides immediate feedback on each integration or delivery.

Related products and services

ECS + Container Service


Inform Developer About Construction and UT Failures
Developer
Submit Code
Source Code Management
Download Code Construction
Webhook Notification
CI server
CD server
Alibaba Cloud Image Service
Docker Image and Compose Template Used for Integration
Alibaba Cloud Container Service
Test Environment
Pre-release Environment
Production Environment
Testing, and Deployment
Image Pulling

Machine learning based on cloud-native technology

Enables rapid application developments with a focus on machine learning

Container Service allows data engineers to easily develop and deploy machine learning applications in heterogeneous computing clusters. Integrated with multiple distributed storage systems, the service supports faster read and write speeds to facilitate the testing, training, and release of data models. You can focus on your core business operations instead of worrying about the deployment and maintenance process.

Benefits

- Ecosystem support

  Supports mainstream deep learning frameworks, such as TensorFlow, Caffe, MXNet, and Pytorch, and offers optimized features of these frameworks.

- Quick start and elastic scaling

  Provides machine learning services for development, training, and inference. Supports the startup of training and inference tasks within seconds, and elastic scaling of GPU resources.

- Easy to use

  Allows you to easily create and manage large-scale GPU clusters and monitor core metrics, such as GPU utilization.

- Deep integration

  Seamless integration with Apsara Stack storage, logging and monitoring, and security infrastructure capabilities.

Related products and services

ECS/EGS/HPC + Container Service + OSS/NAS/CPFS

Microservices architecture

Agile development and deployment to speed up the evolution of business models

In the production environment, you can split your system into microservices and use Apsara Stack image repositories to store these microservice applications. Apsara Stack can schedule, orchestrate, deploy, and implement phased releases of microservice applications while you focus on feature updates.

Benefits

- Load balancing and service discovery

  Forwards layer 4 and layer 7 requests and binds the requests to backend containers.

- Multiple scheduling and disaster recovery policies

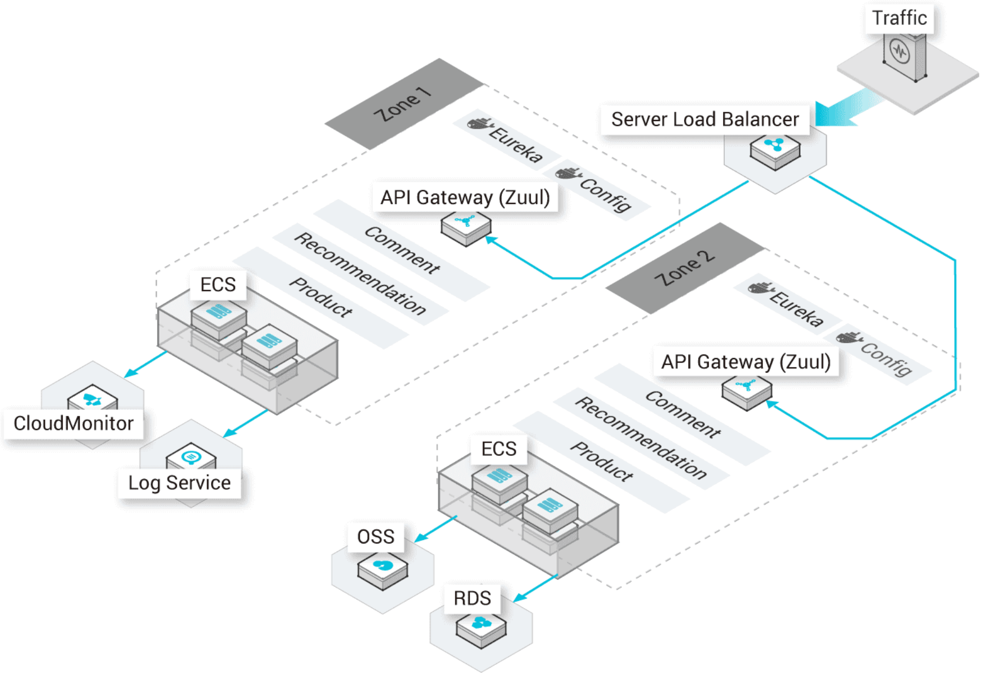
  Supports different levels of affinity scheduling policies, and cross-zone high availability and disaster recovery.

- Microservices monitoring and auto scaling

  Supports microservice and container monitoring, and microservice auto scaling.

Related products and services

ECS + ApsaraDB RDS + OSS + Container Service



Hybrid cloud architecture

Unified O&M of cloud resources

You can centrally manage cloud and on-premises resources in the Container Service console. Containers hide the differences between infrastructures. This enables you to use the same images and orchestration templates to deploy applications in the cloud and on premises.

Benefits

- Application scaling in the cloud

  During peak hours, Container Service can scale up applications in the cloud and forward traffic to the scaled-up resources.

- Disaster recovery in the cloud

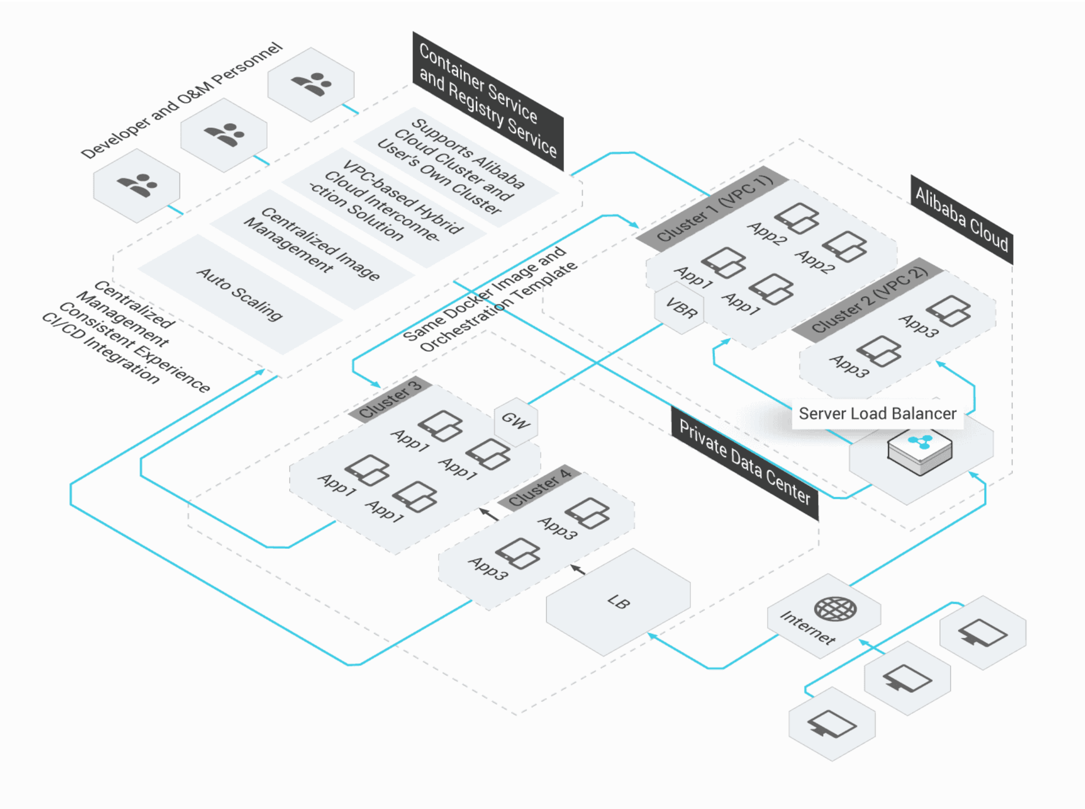
  Business systems can be deployed on premises for service provisioning and in the cloud for disaster recovery.

- On-premises development and testing

  Applications that are developed and tested on premises can be seamlessly released to the cloud.

Related products and services

ECS + VPC + Express Connect



Automatic scaling architecture

Traffic-based scalability

Container Service enables businesses to auto-scale their resources based on traffic. This prevents traffic spikes from bringing down your system and eliminates idle resources during off-peak hours.

Benefits

- Quick response

Container scale-out can be triggered within seconds when traffic reaches the scale-out threshold.

- Auto scaling

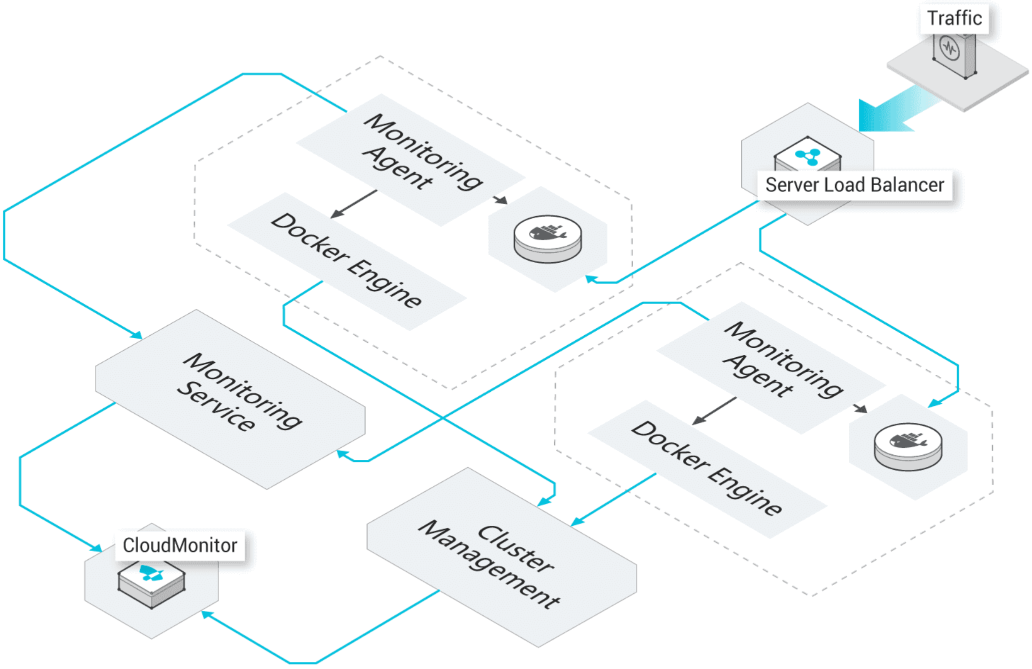
  The scaling process is fully automated without human interference.

- Low cost

  Containers are automatically scaled in when traffic decreases to avoid resource waste.

Related products and services

ECS + CloudMonitor



3.6. Limits

Limits for Kubernetes clusters

Limit	Description
Cluster	• You can create up to 50 clusters across all regions for each account. A cluster can contain up to 40 nodes. To create more clusters or nodes, submit a ticket. • Kubernetes clusters only support Linux containers. • Kubernetes clusters only support VPCs. When creating a Kubernetes cluster, you can either create a new VPC or use an existing one.

Limit	Description
ECS instance	- Only the CentOS operating system is supported. - Limits for adding an existing ECS instance: ◦ The ECS instance to be added must be in the same VPC as the cluster. ◦ The ECS instance to be added must belong to the same account as the cluster.
Cluster scale-in and scale-out	- The number of worker nodes must be within the range of 1 to 5. - Clusters must be manually scaled in or out. Automatic scaling is not supported. - Master nodes in a Kubernetes cluster cannot be scaled out automatically. - Based on the rules of Resource Orchestration Service (ROS), nodes that were created automatically during cluster creation and nodes that were manually added to a cluster will not be removed when you scale in the cluster. Only nodes you added when scaling out the cluster will be removed. Nodes are removed from the cluster in reverse order to when they were added to the cluster during cluster scale-out. Newly added nodes are reclaimed first.

3.7. Terms

cluster

A collection of cloud resources that are required to run containers. Several cloud resources, such as ECS instances, SLB instances, and VPCs, are associated together to form a cluster.

node

A server that has a Docker engine installed and is used to deploy and manage containers. A node can be either an ECS instance or a physical server. The Container Service Agent program is installed on a node and registered to a cluster. The number of nodes in a cluster can be scaled based on your requirements.

container

A runtime instance created from a Docker image. A single node can run multiple containers.

image

A standard packaging format of a containerized application in Docker. An image from the Docker Hub, Alibaba Cloud Container Registry, or your own private registry can be specified to deploy its packaged containerized application. image ID An image ID is a unique identifier composed of the image repository URI and image tag. The latest image tag is used for the image ID by default.

Kubernetes terms

node

A worker server in a Kubernetes cluster. A node can be either a virtual server or a physical server. Pods always run on nodes. kubelet runs on each node in a cluster to manage containers in a pod and ensure that they are running properly.

namespace

A method used in Kubernetes to divide cluster resources between multiple users. By default, Kubernetes starts with three initial namespaces: default, kube-system, and kube-public. Administrators can also create new namespaces as required.

pod

The smallest deployable computing unit that can be created and managed in Kubernetes. A pod is a group of one or more containers that share storage and network resources and a common set of specifications for how to run the containers.

Replication Controller (RC)

A feature that monitors running pods to ensure that a specified number of pod replicas are running at any given time. One or more pod replicas can be specified. If the number of pod replicas is smaller than the specified value, an RC starts new pod replicas. If the number of pod replicas exceeds the specified value, the RC stops the redundant pod replicas.

Replica Set (RS)

The upgraded version of RC. Compared with RCs, RSs support more selector types. RS objects are not used independently, but are used as deployment parameters under ideal conditions.

deployment

An update operation performed on a Kubernetes cluster. Deployment is more widely applied than RS. You can use deployments to create, update, or perform rolling updates for services. A new RS is created when you perform a rolling update for a service. A compound operation is carried out to increase the number of replicas in the new RS to the desired value while decreasing the number of replicas in the original RS to zero. This kind of compound operation is better carried out by a deployment than through RS. We recommend that you do not manage or use the RS created by a deployment.

service

The basic operation unit of Kubernetes. It is an abstraction of real application services. Each service has multiple containers that support it. The Kube-Proxy port and service selector determine whether the service request is forwarded to the back-end container, and a single access interface is displayed externally. Back-end operations are invisible to users.

label

A collection of key-value pairs attached to resource objects. Labels are intended to specify identifying attributes of objects that are meaningful and relevant to users, but do not directly imply semantics to the core system. Labels can be attached to objects at creation time, and subsequently added and modified at any time. Each object can have a set of key/value labels, and each key must be unique for a specified object.

volume

Volumes in Kubernetes clusters are similar to Docker volumes. However, they are different in one key aspect. Docker volumes are used to persist data in Docker containers, while Kubernetes volumes share the same lifetime as the pods that enclose them. The volumes declared in each pod are shared by all containers in the pod. The actual back-end storage technology used is irrelevant when you use Persistent Volume Claim (PVC) logical storage. The specific configurations for Persistent Volume (PV) are completed by storage administrators.

PV and PVC

PVs and PVCs allow Kubernetes clusters to provide a logical abstraction over the storage resources, so that the actual configurations of back-end storage can be ignored by the pod configuration logic, and instead completed by the PV configurators. The relationship between PVs and PVCs is similar to that of nodes and pods. PVs and nodes are resource providers which can vary by cluster infrastructure, and are configured by the administrators of a Kubernetes cluster. PVCs and pods are resource consumers that can vary based on service requirements, and are configured by either the users or service administrators of a Kubernetes cluster.

Ingress

A collection of rules that allow inbound access to cluster services. An Ingress can be configured to provide services with externally-reachable URLs, load balance traffic, terminate SSL, and offer name-based virtual hosting. You can request the Ingress by posting Ingress resources to API servers. An Ingress controller is responsible for fulfilling an Ingress, usually with a load balancer. It can also be used to configure your edge router or additional frontends to help handle the traffic.

Related documents

- Docker glossary
- Kubernetes concepts

4.Auto Scaling (ESS)

4.1. What is ESS?

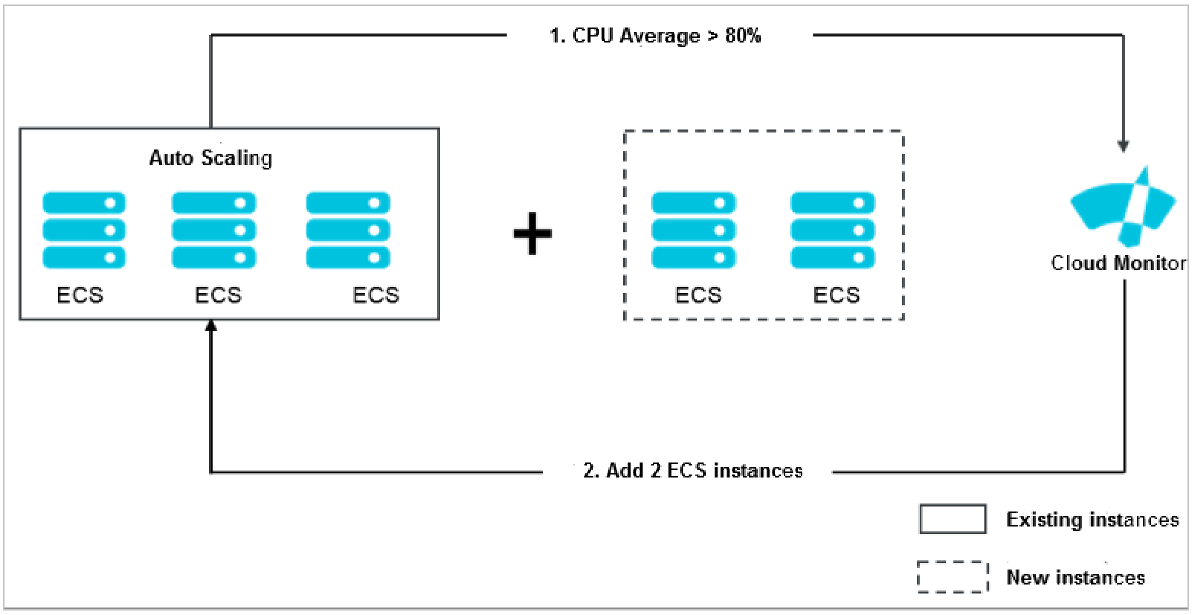
Auto Scaling (ESS) is a management service that automatically adjusts your elastic computing resources based on your business needs and policies.

When business loads increase, ESS automatically adds ECS instances based on the scaling rules that you configured to ensure sufficient computing capabilities. When business loads decrease, ESS automatically removes ECS instances to save costs.

ESS provides the following features:

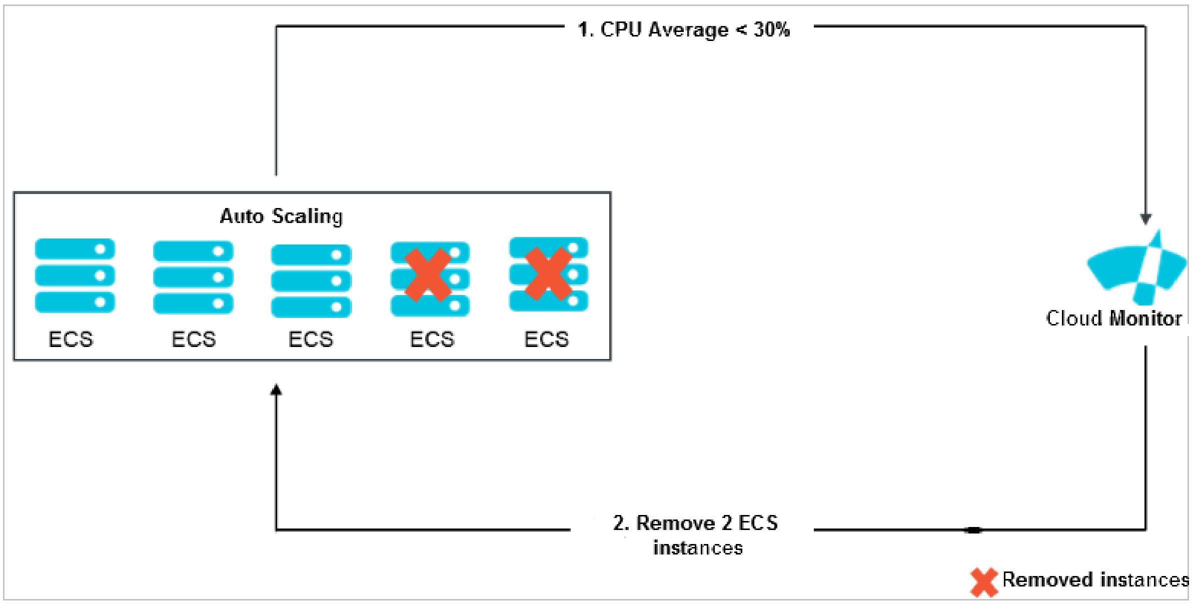
- Scale-out

  When business loads surge above normal loads, ESS automatically increases underlying resources. This helps maintain access speed and ensure that resources are not overloaded. For example, if the CPU utilization of ECS instances exceeds 80%, ESS scales out ECS resources based on the rules that you configured. During the scale-out event, ESS automatically creates and adds ECS instances to the scaling group, and adds the new instances to the backend server groups of the associated SLB instances and the whitelists of the associated ApsaraDB for RDS instances. The following figure shows the implementation of a scale-out event.



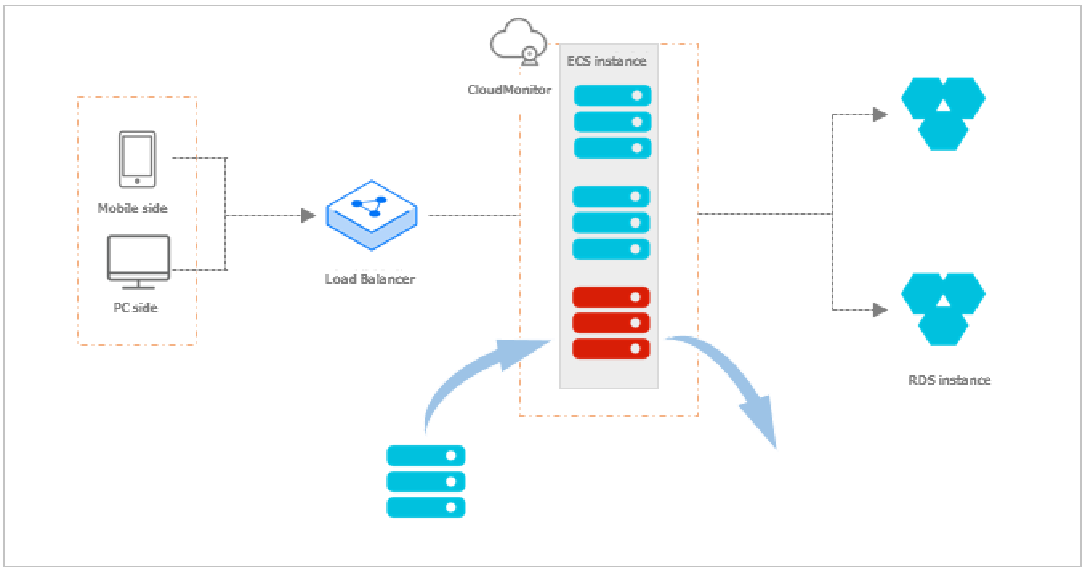
- Scale-in

  When your business loads decrease, ESS automatically releases underlying resources to prevent resource wastes and reduce costs. For example, if the CPU utilization of ECS instances in a scaling group is less than 30%, ESS automatically scales in ECS instances based on the scaling rules that you specified. During the scale-in event, ESS removes ECS instances from the scaling group and also from the backend server groups of the associated SLB instances and the whitelists of the associated ApsaraDB for RDS instances. The following figure shows the implementation of a scale-in event.



- Elastic recovery

  If an ECS instance in a scaling group is not in the Running state, ESS considers the instance to be unhealthy. If an ECS instance is considered unhealthy, ESS automatically releases the instance and creates a new one. This process is called elastic recovery. It ensures that the number of healthy ECS instances in a scaling group will not fall below the minimum number of ECS instances that you specified for the scaling group. The following figure shows the implementation of elastic recovery.



4.2. Benefits

Compared with manually managing ECS instances, ESS can help you reduce the infrastructure and O&M costs. This topic describes the benefits of ESS.

- Automatic scaling of ECS instances

  ESS automatically adds ECS instances during traffic peaks, and removes ECS instances when traffic loads drop. This helps lower infrastructure costs because you pay only for what you actually use.

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

ESS performs real-time monitoring on instances and automatically replaces unhealthy instances. This helps save O&M costs.

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

  ESS is integrated with Server Load Balancer (SLB) and ApsaraDB for RDS (RDS). ESS automatically manages SLB backend servers and RDS whitelists. This helps save O&M costs.

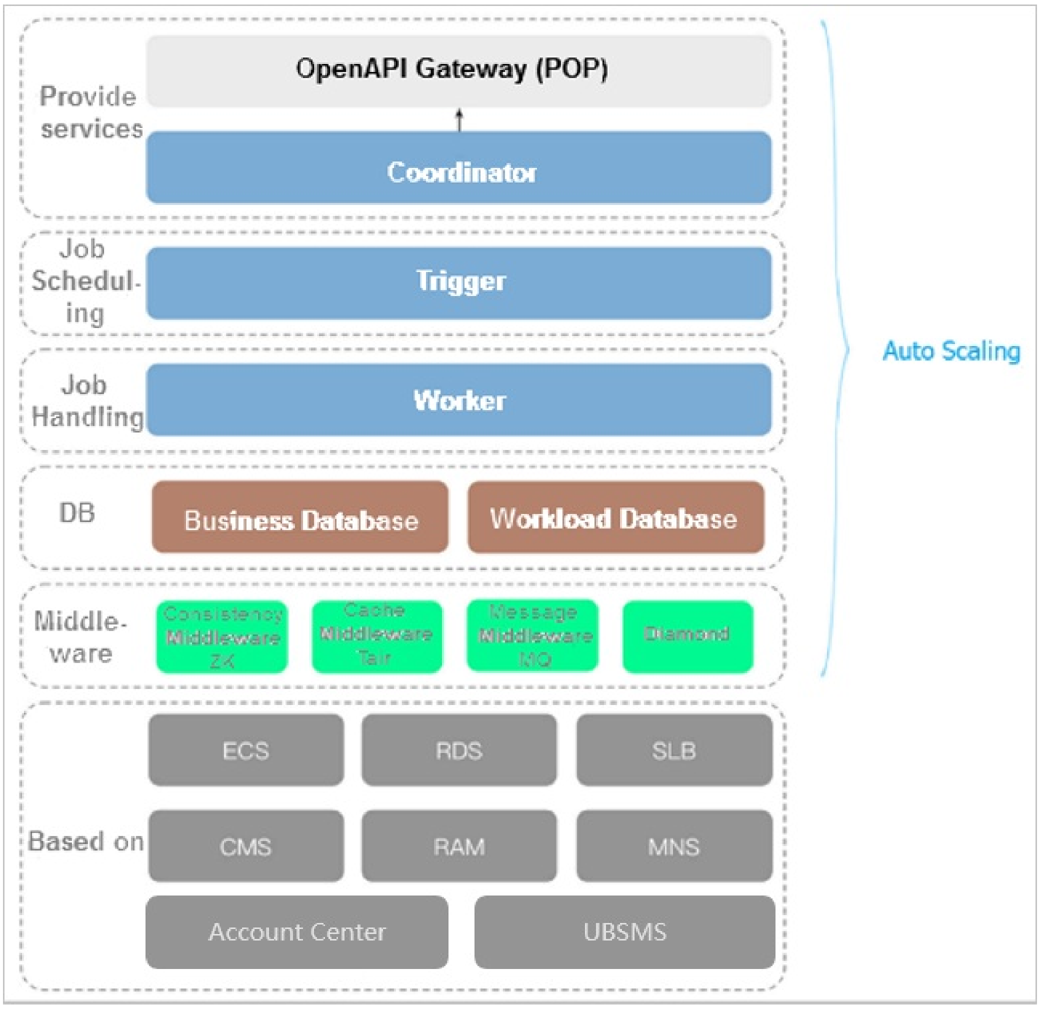
- Various scaling modes for you to mix and match

  ESS allows you to schedule, customize, and fix the minimum number of instances, as well as configure automatic replacement of unhealthy instances. It also provides API operations for you to monitor instances through external monitoring systems.

4.3. Architecture

This topic describes the architecture of Auto Scaling (ESS) and its components.

The following figure shows the ESS architecture.



The following table describes some of the components in the preceding figure.

Component	Description

Component	Description
Open API Gateway	Provides basic services such as authentication and parameter passthrough.
Coordinator	Serves as the ingress of the ESS architecture. It provides external management and control for services, processes API calls, and triggers tasks.
Trigger	Obtains information from health checks of instances and scaling groups, scheduled tasks, and Cloud Monitor to perform tasks scheduling.
Worker	Functions as the core part of ESS. After ESS receives a task, it processes the entire lifecycle of the task, including splitting the task, executing the task, and returning the execution results.
DB	Includes the business database and workload database.
Middleware layer	ZooKeeper: ensures consistency by implementing distributed locks for Server Controller.
	Tair: provides caching services for Server Controller.
	Message Queue (MQ): provides message queuing services of VM statuses.
	Diamond: manages persistent configurations.

4.4. Features

This topic describes the features of ESS.

- Adjust the number of ECS instances in various modes

  > **Note** ESS automatically removes ECS instances from scaling groups. Therefore, these instances cannot be used to store application status information such as sessions and related data such as databases and logs. If applications deployed on these ECS instances require data to be stored, you can store the status information on independent ECS instances, store databases in ApsaraDB for RDS, and store logs in Log Service.

  The scaling modes include:

  - Scheduled mode: configures periodic tasks to add or remove ECS instances at a specific point in time, such as 13:00 every day.
  - Dynamic mode: dynamically adjusts the number of ECS instances based on the monitoring metrics to ensure that the monitoring metric values are within the expected ranges.

  - Custom mode: manually adjusts the number of ECS instances based on your monitoring system statistics.
    - You can manually execute scaling rules.
    - You can manually add or remove ECS instances.
    - After you manually adjust the minimum (MinSize) and maximum (MaxSize) numbers of instances, ESS automatically creates or releases ECS instances to ensure that the number of instances remains between the minimum and the maximum number of instances.
  - Fixed-number mode: maintains a fixed number of healthy ECS instances by specifying the MinSize parameter. This mode can be used to ensure daily business availability.
  - Health mode: automatically removes or releases ECS instances that are considered unhealthy when they are not in the Running state.
  - Multi-mode: combines any of the preceding modes to meet your own business requirements. For example, if you predict that business peak hours are between 13:00 to 14:00, you can configure a scaling mode that creates 20 ECS instances at the scheduled time. If you are not sure whether the actual demand during peak hours will exceed the number of scheduled resources, you can configure another scaling mode to handle unexpected business loads. For example, the actual load requires 40 ECS instances.
- Automatically add or remove ECS instances to or from backend server groups of the associated SLB instances

  ESS automatically adds or removes ECS instances to maintain SLB backend servers and distribute access traffic.
- Automatically add or remove IP addresses of ECS instances to or from whitelists of the associated ApsaraDB for RDS instances

  ESS automatically adds or removes IP addresses of ECS instances to or from the whitelists of the associated ApsaraDB for RDS instances. This helps maintain the whitelists of the associated ApsaraDB for RDS instances and control access of ECS instances to the associated ApsaraDB for RDS instances.

4.5. Scenarios

ESS can be used in the following scenarios:

- Video streaming: Traffic loads surge during holidays and festivals. Cloud computing resources must be automatically scaled out to meet the increased demands.
- Live streaming and broadcast: Traffic loads are ever-changing and difficult to predict. Cloud computing resources must be scaled based on CPU utilization, application load, and bandwidth usage.
- Gaming: Traffic loads increase at 12:00 and from 18:00 to 21:00. Cloud computing resources must be scaled out on a regular basis.

4.6. Limits

This topic describes the limits of ESS.

- ESS does not support vertical scaling. It can only scale the number of ECS instances. The CPU, memory, and bandwidth configurations of ECS instances cannot be automatically adjusted.
- The following table describes the quantity limits that are applied to a scaling group.

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

4.7. Terms

This topic describes the common terms related to Auto Scaling (ESS).

Term	Description
Auto Scaling	Auto Scaling is a management service that automatically adjusts the number of elastic computing resources based on your business demands and policies. It automatically increases ECS instances during high business loads, and automatically releases ECS instances during low business loads.
scaling group	A scaling group is a group of ECS instances that are dynamically scaled based on the configured scenario. You can specify the minimum and maximum numbers of ECS instances in a scaling group, as well as the SLB and Apsara for RDS instances associated with the scaling group.
scaling configuration	Scaling configurations specify the configurations of ECS instances used for automatic scaling.
scaling rule	A scaling rule specifies a specific scaling activity, such as adding or removing N ECS instances.
scaling activity	After a scaling rule is triggered, a scaling activity is executed. A scaling activity shows the changes to the ECS instances in a scaling group.
scaling task	A scaling task is a task that triggers a scaling rule, such as a scheduled task.
cooldown period	The cooldown period indicates a period of time after the completion of a scaling activity in a scaling group. During this period, no other scaling activities can be executed.

5.Resource Orchestration Service (ROS)

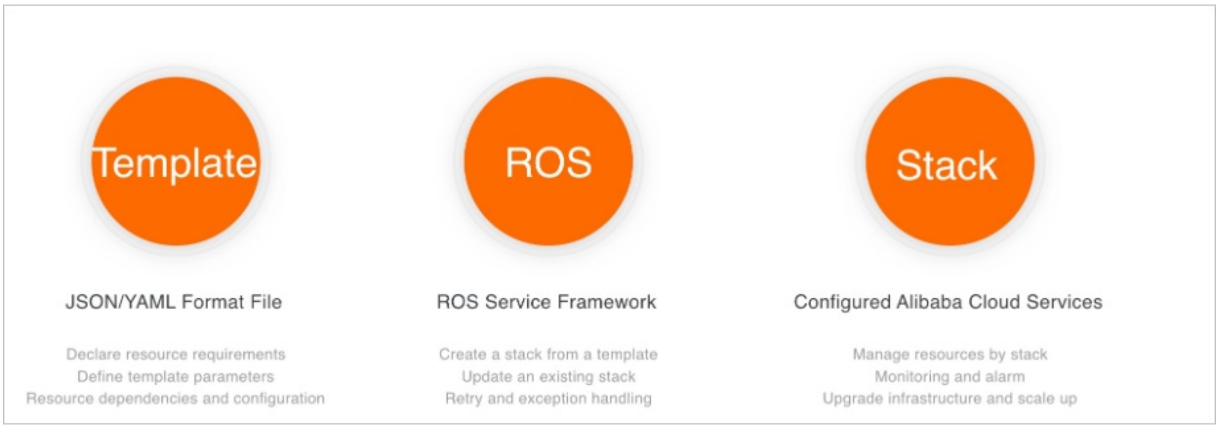
5.1. What is ROS?

Resource Orchestration Service (ROS) is a service provided by Alibaba Cloud to simplify the management of cloud computing resources. You can author stack templates based on the template specifications defined in ROS. Within a template, you can define required cloud computing resources such as ECS and ApsaraDB for RDS instances, and the dependencies between resources. The ROS engine automatically creates and configures all resources in a stack based on a template, making automatic deployment and O&M possible.

An ROS template is a readable, easy-to-author text file. You can directly edit a JSON-formatted template or use the Visual Editor available in the ROS console to edit the template. You can modify templates at any time. You can use version control tools such as SVN and Git to control the template and infrastructure versions. You can use APIs and SDKs to integrate the orchestration capabilities of ROS with your own applications to implement infrastructure as code.

ROS templates are also a standardized way to deliver resources and applications. If you are an independent software vendor (ISV), you can use ROS templates to deliver a holistic system and solution encompassing cloud resources and applications. ISVs can use this method to integrate Alibaba Cloud resources with their own software systems for centralized delivery.

ROS manages a group of cloud resources as a single unit called a stack. A stack is a group of Alibaba Cloud resources. You can create, delete, and clone cloud resources by stack. In DevOps practices, you can use ROS to clone the development, testing, and production environments, as well as migrate and scale out applications.


Template
ROS
Stack
JSON/YAML Format File
Declare resource requirements
Define template parameters
Resource dependencies and configuration
ROS Service Framework
Create a stack from a template
Update an existing stack
Retry and exception handling
Configured Alibaba Cloud Services
Manage resources by stack
Monitoring and alarm
Upgrade infrastructure and scale up

5.2. Benefits

Resource Orchestration Service (ROS) allows you to model and configure your Apsara Stack resources.

After you create a template that defines your required resources (such as ECS and RDS instances), ROS creates and configures these resources based on the template, facilitating resource management. ROS has the following benefits:

Infrastructure as Code

ROS is an Infrastructure as Code (IaC) solution provided by Alibaba Cloud to quickly implement IaC as a key component of DevOps.

Fully managed automation service

ROS is a fully managed service and does not require you to purchase the resources that are used to maintain your templates, allowing you to focus on maintaining the resources of your business and the template specifications. When you need to create multiple projects that are distributed across multiple stacks, managed automation of the creation process enables you to complete tasks faster. We recommend that you use ROS API operations to maintain stacks and use source code versioning software such as Git and SVN for centralized management of templates.

Repeatable deployment

You can use the same templates to deploy resources to the development, test, and production environments. You can set parameters to different values for different environments. For example, you can set the number of ECS instances in the test environment to 2 and the number of ECS instances in the production environment to 20. You can also use the same templates to deploy resources to multiple regions. This improves the efficiency of multi-region deployment.

Standardized deployment

In practice, subtle differences in different environments often lead to complicated management and high costs, prolong troubleshooting time, and interfere with the normal operation of your business. By using ROS for repeated deployment, you can standardize deployment environments, minimize the differences between different environments, and build environment configurations into templates. A rigorous management process similar to code implementation can ensure standardized deployment practices.

Unified authentication, security, and audit

Compared with other similar services, ROS provides better integration with other Apsara Stack services. Integration with Resource Access Management (RAM) provides unified authentication, eliminating the need to establish a separate user authentication system. Operations on all cloud services are called through APIs. You can use ActionTrail to review all O&M operations, including operations on ROS.

5.3. History

This topic describes the development history of Resource Orchestration Service (ROS).

- On December 21, 2015, ROS went into public beta.
- On May 20, 2016, ROS became available for commercial use.

5.4. Limits

This topic describes the limits of Resource Orchestration Service (ROS).

When you use ROS, take note of the following items:

- Each stack can contain up to 200 resources.
- Each user can create up to 50 stacks.
- Each template file can be up to 512 KB in size.

6.Object Storage Service (OSS)

6.1. What is OSS?

Object Storage Service (OSS) is a secure, cost-effective, and highly reliable cloud storage service provided by Alibaba Cloud. It enables you to store a large amount of data in the cloud.

OSS is an immediately available storage solution that has unlimited storage capacity. Compared with user-created server storage, OSS has outstanding advantages in reliability, security, cost-effectiveness, and data processing capabilities. OSS enables you to store and retrieve a variety of unstructured data objects, such as texts, images, audios, and videos over the network at any time.

OSS is an object storage service based on key-value pairs. Files uploaded to OSS are stored as objects in buckets. You can obtain the content of an object based on the object key.

In OSS, you can:

- Create a bucket and upload objects to the bucket.
- Obtain an object URL from OSS to share or download the object.
- Modify the attributes or metadata of a bucket or an object, and configure ACL for the bucket or the object.
- Perform basic and advanced operations in the OSS console.
- Perform basic and advanced operations by using SDKs or calling RESTful API operations in your application.

6.2. Benefits

OSS provides secure, cost-effective, and highly reliable services for storing large amounts of data in the cloud. This topic compares OSS with the traditional user-created server storage to show the benefits of OSS.

Advantages of OSS over user-created server storage

Item	OSS	User-created server storage
Reliability	• Automatically expands capacities without affecting your services. • Automatically stores multiple copies of data for backup.	• Prone to errors due to low hardware reliability. If a disk has a bad sector, data may be lost. • Manual data restoration is complex and requires a lot of time and technical resources.

Item	OSS	User-created server storage
Security	• Provides hierarchical security protection for enterprises. • Provides resource isolation mechanisms for multiple tenants and supports zone-disaster recovery. • Provides various authentication and authorization mechanisms, as well as features such as whitelists, hotlink protection, RAM, and Security Token Service (STS) for temporary access.	• Additional scrubbing devices and black hole policy-related services are required. • A separate security mechanism is required.
Data processing	Provides Image Processing (IMG).	Equipment for data processing must be purchased and deployed separately.

More benefits of OSS

- Ease of use

  Provides standard RESTful API operations (some compatible with Amazon S3 API operations), a wide range of SDKs, client tools, and the console. You can upload, download, retrieve, and manage large amounts of data for websites or mobile applications the way you use regular file systems.

  - The number and size of objects are not limited. You can expand your buckets in OSS.
  - Streaming writes and reads are supported, which is suitable for business scenarios where you must simultaneously read and write videos and other large objects.
  - Lifecycle management is supported. You can configure lifecycle rules to delete expired data in batches.

- Powerful and flexible security mechanisms

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

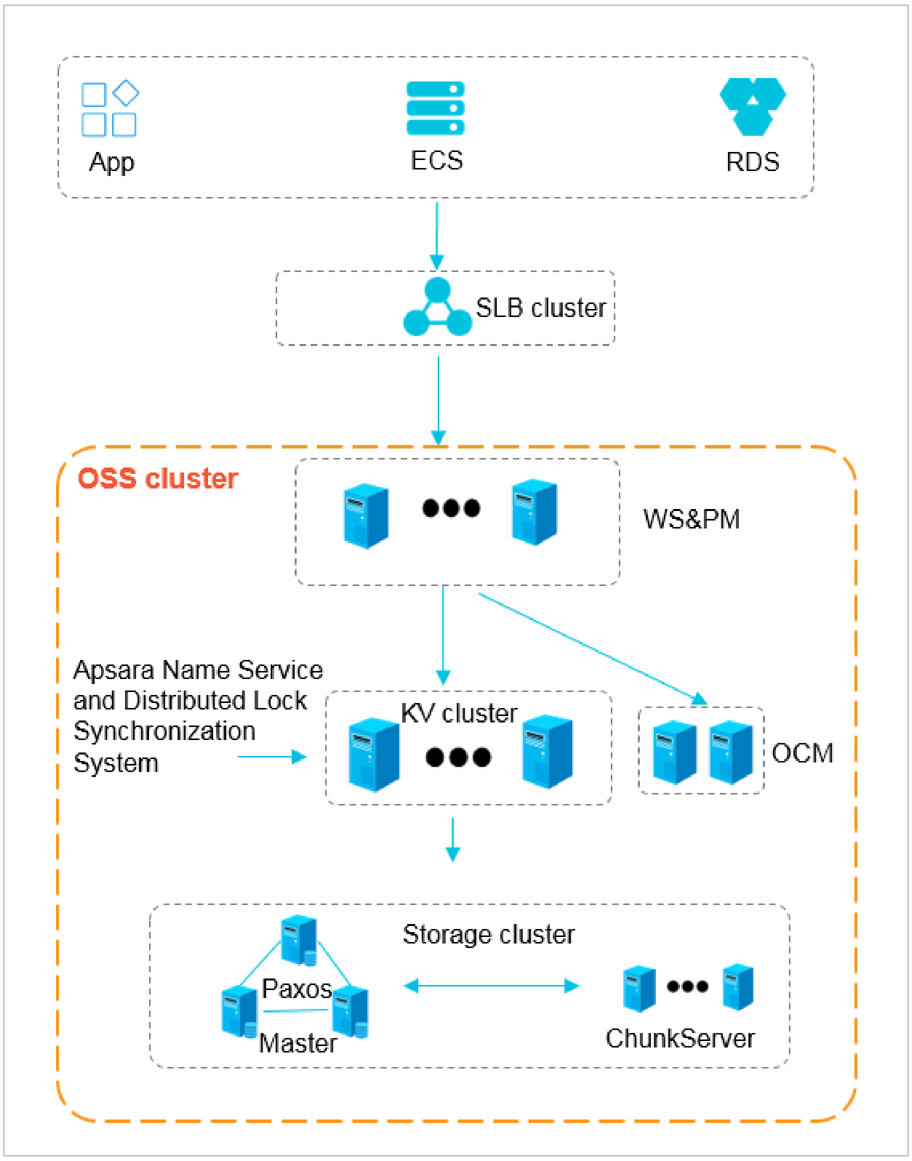
- Rich image processing functions

  Supports format conversion, thumbnails, cropping, watermarking, resizing for objects in formats such as JPG, PNG, BMP, GIF, WebP, and TIFF.

6.3. Architecture

OSS is a storage solution that is built on the Apsara system. It is based on the infrastructure such as Apsara Distributed File System and SchedulerX. The infrastructure provides OSS and other Alibaba Cloud services with importance features such as distributed scheduling, high-speed networks, and distributed storage. The following figure shows the OSS architecture.

OSS architecture



- WS & PM: the protocol layer that receives and authenticates the request sent by using a RESTful protocol. If the authentication is successful, the request is forwarded to KVEngine for further processing. If the authentication fails, an error message is returned.
- KV cluster: used to process structured data, including reading and writing data based on object names. The KV cluster also supports sporadic bursts of requests. When a service has to run on a different physical server due to a change to the service coordination cluster, the KV cluster can coordinate and find the access point.
- Storage cluster: Metadata is stored in the master node. A distributed message consistency protocol of Paxos is adopted between Master nodes to ensure the consistency of metadata. This method ensures efficient distributed storage of and access to objects.

6.4. Terms

This topic describes several basic terms used in OSS.

Object

The basic unit for data operations in OSS. Objects are also known as OSS files. An object is composed of object metadata, object content, and a key. A key can uniquely identify an object in a bucket. Object metadata is a group of key-value pairs that define the properties of an object, such as the last modification time and the object size. You can also assign user metadata to the object.

The lifecycle of an object starts when the object is uploaded, and ends when it is deleted. During the lifecycle, the object cannot be modified. OSS does not support modifying objects. If you want to modify an object, you must upload a new object with the same name as the existing object to replace it.

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

Bucket

A container for OSS objects. Each object in OSS is contained in a bucket. You can configure and modify the attributes of a bucket to manage ACLs and lifecycle rules of the bucket. These attributes apply to all objects in the bucket. Therefore, you can create different buckets to meet different management requirements.

- OSS does not use a hierarchical structure for objects, but instead uses a flat structure. All elements are stored as objects in buckets. However, OSS supports folders as a concept to group objects and simplify management.
- You can create multiple buckets.
- A bucket name must be globally unique within OSS. Bucket names cannot be changed after the buckets are created.
- A bucket can contain an unlimited number of objects.

Strong consistency

A feature requires that object operations in OSS be atomic, which indicates that operations can only either succeed or fail. There are no intermediate states. To ensure that users can access only complete data, OSS does not return corrupted or partial data.

Object-related operations in OSS are highly consistent. For example, when a user receives an upload (PUT) success response, the uploaded object can be read immediately, and copies of the object have been written to multiple devices for redundancy. Therefore, there are no situations where data is not obtained when you perform the read-after-write operation. The same is true for delete operations. After you delete an object, the object and its copies no longer exist.

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

Comparison between OSS and file systems

OSS is a distributed object storage service that stores objects based on key-value pairs. You can retrieve object content based on unique object keys. For example, object name *test1/test.jpg* does not necessarily indicate that the object is stored in a directory named test1. In OSS, *test1/test.jpg* is only a string. There is nothing essentially different between test1/test.jpg and *a.jpg*. Therefore, similar amounts of resources are consumed regardless of which object you access.

A file system uses a typical tree index structure. To access a file named *test1/test.jpg*, you must first access the test1 directory and then search for the *test.jpg* file in this directory. This makes it easy for a file system to support folder operations, such as renaming, deleting, and moving directories because these operations are only performed on directories. However, the performance of a file system depends on the capacity of a single device. The more files and directories that are created in the file system, the more resources and time are consumed.

You can simulate similar folder functions of a file system in OSS, but such operations are costly. For example, if you want to rename the test1 directory as test2, OSS must copy all objects whose names start with test1/ to generate objects whose names start with test2. This operation consumes a large amount of resources. Therefore, we recommend that you do not perform such operations in OSS.

Objects stored in OSS cannot be modified. A specific API operation must be called to append an object, and the generated object is different from objects uploaded by using other methods. To modify even a single byte, you must upload the entire object again. A file system allows you to modify files. You can modify the content at a specified offset location or truncate the end of a file. These features make file systems suitable for more general scenarios. However, OSS supports a large amount of concurrent access, whereas the performance of a file system is subject to the performance of a single device.

We recommend that you do not map operations on OSS objects to file systems because it is inefficient. If you attach OSS as a file system, we recommend that you only add new files, delete files, and read files. You can make full use of OSS advantages, such as the capability to process and store large amounts of unstructured data such as images, videos, and documents.

6.5. Features

This topic lists the common features of OSS.

Before you start to use OSS, we recommend that you have a good understanding of basic terms used in OSS, such as bucket, object, region, and endpoint. For more information, see Terms.

The following table describes features of OSS.

OSS features

Parameter	Feature	Description
Bucket	Create buckets	Before you upload an object to OSS, you must create a bucket to store the object.
	Delete buckets	If you no longer use a bucket, delete it to avoid further fees.
	Modify bucket ACL	OSS supports ACL for access control. You can configure the ACL of a bucket when you create it or modify the ACL of a created bucket.
	Configure static website hosting	You can configure static website hosting for your bucket and access this static website through the bucket domain name.
	Configure hotlink protection	To prevent additional fees caused by unauthorized access to the data in your bucket, you can hotlink protection for your buckets based on the Referer field in HTTP requests.
	Manage CORS	OSS provides cross-origin resource sharing (CORS) over HTML5 to implement cross-origin access.

Parameter	Feature	Description
	Configure lifecycle rules	You can define and manage lifecycle rules for all or a subset of objects in a bucket. You can configure lifecycle rules to manage multiple objects and automatically delete parts.
	Configure server-side encryption	OSS can perform server-side encryption on data uploaded to buckets to secure data.
Object	Upload objects	You can upload any type of objects to a bucket.
	Create folders	You can manage OSS folders the way you manage folders in Windows.
	Search for objects	You can search for objects whose names contain the same prefix in a bucket or folder.
	Obtain object URLs	You can obtain the URL of an object to share or download the object.
	Delete objects	You can delete a single object or multiple objects.
	Delete folders	You can delete a single folder or multiple folders.
	Modify object ACL	You can configure ACL when you upload an object and modify the ACL after you upload the object.
	Manage parts	You can delete all or some parts from a bucket.
API	API	OSS supports RESTful API operations and provides examples.
SDK	SDK	OSS supports development based on SDKs for various programming languages and provides examples.

6.6. Scenarios

This topic describes the application scenarios of OSS.

Massive storage for image, audio, and video applications

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

Dynamic and static content separation for websites and mobile applications

By using the BGP bandwidth, you can download data from OSS with an ultra-low latency.

Offline data storage

OSS provides storage with low cost and high availability. Therefore, you can use OSS to store enterprise data that needs to be archived offline for a long period.

6.7. Limits

This topic describes the limits and performance metrics of OSS.

Item	Limit
Bucket	- You can create a maximum of 100 buckets. - After a bucket is created, its name and region cannot be modified.
Object upload	- Objects larger than 5 GB cannot be uploaded by using the following modes: console upload, simple upload, form upload, or append upload. To upload an object that is larger than 5 GB, you must use multipart upload. The size of an object uploaded by using multipart upload cannot exceed 48.8 TB. - If you upload an object that has the same name of an existing object in OSS, the new object will overwrite the existing object. - OSS traffic is forwarded through SLB and has the following limits: ○ By default, a virtual IP address (VIP) is configured for SLB and the maximum throughput for OSS is 1.25 GB/s. ○ The maximum throughput for each OSS node is 300 MB/s. In scenarios where only stable and frequent write operations are continuously performed, the maximum throughput for each OSS node is 100 MB/s.
Object deletion	- Deleted objects cannot be recovered. - You can delete up to 100 objects at a time in the OSS console. To delete more than 100 objects at a time, you must call an API operation or use an SDK.
Lifecycle	You can configure up to 1,000 lifecycle rules for each bucket.

7.Apsara File Storage NAS

7.1. What is NAS?

Apsara File Storage NAS provides file storage services for compute nodes, such as Elastic Compute Service (ECS) instances and Alibaba Cloud Container Service for Kubernetes (ACK) nodes. NAS supports multiple standard file access protocols. It enables you to use distributed file systems that provide unlimited capacity, parallel shared access, high reliability, and high availability. You can use these file systems without the need to modify your current applications.

After you create a NAS file system and mount target, you can mount the file system on ECS instances or ACK nodes. NAS allows you to access the file system by using the standard Network File System (NFS) protocol. You can also use POSIX APIs to access the file system. Each file system can be mounted on multiple compute nodes to share files and folders.

7.2. Benefits

Apsara File Storage NAS has the following benefits:

- Parallel shared access
- High reliability
- Auto scaling
- High performance
- Easy-to-use

Parallel shared access

A file system can be simultaneously mounted on multiple compute nodes to provide shared access. This access method reduces data replication and synchronization costs.

High reliability

NAS provides reliable data storage. Compared with user-created NAS file systems, NAS file systems greatly reduce maintenance costs and minimize data security risks.

Auto scaling

NAS allows you to respond to business changes in a timely manner. You can scale the capacity of a file system based on your business requirements.

High performance

When your data storage increases, NAS file systems provide a higher throughput to meet your demand. You do not need to purchase high-end NAS storage devices. This reduces a large amount of upfront investment.

Easy-to-use

Apsara File Storage NAS supports the NFSv3 and NFSv4 protocols. You can access file systems by calling standard POSIX API operations, regardless of the types of compute nodes and the location in which file systems reside.

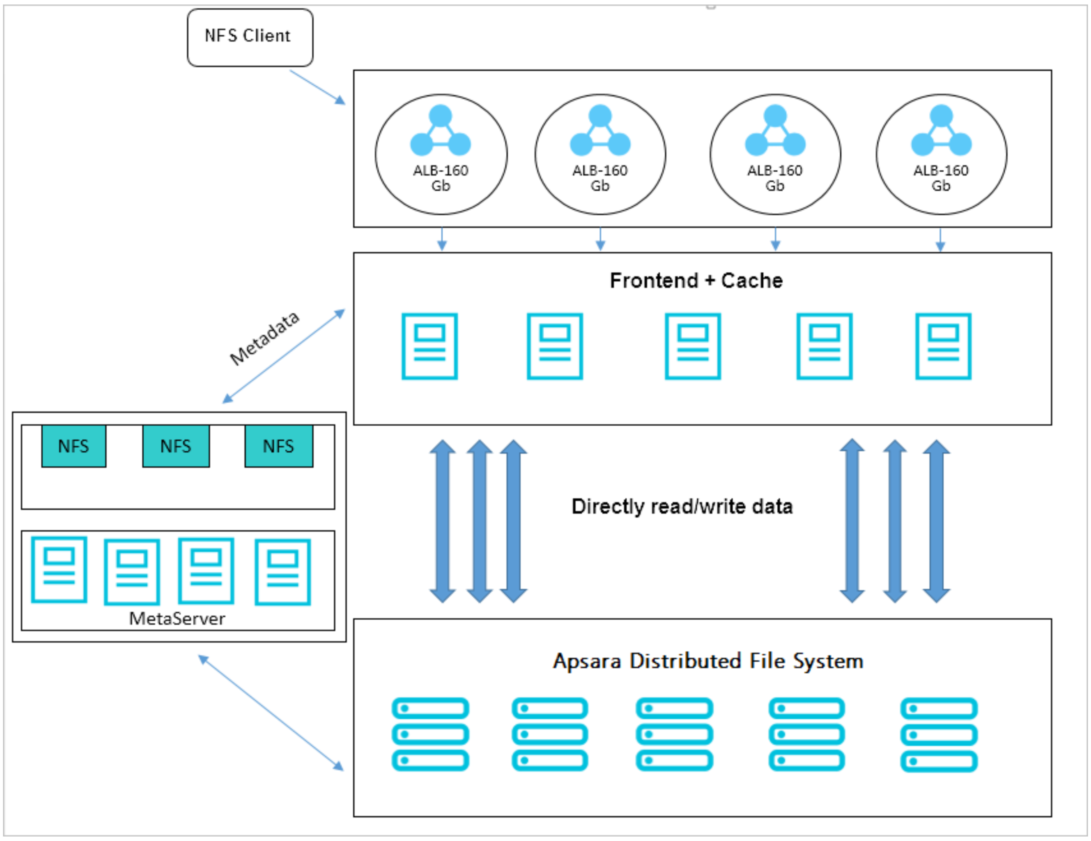
7.3. Architecture

Apsara File Storage NAS is based on Apsara Distributed File System. NAS maintains three copies for each data file across multiple storage nodes. Frontend nodes receive and cache connection requests from NFS clients. Frontend nodes are highly available because they are stateless and distributed.

The metadata of a NAS file system is stored on a MetaServer. When frontend nodes retrieve metadata from the MetaServer by using I/O requests, user data is read from and written to the backend nodes of Apsara Distributed File System.

The system architecture provides separate auto scaling of frontend and backend storage nodes. This ensures high availability, high concurrency, and low latency.

System architecture



7.4. Features

This topic describes the features of Apsara File Storage NAS.

Seamless integration

Apsara File Storage NAS supports the NFSv3 and NFSv4 protocols. NAS also allows you to access data by using standard file system interfaces. Mainstream applications and workloads can be seamlessly integrated with NAS without the need for modification.

Shared access

If you want to access a data source from multiple Elastic Compute Service (ECS) and Elastic Container Instance (ECI) instances, a NAS file system meets this need.

Access control

Apsara File Storage NAS uses multiple security mechanisms to guarantee system data security. These security mechanisms include network isolation based on VPCs, user isolation in classic networks, standard permission control for file systems, access control based on security groups, and RAM user authorization.

Scalable performance

Apsara File Storage NAS provides your applications with optimal storage performance, such as high throughput, high IOPS, and low latency. The storage performance linearly improves as the storage capacity increases. This meets your demands for higher storage performance as your business grows.

7.5. Scenarios

This topic describes the scenarios of Apsara File Storage NAS.

Scenario 1: shared storage and high availability for SLB

For example, assume that your Server Load Balancing (SLB) instance is connected to multiple Elastic Compute Service (ECS) instances. You can store the data of the applications on these ECS instances on a shared NAS file system. This data sharing method ensures high availability of the SLB instance.

Scenario 2: file sharing within an enterprise

For example, the employees of an enterprise need to access the same datasets. The administrator can create a NAS file system and configure different file or directory permissions for users or user groups.

Scenario 3: data backup

For example, you want to migrate your data from a data center to the cloud for backup. You want to use a standard interface to access the cloud storage service. You can back up your data in a NAS file system.

Scenario 4: server logs sharing

For example, you want to store the application server logs of multiple compute nodes to a shared file store. You can store these server logs in a NAS file system for centralized log processing and analysis.

7.6. Limits

- NAS supports the NFSv3 and NFSv4 protocols.
- The following table lists the attributes that are not supported by NFSv4.0 and NFSv4.1, and their client errors.

Protocol	Unsupported attribute	Client error
NFSv4.0	FATTR4_MIMETYPE, FATTR4_QUOTA_AVAIL_HARD, FATTR4_QUOTA_AVAIL_SOFT, FATTR4_QUOTA_USED, FATTR4_TIME_BACKUP, and FATTR4_TIME_CREATE	NFS4ERR_ATTRNOTSUPP

Protocol	Unsupported attribute	Client error
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, and FATTR4_FS_CHARSET_CAP	NFS4ERR_ATTRNOTSUPP

- NFSv4 does not support the following OPs: OP_DELEGPURGE, OP_DELEGRETURN, and NFS4_OP_OPENATTR. The client displays an NFS4ERR_NOTSUPP error.
- NFSv4 does not support Delegation.
- About UID and GID:
  - For NFSv3, if the file UID or GID exists in a Linux local account, the corresponding username and group name is displayed based on the mapping between the local UID and GID. If the file UID or GID does not exist in the local account, the UID and GID is displayed.
  - For NFSv4, if the version of the local Linux kernel is earlier than 3.0, the UIDs and GIDs of all files are displayed as "nobody." If the kernel version is later than 3.0, the display rule is the same as that of NFSv3.

> **Notice** If you use NFSv4 to mount a NAS instance and the Linux kernel version is earlier than 3.0, we recommend that you do not change the owner or group of local files or directories. Such changes can cause the UIDs and GIDs of the files or directories to become "nobody."

- You can mount a NAS instance to up to 10,000 compute nodes.

7.7. Terms

This topic describes the basic terms of Apsara File Storage NAS.

mount target

A mount target is the access address of a NAS file system in a VPC or classic network. Each mount target corresponds to a domain name. To mount a NAS file system to a local directory, you must specify the domain name of the mount target.

permission group

The permission group mechanism is a whitelist mechanism provided by NAS. You can add rules to a permission group of a NAS file system. You can allow users from specified IP addresses or CIDR blocks to access the NAS file system by using different permissions.

Note Each mount target must be associated with a permission group.

authorized object

An authorized object is an attribute of a permission group rule. It specifies the IP address or CIDR block to which the permission group rule is applied. In a VPC, an authorized object can be a single IP address or a CIDR block. In a classic network, an authorized object must be a single IP address. In most cases, this IP address is the internal IP address of an Elastic Compute Service (ECS) instance.

8.Tablestore

8.1. What is Tablestore?

Tablestore is a NoSQL database service independently developed by Alibaba Cloud. Tablestore is a proprietary software program that is certified by the relevant authorities in China. Tablestore is built on the Apsara system of Alibaba Cloud, and can store large amounts of structured data and allow real-time access to these data.

Tablestore provides the following features:

- Offers schema-free data storage. You do not need to define attribute columns before you use them. Table-level changes are not required to add or delete attribute columns. You can configure the time to live (TTL) parameter for a table to manage the lifecycle of data. The expired data is deleted from the table.
- Adopts the triplicate technology to keep three copies of data on three servers across three different racks. A cluster can support single storage type instances (SSD only) or mixed storage type instances (SSD and HDD) to meet different budget and performance requirements.
- Adopts a fully redundant architecture that prevents single points of failure (SPOFs). Tablestore supports smooth online upgrades, hot cluster upgrades, and automatic data migration, which enable you to dynamically add or remove nodes for maintenance without incurring service interruptions. The concurrent read and write throughput and storage capacity can be linearly scaled. Each cluster can have at least 500 hosts.
- Supports highly concurrent read and write operations. Concurrent read and write capabilities can be scaled out as the number of hosts increases. The read and write performance is indirectly related to the amount of data in a single table.
- Supports identity authentication and multi-tenancy. Comprehensive access control and isolation mechanisms are provided to safeguard your data. VPC and access over HTTPS are supported. Provides multiple authentication and authorization mechanisms so that you can define access permissions on individual tables and operations.

8.2. Benefits

Tablestore provides the following benefits:

Scalability

- Tablestore imposes no upper limit on the amount of data that can be stored in tables. When the amount of data increases, Tablestore adjusts partitions to provide more storage space for tables and improve the capability of handling access request bursts.
- Tablestore supports CPUs, disks, memory, and NICs of different specifications in a single-component cluster without affecting cluster running performance. This ensures maximum compatibility with existing devices.

High performance

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

Data reliability

- Tablestore provides high data reliability. It stores multiple data copies and restores data when any of the copies become invalid.
- Tablestore supports automatic fault tolerance for disk failures of servers in a cluster, and supports hot swapping of disks. The failures of a single disk do not affect the overall service. The failures of data disks do not affect the service of the server. If Redundant Array of Independent Disks (RAID) is not created and the system disk is damaged, the server is removed from the cluster. After the hardware failures are fixed, services can be restored within two minutes.
- Tablestore supports full or incremental backup and data recovery from storage.
- Tablestore supports the backup between data clusters in different data centers. The backup process is visualized.

High availability

Tablestore uses automatic failure detection and data migration to shield applications from host- and network-related hardware faults, which provides high availability for your applications.

Ease of management

- Tablestore automatically performs complex O&M tasks, such as the management of data partitions, software and hardware upgrades, configuration updates, and cluster scale-out.
- You can store audit logs to Log Service and download logs from Log Service. This facilitates long-term storage and management of audit logs.

Access security

- Tablestore provides multiple permission management mechanisms. It verifies and authenticates the identity of each application request to prevent unauthorized data access, which improves the data security.
- Tablestore supports the management of data access permissions, including logon permissions, table creation permissions, read and write permissions, and whitelist-related permissions.
- Tablestore allows you to use the Apsara Stack Cloud Management (ASCM) console to manage administrative permissions, including administrator classification. You can use the ASCM console to manage user permissions in a centralized manner. You can manage the access control features of any component within the system. You can also block common users from querying access control details and simplify access control for administrators to improve the usability of access control.

Strong consistency

Tablestore ensures high data consistency for data writes. After a write operation succeeds, three replicas are written to a disk. Applications can read the latest data immediately.

Flexible data models

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

Monitoring integration

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

Multi-tenancy

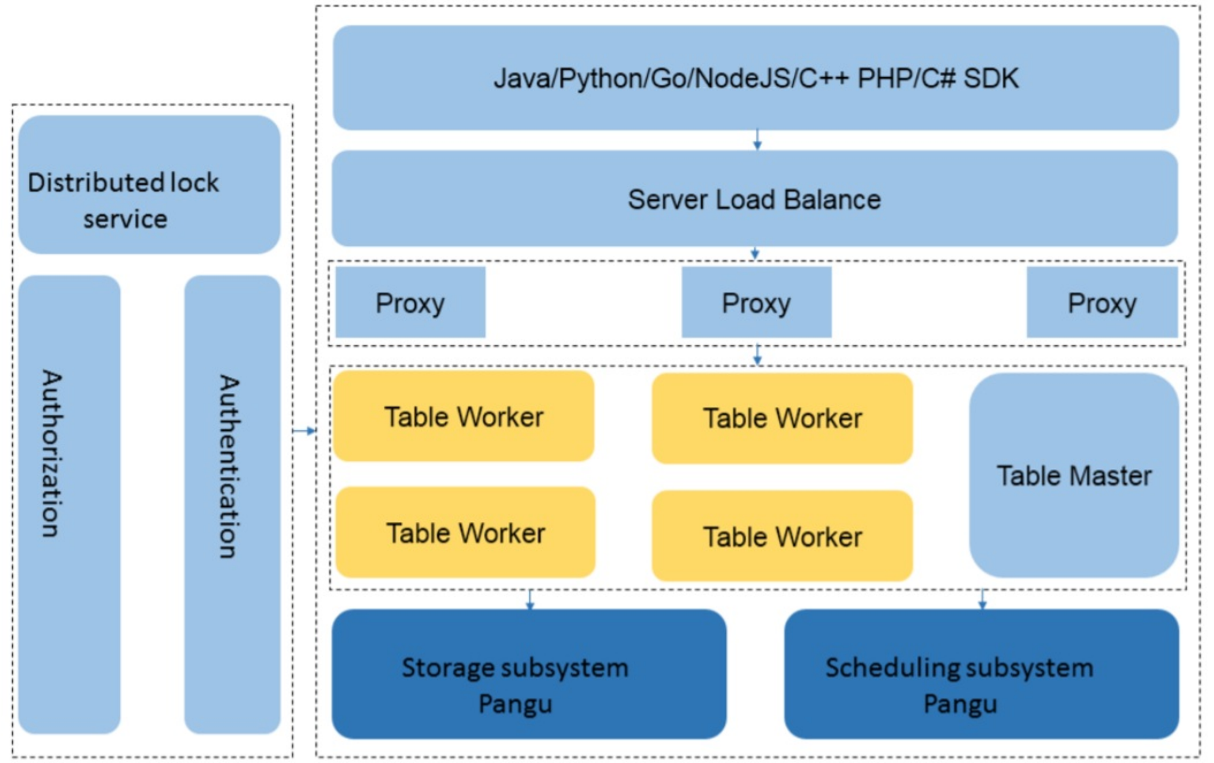
- Isolation: allows tasks of multiple tenants (projects) to be submitted to different queues and run separately. Resources are isolated among tenants.
- Permission: allows you to manage tenants in a centralized manner, dynamically configure and manage tenant resources, isolate resources, view statistics for resource usage, and manage tenants at multiple levels in the console.
- Scheduling: supports multi-tenant scheduling of multiple clusters and multiple resource pools.

8.3. Architecture

This topic describes the Tablestore architecture.

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

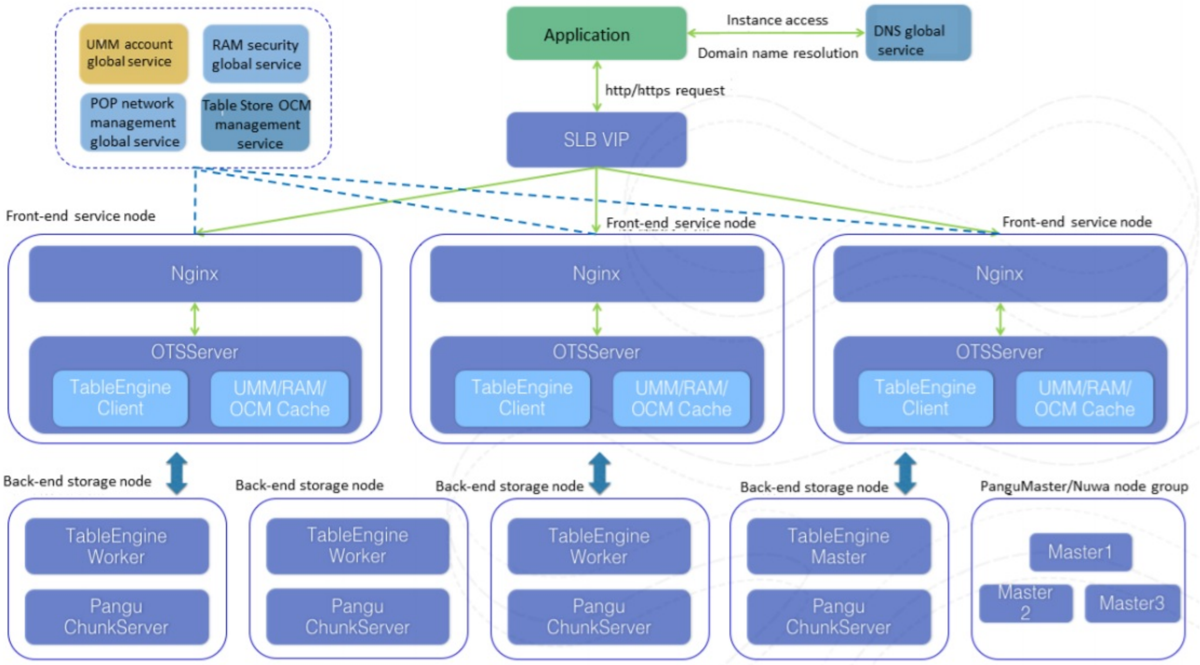
The following figure shows the basic architecture of Tablestore.



- The top layer is the protocol access layer. Server Load Balancer (SLB) distributes user requests to various proxy nodes. The proxy nodes receive requests that are sent by using the RESTful protocol and implement security authentication.
  - If the authentication succeeds, the user requests are forwarded to the corresponding data engine based on the value of the first primary key column for further operations.
  - If the authentication fails, error information is returned to the user.
- Table Worker is the data engine layer that processes structured data. It uses a primary key to search for or store data. Table Worker supports large-scale access request bursts.
- The bottom layer is the persistent storage layer. Apsara Distributed File System is deployed at this layer. Metadata is stored on masters. A distributed message consistency protocol (or Paxos) is

adopted between masters to ensure the metadata consistency. This way, efficient distributed file storage and access are achieved. This method ensures that three copies of data are stored in the system and that the system can recover from any hardware or software fault.

The following figure shows the detailed architecture of Tablestore.



8.4. Scenarios

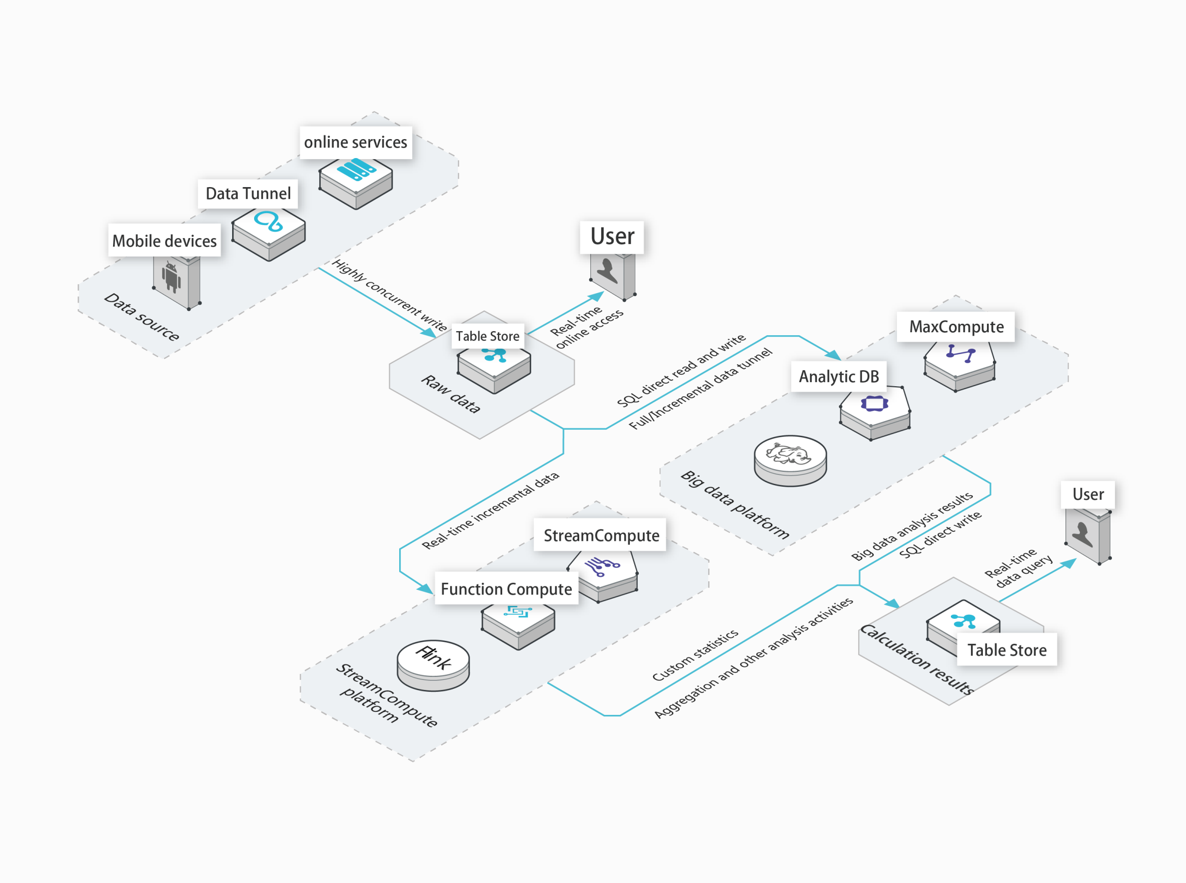
Tablestore can be applied to the following scenarios:

- Scenario 1: Big data storage and analytics

  Tablestore provides cost-effective, highly concurrent, and low-latency storage, and online access to large amounts of data. It provides full and incremental data tunnels and supports direct SQL-based read and write operations on various big data analysis platforms such as MaxCompute. An efficient incremental streaming read interface is provided for easy computing of real-time data streams.

  Tablestore provides the following features:

  - Tablestore supports various big data computing platforms, stream computing services, and real-time computing services.
  - Tablestore provides high-performance and capacity instances to meet the requirements of different business.

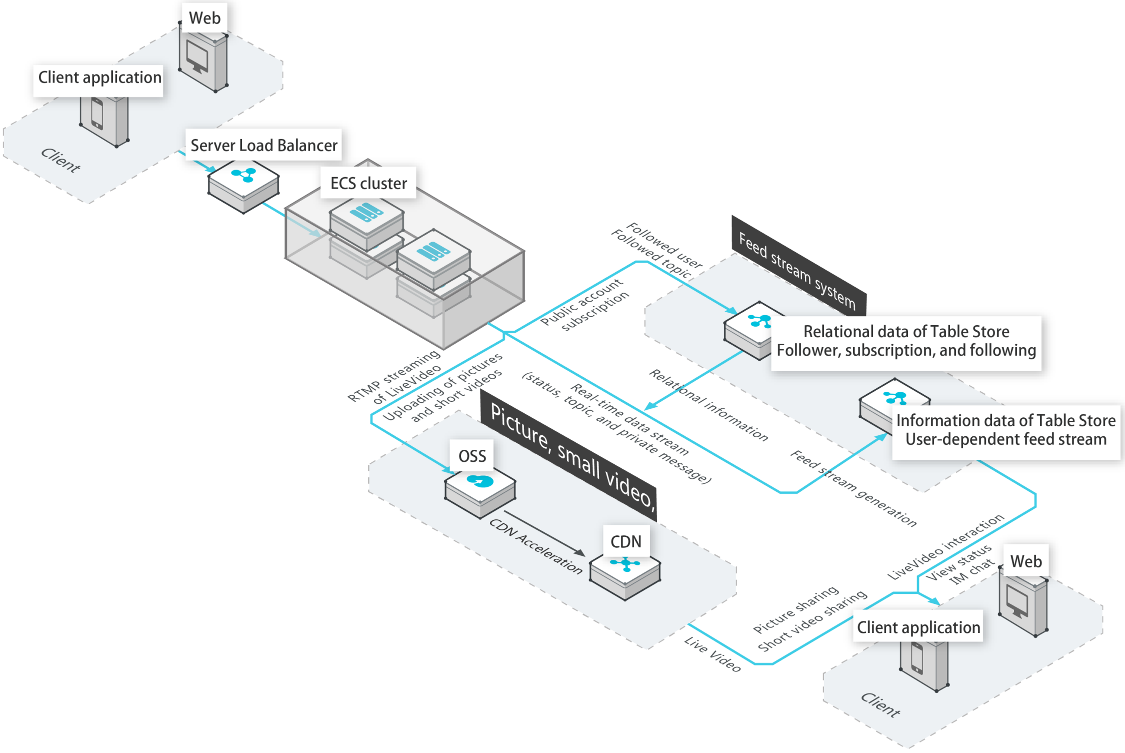


- Scenario 2: Social media feeds on the Internet

  You can use Tablestore to store large amounts of instant messaging (IM) messages and social media feed information such as comments, posts, and likes. The elastic resources available for Tablestore can meet application requirements including handling significant traffic fluctuations, high concurrency, and low latency at relatively low costs.

  Tablestore provides the following features:

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

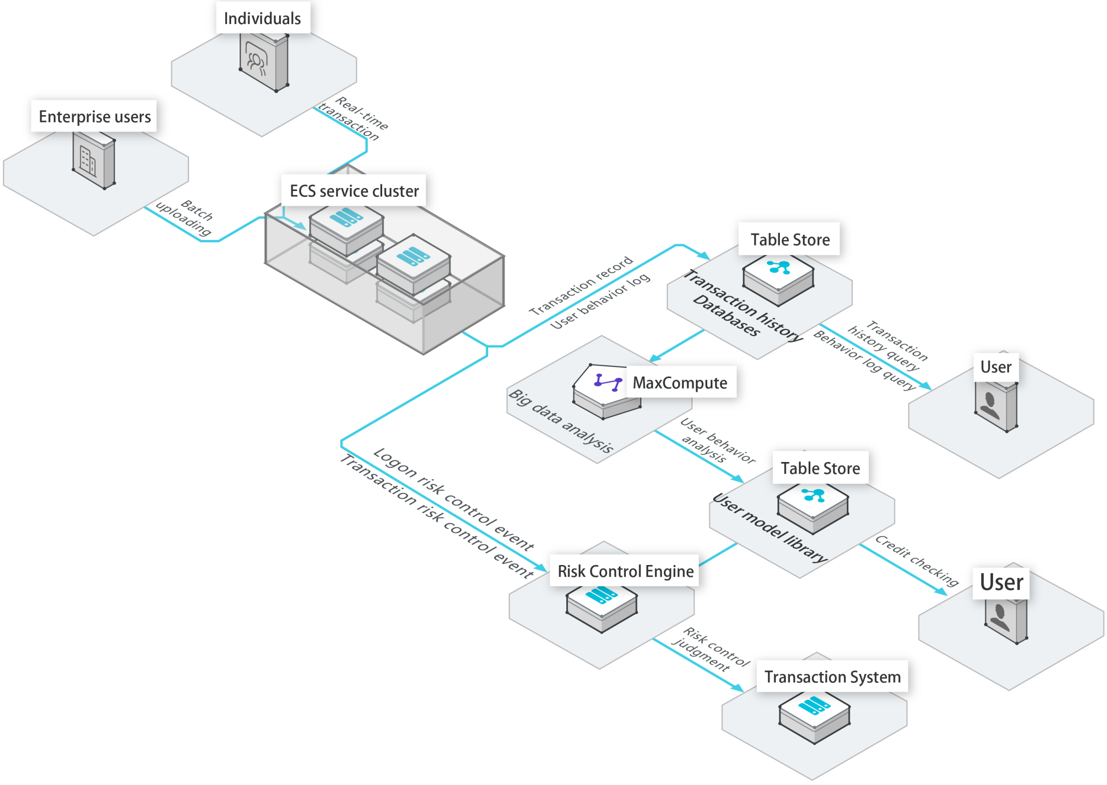


- Scenario 3: Storage and real-time queries of large amounts of transaction records and user models

  Tablestore instances are elastic, low latency, and highly concurrent, which provides optimal running conditions for risk control systems. This helps you control transaction risks. Furthermore, the flexible data structure allows your business model to rapidly evolve to meet market demands.

  Tablestore provides the following features:

  - A table can store full historical transaction records.
  - Data is stored in three copies to ensure high consistency and data security.
  - The schema-free data model allows you to add attribute columns based on your requirements. This allows rapid service development.

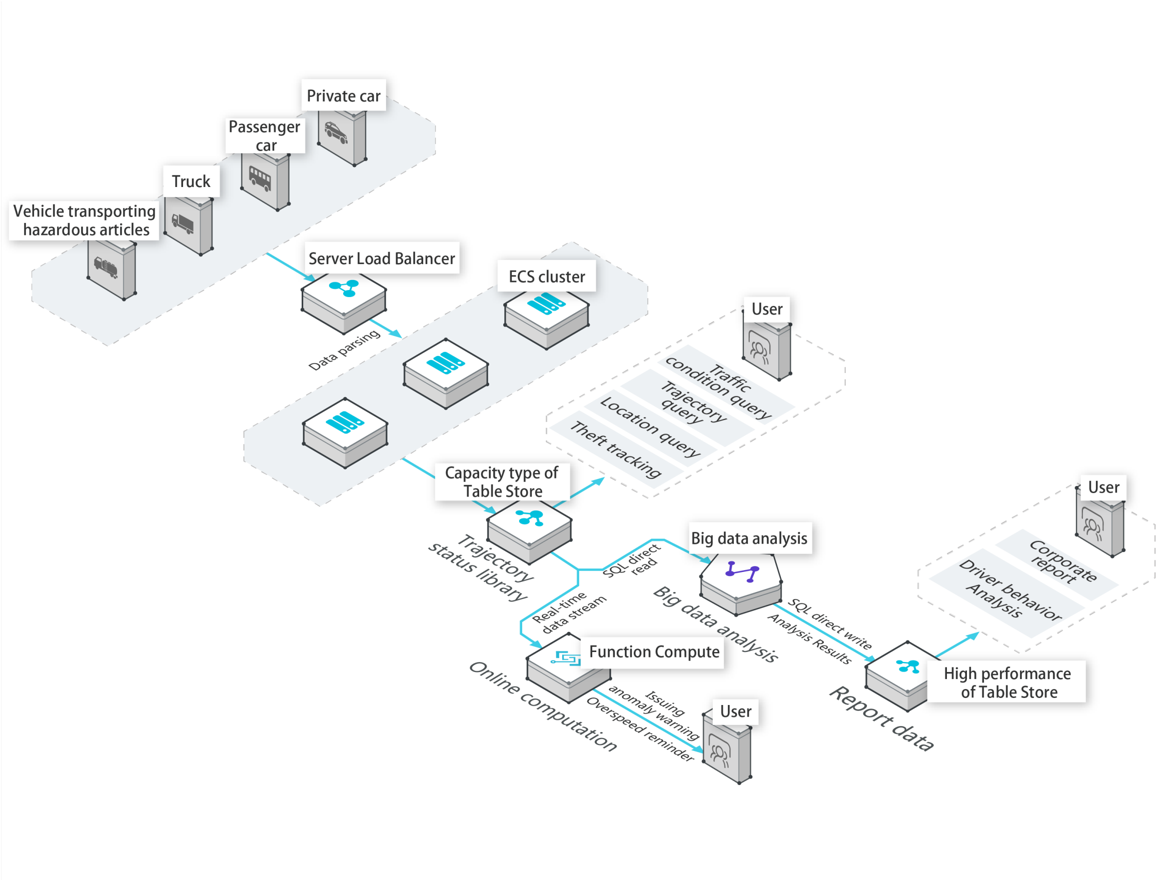


- Scenario 4: Efficient and flexible storage of large amounts of IoV data

  The schema-free data model simplifies access to the data collected from different vehicle-mounted devices. Tablestore can be seamlessly integrated with multiple big data analytics platforms and real-time computing services to implement real-time online queries and business report analysis.

  Tablestore provides the following features:

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

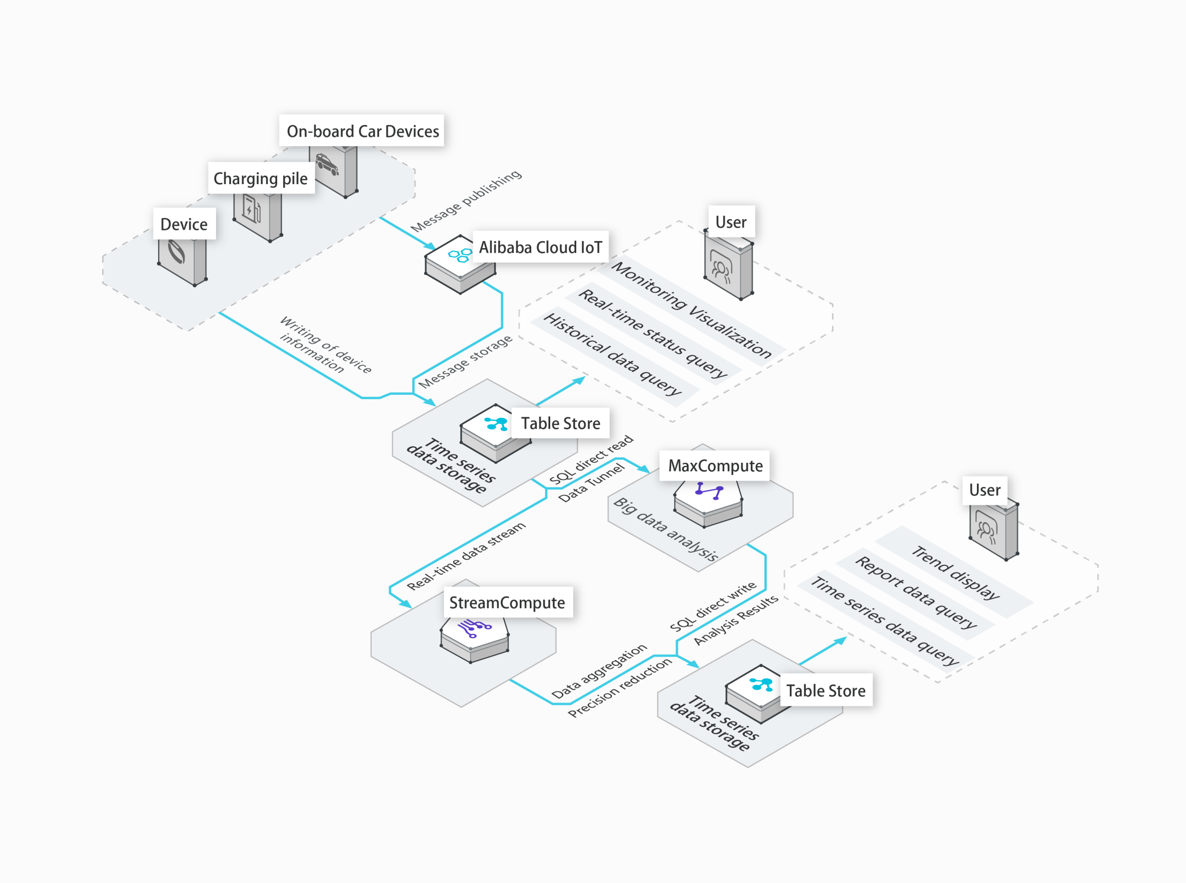


- Scenario 5: Storage of large amounts of IoT data for efficient queries and analysis

  Tablestore can be used to store time series data from IoT devices and monitoring systems. It provides API operations to directly read SQL data and incremental data streams, which allow you to implement offline data analysis and real-time stream computing.

  Tablestore provides the following features:

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

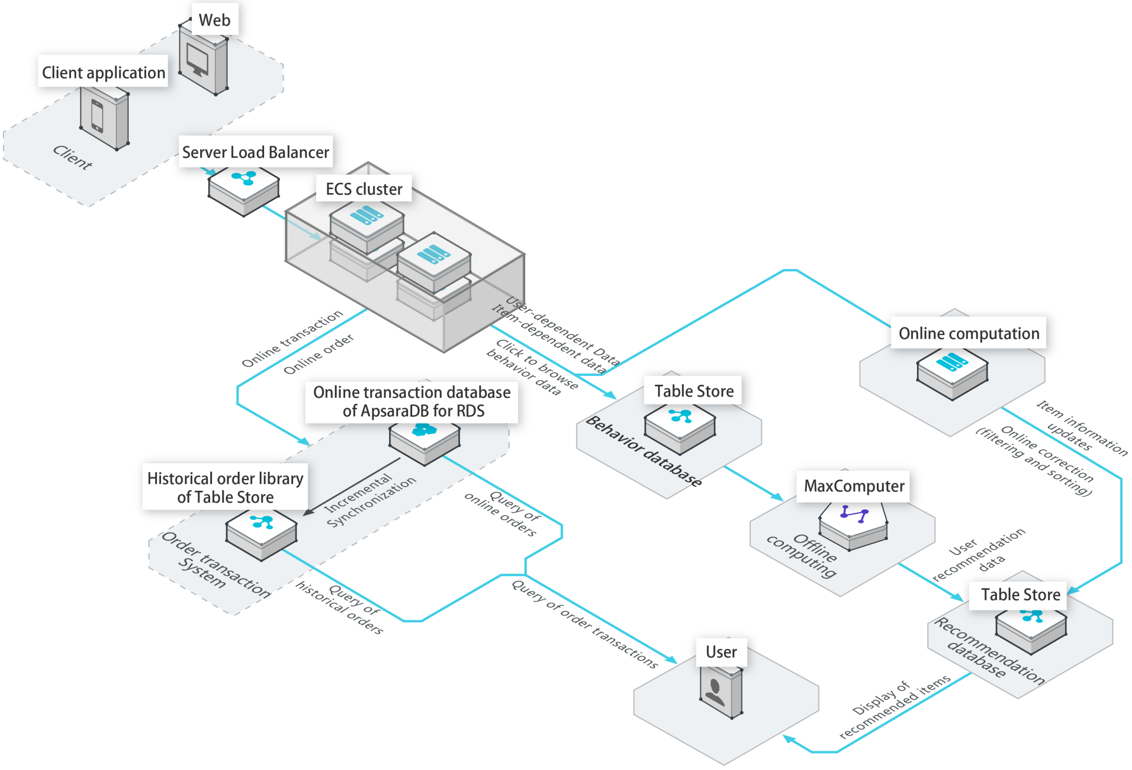


- Scenario 6: Databases for large-scale e-commerce transaction orders and user-specific recommendations

  Tablestore can manage large amounts of historical transaction data and improve access performance. Tablestore can be used together with MaxCompute to implement precision marketing and elastic resource storage. This allows you to handle service requests during peak hours when all users go online.

  Tablestore provides the following features:

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



8.5. Limits

This topic describes the usage limits of Tablestore.

The following table describes the limits on the usage of Tablestore. A part of limits indicate the maximum values that can be used rather than the suggested values. You can tailor table schemas and row sizes to improve performance.

Item	Limit	Description
The number of instances created in an Apsara Stack tenant account	1024	If you need to increase the maximum number of instances, contact an administrator.
The number of tables in an instance	1024	If you need to increase the maximum number of tables, contact an administrator.
The number of columns in a primary key	1~4	A primary key can contain one to four columns.
The size of the value in a STRING primary key column	1 KB	The size of the value in a STRING primary key column cannot exceed 1 KB.
The size of the value in a STRING attribute column	2 MB	The size of the value in a STRING attribute column cannot exceed 2 MB.
The size of the value in a BINARY primary key column	1 KB	The size of the value in a BINARY primary key column cannot exceed 1 KB.

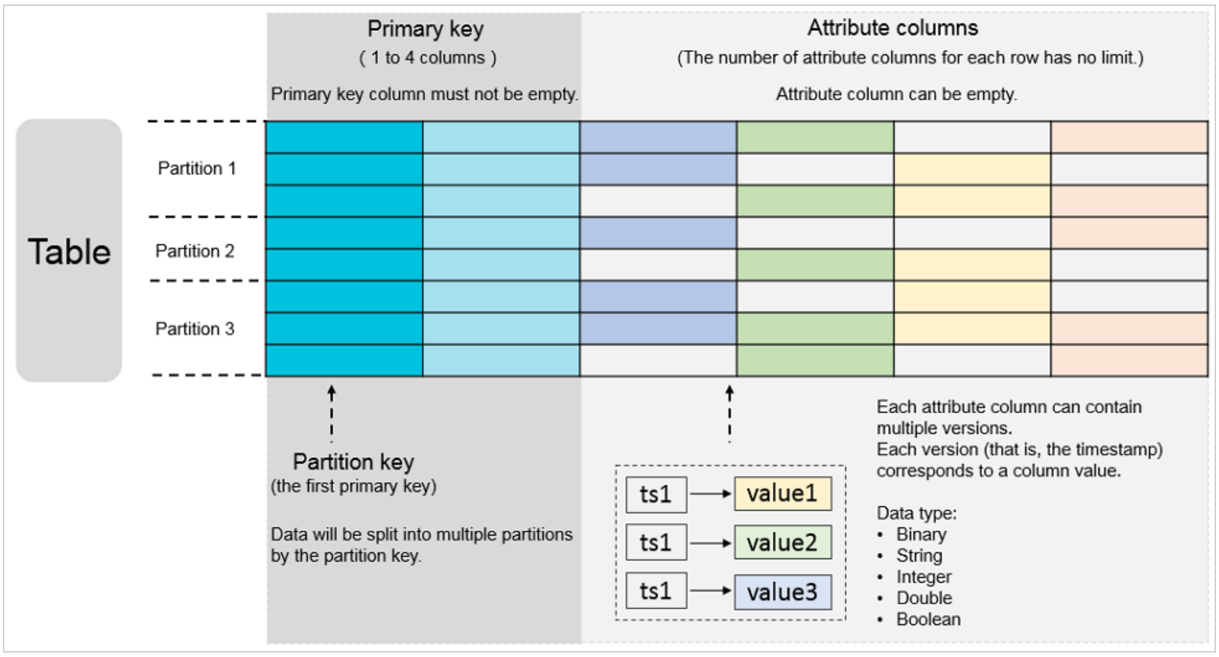
Item	Limit	Description
The size of the value in a BINARY attribute column	2 MB	The size of the value in a BINARY attribute column cannot exceed 2 MB.
The number of attribute columns in a single row	Unlimited	A single row can contain an unlimited number of attribute columns.
The number of attribute columns written by one request	1,024	During a PutRow, UpdateRow, or BatchWriteRow operation, the number of attribute columns written to a single row cannot exceed 1,024.
The data size of a row	Unlimited	The total size of all column names and column values for a row is unlimited.
The number of columns that are specified by the columns_to_get parameter in a read request	0~128	The maximum number of columns obtained from a single row of data in a read request cannot exceed 128.
The number of UpdateTable operations for a table	Upper limit: unlimited Lower limit: unlimited	The frequency of UpdateTable operations for a table is limited.
The frequency of UpdateTable operations for a table	Once every two minutes	The reserved read/write throughput for a table can be adjusted once every two minutes at most.
The number of rows read by one BatchGetRow request	100	None.
The number of rows written by one BatchWriteRow request	200	None.
The size of data written by one BatchWriteRow request	4 MB	None.
Data returned by one GetRange request	5,000 rows or 4 MB	The amount of data returned by a request cannot exceed 5,000 rows or 4 MB. When either of the limits is exceeded, data that exceeds the limits is truncated at the row-level. The data primary key information in the next row is returned.
The data size of an HTTP request body	5 MB	None.

8.6. Terms

This topic describes several basic terms used in Tablestore, including data model, max versions, time to live (TTL), max version offset, primary key and attribute, read/write throughput, region, instance, endpoint, and Serial ATA (SATA).

data model

A data model that consists of tables, rows, primary keys, and attribute columns in Tablestore. The following figure shows an example of a data model.


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

max versions

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

TTL

A data table attribute that indicates the validity period of data in seconds. To save space and reduce costs for data storage, Tablestore deletes any data that exceeds its TTL.

max version offset

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

To prevent the writing of unexpected data, a server checks the versions of attribute columns when the server processes writing requests. If the specified version is earlier than the current writing time minus the max version offset value or later than or equal to the current writing time plus the max version offset value, data fails to be written to the row.

The valid version range of an attribute column: **[Data written time - Valid version offset, Data written time + Valid version offset)**. Data written time is the number of seconds that have elapsed since 00:00:00, 1 January 1970. Versions of the attribute columns are written in milliseconds. A version of an attribute column must fall within the valid version range after the version number is converted to seconds (divide by 1,000).

primary key and attribute

A primary key is the unique identifier of each row in a table. A primary key consists of one to four primary key columns. When you create a table, you must define a primary key. You must specify the name, data type, and sequence of each primary key column. The data type of primary key columns can be only STRING, INTEGER, or BINARY. The size of a STRING or BINARY primary key column cannot exceed 1 KB.

An attribute is the attribute data stored in a row. You can create an unlimited number of attribute columns for each row.

read/write throughput

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

region

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

instance

A logical entity that is used to manage tables in Tablestore. Instances correspond to databases in traditional relational databases. An instance is the basic unit of the Tablestore resource management system. Tablestore allows you to control access and meter resources by instance.

endpoint

The connection URL for each instance. You must specify an endpoint before you perform any operations on Tablestore tables and data.

SATA

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

8.7. Features

8.7.1. Features

This topic describes the basic features of Tablestore, including data partition, load balancing, and automatic recovery from single points of failure (SPOFs).

Tablestore provides the following features:

- Data partition and load balancing

  The first column of a primary key in each row of a table is the partition key. The system splits a table into multiple partitions based on the range of partition key values. These partitions are evenly scheduled across different storage nodes. When the data in a partition exceeds the size limit, the partition is automatically split into two smaller partitions. The data and access loads are distributed across these two partitions. The partitions are scheduled to different nodes. As a result, access loads are distributed to different nodes. This allows single-table data and access loads to scale linearly. A partition is a logical organization of data based on the shared storage mechanism. No migration of physical data is involved when a partition is split. However, this may cause the partition to be unable to provide services for 100 milliseconds.

- Automatic recovery from single points of failure (SPOFs)

Each node in the storage engine of Tablestore provides services for multiple data partitions of different tables. The master node manages partition distribution and scheduling, and monitors the health of each service node. If a service node fails, the master node migrates data partitions from the faulty node to other healthy nodes. Services can recover from SPOF in a short time because migrations are performed on the logical level and do not involve the physical migration of data.

8.7.2. Global secondary index

8.7.2.1. Features

This topic describes the features of global secondary index in Tablestore.

Tablestore provides the following features for you to use global secondary index:

- Supports asynchronous data synchronization between a base table and index tables. Under normal network conditions, the data synchronization can reach single-digit millisecond latency.
- Supports single-column indexes and compound indexes.
- Support covered indexes. Predefined columns are specified in advance in a base table. You can create an index table on any predefined column or primary key column of the base table. You can also specify multiple predefined columns of a base table as attribute columns of an index table or choose not to specify attribute columns. If you specify predefined columns of a base table as the attribute columns of an index table, you can directly query this index table instead of querying the base table to obtain the value of the predefined column. For example, a base table includes the primary key columns PK0, PK1, and PK2 and the predefined attribute columns Defined0, Defined1, and Defined2.
    - You can create an index table on PK2 without specifying an attribute column or specifying Defined0 as an attribute column.
    - You can create an index table on PK1 and PK2 without specifying an attribute column or specifying Defined0 as an attribute column.
    - You can create an index table on PK2, PK1, and PK0 and specify Defined0, Defined1, and Defined2 as attribute columns.
    - You can create an index table on Defined0 without specifying an attribute column.
    - You can create an index table on Define0 and PK1 and specify Defined1 as an attribute column.
    - You can create an index table on Define1 and Define0 without specifying an attribute column or specifying Defined2 as an attribute column.
- Supports sparse indexes. You can specify a predefined column in the base table as an attribute column in the index table. A row will be indexed when all indexed columns exist even if the predefined column is excluded from the row of the base table. However, a row will not be indexed when the row excludes one or more indexed columns. For example, a base table includes the primary key columns PK0, PK1, and PK2 and the predefined columns Defined0, Defined1, and Defined2. You can create an index table on Defined0 and Defined1 and specify Defined2 as an attribute column.
    - The index table includes the rows in the base table that include Defined0 and Defined1 but exclude Defined2.
    - The index table excludes the rows in the base table that includes Defined0 and Defined2 but excludes Defined1.
- Supports the deletion or creation of index tables for an existing base table. An index table can contain the existing data of the base table.

- When you query an index table, the query is not performed on the base table. You must query the base table. The automatic query on the base table after a query on an index table will be supported in later versions.

8.7.2.2. Usage notes

This topic describes the terms, limits, and precautions for global secondary indexes.

Terms

Term	Description
index table	The table created based on indexing of columns from the base table. The data in the index table is read-only.
predefined column	The column you predefine when you create a table. Tablestore uses a schema-free model. You can also specify the data type of the column. You can write an unlimited number of columns to a row. You do not need to specify a fixed number of predefined columns in a schema.
single-column index	The index that is created for a single column.
compound index	The index that is created for multiple columns in a table. A compound index can have indexed columns 1 and 2.
indexed attribute column	The predefined column in a base table that is mapped to non-primary key columns in an index table.
autocomplete	Tablestore automatically adds all primary key columns of the base table to the index table.

Limits

- The index table names must be unique in an instance.
- You can create a maximum of five index tables for a base table. If the limit is reached, the index table fails to be created.
- You can create a maximum of 15 predefined columns for a base table. If the limit is reached, the base table fails to be created.
- An index table can contain a maximum of four indexed columns, which are random combinations of the primary keys and predefined columns of the base table. If the limit is reached, the index table fails to be created.
- An index table can contain a maximum of eight attribute columns. If the limit is reached, the index table fails to be created.
- You can set the data type of an indexed column to STRING, INTEGER, or BINARY. The limits on index columns are the same as those on primary key columns of the base table.
- If an index table contains multiple columns, the size limit on the columns is the same as that on primary key columns of the base table.
- If you specify a column of the STRING or BINARY type as an attribute column of an index table, the limits on attribute columns are the same as those on attribute columns of the base table.
- You cannot create an index table on a table that has the time to live (TTL) parameter configured. If

you want to create index tables on a table that has the TTL parameter configured, use DingTalk to contact technical support.

- You cannot create an index table from a base table that has the max versions parameter configured. If a base table has the max versions parameter configured, index tables fail to be created from the base table. You cannot configure the max versions parameter for a base table that is associated with an index table.
- You cannot customize versions when you write data to a base table that is associated with an index table. Otherwise, the data fails to be written to the base table.
- You cannot use the Stream feature in an index table.
- An indexed base table cannot contain repeated rows that have the same primary key during the same batch write operation. Otherwise, the data fails to be written to the base table.

Usage notes

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

  When you create an index for the Defined0 column, Tablestore generates an index table that has the primary key columns Defined0, PK0, and PK1. When you create an index for the Defined0 and PK1 columns, Tablestore generates an index table that has the primary key columns Defined0, PK1, and PK0. When you create an index for the PK1 column, Tablestore generates an index table that has the primary key columns PK1 and PK0. When you create an index table, you need only to specify the column that you want to index. Tablestore adds the other primary key columns of the central table to the index table. For example, a base table contains the primary key columns PK0 and PK1 and a predefined column Defined0.

  - When you create an index for the Defined0 column, Tablestore generates the index table that has the primary key columns Defined0, PK0, and PK1.
  - When you create an index for the PK1 column, Tablestore generates the index table that has the primary key columns PK1 and PK0.

- You can specify predefined columns as attribute columns in the base table. When you specify a predefined column of the base table as an attribute column of the index table, you can search this index table instead of the base table for the column value. However, this increases storage costs. Otherwise, you must query the base table based on the index table. You can choose between these methods.
- We recommend that you do not specify a column whose values are date or time as the first primary key column of an index table because it may slow down index table updates. We recommend that you hash columns related to the time or date and create indexes for the hashed columns. If you have similar requirements, use DingTalk to contact technical support.
- We recommend that you do not define a column of low cardinality or a column that contains enumerated values as the first primary key column of an index table. For example, the gender column restricts the horizontal scalability of the index table and leads to poor write performance.

8.7.2.3. Scenarios

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

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

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

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

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

Assume that you have the following query requirements:

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

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

```
private static void getRangeFromMainTable(SyncClient client, long cellNumber)
{
  RangeRowQueryCriteria rangeRowQueryCriteria = new RangeRowQueryCriteria(TABLE_NAME);
  // Specify the primary key to start from.
  PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
  startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKeyValue.fromLong(cel
lNumber));
  startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKeyValue.fromLong(0))
;
  rangeRowQueryCriteria.setInclusiveStartPrimaryKey(startPrimaryKeyBuilder.build());
  // Specify the primary key to end with.
  PrimaryKeyBuilder endPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
  endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKeyValue.fromLong(cell
Number));
  endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKeyValue.INF_MAX);
  rangeRowQueryCriteria.setExclusiveEndPrimaryKey(endPrimaryKeyBuilder.build());
  rangeRowQueryCriteria.setMaxVersions(1);
  String strNum = String.format("%d", cellNumber);
  System.out.println("The cell number" + strNum + "makes the following calls:");
  while (true) {
    GetRangeResponse getRangeResponse = client.getRange(new GetRangeRequest(rangeRowQueryCrit
eria));
    for (Row row : getRangeResponse.getRows()) {
      System.out.println(row);
    }
    // If the nextStartPrimaryKey value is not null, continue the read operation.
    if (getRangeResponse.getNextStartPrimaryKey() ! = null) {
      rangeRowQueryCriteria.setInclusiveStartPrimaryKey(getRangeResponse.getNextStartPrimaryKey()
);
    } else {
      break;
    }
  }
}
```

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

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

  `IndexOnBeCalledNumber`:

PK0	PK1	PK2
CalledNumber	CellNumber	StartTime
123456	234567	1532574734
123456	345678	1532574795
123456	345678	1532574861
654321	123456	1532574644

PK0	PK1	PK2
765432	234567	1532574714
345678	456789	1532584054

> ⓘ **Note** Tablestore automatically adds all primary key columns of the central table to the index table. The primary key of the global secondary index consists of the index column and the primary key columns of the base table. Therefore, the global secondary index contains three primary key columns.

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

```
private static void getRangeFromIndexTable(SyncClient client, long cellNumber) {
   RangeRowQueryCriteria rangeRowQueryCriteria = new RangeRowQueryCriteria(INDEX0_NAME);
   // Specify the primary key to start from.
   PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
   startPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_1, PrimaryKeyValue.fromLong(cellNumber));
   startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKeyValue.INF_MIN);
   startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKeyValue.INF_MIN);
   rangeRowQueryCriteria.setInclusiveStartPrimaryKey(startPrimaryKeyBuilder.build());
   // Specify the primary key to end with.
   PrimaryKeyBuilder endPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
   endPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_1, PrimaryKeyValue.fromLong(cellNumber));
   endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKeyValue.INF_MAX);
   endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKeyValue.INF_MAX);
   rangeRowQueryCriteria.setExclusiveEndPrimaryKey(endPrimaryKeyBuilder.build());
   rangeRowQueryCriteria.setMaxVersions(1);
   String strNum = String.format("%d", cellNumber);
   System.out.println("The cell number" + strNum + "was called by the following numbers:");
   while (true) {
      GetRangeResponse getRangeResponse = client.getRange(new GetRangeRequest(rangeRowQueryCriteria));
      for (Row row : getRangeResponse.getRows()) {
         System.out.println(row);
      }
      // If the nextStartPrimaryKey value is not null, continue the read operation.
      if (getRangeResponse.getNextStartPrimaryKey() ! = null) {
         rangeRowQueryCriteria.setInclusiveStartPrimaryKey(getRangeResponse.getNextStartPrimaryKey());
      } else {
         break;
      }
   }
}
```

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

This query specifies `BaseStationNumber` and `StartTime` as conditions. Therefore, you can create a compound index based on the `BaseStationNumber` and `StartTime` columns. `IndexOnBaseStation1` :

PK0	PK1	PK2
BaseStationNumber	StartTime	CellNumber
001	1532574644	123456
001	1532574714	234567
002	1532574795	345678
002	1532574861	345678
003	1532574734	234567
003	1532584054	456789

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

```
private static void getRangeFromIndexTable(SyncClient client,
                                           long baseStationNumber,
                                           long startTime) {
    RangeRowQueryCriteria rangeRowQueryCriteria = new RangeRowQueryCriteria(INDEX1_NAME);
    // Specify the primary key to start from.
    PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
    startPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKeyValue.fromLong(ba
seStationNumber));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKeyValue.fromLong(sta
rtTime));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKeyValue.INF_MIN);
    rangeRowQueryCriteria.setInclusiveStartPrimaryKey(startPrimaryKeyBuilder.build());
    // Specify the primary key to end with.
    PrimaryKeyBuilder endPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
    endPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKeyValue.fromLong(bas
eStationNumber));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKeyValue.INF_MAX);
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKeyValue.INF_MAX);
    rangeRowQueryCriteria.setExclusiveEndPrimaryKey(endPrimaryKeyBuilder.build());
    rangeRowQueryCriteria.setMaxVersions(1);
    String strBaseStationNum = String.format("%d", baseStationNumber);
    String strStartTime = String.format("%d", startTime);
    System.out.println("All called numbers forwarded by the base station" + strBaseStationNum + "that sta
rt from" + strStartTime + "are listed:");
    while (true) {
        GetRangeResponse getRangeResponse = client.getRange(new GetRangeRequest(rangeRowQueryCrit
eria));
        for (Row row : getRangeResponse.getRows()) {
            System.out.println(row);
        }
        // If the nextStartPrimaryKey value is not null, continue the read operation.
        if (getRangeResponse.getNextStartPrimaryKey() ! = null) {
            rangeRowQueryCriteria.setInclusiveStartPrimaryKey(getRangeResponse.getNextStartPrimaryKey()
);
        } else {
            break;
        }
    }
}
```

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

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

```
private static void getRowFromIndexAndMainTable(SyncClient client,
                                                long baseStationNumber,
                                                long startTime,
                                                long endTime,
                                                String colName) {
    RangeRowQueryCriteria rangeRowQueryCriteria = new RangeRowQueryCriteria(INDEX1_NAME);
    // Specify the primary key to start from.
    PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
    PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
    startPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKeyValue.fromLong(ba
seStationNumber));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKeyValue.fromLong(sta
rtTime));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKeyValue.INF_MIN);
    rangeRowQueryCriteria.setInclusiveStartPrimaryKey(startPrimaryKeyBuilder.build());
    // Specify the primary key to end with.
    PrimaryKeyBuilder endPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
    endPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKeyValue.fromLong(bas
eStationNumber));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKeyValue.fromLong(end
Time));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKeyValue.INF_MAX);
    rangeRowQueryCriteria.setExclusiveEndPrimaryKey(endPrimaryKeyBuilder.build());
    rangeRowQueryCriteria.setMaxVersions(1);
    String strBaseStationNum = String.format("%d", baseStationNumber);
    String strStartTime = String.format("%d", startTime);
    String strEndTime = String.format("%d", endTime);
    System.out.println("The duration of calls forwarded by the base station" + strBaseStationNum + "from
" + strStartTime + "to" + strEndTime + "is listed:");
    while (true) {
        GetRangeResponse getRangeResponse = client.getRange(new GetRangeRequest(rangeRowQueryCrit
eria));
        for (Row row : getRangeResponse.getRows()) {
            PrimaryKey curIndexPrimaryKey = row.getPrimaryKey();
            PrimaryKeyColumn mainCalledNumber = curIndexPrimaryKey.getPrimaryKeyColumn(PRIMARY_KEY
_NAME_1);
            PrimaryKeyColumn callStartTime = curIndexPrimaryKey.getPrimaryKeyColumn(PRIMARY_KEY_NAM
E_2);
            PrimaryKeyBuilder mainTablePKBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
            mainTablePKBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, mainCalledNumber.getValue())
;
            mainTablePKBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, callStartTime.getValue());
            PrimaryKey mainTablePK = mainTablePKBuilder.build(); // Specify primary keys for the base table.
            // Query the base table.
            SingleRowQueryCriteria criteria = new SingleRowQueryCriteria(TABLE_NAME, mainTablePK);
            criteria.addColumnsToGet(colName); // Read the Duration column value of the base table.
            // Set the latest version to read.
            criteria.setMaxVersions(1);
            GetRowResponse getRowResponse = client.getRow(new GetRowRequest(criteria));
            Row mainTableRow = getRowResponse.getRow();
            System.out.println(mainTableRow);
        }
        // If the nextStartPrimaryKey value is not null, continue the read operation.
        if (getRangeResponse.getNextStartPrimaryKey() ! = null) {
            rangeRowQueryCriteria.setInclusiveStartPrimaryKey(getRangeResponse.getNextStartPrimaryKey()
);
        } else {
            break;
        }
    }
}
```

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

The following index table is created.

IndexOnBaseStation2 :

PK0	PK1	PK2	Defined0
BaseStationNumber	StartTime	CellNumber	Duration
001	1532574644	123456	60
001	1532574714	234567	10
002	1532574795	345678	5
002	1532574861	345678	100
003	1532574734	234567	20
003	1532584054	456789	200

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

~~~
private static void getRangeFromIndexTable(SyncClient client,
                                           long baseStationNumber,
                                           long startTime,
                                           long endTime,
                                           String colName) {
    RangeRowQueryCriteria rangeRowQueryCriteria = new RangeRowQueryCriteria(INDEX2_NAME);
    // Specify the primary key to start from.
    PrimaryKeyBuilder startPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
    startPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKeyValue.fromLong(ba
seStationNumber));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKeyValue.fromLong(sta
rtTime));
    startPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKeyValue.INF_MIN);
    rangeRowQueryCriteria.setInclusiveStartPrimaryKey(startPrimaryKeyBuilder.build());
    // Specify the primary key to end with.
    PrimaryKeyBuilder endPrimaryKeyBuilder = PrimaryKeyBuilder.createPrimaryKeyBuilder();
    endPrimaryKeyBuilder.addPrimaryKeyColumn(DEFINED_COL_NAME_3, PrimaryKeyValue.fromLong(bas
eStationNumber));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_2, PrimaryKeyValue.fromLong(end
Time));
    endPrimaryKeyBuilder.addPrimaryKeyColumn(PRIMARY_KEY_NAME_1, PrimaryKeyValue.INF_MAX);
    rangeRowQueryCriteria.setExclusiveEndPrimaryKey(endPrimaryKeyBuilder.build());
    // Specify the name of the column to read.
    rangeRowQueryCriteria.addColumnsToGet(colName);
    rangeRowQueryCriteria.setMaxVersions(1);
    String strBaseStationNum = String.format("%d", baseStationNumber);
    String strStartTime = String.format("%d", startTime);
    String strEndTime = String.format("%d", endTime);
    System.out.println("The duration of calls forwarded by the base station" + strBaseStationNum + "from
" + strStartTime + "to" + strEndTime + "is listed:");
    while (true) {
        GetRangeResponse getRangeResponse = client.getRange(new GetRangeRequest(rangeRowQueryCrit
eria));
        for (Row row : getRangeResponse.getRows()) {
            System.out.println(row);
        }
        // If the nextStartPrimaryKey value is not null, continue the read operation.
        if (getRangeResponse.getNextStartPrimaryKey() ! = null) {
            rangeRowQueryCriteria.setInclusiveStartPrimaryKey(getRangeResponse.getNextStartPrimaryKey()
);
        } else {
            break;
        }
    }
}
```
~~~

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

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

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

9.ApsaraDB for RDS

9.1. What is ApsaraDB for RDS?

ApsaraDB for RDS is a stable, reliable, and scalable online database service. Based on the distributed file system and high-performance storage, ApsaraDB for RDS allows you to perform database operations and maintenance with its set of solutions for disaster recovery, backup, restoration, monitoring, and migration.

ApsaraDB for RDS supports four storage engines, which are MySQL, SQL Server, PolarDB, and PostgreSQL. You can create database instances based on these storage engines to meet your business requirements.

RDS MySQL

Originally based on a branch of MySQL, ApsaraDB RDS for MySQL provides excellent performance. It is a tried and tested solution that handled the high-volume concurrent traffic during Double 11. ApsaraDB RDS for MySQL provides basic features, such as whitelist configuration, backup and restoration, Transparent Data Encryption (TDE), data migration, and management for instances, accounts, and databases. ApsaraDB RDS for MySQL also provides the following advanced features:

- **Read-only instance:** In scenarios where ApsaraDB for RDS handles a small number of write requests but a large number of read requests, you can create read-only instances to scale up the reading capability and increase the application throughput.
- **Read/write splitting:** The read/write splitting feature provides an extra read/write splitting endpoint. This endpoint enables an automatic link for the primary instance and all its read-only instances. An application can connect to the read/write splitting endpoint to read and write data. Write requests are automatically distributed to the primary instance while read requests are distributed to read-only instances based on their weights. To scale up the reading capacity of the system, you can add more read-only instances.

RDS SQL Server

ApsaraDB RDS for SQL Server provides strong support for a variety of enterprise applications under the high-availability architecture, has the capability of restoring data to any point in time.

ApsaraDB RDS for SQL Server provides basic features such as whitelist configuration, backup and restoration, transparent data encryption, data migration, and management for instances, accounts, and databases.

PolarDB

PolarDB is a stable, secure, and scalable enterprise-class relational database. Based on PostgreSQL, PolarDB enhances performance, application solutions, and compatibility. It also provides the capability of directly running Oracle applications. You can run various enterprise applications on PolarDB stably at low costs.

PolarDB provides features such as account management, resource monitoring, backup and restoration, and security control. These features are under continuous improvement, and more features are under development to adapt to PolarDB.

RDS PostgreSQL

ApsaraDB RDS for PostgreSQL is an advanced open source database that is fully compatible with SQL and supports a diverse range of data formats such as JSON, IP, and geometric data. In addition to support for features such as transactions, subqueries, multi-version concurrency control (MVCC), and data integrity check, ApsaraDB RDS for PostgreSQL integrates a series of features including high availability, backup, and restoration to ease operations and maintenance loads.

9.2. Benefits

9.2.1. Ease of use

ApsaraDB RDS is a ready-to-use service that provides features such as on-demand upgrade, easy management, high transparency, and high compatibility.

Ready-to-use

You can use API operations to create ApsaraDB RDS instances of your desired instance type.

On-demand upgrade

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

Transparency and compatibility

You can easily use ApsaraDB RDS in the same way as native database engines without the need to acquire new knowledge. ApsaraDB RDS is compatible with your existing programs and tools. Data can be migrated to ApsaraDB RDS by using ordinary import and export tools.

Easy management

You can add, delete, restart, back up, and restore databases by using the Apsara Uni-manager Management Console.

9.2.2. High performance

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

Parameter optimization

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

SQL optimization

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

High-end back-end hardware

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

9.2.3. High security

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

DDoS attack prevention

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

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

Access control

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

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

System security

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

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

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

TDE

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

9.2.4. High reliability

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

Hot standby

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

Multi-copy redundancy

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

Data backup

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

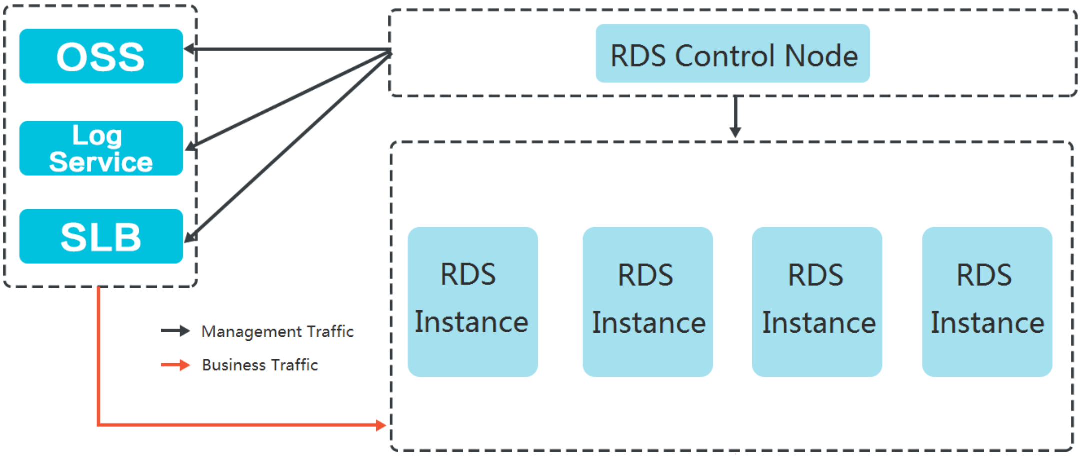
Data recovery

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

9.3. Architecture

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

RDS system architecture



9.4. Features

9.4.1. Data link service

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

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

DNS

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

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

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

SLB

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

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

9.4.2. High-availability service

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

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

Detection

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

Repair

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

For example:

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

Notice

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

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

HA policies

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

There are two service priorities:

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

There are three data replication modes:

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

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

9.4.3. Backup and recovery service

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

ApsaraDB RDS can back up databases anytime and restore them to a point in time based on backup policies, which makes data more traceable.

Backup

The Backup module compresses and uploads data and logs on both the primary and secondary nodes. By default, ApsaraDB RDS uploads backup files to Object Storage Service (OSS). When the secondary node operates normally, backups are always created on the secondary node. This way, the services on the primary node are not affected. When the secondary node is unavailable or damaged, the Backup module creates backups on the primary node.

Recovery

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

- Primary node rollback: rolls back the primary node to a specific point in time when an operation error occurs.
- Secondary node repair: creates a new secondary node to reduce risks when an irreparable fault occurs on the secondary node.

- Read-only instance creation: creates a read-only instance from backup files.

Storage

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

9.4.4. Monitoring service

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

Service

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

Network

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

OS

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

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

Instance

The Instance module collects the following information on RDS instances:

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

9.4.5. Scheduling service

The Resource module implements the scheduling of resources and services.

Resource

The Resource module allocates and integrates underlying RDS resources when you activate and migrate instances. When you use the RDS console or API to create an instance, the Resource module calculates the most suitable host to carry the traffic to and from the instance. This module also allocates and integrates the underlying resources required to migrate RDS instances. After repeated instance creation, deletion, and migration operations, the Resource module calculates the degree of resource fragmentation. It also regularly integrates resources to improve the service carrying capacity.

9.4.6. Migration service

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

The migration service helps you migrate data from on-premises databases to ApsaraDB RDS instances or between ApsaraDB RDS instances.

DTS

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

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

- Schema migration

  DTS migrates the schema definitions of migration objects to the destination instance. Tables, views, triggers, stored procedures, and stored functions can be migrated in this mode.

- Full migration

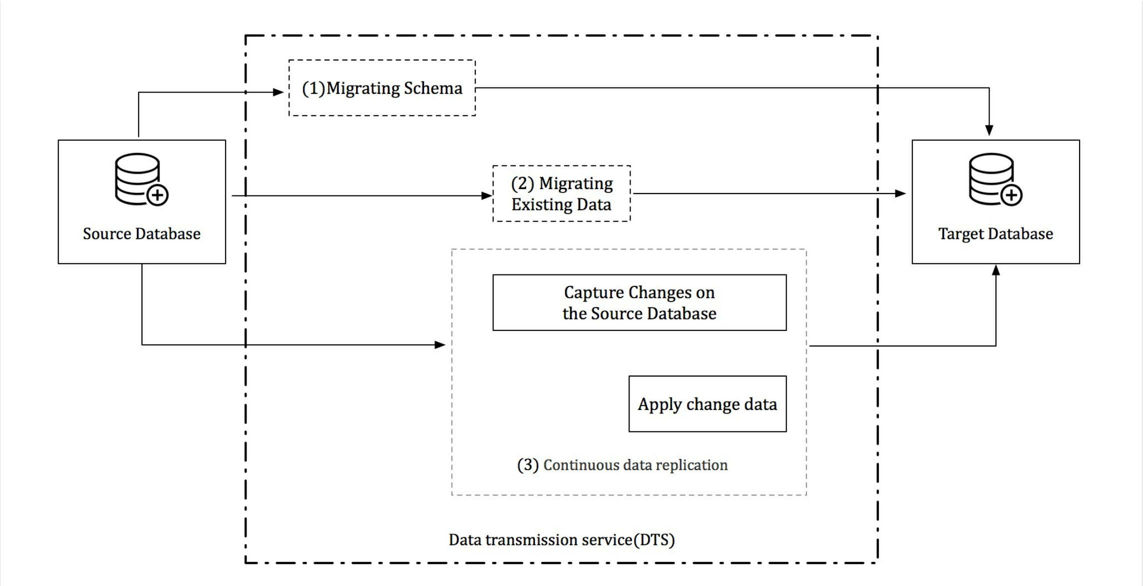
  DTS migrates all data of migration objects from the source database to the destination instance.

  > **Notice** For data consistency purposes, non-transaction tables that do not have primary keys are locked during a full migration. You cannot write data to locked tables. The lock duration is determined by the data volume in the tables. The tables are unlocked only after they are fully migrated.

- Incremental migration

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

  > **Notice** If a DDL operation is performed when data is migrated, schema changes are not synchronized to the destination instance.


(1)Migrating Schema
Source Database
(2) Migrating Existing Data
Target Database
Capture Changes on the Source Database
Apply change data
(3) Continuous data replication
Data transmission service(DTS)

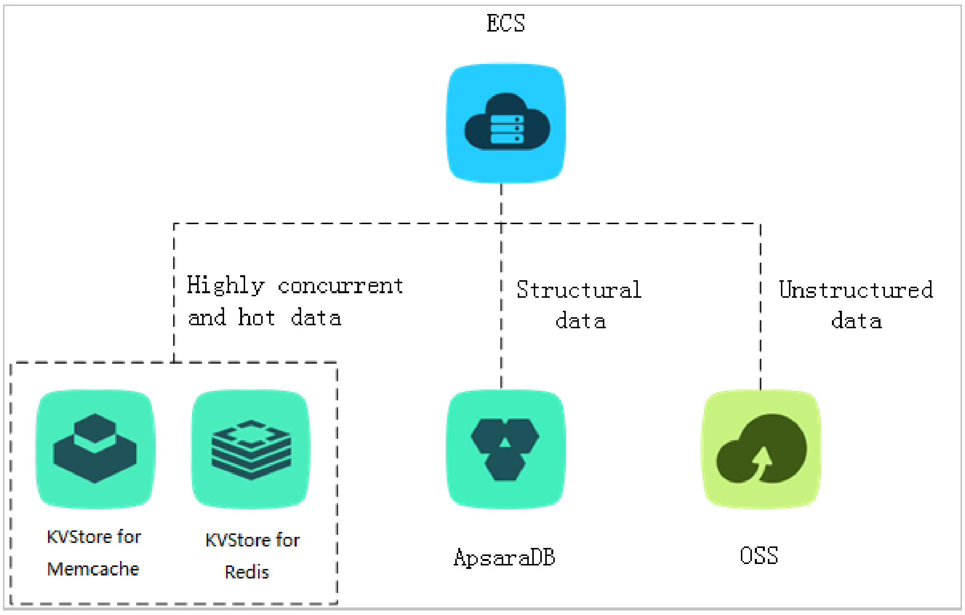
9.5. Scenarios

9.5.1. Diversified data storage

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

You can diversify the storage capabilities of ApsaraDB for RDS through services such as KVStore for Memcache, KVStore for Redis, and OSS, as shown in Diversified data storage.

Diversified data storage



Cache data persistence

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

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

Multi-structure data storage

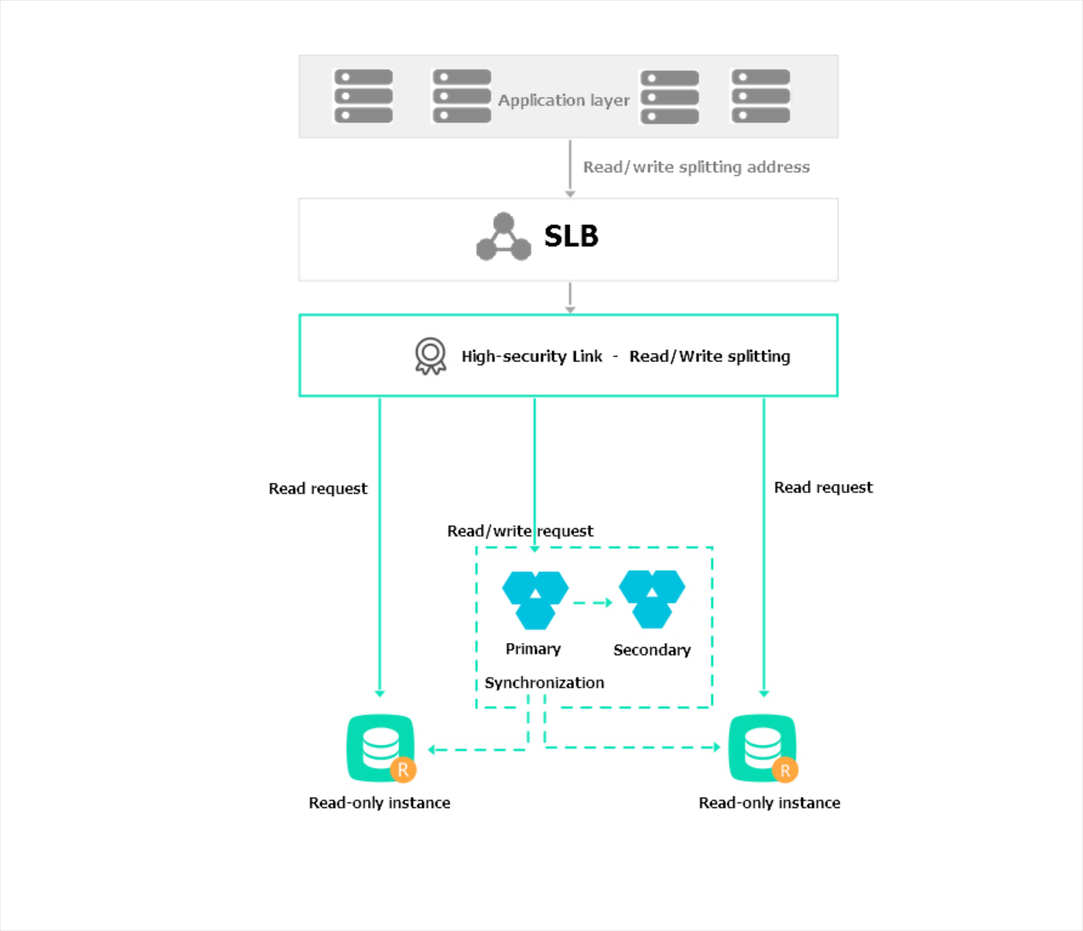
OSS is a secure, reliable, low-cost, and high-capacity storage service from Alibaba Cloud. ApsaraDB for RDS can be used with OSS to implement a multi-type data storage solution. For example, imagine ApsaraDB for RDS and OSS are used together to implement an online forum. Resources such as the images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

9.5.2. Read/write splitting

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

ApsaraDB RDS for MySQL allows you to directly attach read-only instances to ApsaraDB for RDS to reduce read pressure on the primary instance. The primary instance and read-only instances of ApsaraDB RDS for MySQL each have their own connection addresses. The system also offers an extra read/write splitting address after read/write splitting is enabled. This address associates the primary instance with all of its read-only instances for automatic read/write splitting, allowing applications to send all read and write requests to a single address. Write requests are automatically routed to the primary instance, and read requests are routed to each read-only instance based on their weights. You can scale out the processing capability of the system by adding more read-only instances. There is no need to modify applications, as shown in Read/write splitting.

Read/write splitting

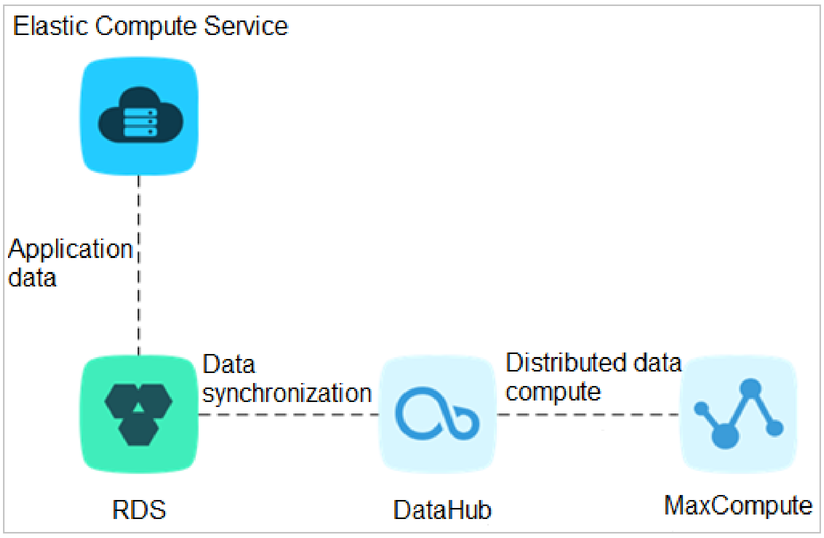


9.5.3. Big data analysis

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

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

Big data analysis diagram



9.6. Limits

9.6.1. Limits on ApsaraDB RDS for MySQL

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

To ensure instance stability and security, ApsaraDB RDS for MySQL has some limits. The Limits on ApsaraDB RDS for MySQL table describes the limits on ApsaraDB RDS for MySQL.

Limits on ApsaraDB RDS for MySQL

Operation	Limit
Database parameter modification	Most database parameters must be modified by using API operations. For security and stability considerations, only specific parameters can be modified.
Root permissions of databases	The root or system administrator permissions are not provided.
Database backup	• Logical backup can be performed by using the command line interface (CLI) or graphical user interface (GUI). • Physical backup can be performed only by using the ApsaraDB RDS console or API operations.
Database restoration	• Logical restoration can be performed by using the CLI or GUI. • Physical restoration can be performed only by using the ApsaraDB RDS console or API operations.
Database import	• Logical import can be performed by using the CLI or GUI. • Data can be imported by using the MySQL CLI or DTS.
ApsaraDB RDS for MySQL storage engine	• Only InnoDB and TokuDB are supported. Due to the inherent shortcomings of the MyISAM engine, some data may be lost. Only some existing instances use the MylSAM engine. MyISAM engine tables in new instances are converted to InnoDB engine tables. • For performance and security considerations, we recommend that you use the InnoDB storage engine. • The Memory engine is not supported. New Memory tables are converted to InnoDB tables.
Database replication	ApsaraDB RDS for MySQL provides dual-node clusters based on a primary/secondary replication architecture. The secondary instances in this replication architecture are hidden and cannot be directly accessed.
Instance restart	Instances must be restarted by using the ApsaraDB RDS console or API operations.

Operation	Limit
Account and database management	ApsaraDB RDS for MySQL uses the ApsaraDB RDS console to manage accounts and databases. ApsaraDB RDS for MySQL also allows you to create a privileged account to manage users, passwords, and databases.
Standard account	● Custom authorization is not supported. ● The ApsaraDB RDS console allows you to manage accounts and databases. ● Instances that support standard accounts also support privileged accounts.
Privileged account	● Custom authorization is supported. ● The ApsaraDB RDS console does not provide interfaces to manage accounts or databases. These operations can be performed only by using code or DMS. ● The privileged account cannot be reverted back to a standard account.

9.7. Terms

Term	Description
Region	The geographical location where the server of your RDS instance resides. You must specify a region when you create an RDS instance. The region of an instance cannot be changed after instance creation. RDS must be used together with ECS and only supports intranet access. Because of this, RDS instances must be located in the same region as their corresponding ECS instances.
Zone	The physical area with an independent power supply and network in a region. Zones in a region can communicate through the intranet. Network latency for resources within the same zone is lower than for those across zones. Faults are isolated between zones. Single zone refers to the case where the three nodes in the RDS instance replica set are all located in the same zone. Network latency is reduced if an ECS instance and its corresponding RDS instance are both deployed in the same zone.
Instance	The most basic unit of RDS. An instance is the operating environment of ApsaraDB for RDS and works as an independent process on a host. You can create, modify, or delete an RDS instance from the RDS console. Instances are mutually independent and their resources are isolated. They do not compete for resources such as CPU, memory, or I/O. Each instance has its own features, such as database type and version. RDS controls instance behavior by using corresponding parameters.
Memory	The maximum amount of memory that can be used by an ApsaraDB for RDS instance.
Disk capacity	The amount of disk space selected when creating an ApsaraDB for RDS instance. Instance data that occupies disk space includes aggregated data as well as data required for normal instance operations such as system databases, database rollback logs, redo logs, and indexing. Ensure that the disk capacity is sufficient for the RDS instance to store data. Otherwise, the RDS instance may be locked. If the instance is locked due to insufficient disk capacity, you can unlock the instance by expanding the disk capacity.
IOPS	The maximum number of read/write operations performed per second on block devices at a granularity of 4 KB.

Term	Description
CPU core	The maximum computing capability of the instance. A single Intel Xeon series CPU core has at least 2.3 GHz of computational power with hyper-threading capabilities.
Number of connections	The number of TCP connections between a client and an RDS instance. If the client uses a connection pool, the connection between the client and RDS instance is a persistent connection. Otherwise, it is a transient connection.

9.8. Instance types

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

ApsaraDB RDS for MySQL instance types

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

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk space	Zone-disaster recovery deployment
High-availability Edition	5.6 or 5.7	Dedicated instance (X4)	mysql.x4.large.2	4 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	Single-data center deployment and dual-data center deployment
			mysql.x4.xlarge.2	8 cores, 32 GB	5,000	9,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x4.2xlarge.2	16 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x4.4xlarge.2	32 cores,128 GB	20,000	36,000	1,000 GB to 3,000 GB (in 5 GB increments)	
		Dedicated host	rds.mysql.st.d13	30 cores, 220 GB	64,000	20,000	1,000 GB to 3,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk space	Zone-disaster recovery deployment
	5.6	Dedicated instance (with high memory)	mysql.x8.medium.4	2 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	Dual-data center deployment
			mysql.x8.large.4	4 cores, 32 GB	5,000	9,000	50 GB to 1,000 GB (in 5 GB increments)	
			mysql.x8.xlarge.4	8 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x8.2xlarge.4	16 cores, 128 GB	20,000	36,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x8.4xlarge.4	32 cores, 256 GB	40,000	72,000	1,000 GB to 3,000 GB (in 5 GB increments)	
			mysql.x4.large.25	4 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk space	Zone-disaster recovery deployment
Enterprise Edition	5.7	Dedicated instance (with high CPU)	mysql.x4.xlarge.25	8 cores, 32 GB	5,000	9,000	500 GB to 3,000 GB (in 5 GB increments)	Single-data center deployment
			mysql.x4.2xlarge.25	16 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x4.4xlarge.25	32 cores, 128 GB	20,000	36,000	1,000 GB to 3,000 GB (in 5 GB increments)	
		Dedicated instance (with	mysql.x8.medium.25	2 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	
			mysql.x8.large.25	4 cores, 32 GB	5,000	9,000	50 GB to 1,000 GB (in 5 GB increments)	
			mysql.x8.xlarge.25	8 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	

Edition	Version	high memory) Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk space	Zone-disaster recovery deployment
			mysql.x8.2xlarge.25	16 cores, 128 GB	20,000	36,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysql.x8.4xlarge.25	32 cores, 256 GB	40,000	72,000	1,000 GB to 3,000 GB (in 5 GB increments)	
		Dedicated instance (X8)	mysqlro.x8.medium.1	2 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	
			mysqlro.x8.large.1	4 cores, 32 GB	5,000	9,000	50 GB to 1,000 GB (in 5 GB increments)	
			mysqlro.x8.xlarge.1	8 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysqlro.x8.2xlarge.1	16 cores, 128 GB	20,000	36,000	500 GB to 3,000 GB (in 5 GB increments)	

Edition	Version	Instance family	Instance type	CPU and memory	Maximum connections	Maximum IOPS	Disk space	Zone-disaster recovery deployment
Read-only instance	5.6 or 5.7	Dedicated instance (X4)	mysqlro.x4.large.1	4 cores, 16 GB	2,500	4,500	50 GB to 1,000 GB (in 5 GB increments)	Single-data center deployment and dual-data center deployment
			mysqlro.x4.xlarge.1	8 cores, 32 GB	5,000	9,000	50 GB to 1,000 GB (in 5 GB increments)	
			mysqlro.x4.2xlarge.1	16 cores, 64 GB	10,000	18,000	500 GB to 3,000 GB (in 5 GB increments)	
			mysqlro.x4.4xlarge.1	32 cores,128 GB	20,000	36,000	500 GB to 3,000 GB (in 5 GB increments)	
		Dedicated host	rds.mysql.st.d13	30 cores, 220 GB	64,000	20,000	1,000 GB to 3,000 GB (in 5 GB increments)	

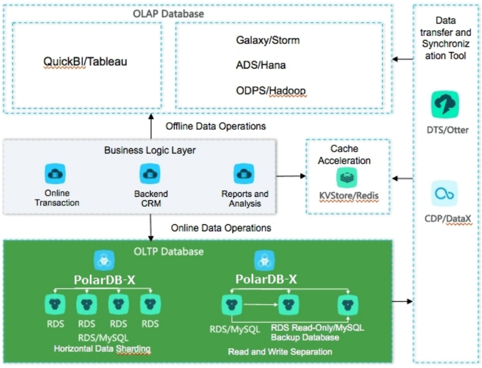
10.Cloud Native Distributed Database PolarDB-X

10.1. What is PolarDB-X?

Cloud Native Distributed Database PolarDB-X (PolarDB-X) is a middleware service independently developed by Alibaba Group for scale-out of single-instance relational databases. It is compatible with Distributed Relational Database Service (DRDS). Compatible with the MySQL protocol, PolarDB-X supports most MySQL data manipulation language (DML) and data definition language (DDL) syntax. It provides the core capabilities of distributed databases, such as database sharding, table sharding, smooth scale-out, configuration changing, and transparent read/write splitting. PolarDB-X features lightweight (stateless), flexibility, stability, and efficiency, and provides you with O&M capabilities throughout the lifecycle of distributed databases.

PolarDB-X is mainly used for operations on large-scale online data. By splitting data in specific business scenarios, PolarDB-X maximizes the operation efficiency, meeting the requirements of online businesses on relational databases.

Product structure diagram of PolarDB-X



Problems solved

- Capacity bottleneck of single-instance databases: As the data volume and access volume increase, traditional single-instance databases encounter great challenges that cannot be completely solved by hardware upgrades. Distributed solutions use multiple instances to work jointly, effectively resolving the bottlenecks of data storage capacity and access volumes.
- Difficult scale-out of relational databases: Due to the inherent attributes of distributed databases, data can be stored to different shards through smooth data migration, supporting the dynamic scale-out of relational databases.

10.2. Benefits

Audit logs

The audit logs will store to Log Service (SLS) automatically and you can download logs from SLS. This facilitates long-term storage and management of audit logs.

Auto scaling

PolarDB-X instances and ApsaraDB RDS for MySQL instances can be dynamically added and removed. This allows you to choose service capabilities in a flexible manner.

High availability of clusters

The multi-node cluster architecture is used. Each component management node in the platform needs to implement a high availability mechanism. Therefore, the failure of a service node in a cluster does not affect the overall operation of the corresponding service instance. Clusters have the adaptive load balancing capability, which ensures service operation in high-concurrency and high-load scenarios without requiring you to adjust cluster parameters.

Backup

The data of PolarDB-X is stored in ApsaraDB RDS for MySQL, which supports full or incremental backup and data recovery from storage. The capability of data cluster backup between data centers is provided, which allows data backup across data centers. The backup process is visualized.

Disaster recovery

Metadatabases support fast switchover for data recovery. Single-node failures do not affect services.

Security and controllability

PolarDB-X supports an account permission system similar to that of ApsaraDB RDS for MySQL, and provides useful features, such as the IP address whitelist and default disabling of high-risk SQL statements. PolarDB-X provides a complete standard API system that can be used by your local management system. Complete product support and architecture services are also available to you.

Isolation

By using the multi-instance approach, PolarDB-X supports multi-tenant parallel execution in a PolarDB-X cluster, and tenant tasks are submitted to queues on different instances for execution. PolarDB-X isolates resources among tenants by using PolarDB-X instances.

Permissions

- PolarDB-X allows you to manage tenants in a centralized manner, dynamically configure and manage tenant resources, isolate resources, view statistics on resource usage, and manage tenants at multiple levels in the console.
- PolarDB-X supports permission management and fine-grained audit of user operations.
- PolarDB-X provides a comprehensive permission authentication and isolation mechanism to ensure your data privacy.
- PolarDB-X allows you to isolate resources by using the multitenancy mode.

Scheduling

PolarDB-X supports multi-tenant scheduling in multiple clusters and resource pools.

10.3. Architecture

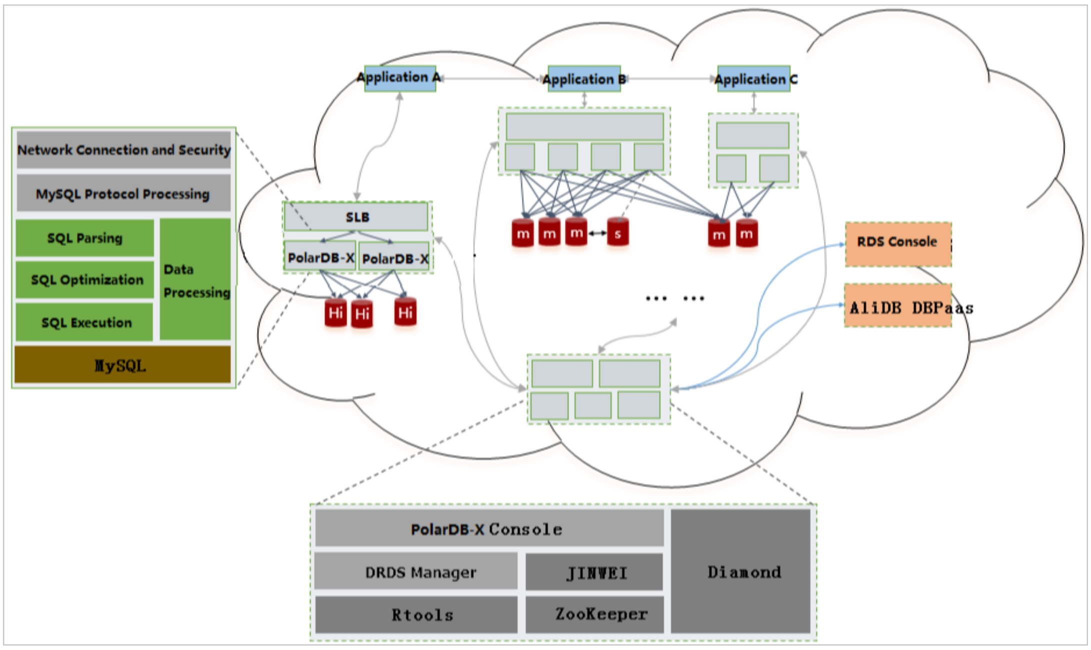
Cloud Native Distributed Database PolarDB-X (PolarDB-X) supports two data output methods: overall output by Apsara Stack and separate output by Alibaba middleware. The two output methods differ in features and dependent components of PolarDB-X.

The following table describes the differences between these two methods.

Item	Overall output by Apsara Stack	Separate output by Alibaba middleware
MySQL	ApsaraDB RDS for MySQL	Alibaba Cloud Database Platform as a Service (DBPaaS)
Load balancing	Centralized Server Load Balancer (Centralized SLB)	Client load balancer (VIPServer)
Special storage support	None	High compression-ratio column store (HiStore)

The following figure shows the system architecture of PolarDB-X.

PolarDB-X system architecture



PolarDB-X Server

PolarDB-X Server is the service layer of PolarDB-X. Multiple PolarDB-X Server nodes form a cluster to provide distributed database services, including read/write splitting, routed SQL execution, result merging, dynamic database configuration, and globally unique ID (GUID).

> **Note** PolarDB-X instances are stateless. Therefore, ApsaraDB RDS for MySQL instances are used for storage. PolarDB-X implements data encryption by using encryption algorithms such as transparent data encryption (TDE) supported by ApsaraDB RDS for MySQL.

ApsaraDB RDS for MySQL (marked by m and s in the figure)

ApsaraDB RDS for MySQL stores data and performs data operations online. It implements high availability by using primary/secondary replication. It also implements dynamic database failover with the primary/secondary switchover mechanism.

You can implement management, monitoring, and alerting within the instance lifecycle in the ApsaraDB RDS for MySQL console.

HiStore

When PolarDB-X outputs data separately (not overall output by Apsara Stack), it uses HiStore as the physical storage. HiStore is a low-cost and high-performance database developed by Alibaba to support column store. By using the column store, knowledge grid, and multiple cores, HiStore provides higher data aggregation and ad hoc query capabilities, with lower costs than row store (such as MySQL).

You can implement management, monitoring, and alerting within the instance lifecycle in the HiStore console.

DBPaaS

When PolarDB-X outputs data separately (not overall output by Apsara Stack), the ApsaraDB RDS for MySQL O&M platform DBPaaS implements management, monitoring, alerting, and resource management in the lifecycle of ApsaraDB RDS for MySQL instances.

Centralized SLB

You do not need to install a client on user instances. SLB is used to distribute your requests. When an instance fails or a new instance is added, SLB ensures that traffic on the bound instances is distributed evenly.

VIPServer

You must install a client on user instances, with a weak dependency on the central controller (interaction is performed only when the load configuration changes). VIPServer is used to distribute your requests. When an instance fails or a new instance is added, VIPServer ensures that traffic on the bound instances is distributed evenly.

Diamond

Diamond is a system responsible for PolarDB-X configuration storage and management. It provides the configuration storage, query, and notification functions. Diamond stores the database source data, sharding rules, and PolarDB-X switch configuration.

Data Replication System

Data Replication System is responsible for data migration and synchronization of PolarDB-X. The core capabilities of this system include full data migration and incremental data synchronization. Its derived features include smooth data import, smooth scale-out, and global secondary index (GSI). Data Replication System requires the support of ZooKeeper and PolarDB-X Rtools.

PolarDB-X console

PolarDB-X Console is designed for business database administrators (DBAs) to isolate resources as required and perform operations, such as instance management, database and table management, read/write splitting configuration, smooth scale-out, monitoring data display, and IP address whitelisting.

DRDS Manager

DRDS Manager is designed for global O&M personnel and DBAs. It provides the PolarDB-X resource management and system monitoring functions:

- Manages all resources on which ApsaraDB RDS for MySQL instances depend, including virtual machines, SLB instances, and domain names.

- Monitors the status of PolarDB-X instances, including queries per second (QPS), active threads, connections, node network I/O, and node CPU utilization.

Rtools

Rtools is the O&M support system of PolarDB-X. It allows you to manage database configuration, read/write weight, connection parameters, database and table topologies, and sharding rules.

10.4. Features

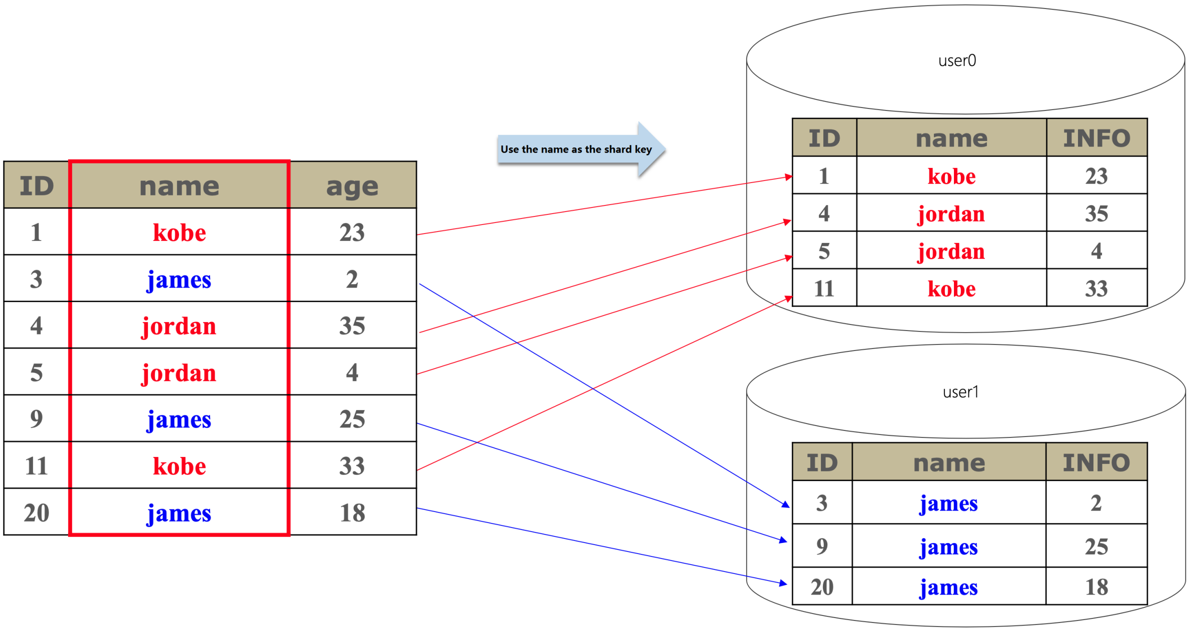
10.4.1. Scalability

Concurrency and storage capacity scalability

The essence of scalability lies in splitting. PolarDB-X distributes data to multiple ApsaraDB RDS for MySQL instances to obtain the distribution of read/write requests and storage through Horizontal partitioning. The PolarDB-X layer is stateless and increases nodes to cope with concurrent SQL loads, which is similar to a business application.

Horizontal partitioning

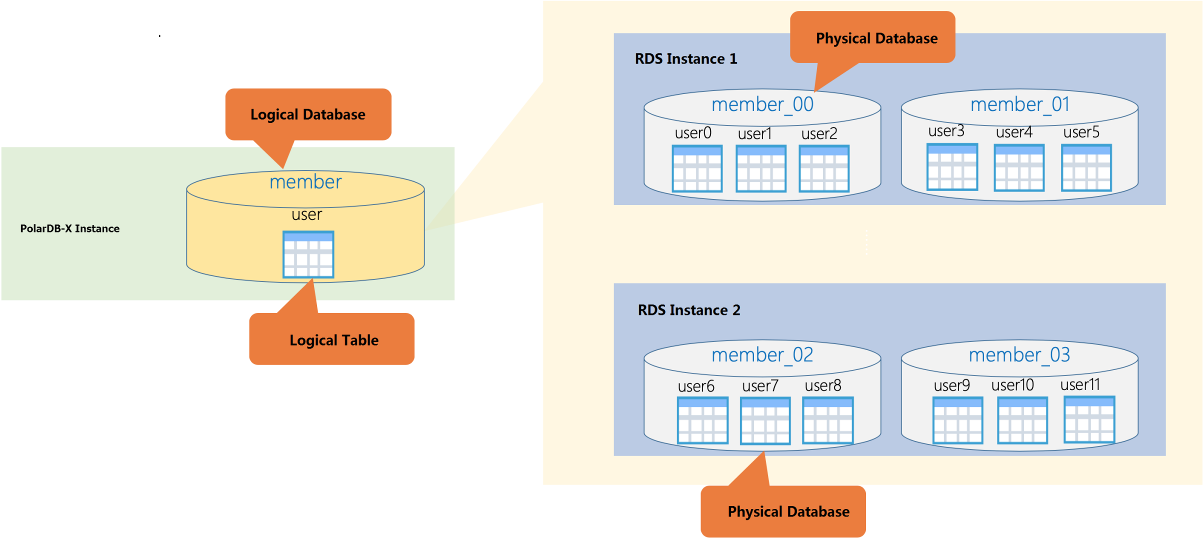
Data is distributed to multiple ApsaraDB RDS for MySQL instances based on certain calculation and routing rules. In fact, PolarDB-X has many algorithms to cope with the loads in various scenarios.



ID	name	age
1	kobe	23
3	james	2
4	jordan	35
5	jordan	4
9	james	25
11	kobe	33
20	james	18

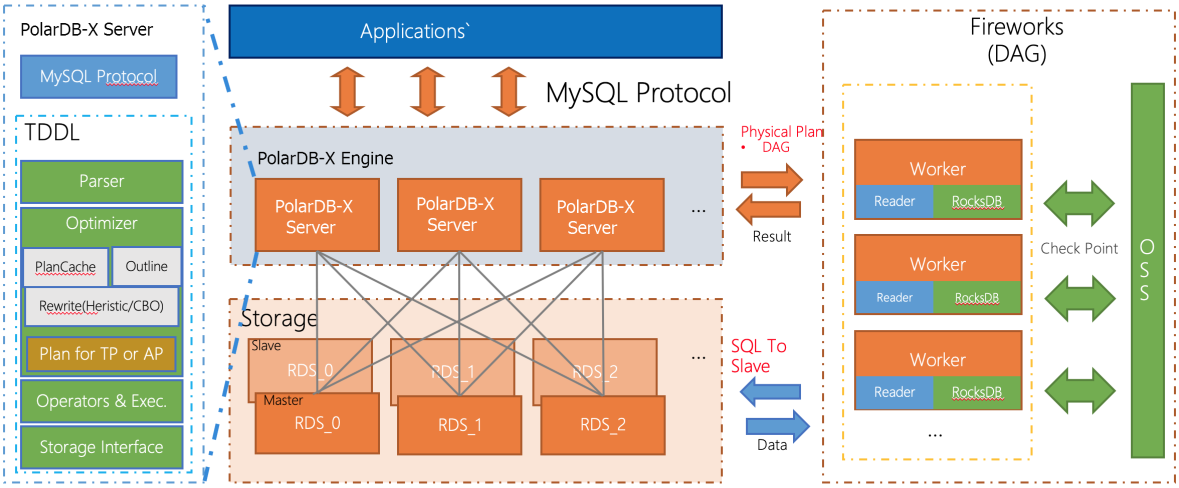
ID	name	INFO
1	kobe	23
4	jordan	35
5	jordan	4
11	kobe	33

ID	name	INFO
3	james	2
9	james	25
20	james	18



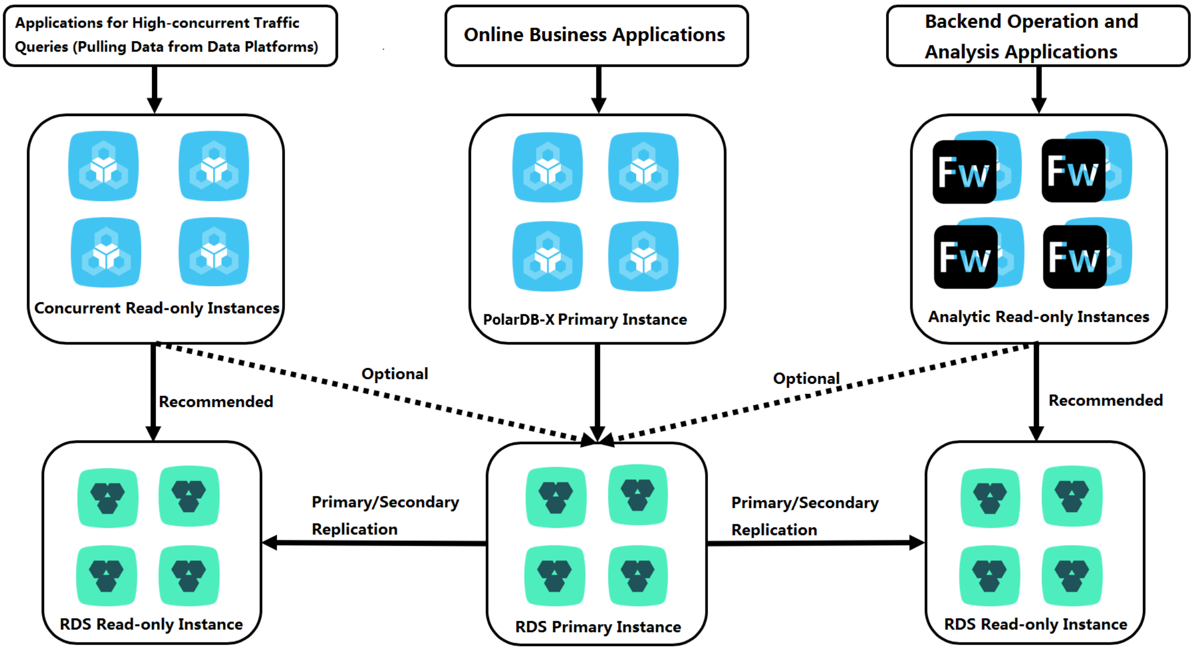
Computing scalability

PolarDB-X often needs to perform complex computing on data far exceeding the capacity of a single instance. These SQL statements include multi-table join, multi-layer nested subqueries, grouping, sorting, and aggregation.



To process complex SQL statements in the online databases, PolarDB-X has expanded the Symmetric Multi-Processing (SMP) and Massively Parallel Processing with Directed Acyclic Graph (MPP&DAG). SMP is fully integrated into the PolarDB-X kernel, while MPP&DAG of PolarDB-Xbuilds a computing cluster that dynamically obtains execution plans for distributed computing at runtime and improves the computing capability by adding nodes.

Currently, thePolarDB-X instances that process data on multiple instances in parallel are provided for businesses in the form of analytic read-only instances.


Applications for High-concurrent Traffic Queries (Pulling Data from Data Platforms)
Online Business Applications
Backend Operation and Analysis Applications
Concurrent Read-only Instances
PolarDB-X Primary Instance
Analytic Read-only Instances
Optional
Recommended
Optional
Recommended
Primary/Secondary Replication
Primary/Secondary Replication
RDS Read-only Instance
RDS Primary Instance
RDS Read-only Instance

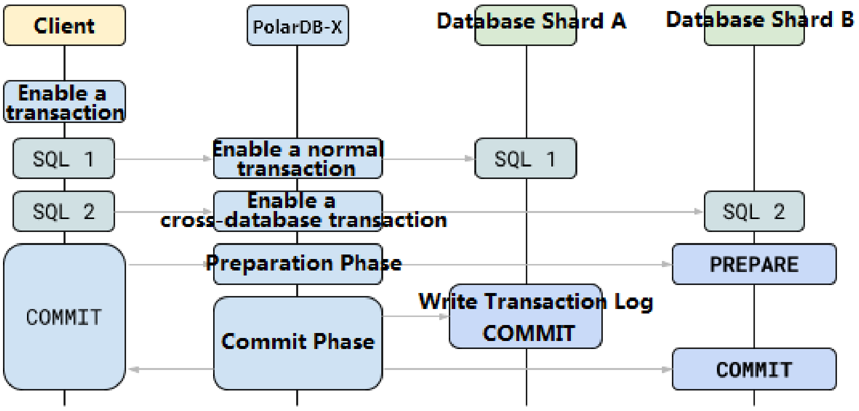
10.4.2. Distributed transactions

Distributed transactions use Two-Phase Commit (2PC) to ensure the atomicity and consistency of transactions.

A 2PC transaction is divided into the PREPARE phase and the COMMIT phase.

- In the PREPARE phase, data nodes prepare all the resources required for committing transactions, such as locking and logging.
- In the COMMIT phase, data nodes commit transactions.

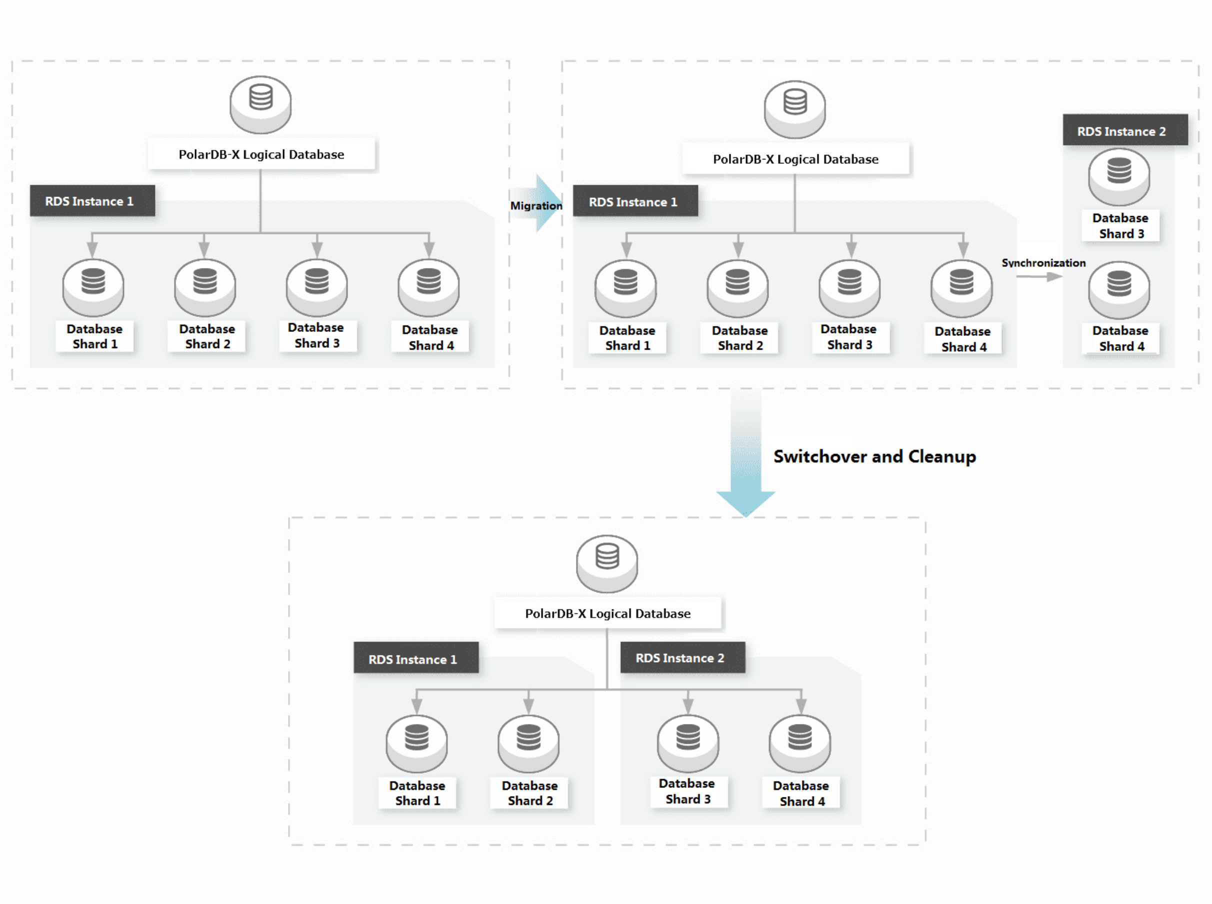
When you commit a distributed transaction, the PolarDB-X server, as a transaction manager, sends a COMMIT request to each data node only after all data nodes (MySQL servers) have their resources ready in PREPARE phase.



10.4.3. Smooth scale-out

When the underlying storage of the logical database reaches the physical bottleneck, for example, when the remaining disk space is about 30%, you can smoothly scale it out to improve the performance.

Smooth scale-out is an online horizontal expansion method. It smoothly migrates the original database shards to the new ApsaraDB RDS for MySQL instances and increases the overall data storage capacity by adding ApsaraDB RDS for MySQL instances, which reduces the pressure on each RDS instance to process data.



10.4.4. Read/write splitting

When a primary ApsaraDB RDS for MySQL instance is heavily loaded with many read requests, you can use the read/write splitting function of PolarDB-X to distribute the read traffic, which reduces the read pressure on the primary ApsaraDB RDS for MySQL instance.

The read/write splitting function of PolarDB-X is transparent to applications. The read traffic can be distributed to the primary ApsaraDB RDS for MySQL instance and multiple ApsaraDB RDS for MySQL read-only instances according to the read weight set in the PolarDB-X console, without changing any code of the application. All the write traffic is distributed to the primary ApsaraDB RDS for MySQL instance.

After read/write splitting is set, real-time strong consistency can be implemented when data is read from the primary ApsaraDB RDS for MySQL instance. Data on the read-only instances is replicated asynchronously from the primary ApsaraDB RDS for MySQLS instance, with a millisecond-level latency, therefore real-time strong consistency cannot be implemented when data is read from read-only ApsaraDB RDS for MySQL instances. For SQL statements which require real time and strong consistency for reading data, specify the primary ApsaraDB RDS for MySQL instance to execute these statements through hints of PolarDB-X.

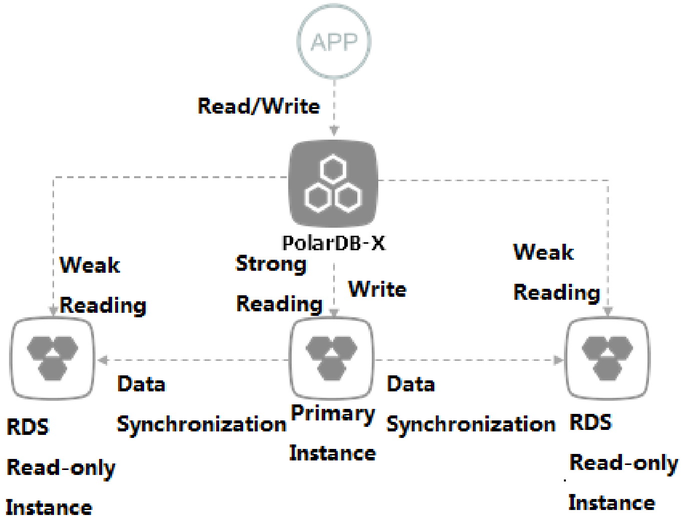
Read/write splitting in non-partition mode

In non-partition mode, PolarDB-X can implement read/write splitting without horizontal partitioning. When you create a PolarDB-X database in the PolarDB-X console, after you select an ApsaraDB RDS for MySQL instance, you can directly import a database in the instance to PolarDB-X for read/write splitting. In this case, you do not need to migrate data, but you also cannot perform horizontal partitioning on tables in the PolarDB-X database.

Support for transactions by read/write splitting

Read/write splitting is valid only for read requests (query requests) that are not in explicit transactions (transactions that need to be explicitly committed or rolled back). Write requests and read requests (including read-only transactions) in explicit transactions are executed in the primary instance and are not distributed to read-only instances.

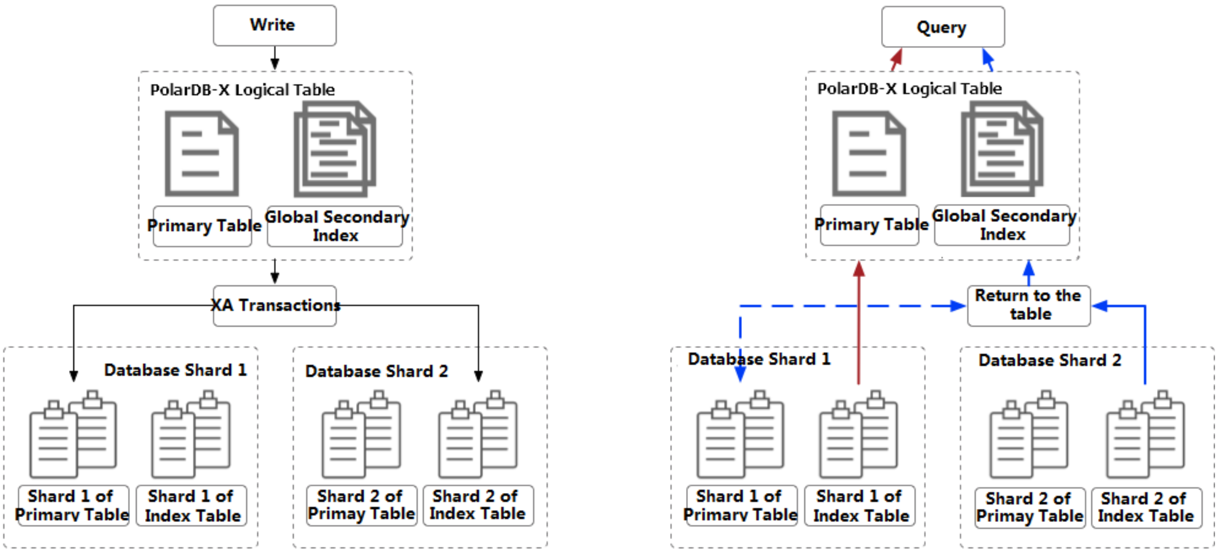
- Common SQL statements for read requests include SELECT, SHOW, EXPLAIN, and DESCRIBE.
- Common SQL statements for write requests include INSERT, REPLACE, UPDATE, DELETE, and CALL.



Read/Write Splitting

10.4.5. Global secondary index

Global secondary indexes of PolarDB-X allow users to add shard dimensions as needed and provides globally unique constraints. Each global secondary index corresponds to an index table and uses XA transactions to ensure strong data consistency between primary tables and index tables.



The global secondary indexes of PolarDB-X provide the following capabilities:

- Add dimensions for sharding.
- Support globally unique indexes.
- Provide XA transactions to ensure strong data consistency between primary tables and index tables.
- Support overwrite columns to reduce overheads from querying the primary table.
- Support Online Schema Change, so the primary table remains unlocked when a global secondary index is added.
- Uses hints to specify indexes to automatically determine whether to query the primary table.

FAQ

Q: What problems can global secondary indexes solve?

A: If the queried dimension is different from the dimension for sharding of a logical table, cross-shard queries are initiated. As cross-shard queries increase, performance problems such as slow query and connection pool exhaustion may occur. Global secondary indexes reduce cross-shard queries and eliminates performance bottlenecks by adding dimensions for sharding. When creating a global secondary index, you need to select a shard key that is different from that of the primary table.

Q: What is the relationship between a global secondary index and a local secondary index?

A:

- A local secondary index stores data rows and corresponding index rows on the same shard in a distributed database. In PolarDB-X, it specifically refers to a MySQL secondary index of a physical table.
- A global secondary index stores data rows and corresponding index rows on different shards, which is different from a local secondary index. A global secondary index quickly determines the data shards involved in the query.
- When PolarDB-X distributes queries to a single shard through a global secondary index, the local secondary index of the shard can improve the performance of the query within the shard.

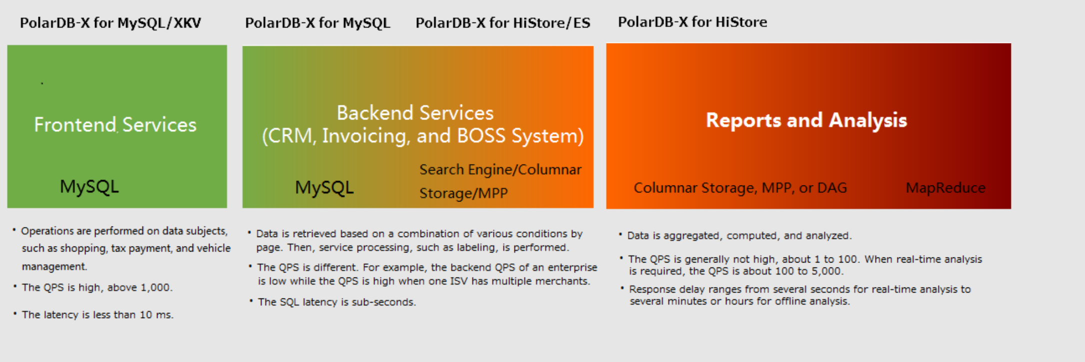
10.5. Scenarios

This topics describes the typical scenarios of PolarDB-X.

PolarDB-X is suitable for businesses that feature high concurrency and low latency in the frontend. It partitions data in specific business scenarios and provides distributed secondary indexes, enabling business databases to keep a high upper limit for queries per second (QPS).

PolarDB-X is trying to support Alibaba columnar databases to meet the needs of the huge-volume storage with low costs, efficient data aggregation, and ad hoc queries.

PolarDB-X scenarios



- Operations are performed on data subjects, such as shopping, tax payment, and vehicle management.
- The QPS is high, above 1,000.
- The latency is less than 10 ms.

- Data is retrieved based on a combination of various conditions by page. Then, service processing, such as labeling, is performed.
- The QPS is different. For example, the backend QPS of an enterprise is low while the QPS is high when one ISV has multiple merchants.
- The SQL latency is sub-seconds.

- Data is aggregated, computed, and analyzed.
- The QPS is generally not high, about 1 to 100. When real-time analysis is required, the QPS is about 100 to 5,000.
- Response delay ranges from several seconds for real-time analysis to several minutes or hours for offline analysis.

The following examples are business scenarios for your reference:

- Customer-oriented Internet applications to carry out the business for users (PolarDB-X for MySQL).
- Data businesses that feature high concurrency and low latency in the frontend, such as the bank and hospital counter businesses, Internet of Vehicles (IOV) data operations, tracing, and fuel consumption curves (PolarDB-X for MySQL).
- Storage and aggregation analysis of archived data that is unchangeable (including historical data), such as completed orders, logs, and operation and behavior records (PolarDB-X for HiStore).

10.6. Limits

This topic introduces the restrictions of using PolarDB-X.

Item	Limit
Table shard size	We recommend that a table shard contain a maximum of five million records.
Table shard quantity	Theoretically, the number of table shards in each database shard is not restricted, but depends on the hardware of the PolarDB-X server.
Default database shard quantity for a single ApsaraDB RDS for MySQL instance	8, which cannot be changed.
Distributed JOIN	PolarDB-X supports most JOIN semantics, but also has some restrictions on complex JOIN semantics. For example, JOIN operations between large tables may result in performance or system unavailability due to the high cost and slow speed. Therefore, prevent it whenever possible.

10.7. Terms

This topic defines and analyzes the terms related to PolarDB-X.

Term	Description
PolarDB-X	PolarDB-X is developed by Alibaba. It is a distributed relational databases middleware that is highly compatible with the MySQL protocol and syntax.
PolarDB-XServer (PolarDB-X server node)	PolarDB-X Server is a core component of PolarDB-X. It provides the SQL statement parsing, optimization, routing, and result aggregation functions.
PolarDB-X instance	A PolarDB-X instance is a distributed database server cluster that consists of a group of PolarDB-X server nodes. Each server node is stateless and processes SQL requests.
Specifications of PolarDB-X instances	The specifications of PolarDB-X instances reflect the processing capability of PolarDB-X. Each type provides different CPU and memory resources. Instances with higher specifications provide higher processing capabilities. For example, in a standard PolarDB-X test scenario, an instance with an 8-core CPU and 16 GB of memory has twice the capability of an instance with a 4-core CPU and 8 GB of memory.
Instance upgrade and downgrade	PolarDB-X can adjust the processing capability by upgrading or downgrading instance specifications.
Horizontal partitioning (sharding)	The process that splits a single-instance database into multiple physical database shards, partitions and distributes table data from the single-instance into multiple physical table shards according to sharding rules, and then stores the table shards on different database shards.
Sharding rule	A rule used to partition a logical database table into multiple physical table shards during horizontal partitioning.
Shard key	A database field that generates sharding rules during horizontal partitioning.
Database shard	After the horizontal partitioning of PolarDB-X is complete, data in the logical database is stored in multiple physical storage instances. The physical database in each storage instance is a database shard.
Table shard	After the horizontal partitioning of PolarDB-X is complete, a physical data table in each database shard is called a table shard.
Logical SQL statement	The SQL statement sent by an application to PolarDB-X.
Physical SQL statement	The statement sent to ApsaraDB RDS for MySQL for execution after PolarDB-X parses a logical SQL statement.
Transparent read/write splitting	When a single storage node of PolarDB-X encounters an access bottleneck, you can add read-only instances to share the load on the primary instance.PolarDB-X You do not need to modify application code for the read/write splitting function, so it is called transparent read/write splitting.

Term	Description
Non-partition mode	PolarDB-X supports the extension of database service capabilities through transparent read/write splitting without horizontal partitioning. This is called the non-partition mode.
Smooth scale-out	PolarDB-X can scale out the database by adding storage instance nodes. Smooth scale-out does not affect access to original data.
Broadcast of small tables	PolarDB-X stores tables with small data volumes and infrequent updates in single table mode, which are called small tables. The solution which copies a small table to database shards related to it by JOIN statements through data synchronization to improve the JOIN efficiency, is called broadcast or replication of small tables.
Full table scan	In database partition mode, if no shard key is specified in the SQL statement, PolarDB-X executes the SQL statement on all table shards, merges the results and returns them. This process is called full table scan. To prevent impact on performance, we recommend that you do not perform a full table scan.
PolarDB-X sequence	A PolarDB-X sequence (a 64-digit number of the BIGINT data type in MySQL) aims to ensure that the data (for example, PRIMARY KEY and UNIQUE KEY) in the defined unique field is globally unique and in ordered increments.
PolarDB-X hint (PolarDB-X custom annotations)	A custom hint provided by PolarDB-X to specify certain special actions. It uses related syntax to control the SQL execution to optimize SQL statements.

11.AnalyticDB for MySQL

11.1. What is AnalyticDB for MySQL?

AnalyticDB for MySQL is a real-time online analytical processing (RT-OLAP) service that is developed by Alibaba Cloud to analyze large amounts of data at high concurrency. AnalyticDB for MySQL can analyze hundreds of billions of data records across multiple dimensions within milliseconds and provide you with data-driven insights into your business.

> **Note** OLAP systems are often compared with online transaction processing (OLTP) systems. OLAP systems are ideal for scenarios that require complex multidimensional queries and analytics on large amounts of data. The OLAP model is commonly adopted in analytical databases. OLTP systems are suitable for transaction processing, and ensure strong atomicity and consistency in data manipulation. The OLTP model supports frequent INSERT and UPDATE operations and is often used for relational database management systems, such as MySQL and Microsoft SQL Server.

AnalyticDB for MySQL is an RT-OLAP system that offers the following benefits:

- Compatible with MySQL, business intelligence (BI) tools, and extract, transform, and load (ETL) tools for easy, cost-effective, and efficient analysis and integration of data.
- Uses relational models to store data and provides SQL statements to flexibly compute and analyze data. You do not need to create a data model in advance.
- Uses distributed computing technologies to provide excellent real-time computing capabilities. When AnalyticDB for MySQL processes tens of billions of data records or more, its performance can match or even surpass that of multidimensional online analytical processing (MOLAP) systems. AnalyticDB for MySQL can compute tens of billions of data records within several hundred milliseconds. You can then explore large amounts of data without constraints, instead of viewing data reports based on a predefined logic.
- Computes hundreds of billions of data records in real time. AnalyticDB for MySQL uses all data generated in your business system for data analysis, rather than sampling a portion of the data. This maximizes the effectiveness of analysis results.
- Supports a large number of concurrent queries and ensures high system availability through dynamic multi-copy storage and computing technology. Therefore, AnalyticDB for MySQL can serve as a backend system for various products, including user-facing and enterprise-facing products. AnalyticDB for MySQL is used in Internet business systems that have hundreds of thousands to tens of millions of users, such as Data Cube, Taobao Index, Kuaidi Dache, Alimama DMP, and Taobao Groceries.

AnalyticDB for MySQL is a real-time computing system that provides rapid and flexible online data analysis and computation.

11.2. Benefits

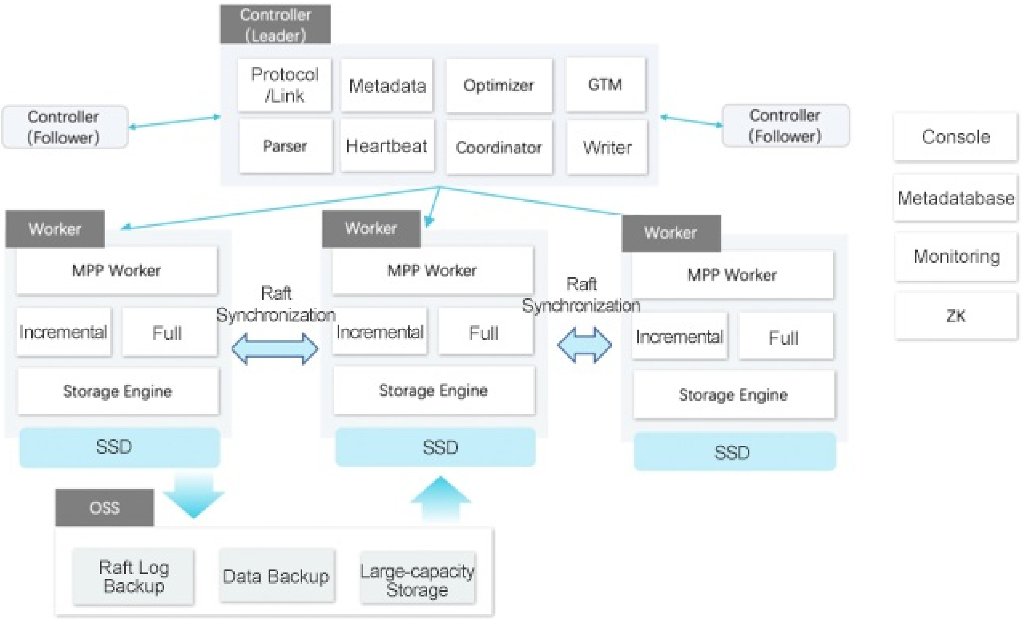
This topic describes the benefits of AnalyticDB for MySQL.

Benefit	Description
Computation for large amounts of data	Computes trillions of data records or petabytes of data in a single table.

Benefit	Description
Full data analysis	Performs analysis on full data, which delivers more representative results compared with sampled data.
Rapid query response	Provides multidimensional pivoting for tens of billions of data entries within milliseconds.
High concurrency and availability	Supports a large number of concurrent queries and ensures high system availability through dynamic multi-copy storage and computing technology. Therefore, AnalyticDB for MySQL can serve as a backend system for various products, including user-facing and enterprise-facing products.
Flexible query method	Provides SQL statements to perform multidimensional analysis, pivoting, and filtering for large amounts of data in a flexible manner.
Parallel data import through multiple channels	Supports both online and offline channels to import data. The import performance improves linearly with the cluster size.
Refined security mechanism	Supports column-based permission management and finer-grained user operation audits. AnalyticDB for MySQL ensures data security by using a public/private key mechanism.
High compatibility	Provides full compatibility with MySQL protocols (including data element information), inherent compatibility with commercial analysis tools and applications, and built-in support for fast data access from various types of data sources. This greatly reduces the cost of access to business systems and software.
Isolation	Supports multi-tenant isolation. Each cluster is provisioned with exclusive resources such as CPU, memory, I/O resources, and disks.
Permissions	Supports centralized management of multiple tenants in the console. For example, you can dynamically configure and manage tenant resources, isolate resources, view statistics for resource usage, and manage tenants at multiple levels.

11.3. Architecture

AnalyticDB for MySQL is a distributed real-time computing system based on the MPP architecture. AnalyticDB for MySQL is built on top of the Apsara system and incorporates distributed retrieval technology. AnalyticDB for MySQL consists of the underlying infrastructure, compute nodes, controllers, and storage nodes.



Underlying infrastructure

The underlying infrastructure consists of the following parts:

- Apsara system: used for logical isolation, persistence data storage, and to construct schemas and indexes.
- Metadatabase: refers to ApsaraDB for RDS or Tablestore that is used to store metadata of AnalyticDB for MySQL.

  Note Metadata is not involved in actual computations.

- Apache ZooKeeper module: performs distributed coordination among components.

Controllers

A controller is used to control the allocation of database resources in compute nodes and the distribution of computing resources. The controller can also manage compute nodes and tasks that run in the database background. A controller consists of multiple modules:

- Alibaba Cloud Server Load Balancer (SLB): groups controllers and implements load balancing among controllers.
- Client access manager.
- SQL parser.
- AnalyticDB for MySQL console.

AnalyticDB for MySQL supports the following clients, drivers, programming languages, and middleware:

- Clients that support MySQL 5.1, 5.5, or 5.6 protocols and drivers: MySQL 5.1.x Connector/J, MySQL 5.3.x Connector/ODBC, and MySQL 5.1.x, 5.5.x, or 5.6.x client.
- Programming languages: JAVA, Python, C/C++, Node.js, PHP, and R (RMySQL).

- Middleware: Websphere Application Server 8.5, Apache Tomcat, and JBoss.

Compute nodes

Compute nodes run computing tasks that are issued by controllers to read, filter, merge, and compute data.

Storage nodes

Storage nodes write data, save data to disk storage, and replicate data between nodes. Storage nodes support data backup and restoration.

11.4. Features

11.4.1. DDL

AnalyticDB for MySQL provides DDL statements to manage databases.

- Allows you to view all databases on which you have permissions by using SHOW DATABASES.
- Allows you to create tables and modify table attributes.
- Allows you to add columns to a table.
- Allows you to modify indexes.
- Allows you to create and delete views.

> **Note** AnalyticDB for MySQL allows you to create and delete database clusters only in the ASCM console. DDL statements are not supported for these operations.

11.4.2. DML

11.4.2.1. SELECT

AnalyticDB for MySQL is more than 95% compatible with standard MySQL queries. You can use SELECT statements to query data.

- Supports column mapping methods such as expressions, functions, aliases, column names, and CASE WHEN.
- Supports clauses such as FROM <table name> AS <alias> and JOIN <table name> AS <alias>.

  Supports joins between tables, including LEFT JOIN, RIGHT JOIN, FULL JOIN, and OUTER JOIN.
- Supports WHERE clauses combined with AND and OR operators, function expressions, or BETWEEN and IS operators.
- Supports GROUP BY operations for multiple columns and aliases generated from column mapping expressions such as CASE WHEN. Common aggregate functions are supported.
- Supports ORDER BY operations for expressions and columns in either ascending or descending order.

  Supports HAVING operations.
- Supports subqueries.
- Supports COUNT(DISTINCT) operations.
- Supports constant columns.

- Supports operators such as UNION, UNION ALL, MINUS, and INTERSECT.

11.4.2.2. INSERT, DELETE, and UPDATE

AnalyticDB for MySQL allows you to use INSERT, DELETE, and UPDATE statements to update data in databases.

- Allows you to perform INSERT, DELETE, and UPDATE operations on real-time tables that have primary keys defined.
- Provides multiple mechanisms to prevent written data from being lost. Both REPLACE INTO/INSERT OVERWRITE and INSERT IGNORE INTO statements are supported.
  - `REPLACE INTO` can be used to overwrite data in a table. The statement checks whether the data that you want to write exists in the table based on the primary key. If yes, the statement deletes the row and inserts new data. Otherwise, the statement inserts the new data into the table.
  - `INSERT INTO` can be used to insert data into a table. If an entry has the same primary key as an existing entry, the entry is not inserted. This statement is equivalent to `INSERT IGNORE INTO`.
- Provides the `INSERT INTO...SELECT FROM` statement.

11.4.3. Resource management

AnalyticDB for MySQL uses elastic compute units (ECUs) to manage resources. AnalyticDB for MySQL provides distributed resource scheduling by using the Apsara system and the underlying technology.

AnalyticDB for MySQL provides each database cluster with independent controllers, compute nodes, and storage nodes. You can specify ECU specifications to control the resource usage of controllers, compute nodes, and storage nodes. The ECU specifications determine the number of CPU cores, dedicated memory size, and disk size of an ECU.

11.4.4. Permissions and authorization

AnalyticDB for MySQL supports the standard MySQL permission model.

- Allows you to configure access control lists (ACLs) to control access to databases, tables, and columns.
- Allows privileged accounts to grant permissions to standard accounts.
- Allows you to configure IP address whitelists to allow access from authorized clients.
- Provides role-based permissions.
- Provides the ADD USER statement to add users and the REMOVE USER statement to remove users.

  Provides the GRANT statement to grant permissions and the REVOKE statement to revoke permissions.
- Provides the SHOW GRANTS ON statement to view user permissions on each object.
- Provides the LIST USERS statement to view all authorized users.
- Privileged account: the account that is created in the console after a cluster is created. A privileged account has permissions to create and grant permissions to standard accounts.
- Standard account: an account that can perform data definition language (DDL) and data manipulation language (DML) operations on databases after the account is authorized.

11.4.5. Metadata

This topic describes the metadata stored in AnalyticDB for MySQL.

- TABLES: stores the basic information of all tables that belong to a database cluster, including tables created by users and system metadata tables.
- USER_PRIVILEGES: stores authorization information about users that belong to a database cluster.
- COLUMNS: stores the definition of each field for all tables that belong to a database cluster.

11.4.6. Data import and export

This topic describes the data import and export methods supported by AnalyticDB for MySQL.

AnalyticDB for MySQL supports the following data import methods:

- Allows you to write data by using synchronization tools, such as Kettle.
- Allows you to import data from CSV and text files, including files that use multiple delimiters, by using the LOAD DATA statement.
- Allows you to import data from Object Storage Service (OSS) or MySQL by using external tables.

AnalyticDB for MySQL provides the following methods to export data in parallel:

- Allows you to query tables by using the SELECT statement.
- Allows you to export data to OSS or MySQL by using external tables.

11.5. Scenarios

11.5.1. Overview

This topic describes the typical industries where AnalyticDB for MySQL is applicable.

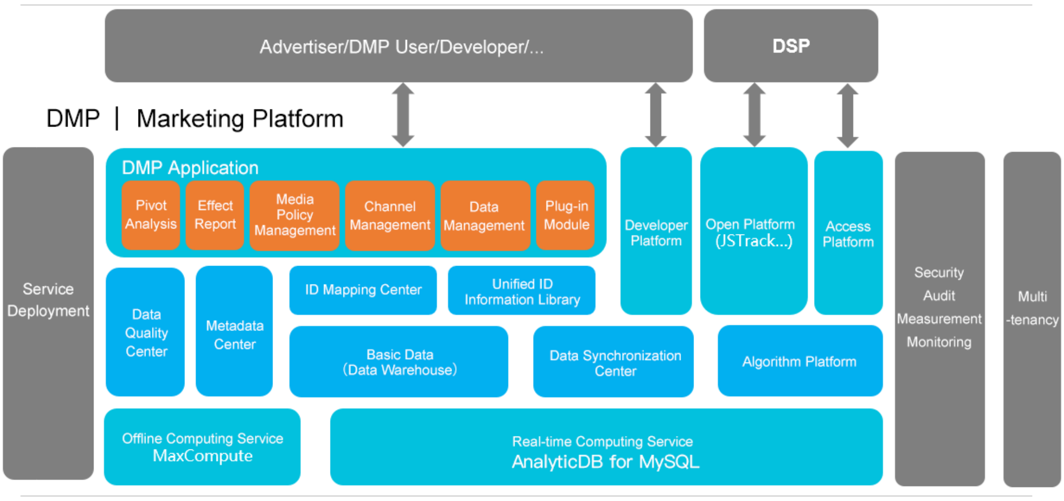
The following table lists the common scenarios of AnalyticDB for MySQL.

Scenario	Description
E-commerce industry	Analytical customer relationship management (A-CRM), popular product selection, automated operations, and stock keeping unit (SKU) combination analysis
O2O	Data analysis, customer relationship management (CRM) system, and geo-fencing system
Advertising industry	Digital marketing and Mobile-Digital Media Player (M-DMP) system
Financial industry	Real-time multidimensional data analysis, transaction flow query system, and reporting system
Security	Crowd analysis, potential key element mining, relational network analysis, and detail query
Traffic management	Checkpoint-related vehicle data analysis and determination

Scenario	Description
Logistics and IoT	Internet of Vehicles (IoV) data analysis, enterprise security data analysis, sensor data storage and retrieval, and real-time logistics data warehousing

11.5.2. Alimama DMP

The following figure shows the application architecture of Alimama Data Management Platform (DMP).

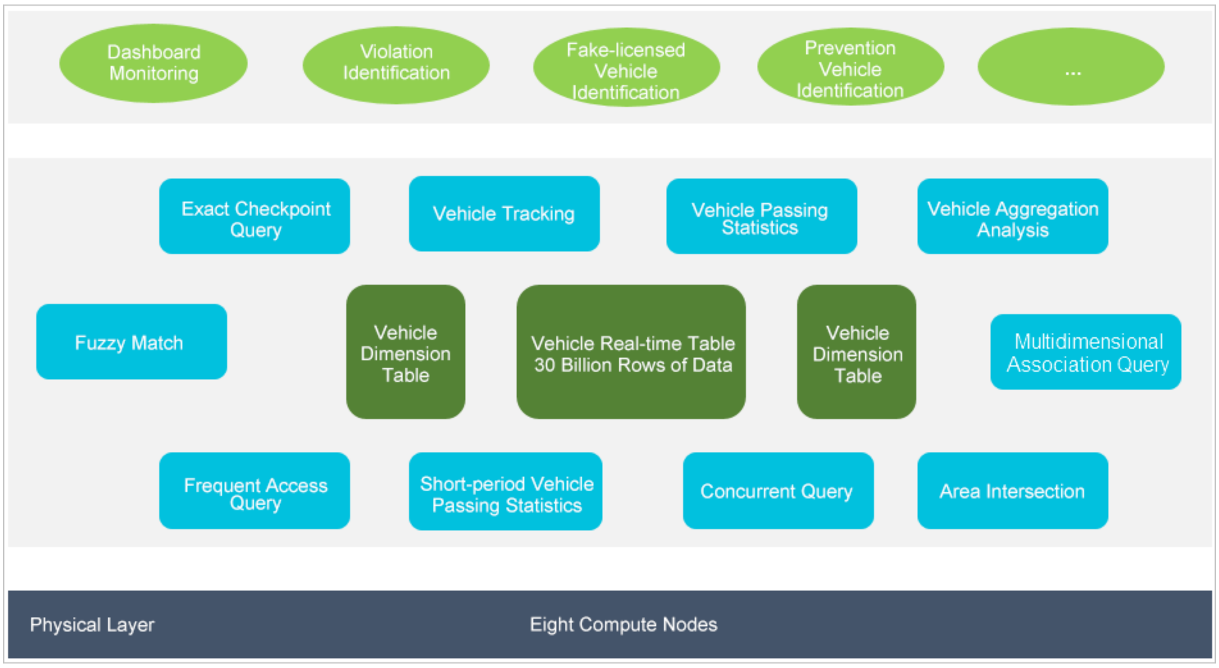


Big data services play core roles in the Alimama DMP system:

- MaxCompute performs user data scrubbing and tag mining.
- AnalyticDB for MySQL provides advertisers with a way to analyze big data and compute crowd management data. AnalyticDB for MySQL can export large amounts of specified user group data to the KVStore system, which supports faster query.
- The targeting engine serves the demand-side platform (DSP) based on the KVStore data.

11.5.3. Traffic police application

The following figure shows an example of traffic police application.


Dashboard Monitoring
Violation Identification
Fake-licensed Vehicle Identification
Prevention Vehicle Identification
...
Exact Checkpoint Query
Vehicle Tracking
Vehicle Passing Statistics
Vehicle Aggregation Analysis
Fuzzy Match
Vehicle Dimension Table
Vehicle Real-time Table 30 Billion Rows of Data
Vehicle Dimension Table
Multidimensional Association Query
Frequent Access Query
Short-period Vehicle Passing Statistics
Concurrent Query
Area Intersection
Physical Layer
Eight Compute Nodes

The traffic police business system has the following features:

- Large amounts of data: Thirty to fifty billion tables record the vehicles that pass through various checkpoints around the city. These tables can be stored for up to six months and consume up to 30 TBs of data.
- Rapid increase: Ten million rows of data are added every day in a municipal system.
- Complex query: Multiple departments need to query data by using diversified methods. Business application query applies to multiple scenarios, such as single table query, multi-table query (JOIN), fuzzy search (LIKE), trace analysis (IN), area intersection (INTERSECT), vehicle quantity query in a short period of time (HAVING COUNT), and multi-user query. Complex queries have high requirements for table schema design, memory usage, CPU utilization, and query concurrency.

11.6. Limits

This topic describes the naming conventions and limits for objects in AnalyticDB for MySQL.

Object	Naming convention	Limit
Database name	A database name can be up to 64 characters in length and can contain letters, digits, and underscores (_). It must start with a lowercase letter and cannot contain consecutive underscores (_).	Do not use analyticdb as the database name. The name analyticdb is reserved for a built-in database.

Object	Naming convention	Limit
Table name	A table name must be 1 to 127 characters in length and can contain letters, digits, and underscores (_). It must start with a letter or underscore (_).	- A table name cannot contain single quotation marks ('), double quotation marks (''), exclamation points (!), or spaces. - A table name cannot be SQL reserved keywords.
Column name	A column name must be 1 to 127 characters in length and can contain letters, digits, and underscores (_). It must start with a letter or underscore (_).	- A column name cannot contain single quotation marks ('), double quotation marks (''), or spaces. - A column name cannot be SQL reserved keywords.
Account name	An account name must be 2 to 16 characters in length, and can contain lowercase letters, digits, and underscores (_). It must start with a lowercase letter and end with a lowercase letter or digit.	None
Password	A password must be 8 to 32 characters in length and must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. Special characters include ! @ # $ % ^ & * ( ) _ + - =	None

11.7. Terms

This topic describes the terms used in AnalyticDB for MySQL.

database cluster

A warehouse used to organize, store, and manage data. Database clusters are the basic unit used to isolate tenants. Each database cluster has independent computing resources, user permissions, and user quotas.

database account

An account used in AnalyticDB for MySQL. It can be a privileged or standard account. You can create a privileged account after you create an AnalyticDB for MySQL cluster as an administrator. You can create a standard account by using SQL statements after you log on to an AnalyticDB for MySQL cluster as an administrator. A privileged account can grant permissions to different departments. User operations can be audited in fine granularity.

table

AnalyticDB for MySQL supports standard relational table models.

column

Table data in AnalyticDB for MySQL is stored in columns. A column has the following features:

- Supports standard MySQL data types, such as BOOLEAN, TINYINT, SMALLINT, INT, BIGINT, FLOAT, DOUBLE, VARCHAR, DATE, and TIMESTAMP.
- Supports automatic creation and manual deletion of full table indexes.

index

AnalyticDB for MySQL automatically creates indexes for all columns. If a column does not need an index, you can execute the DISABLE INDEX statement on the column to delete its index.

primary key

AnalyticDB for MySQL allows you to specify a primary key for a table. When you execute an INSERT, UPDATE, or DELETE statement, AnalyticDB for MySQL can use the primary key to identify unique entries.

> **Note** A primary key in AnalyticDB for MySQL is only used to identify unique entries. You cannot modify the primary key. If you want to modify the primary key, you must create a table.

12.AnalyticDB for PostgreSQL

12.1. What is AnalyticDB for PostgreSQL?

AnalyticDB for PostgreSQL is a distributed analytic database service that leverages the massively parallel processing (MPP) architecture, where each instance is composed of multiple compute nodes. AnalyticDB for PostgreSQL provides MPP warehousing services that support horizontal scaling of storage and compute capabilities, online analysis of petabytes of data, and offline processing of Extract, Transform, and Load (ETL) tasks.

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

- Supports the SQL:2003 standard, OLAP aggregate functions, views, Procedural Language for SQL (PL/SQL), user-defined functions (UDFs), and triggers. AnalyticDB for PostgreSQL is partially compatible with the Oracle syntax.
- Supports horizontal scaling of storage and compute capabilities based on the MPP architecture. AnalyticDB for PostgreSQL also supports range and list partitioning.
- Supports row store, column store, and multiple indexes. AnalyticDB for PostgreSQL also supports multiple compression methods based on column store to reduce storage costs.
- Supports standard database isolation levels and distributed transactions to ensure data consistency.
- Provides the vector computing engine and the CASCADE-based SQL query optimizer to ensure high-performance SQL analysis.
- Uses a primary/secondary architecture to ensure dual-copy data storage and service availability when O&M and management nodes fail.
- Provides online scaling, system monitoring, and disaster recovery to reduce O&M costs.
- Allows you to store audit logs in Log Service and download logs from Log Service. This facilitates long-term storage and management of audit logs.
- Provides automatic fault tolerance for hard disk failures of servers in a cluster, and supports hot swapping of hard disks. In the event of a hard disk failure, services can be restored within two minutes.
- Provides permission management and fine-grained audit of user operations.
- Provides a comprehensive permission authentication and isolation mechanism to ensure your data privacy.
- Supports multi-tenancy for resource isolation.
- Supports multi-tenant parallel execution on a cluster by using multiple instances. Tasks from tenants are submitted to queues on different instances for execution. Resources of each AnalyticDB for PostgreSQL instance are isolated among tenants.
- Allows you to configure and manage tenants in a dynamic and centralized manner. You can also isolate resources and query usage statistics of resources. Management of multi-level tenants is supported.
- Supports multi-tenant scheduling of multiple clusters and resource pools.
- Metadatabases support fast switchover for disaster recovery. In the event of a failure, services can be restored within one minute.

AnalyticDB for PostgreSQL supports mainstream connection pools such as PgBouncer and pgpool-II.

- PgBouncer: a lightweight connection pool for PostgreSQL.

- pgpool-II: a connection pool that provides abundant features such as connection pooling, load balancing, and query caching. pgpool-II also supports automatic retrying of tasks when network exceptions and lock conflicts occur.

12.2. Benefits

This topic describes the benefits of AnalyticDB for PostgreSQL.

- Real-time analysis

  Built on the MPP architecture that supports horizontal scaling and can respond to queries on petabytes of data within seconds. AnalyticDB for PostgreSQL supports the leading vector computing feature and intelligent indexes of column store. It also supports the CASCADE-based SQL query optimizer to enable complex queries without the need for tuning.

- Stability and reliability

  Provides ACID properties for distributed transactions. Transactions are consistent across nodes and all data is synchronized between primary and secondary nodes. AnalyticDB for PostgreSQL supports distributed deployment and provides transparent monitoring, switching, and restoration to secure your data infrastructure.

- Easy to use

  Supports rich SQL syntax and functions, Oracle functions, stored procedures, user-defined functions (UDFs), and isolation levels of transactions and databases. You can use popular business intelligence (BI) software and ETL tools online.

- Ultra-high performance

  Supports row store, column store, and multiple indexes. The vector engine provides high-performance analysis and computing capabilities. The CASCADE-based SQL optimizer enables complex queries without the need for tuning. AnalyticDB for PostgreSQL supports high-performance parallel import of data from OSS.

- Flexible scalability

  Enables you to scale out compute nodes as well as CPU, memory, and storage resources on demand to improve OLAP performance.

  Supports transparent OSS operations. OSS offers a larger storage capacity for cold data that does not require online analysis.

  Supports online scaling to add, remove, modify, and query data during data redistribution.

- Resource isolation

  Supports multi-tenant parallel execution on a cluster by using multiple instances. Tasks from tenants are submitted to queues on different instances for execution. Resources of each AnalyticDB for PostgreSQL instance are isolated among tenants.

- Permission management

  Allows you to configure and manage tenants in a dynamic and centralized manner. You can also isolate resources and query usage statistics of resources. Management of multi-level tenants is supported.

- Resource scheduling

  Supports multi-tenant scheduling of multiple clusters and resource pools.

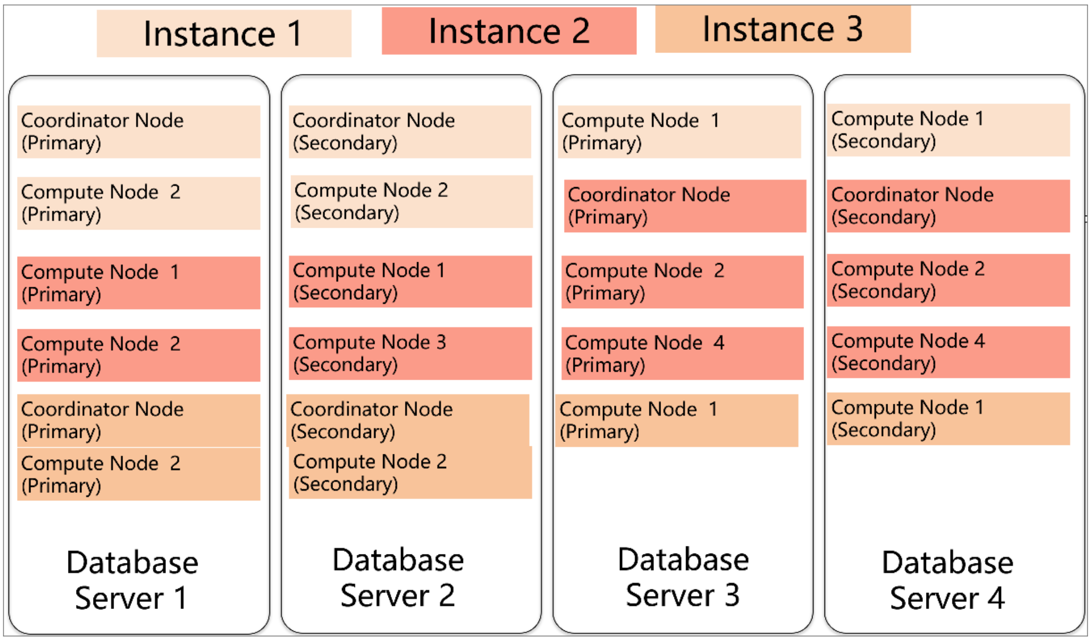
12.3. Architecture

This topic describes the architecture of AnalyticDB for PostgreSQL.

Physical architecture of a cluster

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

Physical cluster architecture


Instance 1
Instance 2
Instance 3
Coordinator Node (Primary)
Compute Node 2 (Primary)
Compute Node 1 (Primary)
Compute Node 2 (Primary)
Coordinator Node (Primary)
Compute Node 2 (Primary)
Database Server 1
Coordinator Node (Secondary)
Compute Node 2 (Secondary)
Compute Node 1 (Secondary)
Compute Node 3 (Secondary)
Coordinator Node (Secondary)
Compute Node 2 (Secondary)
Database Server 2
Compute Node 1 (Primary)
Coordinator Node (Primary)
Compute Node 2 (Primary)
Compute Node 4 (Primary)
Compute Node 1 (Primary)
Database Server 3
Compute Node 1 (Secondary)
Coordinator Node (Secondary)
Compute Node 2 (Secondary)
Compute Node 4 (Secondary)
Compute Node 1 (Secondary)
Database Server 4

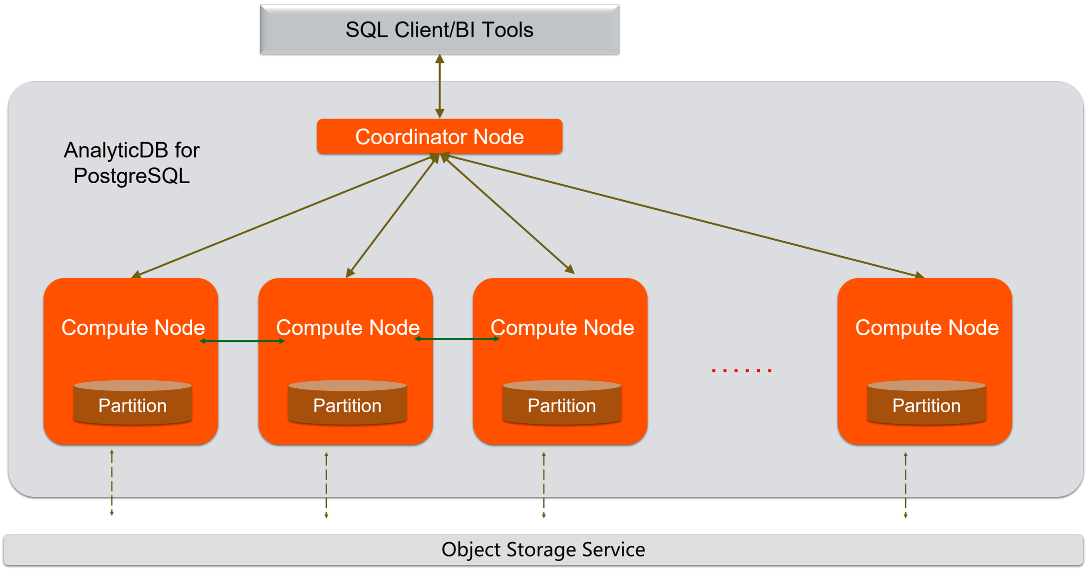
You can create multiple instances in a physical cluster of AnalyticDB for PostgreSQL by using the management and control system. Each instance consists of a coordinator node and multiple compute nodes.

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

Logical architecture of an instance

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

Logical architecture of an instance



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

12.4. Specifications

This topic describes the specifications of AnalyticDB for PostgreSQL.

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

- SSD storage: provides excellent I/O capabilities and high analysis performance.
- HDD storage: provides large storage capacity at a low cost.

Specifications

The following table lists the supported specifications of compute nodes in AnalyticDB for PostgreSQL.

Storage type	Number of cores per node	Memory	Available storage space	Total dual-copy storage space	Description
High-performance SSD	1	8GB	80GB	160GB	These specifications are recommended for low-concurrency scenarios that require less than 5 concurrent queries and less than 32 nodes. These specifications are available for 4 to 32 nodes per instance.

Storage type	Number of cores per node	Memory	Available storage space	Total dual-copy storage space	Description
High-performance SSD	4	32GB	320GB	640GB	These specifications are recommended for high-performance SSD storage and available for 8 to 2,048 nodes per instance.
High-capacity HDD	2	16GB	1TB	2TB	These specifications are recommended for low-concurrency scenarios that require less than 5 concurrent queries and less than 8 nodes. These specifications are available for 4 to 32 nodes per instance.
High-capacity HDD	4	32GB	2TB	4TB	These specifications are recommended for high-capacity HDD storage and available for 8 to 2,048 nodes per instance.

12.5. Features

This topic describes the features of AnalyticDB for PostgreSQL.

Distributed architecture

AnalyticDB for PostgreSQL is built on the massively parallel processing (MPP) architecture. Data is distributed evenly across nodes by hash value or RANDOM function, and is analyzed and computed in parallel. Storage and compute capabilities are scaled out by adding nodes to ensure a quick response as the data volume increases.

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

High-performance data analysis

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

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

High-availability service

AnalyticDB for PostgreSQL builds a system for automatic monitoring, diagnosis, and troubleshooting based on the Apsara system. This helps reduce O&M costs.

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

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

Data synchronization methods and tools

You can use Data Transmission Service (DTS) or DataWorks Data Integration to synchronize data from MySQL or PostgreSQL databases to AnalyticDB for PostgreSQL. You can use popular Extract, Transform, and Load (ETL) tools to import ETL data to and schedule jobs in AnalyticDB for PostgreSQL databases. You can also use standard SQL syntax to query data from formatted files stored in OSS by using foreign tables in real time.

AnalyticDB for PostgreSQL supports popular business intelligence (BI) reporting tools such as Quick BI, DataV, Tableau, and FineReport. It also supports ETL tools, including Informatica and Kettle.

Data security

AnalyticDB for PostgreSQL supports the configuration of whitelists. You can add up to 1,000 IP addresses of servers to a whitelist to allow access to your instance and control risks from access sources. AnalyticDB for PostgreSQL also supports Anti-DDoS to monitor inbound traffic in real time. When large amounts of malicious traffic is identified, the traffic is scrubbed by means of IP filtering. If traffic scrubbing is insufficient, blackhole filtering is triggered.

The pgcrypto extension allows you to encrypt columns or tables by using cryptography functions that use encryption algorithms. Algorithms include Message-Digest Algorithm 5 (MD5), Secure Hash Algorithm 1 (SHA-1), SHA-224, SHA-256, SHA-384, SHA-512, Blowfish, Advanced Encryption Standard 128 (AES-128), AES-256, Raw Encryption, Pretty Good Privacy (PGP) symmetric keys, and PGP public keys.

Supported SQL features

- Supports row store and column store.
- Supports multiple indexes, including B-tree indexes, bitmap indexes, and hash indexes.
- Supports distributed transactions and standard isolation levels to ensure data consistency among nodes.
- Supports character, date, and arithmetic functions.
- Supports stored procedures, user-defined functions (UDFs), and triggers.
- Supports views.
- Supports range partitioning, list partitioning, and the definition of multi-level partitions.
- Supports partition tables and diverse partition-related operations such as add, drop, rename, clear, swap, and split.
- Supports multiple data types. The following table provides a list of data types and their information.

Data type	Alias	Storage size	Range	Description
bigint	int8	8 bytes	-9223372036854775808 to 9223372036854775807	An integer within a large range.
bigserial	serial8	8 bytes	1 to 9223372036854775807	A large auto-increment integer.
bit [(n)]	None	n bits	A bit string constant	A bit string with a fixed length.
bit varying [(n)]	varbit	A bit string with a variable length.	A bit string constant	A bit string with a variable length.
boolean	bool	1 byte	true/false, t/f, yes/no, y/n, 1/0	A boolean value (true or false).
box	None	32 bytes	((x1,y1),(x2,y2))	A rectangular box on a plane, which is not allowed in a column that is used as the distribution key.
bytea	None	1 byte + binary string	Sequence of octets	A binary string with a variable length.
character [(n)]	char [(n)]	1 byte + n	A string up to n characters in length	A blank-padded string with a fixed length.
character varying [(n)]	varchar [(n)]	1 byte + string size	A string up to n characters in length	A string with a limited variable length.
cidr	None	12 or 24 bytes	None	IPv4 and IPv6 networks.
circle	None	24 bytes	<(x,y),r> (center and radius)	A circle on a plane, which is not allowed in distribution key columns.
date	None	4 bytes	4713 BC - 294,277 AD	Calendar date (year, month, day).
decimal [(p, s)]	numeric [(p, s)]	variable	No limits	User-specified precision, which is exact.
double precision	float8	8 bytes	Precise to 15 decimal digits	Variable precision, which is inexact.
	float			
inet	None	12 or 24 bytes	None	IPv4 and IPv6 hosts and networks.
integer	int, int4	4 bytes	-2.1E+09 to +2147483647	An integer in typical cases.

Data type	Alias	Storage size	Range	Description
interval [(p)]	None	12 bytes	-178000000 years - 178000000 years	A time range.
json	None	1 byte + json size	JSON string	A string with an unlimited variable length.
lseg	None	32 bytes	((x1,y1),(x2,y2))	A line segment on a plane, which is not allowed in distribution key columns.
macaddr	None	6 bytes	None	A Media Access Control (MAC) address.
money	None	8 bytes	-92233720368547758.08 to +92233720368547758.07	Currency amount.
path	None	16+16n bytes	[(x1,y1),...]	A geometric path on a plane, which is not allowed in distribution key columns.
point	None	16 bytes	(x,y)	A geometric point on a plane, which is not allowed in distribution key columns.
polygon	None	40+16n bytes	((x1,y1),...)	A closed geometric path on a plane, which is not allowed in a column that is used as the distribution key.
real	float4	4 bytes	Precise to 6 decimal digits	Variable precision, which is inexact.
serial	serial4	4 bytes	1 to 2147483647	An auto-increment integer.
smallint	int2	2 bytes	-32768 to 32767	An integer within a small range.
text	None	1 byte + string size	A string with a variable length	A string with an unlimited variable length.
time [(p)] [without time zone]	None	8 bytes	00:00:00[.000000] - 24:00:00[.000000]	The time of a day without the time zone.
time [(p)] with time zone	timetz	12 bytes	00:00:00+1359 - 24:00:00-1359	The time of a day with the time zone.

Data type	Alias	Storage size	Range	Description
timestamp [(p)] [without time zone]	None	8 bytes	4713 BC - 294,277 AD	The date and time without the time zone.
timestamp [(p)] with time zone	timestamptz	8 bytes	4713 BC - 294,277 AD	The date and time with the time zone.
xml	None	1 byte + xml size	Variable-length XML string	A string with an unlimited variable length.

12.6. Scenarios

This topic describes the common scenarios in which AnalyticDB for PostgreSQL can be used.

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

- Extract, Transform, and Load (ETL) for offline data processing

  AnalyticDB for PostgreSQL provides the following benefits that make it ideal to optimize complex SQL queries as well as aggregate and analyze large amounts of data:

  - Supports standard SQL syntax, OLAP window functions, and stored procedures.
  - Provides the CASCADE-based SQL query optimizer to enable complex queries without the need for tuning.
  - Built on the MPP architecture that supports horizontal scaling of storage and compute capabilities to analyze and process petabytes of data.
  - Provides column store-based high-performance aggregation of large tables at a high compression ratio to maximize storage capacity.

- Online high-performance query

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

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

- Multi-model data analysis

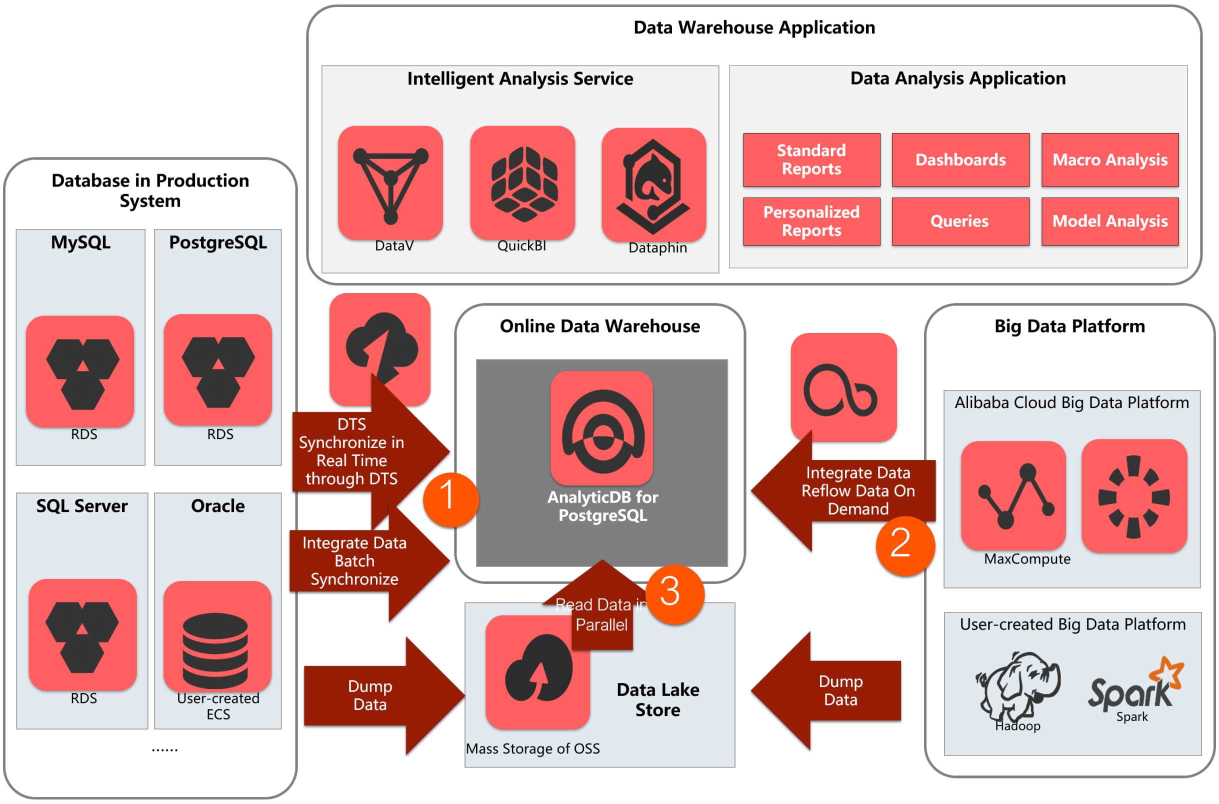
  AnalyticDB for PostgreSQL provides the following benefits for processing unstructured data from a variety of sources:

  - Supports the PostGIS extension for geographic data analysis and processing.

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

Typical scenarios

AnalyticDB for PostgreSQL is applicable to the following scenarios:


Data Warehouse Application
Intelligent Analysis Service
DataV
QuickBI
Dataphin
Data Analysis Application
Standard Reports
Dashboards
Macro Analysis
Personalized Reports
Queries
Model Analysis
Database in Production System
MySQL
RDS
PostgreSQL
RDS
SQL Server
RDS
Oracle
User-created ECS
......
DTS Synchronize in Real Time through DTS
1
Integrate Data Batch Synchronize
Dump Data
Online Data Warehouse
AnalyticDB for PostgreSQL
Read Data in Parallel
3
Data Lake Store
Mass Storage of OSS
Integrate Data Reflow Data On Demand
2
Dump Data
Big Data Platform
Alibaba Cloud Big Data Platform
MaxCompute
User-created Big Data Platform
Hadoop
Spark

- Data warehousing service

  Data Transmission Service (DTS) can synchronize data in real time in production system databases such as ApsaraDB RDS for MySQL, ApsaraDB RDS for PostgreSQL, and PolarDB as well as traditional databases such as Oracle and SQL Server. Data can also be batch synchronized to AnalyticDB for PostgreSQL by using Data Integration. AnalyticDB for PostgreSQL supports ETL operations on large amounts of data. You can also use DataWorks to schedule these tasks. AnalyticDB for PostgreSQL also provides high-performance online analysis capabilities and can use Quick BI, DataV, Tableau, and FineReport for report presentation and real-time query.

- Big data analytics platform

  You can use Data Integration or OSS to import large amounts of data from MaxCompute, Hadoop, and Spark to AnalyticDB for PostgreSQL for high-performance analysis, processing, and exploration.

- Data lake analytics

  AnalyticDB for PostgreSQL can use foreign tables to access the large amounts of data stored in OSS in parallel and build an Alibaba Cloud data lake analytics platform.

12.7. Introduction to AnalyticDB for PostgreSQL V6.0

AnalyticDB for PostgreSQL V6.0 is an online database service developed based on the kernel of open source Greenplum Database 6.0. AnalyticDB for PostgreSQL V6.0 improves the capabilities to process concurrent transactions for real-time data warehouses.

Kernel upgrade

The Greenplum kernel is updated from 4.3 to 6.0, and the PostgreSQL kernel is updated from 8.2 to 9.4. AnalyticDB for PostgreSQL V6.0 has the following new features:

- JSONB: supports the JSON and JSONB storage types to provide more JSON processing functions and higher processing efficiency.
- UUID: supports the UUID data type.
- GIN and SP-GiST indexes: allow high-performance fuzzy match and retrieval.
- Fine-grained permission control: supports schema-level and column-level permission control and authorization.
- Efficient VACUUM statements: When you execute VACUUM statements to release space, locked pages are skipped and vacuumed at a later time to reduce blocking.
- DBLINK: supports cross-database queries.
- Recursive common table expression (CTE): processes hierarchical or tree-structured data to facilitate multi-level recursive queries.
- PL/SQL enhancement:
    - Supports the RETURN QUERY EXECUTE statement to dynamically execute SQL statements.
    - Supports anonymous blocks.

HTAP improvement

AnalyticDB for PostgreSQL V6.0 introduces the global deadlock detection mechanism to check and unlock global deadlocks by dynamically collecting and analyzing lock information. Updates and modifications to heap tables can only be completed by using fine-grained row locks. AnalyticDB for PostgreSQL V6.0 supports concurrent update, delete, and query operations to improve the concurrency and throughput of the system. AnalyticDB for PostgreSQL V6.0 optimizes transaction locks to reduce lock competition at the beginning and end of transactions. AnalyticDB for PostgreSQL V6.0 provides high-performance OLAP analysis and high-throughput transaction processing features.

OLAP features

- Replicated tables are supported. For dimension tables in data warehouses, replicated tables can be created by using DISTRIBUTED REPLICATED clauses. This reduces data transmission and improves query efficiency.
- The Zstandard compression algorithm is supported. This algorithm offers compression and decompression performance three times better than the zlib algorithm.

13.KVStore for Redis

13.1. What is KVStore for Redis?

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

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

13.2. Benefits

High performance

- Supports cluster instances with the memory capacity of 128 GB or larger. These instances can meet large capacity and high performance requirements.
- Supports master-replica instances with a maximum memory capacity of 32 GB. These instances can meet common capacity and performance requirements.
- Supports CPUs, hard disks, memory, and network interface controllers (NICs) of different specifications in a cluster without affecting the running performance of the cluster. This ensures maximum compatibility with your existing devices.

Elastic scaling

- Easy scaling: You can expand the instance storage capacity with only a few clicks in the console based on your business requirements.
- Online scaling: You can expand the instance storage capacity without service interruption.

Resource isolation

- Supports instance-level resource isolation among different instances. This ensures stability of individual services.
- Supports multi-tenant isolation to ensure that each instance can use exclusive resources, such as CPU, memory, I/O resources, and disks.
- Supports multi-tenant parallel execution on a cluster by using multiple instances. Tasks from tenants are submitted to queues on different instances for execution. KVStore for Redis isolates resources among tenants based on different instances.

Data security

- Data persistence: Based on the hybrid storage of memory and disks, KVStore for Redis provides high-speed data read/write capability and enables data persistence. KVStore for Redis also allows you to load data from a persistent database into a cache database.
- Master-replica backup and failovers: KVStore for Redis backs up data on both a master node and replica nodes. It also supports the failover feature to prevent data loss.
- Access control: KVStore for Redis supports password authentication to ensure secure and reliable

access to databases.

- Data transmission encryption: KVStore for Redis supports encryption based on Secure Sockets Layer (SSL) and Secure Transport Layer (TLS) to secure data transmission.

High availability

- Master-replica architecture: Each instance runs in this architecture to eliminate the risk of single points of failure (SPOFs) and ensure high availability.
- Automatic failure detection and recovery: The system automatically detects hardware failures and performs a failover within a few seconds after a failure occurs. This minimizes the adverse impact caused by unexpected hardware failures.
- Supports automatic fault tolerance for server hard disk failures in a cluster, and supports hot swapping of hard disks. In case of a hard disk failure, services can be recovered within two minutes.

Easy-to-use

- Out-of-the-box service: KVStore for Redis requires no setup or installation. After you purchase the service, you can immediately use it to ensure efficient deployment of your workloads.
- Compatible with open source Redis: KVStore for Redis is compatible with Redis commands. You can use all Redis clients to connect to KVStore for Redis and manage databases.
- Support for multiple commands in each query.

Permission management

- Supports data access permissions management, such as the logon permission, table creation permission, read and write permission, and whitelist control permission.
- Allows you to log on to the KVStore for Redis console to manage permissions on access control, including administrative rights settings.
- KVStore for Redis provides a unified permission management feature. This feature allows you to manage various permissions for each component of the system in the KVStore for Redis console. This isolates common users from internal permission management details, simplifies the permission management for administrators, and improves the user experience of permission management.
- Allows you to manage multiple tenants in a centralized manner in the console. For example, you can dynamically configure and manage tenant resources, isolate resources, view statistics on resource usage, and manage tenants at multiple levels.

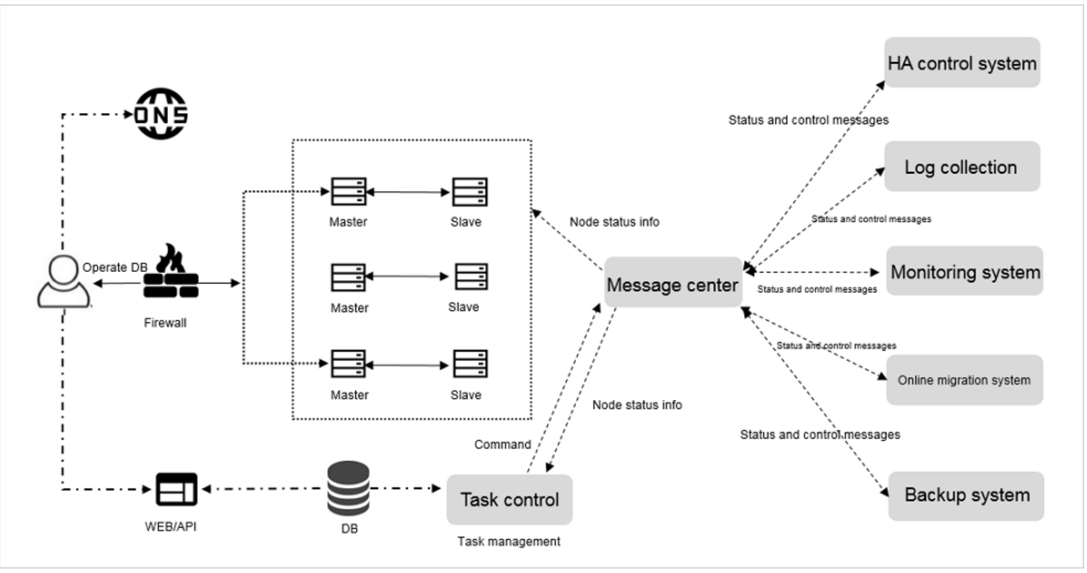
Scheduling

- Supports multi-cluster scheduling, multi-resource pool scheduling, and multi-tenant scheduling.

13.3. Architecture

The architecture of KVStore for Redis is as shown in Architecture diagram.

Architecture diagram



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

- **HA control system**

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

- **Log collection**

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

- **Monitoring system**

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

- **Online migration system**

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

- **Backup system**

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

- **Task Control**

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

13.4. Features

- High-availability technology ensures service stability

The system synchronizes data between the master node and replica node in real time. If the master node fails, the system automatically fails over to the replica node and restores services within a few seconds to ensure high availability of services. The replica node takes over the role of the master node. During the failover process, you can manage your workloads without service interruption.

A cluster instance runs in a distributed architecture. Each node of the instance uses a master-replica high-availability structure to automatically perform failovers and disaster recovery. This mechanism ensures high availability of the KVStore for Redis service.

- Easy backup and recovery support custom backup policies

  You can manually back up data at any time in the console. You can also customize automatic backup policies. The system retains backups for up to seven days. You can use the backups to restore data with only a few clicks to minimize data loss caused by user errors.

- Multiple network security protections secure your data

  A Virtual Private Cloud (VPC) network isolates network transmission at the transport layer. The anti-DDoS services monitor and mitigate DDoS attacks. KVStore for Redis supports up to 1,000 IP addresses or Classless Inter-Domain Routing (CIDR) blocks in each whitelist to prevent malicious logon attempts.

- Kernel optimization avoids vulnerability exploits

  Alibaba Cloud has performed in-depth engine optimization for the Redis source code to prevent running out of memory, fix security vulnerabilities, and protect your business.

- Elastic scaling eliminates capacity and performance bottlenecks

  KVStore for Redis supports multiple memory specifications. You can upgrade the memory capacity to support your increasing workloads.

  The cluster architecture enables elastic scaling of storage space and throughput performance of the database service. This eliminates performance bottlenecks.

- Multiple instance types support flexible configuration changes

  The standalone cache architecture and master-replica storage architecture are available. You can modify instance configurations to meet different application scenarios.

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

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

- Visualized management simplifies operations and maintenance

  The KVStore for Redis console is a visualized management platform. Certain frequent and risky operations, such as instance cloning, backup, and data recovery, can be completed with only a few mouse clicks in the console.

- Automatic engine version upgrades prevent software flaws

  KVStore for Redis instances automatically upgrade engine versions to fix flaws at the earliest opportunity and keep your system up to date. This simplifies version management.

- Custom parameter configurations support individual requirements

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

- Asynchronous replication

The global replica instance synchronizes data asynchronously among child instances to minimize the impact on the service performance.

- A redundancy design is adopted for each system component to eliminate the risk of single points of failure.

13.5. Scenarios

Game industry applications

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

Scenario 1: KVStore for Redis works as a storage database

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

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

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

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

Live video applications

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

Two-node hot standby ensures high availability

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

Cluster editions eliminate the performance bottleneck

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

Easy scaling relieves pressure at peak hours

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

E-commerce industry applications

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

Scenario 1: rapid online sales promotion systems

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

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

Scenario 2: counter-based inventory management systems

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

13.6. Limits

Item	Description
List data type	The number of lists is not limited. The size of each element is 512 MB or less. We recommend that the number of elements in a list is less than 8,192. The value length is 1 MB or less.
Set data type	The number of sets is not limited. The size of each element is 512 MB or less. We recommend that the number of elements in a set is less than 8,192. The value length is 1 MB or less.
Sorted set data type	The number of sorted sets is not limited. The size of each element is 512 MB or less. We recommend that the number of elements in a sorted set is less than 8,192. The value length is 1 MB or less.
Hash data type	The number of fields is not limited. The size of each element in a hash table is 512 MB or less. We recommend that the number of elements in a hash table is less than 8,192. The value length is 1 MB or less.
Number of databases (DBs)	Each instance supports 256 DBs.
Supported Redis commands	For more information, see the "**Supported Redis commands**" topic of *KVStore for Redis User Guide* .
Monitoring and alerts	KVStore for Redis does not provide capacity alerts. You have to configure this feature in CloudMonitor. We recommend that you set alerts for the following metrics: instance faults, instance failover, connection usage, failed operations, capacity usage, write bandwidth usage, and read bandwidth usage.
Expired data deletion policies	• Active expiration: the system periodically detects and deletes expired keys in the background. • Passive expiration: the system deletes expired keys when you access these keys.
Idle connection recycling mechanism	KVStore for Redis does not actively recycle idle connections to KVStore for Redis. You can manage the connections.
Data persistence policy	KVStore for Redis uses the AOF_FSYNC_EVERYSEC policy, and runs the fsync command at a one-second interval.

13.7. Terms

Redis

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

Instance ID

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

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

Master-replica instance

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

High-performance cluster instance

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

Connection address

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

Eviction policy

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

DB

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

13.8. Instance types

> ? **Note** The maximum bandwidth includes the maximum upstream bandwidth and the maximum downstream bandwidth.

Standard dual-replica edition

Standard plan

Type	Service code	Maximum connections	Maximum bandwidth (MB)	CPU	Description	Zone-disaster recovery deployment
1 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredissdb1g.1db.0rodb.4proxy.default	10,000	10	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
4 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredissdb4g.1db.0rodb.4proxy.default	10,000	24	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
8 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredissdb8g.1db.0rodb.4proxy.default	10,000	24	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
16 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredissdb16g.1db.0rodb.4proxy.default	10,000	32	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region
32 GB standard primary/secondary edition for zone-disaster recovery	redis.logic.sharding.drredissdb32g.1db.0rodb.4proxy.default	10,000	32	1-core	Primary/secondary instance for zone-disaster recovery	Deployed across two zones in one region

Premium plan

Type	Service code	Maximum connections	Maximum bandwidth (MB)	CPU	Description	Zone-disaster recovery deployment
1 GB advanced primary/secondary edition	redis.master.small.special2x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone
2 GB advanced primary/secondary edition	redis.master.mid. special2x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone
4 GB advanced primary/secondary edition	redis.master.stand. special2x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone
8 GB advanced primary/secondary edition	redis.master.large.special1x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone
16 GB advanced primary/secondary edition	redis.master.2xlarge.special1x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone
32 GB advanced primary/secondary edition	redis.master.4xlarge. special1x	20,000	48	1-core	Primary/secondary instance	Deployed in one zone

Cluster edition

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

Type	Service code	Maximum connections	Maximum bandwidth (MB)	CPU	Description
64 GB cluster edition	redis.sharding.large.default	80,000	384	8-core	High-performance cluster instance
128 GB cluster edition	redis.sharding.2xlarge.default	160,000	768	16-core	High-performance cluster instance
256 GB cluster edition	redis.sharding.4xlarge.default	160,000	768	16-core	High-performance cluster instance

Cluster edition for zone-disaster recovery

Type	Service code	Maximum connections	Maximum bandwidth (MB)	CPU	Description
16 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredismdb16g.8db.0rodb.8proxy.default	80,000	384	8-core	Cluster instance for zone-disaster recovery
32 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredismdb32g.8db.0rodb.8proxy.default	80,000	384	8-core	Cluster instance for zone-disaster recovery
64 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredismdb64g.8db.0rodb.8proxy.default	80,000	384	8-core	Cluster instance for zone-disaster recovery
128 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredismdb128g.16db.0rodb.16proxy.default	160,000	768	16-core	Cluster instance for zone-disaster recovery

Type	Service code	Maximum connections	Maximum bandwidth (MB)	CPU	Description
256 GB cluster edition for zone-disaster recovery	redis.logic.sharding.drredismdb256g.16db.0rodb.16proxy.default	160,000	768	16-core	Cluster instance for zone-disaster recovery

14.ApsaraDB for MongoDB

14.1. What is ApsaraDB for MongoDB?

ApsaraDB for MongoDB is a high-performance distributed data storage service. It is a stable, reliable, and resizable database service fully compatible with MongoDB protocols. ApsaraDB for MongoDB offers a full range of database solutions, such as disaster recovery, backup, restoration, monitoring, and alerts.

ApsaraDB for MongoDB supports the following features:

- Automatically creates a three-node MongoDB replica set that encapsulates advanced functions such as disaster recovery and failover.
- Supports quick database backup and restoration. You can easily perform standard database backup and database rollback operations in the ApsaraDB for MongoDB console.
- Provides over 20 performance monitoring metrics and sends alerts. This helps you learn about the performance status of your database.
- Provides visual data management tools for convenient operations and maintenance.

14.2. Benefits

- High availability:
  - The three-node replica set architecture that delivers extremely high service availability

    ApsaraDB for MongoDB uses a high-availability architecture of three-node replica sets. The three data nodes are located on different physical servers and can automatically synchronize data. The primary and secondary nodes provide services. When the primary node fails, the system automatically selects a new primary node. When the secondary node fails, another secondary node takes over the services.
  - Automatic backup and quick data restoration

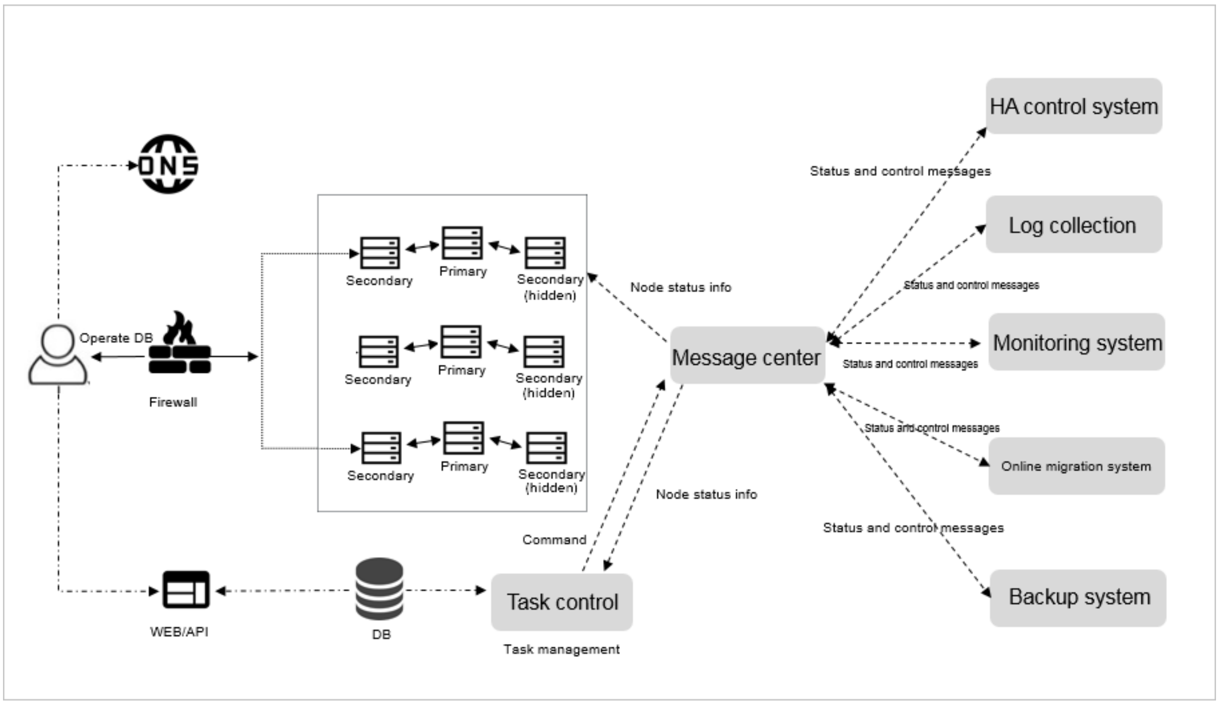
    Data is automatically backed up and uploaded to Object Storage Service (OSS) on a daily basis. This improves data disaster recovery capabilities and reduces disk space consumption. Backup files can be used to restore instance data to their original instance. This prevents irreversible effects on service data caused by incorrect operations and other errors.
- Data backup: ApsaraDB for MongoDB allows you to perform full or incremental backup and restore data from storage. Clusters can be backed up between data centers. The backup process is visually presented.
- High security:
  - Anti-DDoS protection: ApsaraDB for MongoDB filters inbound traffic. When DDoS attacks are identified, the source IP addresses will be scrubbed. If scrubbing fails, blackhole filtering is triggered.
  - IP whitelist configuration: You can configure up to 1,000 IP addresses to connect to an ApsaraDB for MongoDB instance. The IP whitelist can directly control risks at the source.
- Audit log management: You can store audit logs in Log Service automatically and download logs from Log Service. This facilitates long-term storage and management of audit logs.
- Ease of use: ApsaraDB for MongoDB provides various performance monitoring features. ApsaraDB for MongoDB provides real-time monitoring information about the CPU utilization, connections, and disk

usage of your instances and sends alerts to keep you informed of the status of your instances.

- Scalability: ApsaraDB for MongoDB supports three-node replica sets that can be scaled out. You can change the configuration of your instance if the current configuration cannot meet performance requirements or is unsuitable for your business needs. The configuration change process is completely transparent and does not affect your business.
- Isolation: ApsaraDB for MongoDB uses multiple instances and implements tasks for multiple tenants within a cluster at the same time. The tasks of the tenants are implemented in different instances. ApsaraDB for MongoDB instances are separated to isolate resources between different tenants.
- Permission management: ApsaraDB for MongoDB allows you to configure and manage tenants in a dynamic and centralized manner. You can also isolate resources and query usage statistics of resources. Management of multi-level tenants is supported.
- Scheduling: ApsaraDB for MongoDB schedules resources from multiple resource pools for tenants within different clusters.

14.3. Architecture

ApsaraDB for MongoDB provides a three-node replica set. You can directly use the primary or secondary node. The following figure shows the system architecture.



- **HA control system**: This module checks the high availability of an instance. You can use this module to detect and monitor the running status of ApsaraDB for MongoDB instances. If the system detects that the primary node instance is unavailable, it switches over from the primary node to the secondary node to ensure the high availability of MongoDB instances.
- **Log collection**: This module collects MongoDB running logs, including the slow query log and access log of an instance.
- **Monitoring system**: This module collects performance monitoring information about MongoDB instances, including basic metrics, disk capacity, network requests, and the number of operations that you have performed.

- **Online migration system**: When an error occurs with the physical server that hosts the instances, this module recreates an instance on the fly based on the backup files stored in the backup system. This ensures the continuity of your business.
- **Backup system**: This module generates ApsaraDB for MongoDB instance backups and stores the backup files in OSS. Currently, this module allows you to configure custom backup settings and create temporary backups. Files are retained for seven days.
- **Task control**: ApsaraDB for MongoDB supports multiple management operations, such as creating instances, changing instance configurations, and backing up instances. This module controls and tracks these tasks and troubleshoots errors based on your commands.

14.4. Features

Flexible architecture

ApsaraDB for MongoDB provides a three-node replica set. You can directly use the primary and secondary nodes. If the system detects that the primary node is unavailable, it switches over to the secondary node to ensure that ApsaraDB for MongoDB instances remain available.

Elastic scaling

- Quick scaling of storage capacity: You can adjust the storage capacity of an instance in the ApsaraDB for MongoDB console based on business requirements.
- Storage capacity adjustment on the fly: You can adjust the storage capacity of an instance on the fly. This ensures the continuity of your business.

Data security

- Automatic backup:
  - ApsaraDB for MongoDB allows you to set a time interval for backups to be created on a regular basis.
  - You can flexibly set the backup start time based on off-peak hours of your business.
  - Backup files are retained for free for up to seven days.
- Temporary backup:
  - You can create temporary backups.
  - Backup files are retained for free for up to seven days.
- Data restoration: You can use backup files to overwrite existing data and restore instances to a previous state.
- Backup file download: ApsaraDB for MongoDB retains your backup files for free for up to seven days. During this period, you can log on to the ApsaraDB for MongoDB console and download the backup files to a local device.
- Creation of instances from backup sets: You can create an instance in the ApsaraDB for MongoDB console by using backup files. This helps you complete the deployment process with ease.
- IP whitelist configuration: ApsaraDB for MongoDB can filter IP addresses that access your instance. You can log on to the ApsaraDB for MongoDB console and configure a whitelist of up to 1,000 IP addresses. This provides a highly secure access environment.
- Multi-layer network security protection:
  - VPCs are isolated at the TCP layer.

- Anti-DDoS can monitor and block DDoS attacks in real time.
- You can add up to 1,000 IP addresses to the whitelist.

Intelligent operations and maintenance

- Monitoring platform

  This platform provides real-time monitoring information about the CPU utilization, connections, and disk usage of your instances and sends alerts to keep you informed of the status of your instances.

- Graphical O&M platform

  This platform allows you to perform frequent and high-risk operations, such as instance cloning, backup, and data restoration.

- Database kernel version management

  This feature performs upgrades and fixes exceptions. It also optimizes ApsaraDB for MongoDB parameter configurations and maximizes the utilization of system resources.

High availability of clusters

A multi-node cluster architecture is used. Each component management node in the platform implements a high availability mechanism so that the failure of a service node within a cluster does not affect the overall operation of the corresponding service instance.

14.5. Scenarios

- **Businesses that require read/write splitting**

  ApsaraDB for MongoDB uses a high-availability architecture that features a three-node replica set. These three data nodes are located on different physical servers. The secondary and standby nodes automatically synchronize data from the primary node. Services are provided by the primary and secondary nodes. These two nodes have separate domain names and collaborate with MongoDB drivers to distribute read requests.

- **Businesses that require flexibility**

  As a schema-free database, ApsaraDB for MongoDB is particularly suitable for startup businesses because it does not require you to change table schema. You can store data with fixed structures in ApsaraDB for RDS databases, business data with flexible structures in ApsaraDB for MongoDB databases, and frequently accessed data in KVStore for Memcache databases or KVStore for Redis databases. This helps you store data efficiently and reduce costs.

- **Mobile applications**

  ApsaraDB for MongoDB supports two-dimensional space indexes. Therefore, it can provide support for location-based mobile application services. ApsaraDB for MongoDB adopts a dynamic storage method that is suitable for storing heterogeneous data from multiple systems. This satisfies the needs of mobile applications.

- **IoT applications**

  ApsaraDB for MongoDB provides excellent performance and an asynchronous data writing function. In special scenarios, it can provide in-memory database performance. This makes it extremely suitable for IoT writing scenarios with high concurrency. The MapReduce feature of ApsaraDB for MongoDB can aggregate and analyze large amounts of data.

- **Core log systems**

In scenarios where data is asynchronously written to disks, ApsaraDB for MongoDB can provide excellent data insertion performance and processing capabilities of an in-memory database. ApsaraDB for MongoDB allows you to create secondary indexes for dynamic queries. It can use the MapReduce aggregation framework to perform multidimensional data analysis.

14.6. Limits

Procedure	Limit
Create a database replica	The system automatically creates a three-node replica set. ApsaraDB for MongoDB provides a primary node, a secondary node, and a hidden standby node for each replica set. You cannot create secondary nodes.
Restart a database	You must restart instances in the ApsaraDB for MongoDB console.

14.7. Terms

Term	Description
Region	The geographical location of the server on which the ApsaraDB for MongoDB instance runs. You must specify a region when you create an ApsaraDB for MongoDB instance. The region cannot be changed after the instance is created. When you create an ApsaraDB for MongoDB instance, you must use it with an Alibaba Cloud ECS instance. You can access ApsaraDB for MongoDB instances through internal networks. Make sure that the region of an ApsaraDB for MongoDB instance is the same as that of the corresponding ECS instance.
Instance	An ApsaraDB for MongoDB instance. An instance is the basic unit of ApsaraDB for MongoDB services that you create. An instance is the operating environment for ApsaraDB for MongoDB and exists as a separate process on a host. You can create, modify, and delete an instance in the ApsaraDB for MongoDB console. Instances are independent and their resources are isolated. An instance does not consume resources such as CPU, memory, or I/O of another instance. Each instance has its own features, such as database type and version. The system has parameters to control instance behaviors.
Memory	The maximum memory that an instance can use.
Disk capacity	The disk size that you select when you create an instance. The disk capacity occupied by the instance includes datasets and the space required for normal instance operations, such as the system database, database rollback log, redo log, and indexes. Make sure that the ApsaraDB for MongoDB instance has sufficient disk space to store data. Otherwise, the instance may be locked. If an instance is locked due to insufficient disk capacity, you can purchase a larger disk to unlock the instance.
IOPS	The maximum number of read or write operations performed on block devices per second. Each operation consumes 4 KB.

Term	Description
CPU core	The maximum computing capability of the instance. A single core CPU has a minimum of 2.3 GHz hyper-threading (Intel Xeon series Hyper-Threading) computing power.
Connections	The TCP connections between clients and ApsaraDB for MongoDB instances. If a client uses a connection pool, the connections between the client and instance are persistent connections. Otherwise, they are short connections.
Mongos	The routing service that processes requests. All requests must be coordinated through mongos that serves as a request distribution center and forwards data requests to the corresponding shard server. You can use multiple mongos to process requests. If one fails, other mongos can continue to process the requests.
Config server	The servers that store all database metadata configurations, including routers and shards. mongos does not store but caches shard server information and data routing information in memory. The information is stored on config servers. When you start mongos for the first time or shut it down and then restart it, it automatically loads configuration information from config servers. mongos updates the cache when there are metadata changes. This ensures that mongos can always obtain the correct routing information. Config servers store metadata of shards and routers and have high requirements for service availability and data reliability. Therefore, ApsaraDB for MongoDB uses a three-node replica set to ensure the reliability of the config servers.

14.8. Instance specifications

ApsaraDB for MongoDB replica set specifications

Type	Specification	Code	Maximum connections	Maximum IOPS
General specifications	1 Core - 2 GB	dds.mongo.mid	500	1,000
	2 Core - 4 GB	dds.mongo.standard	1,000	2,000
	4 Core - 8 GB	dds.mongo.large	2,000	4,000
	8 Core - 16 GB	dds.mongo.xlarge	4,000	8,000
	8 Core - 32 GB	dds.mongo.2xlarge	8,000	14,000
	16 Core - 64 GB	dds.mongo.4xlarge	16,000	16,000
Dedicated specifications	2 Core - 16 GB	mongo.x8.medium	2,500	4,500
	4 Core - 32 GB	mongo.x8.large	5,000	9,000
	8 Core - 64 GB	mongo.x8.xlarge	10,000	18,000

Type	Specification	Code	Maximum connections	Maximum IOPS
	16 Core - 128 GB	mongo.x8.2xlarge	20,000	36,000
	32 Core - 256 GB	mongo.x8.4xlarge	40,000	72,000
Dedicated physical machine	60 Core - 440 GB	dds.mongo. 2xmonopolize	100,000	100,000

15.ApsaraDB for OceanBase

15.1. What is ApsaraDB for OceanBase?

ApsaraDB for OceanBase is a financial-grade, distributed relational database service that features high performance, high availability, and high scalability. It supports active geo-redundancy and geo-disaster recovery to ensure high availability. It also supports high scalability to meet the increasing business requirements.

These features of ApsaraDB for OceanBase help you handle the challenges that are brought by rapid business growth. ApsaraDB for OceanBase also provides scalable and low latency database services in high throughput scenarios. This ensures improved user experience. For example, during the Double 11 in 2017, ApsaraDB for OceanBase handled all the transactions and payment requests. The maximum number of transactions that were made on Alipay reached 256,000 per second. The maximum number of processed requests per second reached 42 million. ApsaraDB for OceanBase accelerates the development of Internet finance.

The distributed engine of ApsaraDB for OceanBase uses the Paxos protocol and maintains multiple replicas. For the Paxos protocol, transactions can be committed only after they are approved by a majority of the acceptors. The Paxos protocol and multiple-replica design allow ApsaraDB for OceanBase to offer high availability and disaster recovery capabilities. ApsaraDB for OceanBase can help you achieve zero downtime. ApsaraDB for OceanBase supports high-availability architectures, such as active geo-redundancy and geo-disaster recovery. You can deploy the ApsaraDB for OceanBase service across data centers, regions, or continents. ApsaraDB for OceanBase provides financial-grade availability features and ensures strong consistency of transactions.

ApsaraDB for OceanBase is similar to an in-memory database and adopts a read/write splitting architecture. To ensure high efficiency for the storage engine, ApsaraDB for OceanBase stores baseline data in solid-state drives (SSDs) and stores incremental data in memory. This ensures that ApsaraDB for OceanBase offers high performance services. ApsaraDB for OceanBase is a cloud-based database service that supports multi-tenant data isolation. Each cluster of ApsaraDB for OceanBase can provide services for multiple tenants. The tenants are isolated so that they are not affected by each other.

ApsaraDB for OceanBase is compatible with most of the MySQL 5.6 features. This allows you to migrate MySQL-based services to ApsaraDB for OceanBase based on zero or small code modifications. This improves the efficiency of developing applications and migrating services. In ApsaraDB for OceanBase, you can create partitioned tables and use subpartitions. This serves as an alternative to MySQL sharding solutions. The ApsaraDB for OceanBase console provides an easy way for you to manage complex databases. For example, you can use the console to upgrade or downgrade instances, view performance data, and view optimization suggestions.

15.2. Benefits

Low costs

ApsaraDB for OceanBase requires lower hardware costs than traditional relational databases to provide the same services. Examples of the traditional relational databases include MySQL databases.

Category	Database operation	Hardware

Category		Database operation			Hardware		
Sub-category		Disk read and write mode	Hardware requirements	Cost	Roles of primary and secondary databases	Available servers/total servers	Cost
Database type	MySQL	Random reads and writes	Read-write solid-state drives (SSDs)	1	The primary database processes read and write requests. The secondary database is used for disaster recovery.	2/4	1
	OceanBase	Random read-only operations	Read-intensive SSDs	2/3	ApsaraDB for OceanBase adopts an architecture where each cluster is deployed across multiple zones. For example, a cluster can be deployed in three zones. The nodes in two of the three zones process read and write requests, and the nodes in the other zone is used for disaster recovery.	2/3	3/4

High scalability

ApsaraDB for OceanBase is a distributed relational database service that uses computers as independent nodes. Data is distributed across the nodes based on the availability of each node and the disaster recovery requirements. If the volume of data increases, ApsaraDB for OceanBase automatically add nodes to meet the increasing business requirements.

Service continuity

ApsaraDB for OceanBase automatically removes faulty nodes. The services that were running on the faulty nodes are switched to the other nodes. If a data center fails, ApsaraDB for OceanBase switches the services that were running on the faulty data center to the nodes in another data center. The switchover occurs in a short period to ensure service continuity.

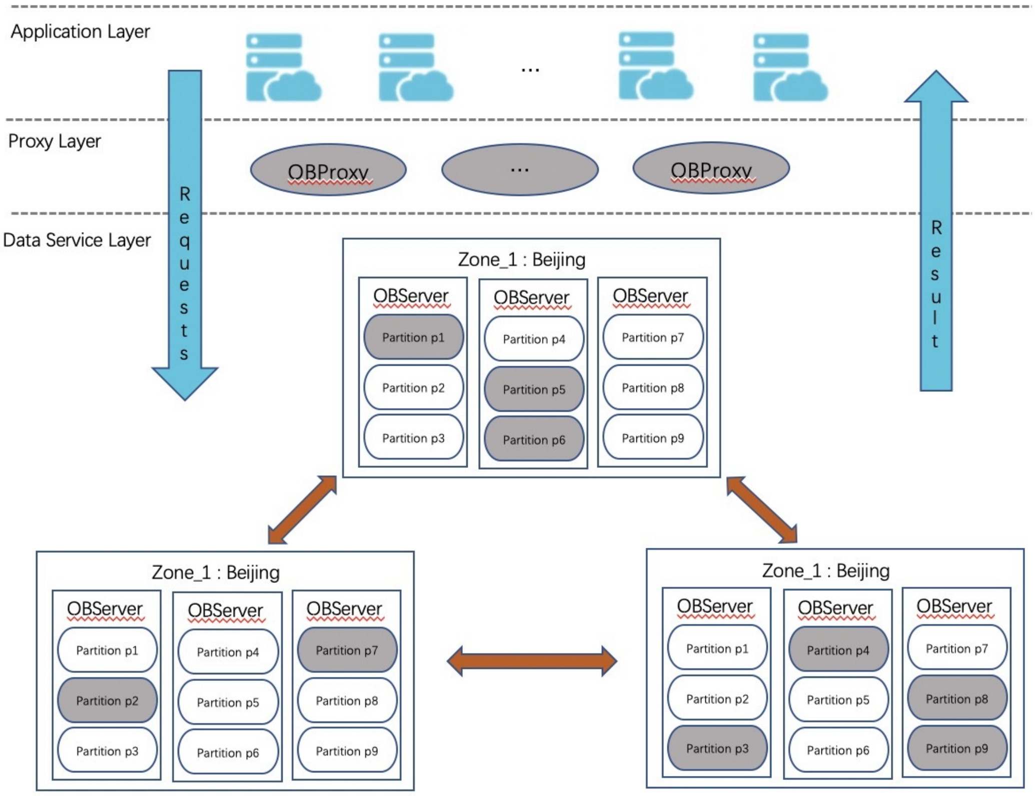
Zero data loss

In ApsaraDB for OceanBase, each time a transaction is committed, the corresponding log entries are synchronized in real time across at least three nodes for persistent storage. If an unrecoverable error occurs on a node, ApsaraDB for OceanBase allows you to restore each completed transaction by using the log entries from the other nodes. This ensures financial-grade data reliability.

15.3. Architecture

ApsaraDB for OceanBase architecture shows the architecture of ApsaraDB for OceanBase.

ApsaraDB for OceanBase architecture



In the figure, the data service layer represents an ApsaraDB for OceanBase database cluster. The cluster is deployed across three zones and each zone hosts multiple physical servers. Physical servers are also known as OBServers in ApsaraDB for OceanBase. ApsaraDB for OceanBase uses a shared-nothing distributed architecture. In this architecture, each node is independent and provides the same features.

Data distribution

ApsaraDB for OceanBase distributes data across OBServers for storage. However, ApsaraDB for OceanBase provides services as an entire database. You do not need to concern yourself with the details about data shards or storage locations. This distinguishes ApsaraDB for OceanBase from traditional databases. The system locates the data to be accessed based on your request and forwards the request to an appropriate OBServer for processing.

In ApsaraDB for OceanBase, data is distributed across OBServers in one zone and data replicas are stored in other zones. The preceding figure shows that the data in the ApsaraDB for OceanBase database cluster has three replicas. Each of the three replicas is stored in a zone of the database cluster. The database cluster is deployed in three zones to provide services.

ApsaraDB for OceanBase uses the Paxos distributed election algorithm to ensure high availability of the system. Each partition in the cluster has at least three replicas that are distributed across different zones. Log entries are synchronized among partition replicas based on the Paxos protocol. Each partition and its replicas consist of an independent Paxos replica group. In each replica group, one replica serves as the leader and the other replicas serve as followers. Write requests for each partition are automatically routed to the leader replica of the partition. In the preceding figure, leader partitions are highlighted in gray. Leader partitions can be distributed across OBServers. In this scenario, write requests for partitions are routed to different OBServers. This improves parallel processing of data writes, implements multi-point data writes, and therefore improves system performance.

Cluster roles

The following table describes the components and the roles in a cluster.

Component	Role	Feature	Description
Zones (Multiple OBServers can be deployed in each zone.)	Root service node	Each zone has an OBServer that serves as the root service node. The root service node provides a wide range of features. For example, you can use the node to manage clusters, data distribution, major freeze operations, and system bootstrapping.	None.
	Partition service node	Each OBServer can serve as a partition service node. The partition service node manages data partitions.	Each partition service node has a Structured Query Language (SQL) engine and a storage engine.

Storage engine

ApsaraDB for OceanBase is similar to an in-memory database. In ApsaraDB for OceanBase, data is divided into baseline data and incremental data. Baseline data is stored in solid state disks and incremental data is stored in memory. This ensures that operations on hotspot data and transactions are performed in memory. In this scenario, ApsaraDB for OceanBase can offer nearly the same transaction processing performance as in-memory databases.

If the incremental data in memory exceeds the specified threshold, major freeze operations are triggered. The major freeze operations combine incremental data and baseline data into new baseline data. To eliminate the impacts of major freeze operations on your services, you can perform polling major freeze operations. To be more specific, when you merge the data of a replica, you can switch the requests for the replica to another zone that is not involved in the major freeze operation. After the major freeze operation is complete, you can switch the requests back to the replica and proceed to merge the data of other replicas. This prevents your services from being affected by daily major freeze operations.

SQL engine

ApsaraDB for OceanBase uses an SQL engine based on the underlying distributed system. The SQL engine includes an SQL parser, SQL rewriter, cost-based SQL optimizer, and SQL executor. The SQL engine is optimized to improve parallel processing based on the features of the distributed system.

ApsaraDB for OceanBase optimizes the processing of distributed transactions by adding a CLEAR phase to the two-phase commit procedure. Theoretically, the transaction processing ends after the coordinator completes the commit operation. In the actual implementation, after the transaction is committed, the coordinator and the participants may still need to query the status or perform retries. This is because transmission errors such as network packet losses or server exceptions may occur. In this scenario, the CLEAR phase ensures that data structures are deleted only after each state machine reaches the expected state.

Multitenancy support

ApsaraDB for OceanBase uses the Database-as-a-Service (DBaaS) model. Each ApsaraDB for OceanBase cluster serves multiple services. Each service has one or more tenants. Tenants are isolated from each other. This allows you to configure the resources that each tenant can use, such as the number of sessions, CPUs, memory, and disk input/output operations per second (IOPS). If a tenant consumes more resources than the allocated resources, the system automatically performs graceful service degradation for the tenant. This ensures that the other tenants are not affected. One of the tenants serves as the system tenant. The system tenant can access all the system tables and background services. You can implement multitenancy at multiple layers. For example, you can use virtual machines at the operating system level. You can also use control groups (cgroups) to implement lightweight isolation or use Java virtual machines to implement isolation. ApsaraDB for OceanBase implements multi-tenant data isolation at the underlying layer of databases. This method allows you to implement isolation within a single tenant and reduce overheads. For example, ApsaraDB for OceanBase implements automatic throttling on large requests for your services. This ensures that small requests are not affected.

Smart access proxy: OBProxy

The OBProxy is a high-performance and easy-to-maintain reverse proxy server. The OBProxy serves as both a server and a client. As a server, the OBProxy receives SQL requests from clients and forwards the data that is returned by OBServers to clients. As a client, the OBProxy forwards requests from clients to OBServers and receives the data that is returned by OBServers. During data forwarding, the OBProxy functions as a data pipeline that is transparent to clients.

Your requests are first sent to the OBProxy in the system. Based on the requested data, the OBProxy parses the SQL code to locate the destination partitions, and then forwards the SQL requests to the servers that store the destination partitions.

High availability

In ApsaraDB for OceanBase, the underlying technologies such as the distributed election and the multi-active database architecture are implemented based on the Paxos protocol. The Paxos protocol supports multiple active nodes. If a node fails, you can restore the services on the faulty node within several seconds by switching the services to other running nodes. During this process, no data loss occurs.

15.4. Features

15.4.1. Overview

ApsaraDB for OceanBase provides a wide range of relational database services. You can call SQL API operations to access and manage the data that is stored in ApsaraDB for OceanBase instances.

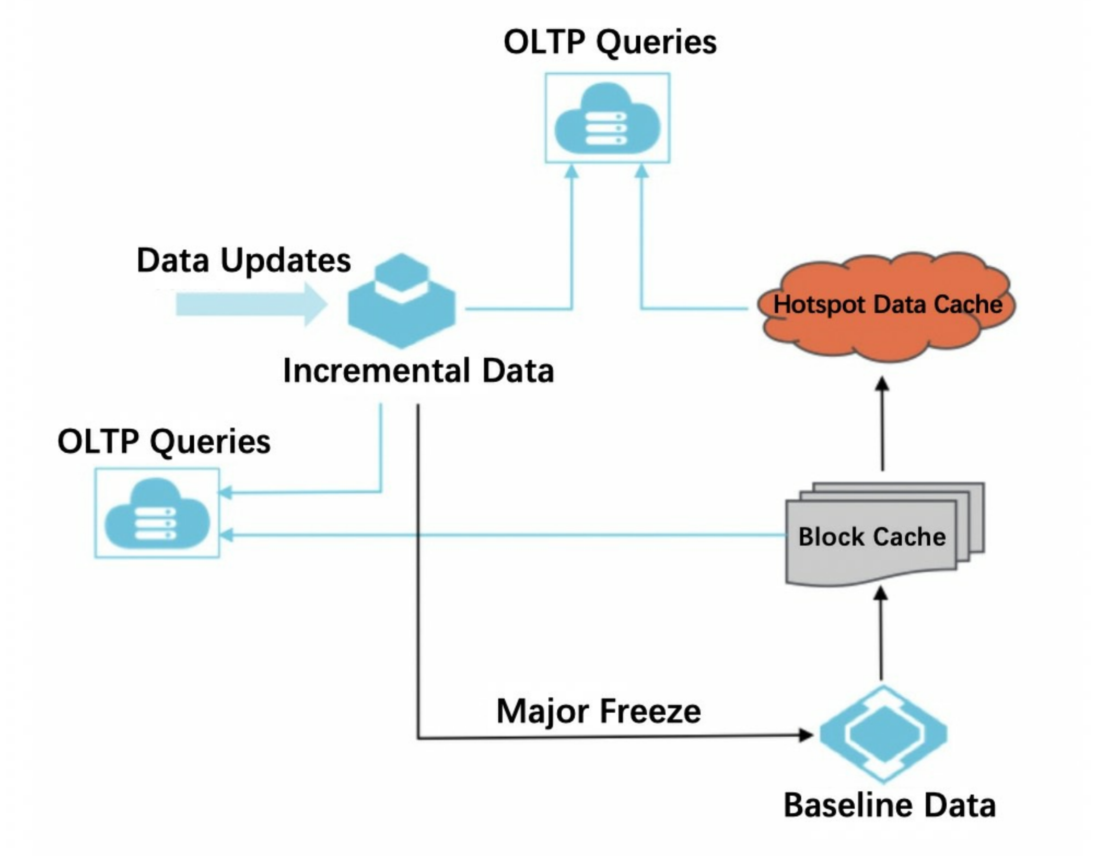
15.4.2. High-efficiency storage engine

ApsaraDB for OceanBase uses a shared-nothing distributed architecture. In this architecture, each OBServer is independent and provides the same features. The partitions that each OBServer manages are different from the partitions that each of the other OBServers manages.

The storage engine for a single OBServer uses a read/write splitting architecture. The storage engine uses MemTables to store updated dynamic data in the memory. The storage engine uses sorted string tables (SSTables) to store baseline data in disks.

All the data for each partition is stored on an OBServer, such as baseline data, incremental data, and transaction log records. Therefore, each data read and write operation for a partition is performed on only one OBServer. If you need to write data to partitions on multiple OBServers, you can perform concurrent write operations.

High-efficiency storage engine in ApsaraDB for OceanBase



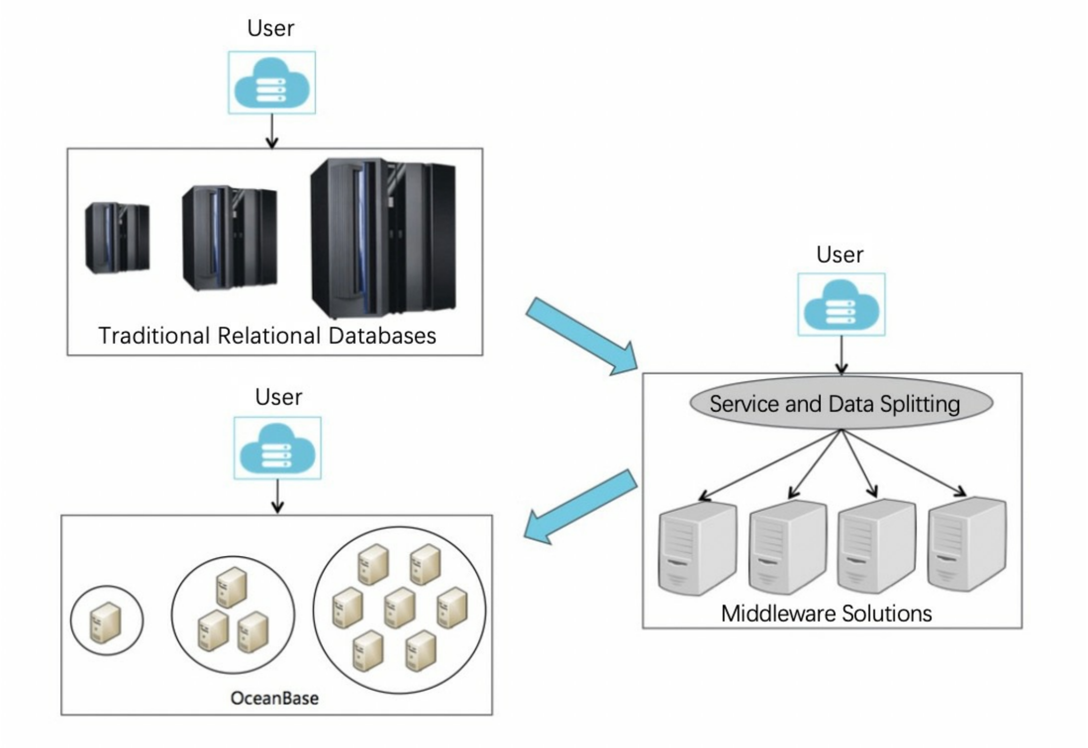
The storage engine uses a read/write splitting architecture. This architecture offers diverse benefits. For example, you can compress large amounts of static baseline data to reduce storage costs. This architecture also eliminates your concerns on row cache expiration that is caused by data writes. The disadvantage is that the architecture results in complex procedures of data reads. To reduce the complexity, the system must merge data in real time. However, this may compromise system performance. To resolve this issue, ApsaraDB for OceanBase provides various optimization methods. For example, the Bloom filter cache filter outs the rows that do not exist. The Bloom filter cache executes the INSERT ROW statement to check whether the row exists. If the row does not exist, the I/O operation for this row is not required. The system preferably reads the updated data from active MemTables. If the system retrieves the required columns from the active MemTables, the system does not need to read baseline data or merge incremental data with baseline data.

Incremental data is written in the memory. After the amount of the stored incremental data reaches a specified threshold, the system merges the incremental data and the baseline data to generate new baseline data. This process is known as major freeze. Major freeze operations cause additional loads. This may affect the requests from clients. To resolve this issue, you can use a rotated policy to perform polling major freeze operations. To be specific, if you need to merge incremental data with baseline data in multiple data centers of ApsaraDB for OceanBase, you can switch the requests for one of the data centers to another data center. After the major freeze operation is complete in the original data center, you can switch the requests back to the original data center. This policy allows you to eliminate the impact of major freeze operations on your business. You can use the rotated policy during system upgrades and maintenance. Before you upgrade a version, you can switch the requests from one OBServer to another OBServer. After the upgrade is complete, you can perform a phased switchover to switch the requests back to the original OBServer based on the percentages of requests. This way, you can immediately perform a rollback after a failure occurs during the switchover. This prevents data losses.

15.4.3. High scalability

Scalability comparison shows that ApsaraDB for OceanBase provides higher scalability at the database level than traditional relational databases.

Scalability comparison



ApsaraDB for OceanBase uses a distributed architecture that allows you to scale in or out your ApsaraDB for OceanBase services in an easy way. In addition, the scaling is transparent to the service users. ApsaraDB for OceanBase provides advanced features such as dynamic load balancing in each cluster, distributed queries across servers, and global indexing. These features improve the scalability of ApsaraDB for OceanBase services.

ApsaraDB for OceanBase allows you create partitioned tables and subpartitions. This serves as an alternative to MySQL sharding solutions.

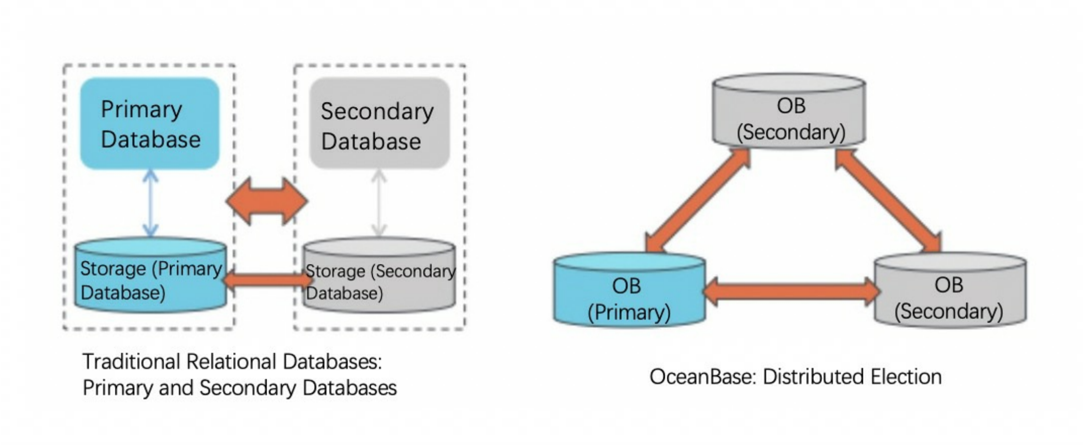
15.4.4. High availability

Availability comparison shows that traditional relational databases such as MySQL implement high availability by using a primary/secondary architecture. In the architecture, the primary databases provide services, and synchronize log entries to the secondary databases in real time. For performance reasons, the full sync mode is not used in production environments to synchronize log entries. If the full sync mode is used, clients receive responses only after transactions are performed in the primary databases and are synchronously replicated to the secondary databases. Therefore, if services are switched from faulty primary databases to secondary databases, data losses may occur. In this scenario, the recovery point objective (RPO) is not zero. If errors occur, switchovers require third-party tools or manual interventions in most cases. This results in a high recovery time objective (RTO).

Compared with traditional relational databases, ApsaraDB for OceanBase uses at least three servers. Each data record is stored in more than 50% of all the servers. For example, if three servers are used, each data record must be stored in two of the three servers. Each write transaction is valid only if the transaction is stored in more than 50% of all the servers. Therefore, no data loss occurs if only a minority of the servers fail. This ensures that a zero RPO can be achieved.

In addition, ApsaraDB for OceanBase uses the Paxos protocol at the underlying layer to ensure high availability. If the primary database fails, a new primary database is automatically elected by the remaining servers based on the Paxos protocol. This ensures automatic switchovers and service continuity.

Availability comparison


Primary Database
Secondary Database
Storage (Primary Database)
Storage (Secondary Database)
Traditional Relational Databases: Primary and Secondary Databases
OB (Secondary)
OB (Primary)
OB (Secondary)
OceanBase: Distributed Election

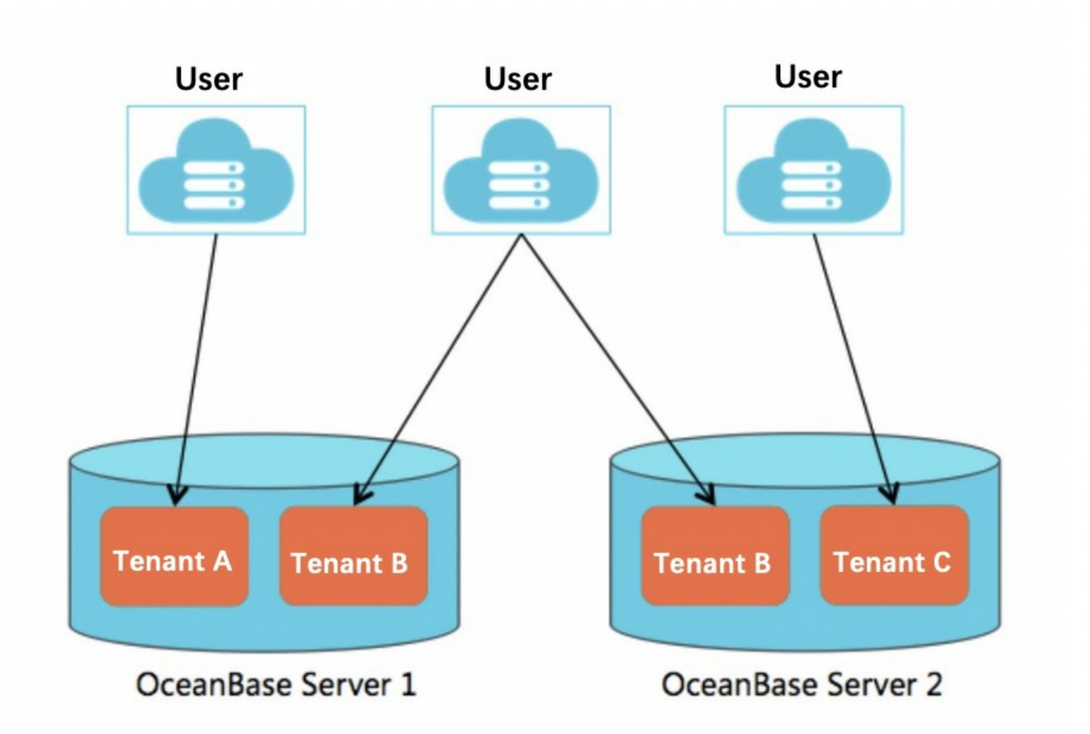
ApsaraDB for OceanBase retains multiple replicas and uses the Paxos protocol. This allows you to deploy ApsaraDB for OceanBase across data centers in different regions and implement high-availability features such as active geo-redundancy. ApsaraDB for OceanBase supports the following typical deployment solutions: Three Data Centers Across Two Regions and Five Data Centers Across Three Regions. This allows ApsaraDB for OceanBase to meet the various business requirements for disaster recovery across data centers and zones.

15.4.5. Multi-tenant data isolation

ApsaraDB for OceanBase is a cloud database service. Multitenancy architecture in ApsaraDB for OceanBase shows that ApsaraDB for OceanBase implements a multitenancy architecture at the underlying layer. In the architecture, resources are isolated between tenants. Based on this architecture, each ApsaraDB for OceanBase cluster can serve multiple services. Each service has one or more tenants. Tenants are isolated from each other. This allows you to configure the resources that each tenant can use.

If a tenant consumes more resources than the allocated resources, the system automatically performs graceful service degradation for the tenant. This ensures that the other tenants are not affected.

Multitenancy architecture in ApsaraDB for OceanBase



15.4.6. Custom components

ApsaraDB for OceanBase provides a wide range of custom components. The custom components help you implement additional product features and improve operations and maintenance (O&M) capabilities. ApsaraDB for OceanBase includes the following major components: OceanBase Cloud Platform (OCP), OBProxy, backup and restore tool, and historical database platform.

OCP is an ApsaraDB for OceanBase platform that you can use to manage database clusters. OCP provides features such as resource and capacity management, cluster and instance lifecycle management, performance monitoring and alerting based on real-time computing, and API management. OCP is an end-to-end platform that you can use to manage ApsaraDB for OceanBase databases and perform O&M tasks. You can also use the API-related features that are provided by OCP to customize management tools and platforms based on your business requirements.

The OBProxy is a reverse proxy server for ApsaraDB for OceanBase clusters. To access ApsaraDB for OceanBase clusters, applications can use MySQL drivers to connect to the OBProxy. The OBProxy routes statements to OBServers. The OBProxy also allows the distributed architecture of ApsaraDB for OceanBase to be transparent to frontend applications. The routing feature improves the execution performance of online transaction processing (OLTP) statements. If the distributed architecture is transparent to services, you can reduce the impacts of transient connections, node failures, and other events on your business.

The backup and restore tool allows you to back up and restore data in ApsaraDB for OceanBase clusters based on the following granularities: clusters and tenants. This tool serves as a supplement to the multi-replica mechanism to ensure data security. The method of data storage in ApsaraDB for OceanBase allows you to back up and restore persistent baseline data in solid-state disks and incremental data in memory. The method also allows you to restore data to a specified point in time. You can use Object Storage Service (OSS) or the local storage method to store your backup data.

The historical database platform provides an end-to-end solution for data storage and archiving. The platform provides an easy way for you to configure migration tasks and specify rules to migrate data from online databases to historical database clusters of ApsaraDB for OceanBase. The historical database clusters are cost-effective and require only about 10% costs of the online databases. Examples of the online databases include ApsaraDB for OceanBase, MySQL, and Oracle databases. This easy-to-use platform allows you to configure throttling rules for migration, verify data integrity and accuracy after migration, and delete online data after data verification.

15.5. Scenarios

15.5.1. Overview

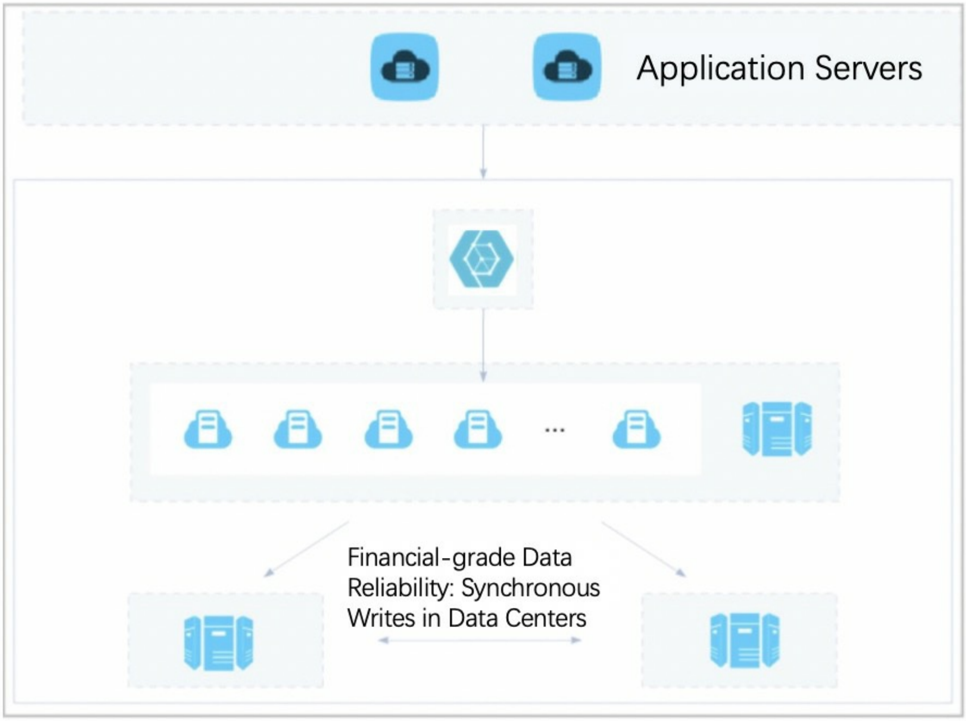
ApsaraDB for OceanBase provides high performance services and financial-grade data reliability. It uses a distributed architecture that allows you to improve storage capabilities based on your business requirements. The other topics in the chapter describe the application scenarios of ApsaraDB for OceanBase.

15.5.2. Financial-grade data reliability

The financial industry has high requirements for data reliability. In ApsaraDB for OceanBase, each time a transaction is committed, the corresponding log entries are synchronized in real time across data centers for persistent storage. If a disaster occurs in a data center, you can restore each completed transaction by using the log entries from other data centers. This ensures financial-grade data reliability.

Architecture shows the architecture of ApsaraDB for OceanBase.

Architecture

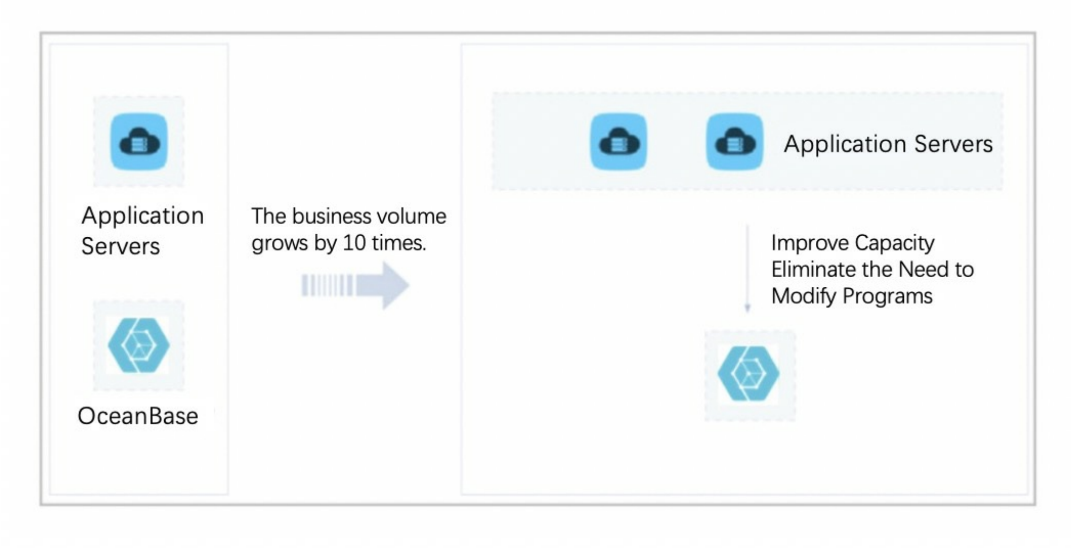


15.5.3. Fast business growth

Fast business growth brings a large number of challenges for databases. ApsaraDB for OceanBase helps you handle these challenges and allows the databases to meet the increasing business requirements. ApsaraDB for OceanBase is a distributed relational database service that uses computers as independent nodes. Data is distributed across the nodes based on the availability and the capacity of each node. If the volume of data increases, ApsaraDB for OceanBase automatically add nodes to meet the increasing business requirements.

Architecture shows the architecture of ApsaraDB for OceanBase.

Architecture



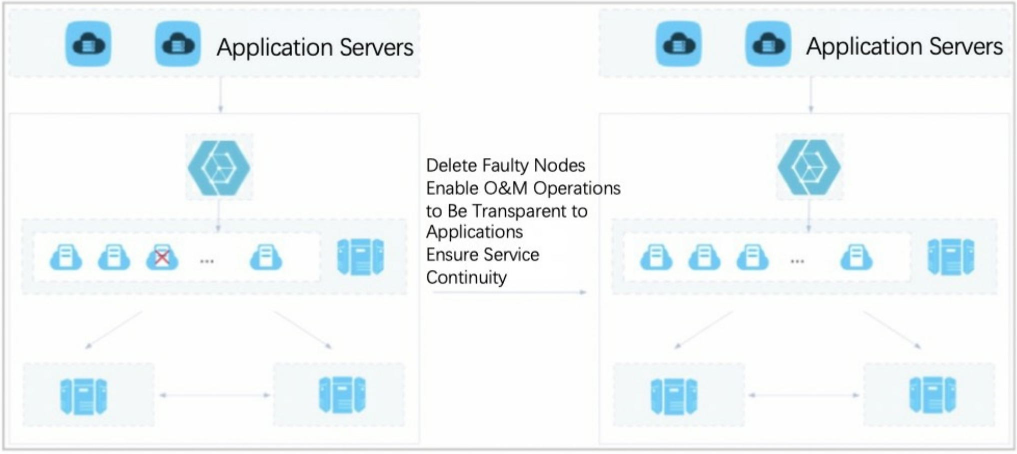
15.5.4. Service continuity

Enterprises need to provide uninterrupted services to ensure a smooth user experience.

Based on a distributed cluster architecture, ApsaraDB for OceanBase automatically removes faulty nodes. You can use the corresponding data replicas that are stored on other nodes to ensure service continuity. If a data center fails, ApsaraDB for OceanBase switches the services that were running on the faulty data center to the nodes in another data center. The switchover occurs in a short period to ensure service continuity.

Architecture shows the architecture of ApsaraDB for OceanBase.

Architecture


Application Servers
Application Servers
Delete Faulty Nodes
Enable O&M Operations to Be Transparent to Applications
Ensure Service Continuity

15.6. Limits

Limits on database features

- Stored procedures, triggers, cursors, and user-defined functions are not supported.
- Some data types such as ENUM and SET are not supported.
- Temporary tables are not supported.
- The INSERT statement cannot be executed to write data to more than one partition.
- The LOAD DATA statement is not supported.
- Global indexing is not supported.

Limits on database usage

- In some scenarios, the leader replicas of partitions in partitioned tables are distributed across OBServers. In these scenarios, if a query needs to retrieve data from multiple partitions, strong consistency is not supported for data reads. Therefore, the following SQL hint must be added: /*+READ_CONSISTENCY(weak)*/.
- In the hint of your SQL code, you can specify the number of concurrent threads. The maximum number of concurrent threads is 128.
- Each OBServer can store a maximum of 80,000 partitions. We recommend that you store no more than 50,000 partitions in each OBServer.
- The following table describes the limits for database and table names.

Item	OceanBase
Maximum length of a database name	64
Maximum length of a table name	64
Maximum length of a column name	128
Maximum length of a view name	64

Item	OceanBase
Maximum length of an alias	255
Maximum length of a row	32M
Maximum length of a table	512
Maximum length of a primary key	64
Maximum length of a primary key column	The maximum length varies based on data types.

- The following table describes the limits for data types.

Item	OceanBase
Maximum length of a CHAR column	255
Maximum length of a VARCHAR column	256K
Maximum length of a BINARY column	255
Maximum length of a VARBINARY column	65K

- The following table describes the limits for indexes.

Item	OceanBase
Maximum length of an index name	64
Maximum number of columns for an index	64
Maximum length of an index column	The maximum length varies based on data types.
Maximum length of an index	32M
Maximum number of indexes for a table	128

- The following table describes the limits for partitioned tables.

Feature	OceanBase
Global indexing	Not supported. You must add the local modifier when you create indexes.
KEY	Supported. MurmurHash functions are used to implement hashing.
RANGE COLUMNS partitions	Supported. You can use DATE, DATETIME, and CHAR columns to implement partitioning. You can also use integer columns, such as INT and BIGINT.
List partitioning	Not supported.

Feature	OceanBase
LIST COLUMNS	Not supported.
Interval partitioning	Not supported.
linear key	Not supported.
linear hash	Not supported.
RANGE	Supported. You can use ABS, CEIL, CEILING, and DATE_DIFF functions to implement partitioning.
Subpartitioning	Supported. You can create only RANGE COLUMNS subpartitions for hash or key partitions.
Generated columns as partition keys	Supported. You can use only SUBSTRING functions to create generated columns.
Cross-partition updates	Not supported.
add/drop partition	Not supported.

15.7. Terms

Baseline data

In ApsaraDB for OceanBase, baseline data is the data snapshots that are created before a specific time point. Baseline data is stored as static data in the SSTable format for persistent storage.

Incremental data

In ApsaraDB for OceanBase, incremental data is the data that is updated after each major freeze operation. Incremental data is stored in memory in the format of B-trees and hash tables.

Tenant

A tenant is a container for various database objects and resources. Each tenant corresponds to a database service instance in ApsaraDB for OceanBase. ApsaraDB for OceanBase is a database service that implements a multitenancy architecture. Each ApsaraDB for OceanBase cluster can host more than one database service instance. Database service instances are isolated from each other. Each of the database service instances corresponds to a tenant. Each tenant has a set of compute and storage resources and provides complete and independent database services.

OBServer

An OBServer is a service process of ApsaraDB for OceanBase. In most cases, an OBServer has exclusive use of a physical server. Therefore, an OBServer is also used as an equivalent to the physical server where it resides. In ApsaraDB for OceanBase, each OBServer is uniquely identified by the IP address and the service port.

OLAP

Online analytical processing (OLAP) is an approach that supports complex analysis and provides intuitive query results. OLAP is one of the major analysis tools that are used in data warehouse systems. The major goal of OLAP is to support decision-making.

OLTP

Online transaction processing (OLTP) is an approach to providing quick responses to user operations. OLTP supports transaction-oriented applications. In the OLTP approach, the user data that is received by frontend applications is immediately transferred to compute centers for processing. The compute centers return the processing result in a short period to ensure quick responses to user operations.

RootServer

A RootServer is a primary server that manages clusters, replicas, and data distribution, and provides listener services.

Zone

A zone is a physical area in a region. The area has an independent power supply and network. In most cases, one zone is deployed in each data center. Each ApsaraDB for OceanBase cluster is deployed in one or more zones. In most cases, data replicas are distributed across zones to ensure data security and high availability. This prevents a single point of failure from affecting the database services of ApsaraDB for OceanBase. One or more physical servers can be deployed in each zone.

16.Data Transmission Service (DTS)

16.1. What is DTS?

Data Transmission Service (DTS) is a data service that is provided by Alibaba Cloud. DTS supports data transmission between various types of data sources, such as relational databases and big data systems.

Features

DTS has the following advantages over traditional data migration and synchronization tools: high compatibility, high performance, security, reliability, and ease of use. DTS allows you to simplify data transmission and focus on business development.

Feature	Description
Data migration	You can use DTS to migrate data between homogeneous and heterogeneous data sources. This feature applies to the following scenarios: data migration to Alibaba Cloud, data migration between instances within Alibaba Cloud, and database splitting and scale-out.
Data synchronization	You can use DTS to synchronize data between data sources. This feature applies to the following scenarios: disaster recovery, data backup, load balancing, cloud BI systems, and real-time data warehousing.
Change tracking	You can use DTS to track data changes from user-created MySQL databases, ApsaraDB RDS for MySQL instances, Cloud Native Distributed Database PolarDB-X instances (formerly known as DRDS), and user-created Oracle databases in real time. This feature applies to the following scenarios: cache updates, business decoupling, asynchronous data processing, synchronization of heterogeneous data, and synchronization of extract, transform, and load (ETL) operations.

16.2. Benefits

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

Diverse transmission methods

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

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

High performance

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

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

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

High security and reliability

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

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

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

Easy-to-use

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

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

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

16.3. Environment requirements

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

- PF51.*
- PV52P2M1.*
- DTS_E.*
- PF61.*
- PF61P1.*
- PV62P2M1.*
- PV52P1.*
- Q5F53M1.*
- PF52M2.*
- Q41.*
- Q5N1.22
- Q5N1.2B
- Q46.22
- Q46.2B
- W41.22
- W41.2B
- W1.22
- W1.2B

- W1.2C
- D13.12

Use the following operating system:

AliOS7U2-x86-64

Notice

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

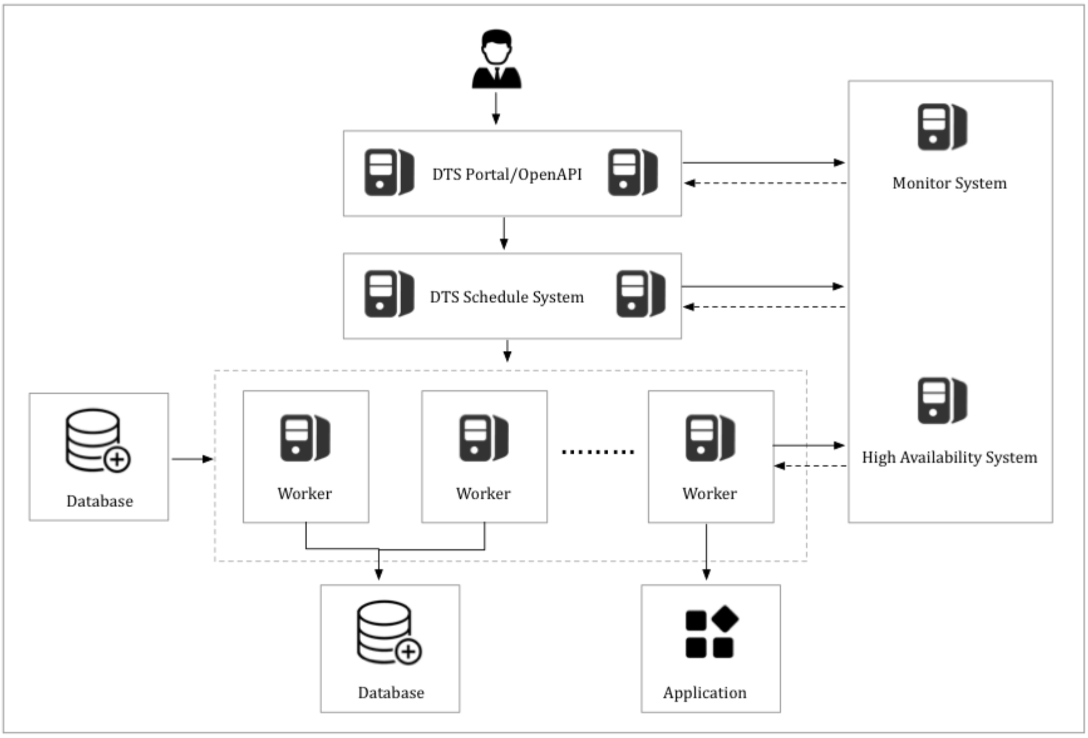
  If the space of the hard disk where the */apsara* directory resides is smaller than 2 TB, tasks may fail to run and errors may occur. In this case, DTS cannot restore failed tasks or pull data properly.

16.4. Architecture

System architecture

System architecture shows the system architecture of DTS.

System architecture



- **High availability**

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

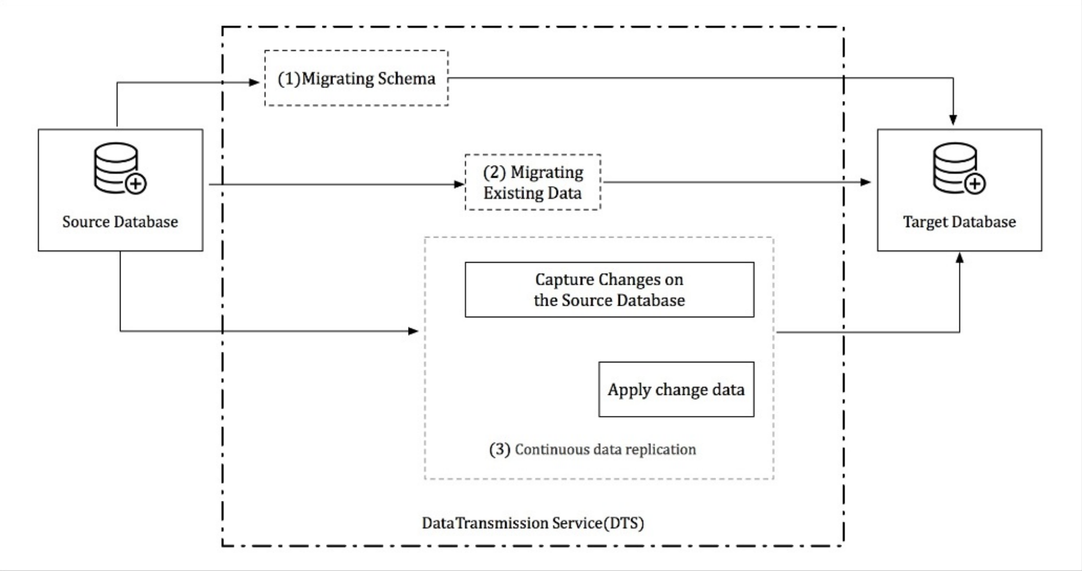
- **Monitor changes in the data source IP address**

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

Data migration process

Data migration workflow shows how data migration works.

Data migration workflow


(1)Migrating Schema
Source Database
(2) Migrating Existing Data
Target Database
Capture Changes on the Source Database
Apply change data
(3) Continuous data replication
DataTransmission Service(DTS)

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

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

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

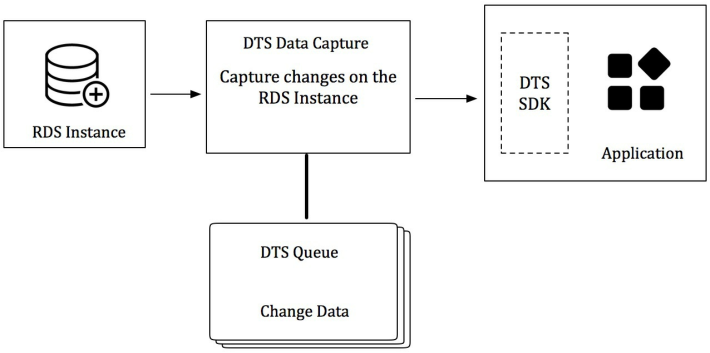
Full data migration takes a longer time. In this process, new data is continuously written into the source instance. To ensure data consistency, DTS starts the incremental data pulling module before full data migration. This module pulls the incremental data from the source instance and then parse, encapsulate,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

When full data migration is complete, DTS starts the incremental data playback module. The module retrieves the incremental data from the incremental data pulling module. After reverse parsing, filtering, and encapsulation, the data is synchronized to the destination instance. Eventually, data is synchronized between the source and destination instances in real time.

Data subscription process

Data subscription workflow shows how data subscription works.

Data subscription workflow



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

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

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

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

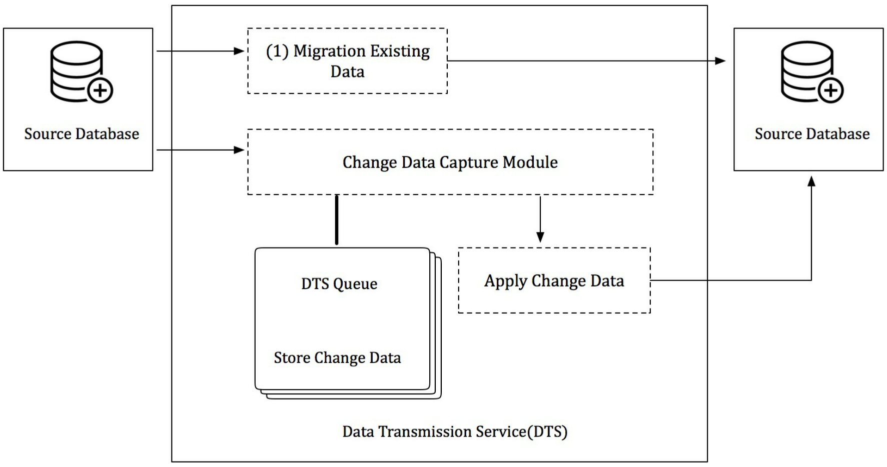
To ensure the high availability of the data pulling module, the DTS disaster recovery module restarts the data pulling module on a healthy service node once an exception is detected in the data pulling module.

The DTS server ensures the high availability of downstream consumption SDKs. If you start multiple consumption SDKs for the same subscription channel, the server pushes the incremental data to only one SDK at a time. If the consumption encounters an exception, the service end selects another consumption process from other healthy downstream nodes to push data to that consumption process. In this way, the high availability of downstream consumption processes can be guaranteed.

Real-time synchronization workflow

Real-time synchronization workflow shows how real-time synchronization works.

Real-time synchronization workflow


Source Database
(1) Migration Existing Data
Source Database
Change Data Capture Module
DTS Queue
Store Change Data
Apply Change Data
Data Transmission Service(DTS)

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

To create a synchronization channel, follow these steps:

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

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

- Data reading module

  The data reading module reads raw data from the source instance and makes the data locally persistent by parsing, filtering, and formatting it. The data reading module connects to the source instance using the database protocol and reads the incremental data from the source instance. For example, the data reading module connects to an RDS for MySQL instance using the binlog dump command.

- Data playback module

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

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

16.5. Features

16.5.1. Data migration

You can use Data Transmission Service (DTS) to migrate data between various types of data sources. This feature applies to the following scenarios: data migration to Alibaba Cloud, data migration between instances within Alibaba Cloud, and database splitting and scale-out. DTS supports data migration between homogeneous and heterogeneous data sources. DTS provides the following extract, transform, and load (ETL) features: object name mapping and data filtering.

Supported databases

Source database	Destination database	Migration type
	User-created MySQL database Version 5.1, 5.5, 5.6, 5.7, or 8.0	● Schema migration ● Full data migration ● Incremental data migration

Source database	Destination database	Migration type
• User-created MySQL database Version 5.1, 5.5, 5.6, 5.7, or 8.0 • ApsaraDB RDS for MySQL All versions	ApsaraDB RDS for MySQL All versions	• Schema migration • Full data migration • Incremental data migration
	PolarDB-X (formerly known as DRDS) All versions	• Full data migration • Incremental data migration
	User-created Oracle database (RAC or non-RAC architecture) Version 9i, 10g, 11g, 12c, 18c, or 19c	• Full data migration • Incremental data migration
	User-created Kafka database Versions 0.1 to 2.0	• Schema migration • Full data migration • Incremental data migration
User-created SQL Server database Version 2005, 2008, 2008 R2, 2012, 2014, 2016, or 2017 **Note** • DTS does not support SQL Server clusters or SQL Server Always On availability groups (AOAGs). • If the version of the source database is 2005, incremental data migration is not supported.	• User-created SQL Server database Version 2005, 2008, 2008 R2, 2012, 2014, 2016, or 2017 **Note** DTS does not support SQL Server clusters or SQL Server Always On availability groups (AOAGs). • ApsaraDB RDS for SQL Server Version 2008, 2008 R2, 2012, 2014, 2016, or 2017	• Schema migration • Full data migration • Incremental data migration

Source database	Destination database	Migration type
User-created Oracle database (RAC or non-RAC architecture) Version 9i, 10g, 11g, 12c, 18c, or 19c	User-created Oracle database (RAC or non-RAC architecture) Version 9i, 10g, 11g, 12c, 18c, or 19c	● Schema migration ● Full data migration ● Incremental data migration
	PolarDB Version 9.3, 9.6, 10, or 11	● Schema migration ● Full data migration ● Incremental data migration
	User-created MySQL database Version 5.1, 5.5, 5.6, 5.7, or 8.0	● Schema migration ● Full data migration ● Incremental data migration
	ApsaraDB RDS for MySQL All versions	● Schema migration ● Full data migration ● Incremental data migration
	PolarDB-X All versions	● Full data migration ● Incremental data migration
	AnalyticDB for MySQL Version 2.0 or 3.0	● Schema migration ● Full data migration ● Incremental data migration

Source database	Destination database	Migration type
- User-created PostgreSQL database Version 9.4, 9.5, 9.6, or 10.x - ApsaraDB RDS for PostgreSQL Version 9.4 or 10	- User-created PostgreSQL database Version 9.4, 9.5, 9.6, or 10.x - ApsaraDB RDS for PostgreSQL Version 9.4 or 10	- Schema migration - Full data migration - Incremental data migration
PolarDB Version 9.3, 9.6, 10, or 11	User-created Kafka database Versions 0.1 to 2.0	Incremental data migration
	User-created Oracle database (RAC or non-RAC architecture) Version 9i, 10g, 11g, 12c, 18c, or 19c	- Full data migration - Incremental data migration
	PolarDB Version 9.3, 9.6, 10, or 11	- Schema migration - Full data migration - Incremental data migration
User-created Redis database Version 2.8, 3.0, 3.2, 4.0, or 5.0	User-created Redis database Version 2.8, 3.0, 3.2, 4.0, or 5.0	- Full data migration - Incremental data migration
User-created MongoDB database Version 3.0, 3.2, 3.4, 3.6, 4.0 or 4.2	User-created MongoDB database Version 3.0, 3.2, 3.4, 3.6, 4.0 or 4.2	- Full data migration - Incremental data migration

Online migration

DTS uses online migration. You must configure the source instance, destination instance, and objects to be migrated. DTS automatically completes the entire data migration process. You can select all of the supported migration types to minimize the impact of online data migration on your services. However, you must ensure that DTS servers can connect to both the source and destination instances.

Data migration types

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

- Schema migration: DTS migrates schemas from the source instance to the destination instance.
- Full data migration: DTS migrates historical data from the source instance to the destination instance.
- Incremental data migration: DTS synchronizes incremental data that is generated during data migration from the source instance to the destination instance. You can select schema migration, full data migration, and incremental migration to ensure service continuity.

ETL features

Data migration supports the following ETL features:

- Object name mapping: You can change the names of the columns, tables, and databases that are migrated to the destination database.
- Data filtering: You can use SQL conditions to filter the required data in a specific table. For example, you can specify a time range to migrate only the latest data.

Alerts

If an error occurs during data migration, DTS immediately sends an SMS alert to the task owner. This allows the owner to handle the error at the earliest opportunity.

Migration task

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

The table Statuses of a migration task describes the statuses of a migration task.

Statuses of a migration task

Status	Description	Available operation
Not Started	The migration task is configured but no precheck is performed.	• Run a precheck • Delete the migration task
Prechecking	A precheck is being performed but the migration task is not started.	Delete the migration task

Status	Description	Available operation
Passed	The migration task has passed the precheck but has not been started.	• Start the migration task • Delete the migration task
Migrating	The task is migrating data.	• Pause the migration task • Stop the migration task • Delete the migration task
Migration Failed	An error occurred during data migration. You can identify the point of failure based on the progress of the migration task.	Delete the migration task
Paused	The migration task is paused.	• Start the migration task • Delete the migration task
Completed	The migration task is completed, or you have stopped data migration by clicking **End**.	Delete the migration task

16.5.2. Data synchronization

You can use Data Transmission Service (DTS) to synchronize data between two data sources. This feature applies to various scenarios, such as data backup, disaster recovery, active geo-redundancy, cross-border data synchronization, load balancing, cloud BI systems, and real-time data warehousing.

Supported databases

Source database	Destination database	Initial synchronization type	Synchronization topology
- User-created MySQL database 5.1, 5.5, 5.6, and 5.7 - RDS MySQL 5.6 and 5.7	User-created MySQL database 5.1, 5.5, 5.6, and 5.7	Initial schema synchronization Initial full data synchronization	One-way synchronization Two-way synchronization
	RDS MySQL 5.6 and 5.7	Initial schema synchronization Initial full data synchronization	One-way synchronization Two-way synchronization
	AnalyticDB for MySQL 2.0 and 3.0	Initial schema synchronization Initial full data synchronization	One-way synchronization
	AnalyticDB for PostgreSQL 4.3 and 6.0	Initial schema synchronization Initial full data synchronization	One-way synchronization
	Datahub	Initial schema synchronization	One-way synchronization
	MaxCompute	Initial schema synchronization Initial full data synchronization	One-way synchronization
Cloud Native Distributed Database PolarDB-X (formerly	Cloud Native Distributed Database PolarDB-X	Initial full data synchronization	One-way synchronization
	Datahub	Initial schema synchronization	One-way synchronization

known as DRDS) Source database	Destination database	Initial synchronization type	Synchronization topology
	AnalyticDB for MySQL 2.0 and 3.0	Initial schema synchronization Initial full data synchronization	One-way synchronization

Objects to be synchronized

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

Synchronization tasks

A synchronization task is a basic unit of data synchronization. To synchronize data between two instances, you must create a synchronization task in the DTS console.

The following table describes the statuses of a synchronization task when you create and run the task.

Task statuses

Task status	Description	Available operation
Prechecking	A precheck is being performed before the synchronization task is started.	• View the configurations of the synchronization task • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts

Task status	Description	Available operation
Precheck Failed	The synchronization task has failed the precheck.	• Run a precheck • View the configurations of the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts
Not Started	The synchronization task has passed the precheck but has not been started.	• Run a precheck • Start the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts
Performing Initial Synchronization	Initial synchronization is being performed.	• View the configurations of the synchronization task • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts

Task status	Description	Available operation
Initial Synchronization Failed	The task has failed during initial synchronization.	• View the configurations of the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts
Synchronizing	The task is synchronizing data.	• View the configurations of the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Pause the synchronization task • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts
Synchronization Failed	An error occurred during synchronization.	• View the configurations of the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Start the synchronization task • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts

Task status	Description	Available operation
Paused	The synchronization task is paused.	• View the configurations of the synchronization task • Modify the objects to be synchronized • Modify the synchronization speed • Start the synchronization task • Delete the synchronization task • Replicate the configurations of the synchronization task • Configure monitoring and alerts

Advanced features

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

- Add or remove the objects to be synchronized

  You can add or remove the required objects when a task is synchronizing data.

- View and analyze the synchronization performance

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

- Monitor synchronization tasks

  DTS allows you to monitor the status of synchronization tasks. If the threshold for synchronization delay is reached, you will receive an alert. You can set the alert threshold based on the sensitivity of your businesses to synchronization delays.

16.5.3. Change tracking

You can use Data Transmission Service (DTS) to track data changes from databases in real time. This feature applies to the following scenarios: cache updates, business decoupling, asynchronous data processing, synchronization of heterogeneous data, and synchronization of extract, transform, and load (ETL) operations.

Supported databases

- User-created MySQL database or ApsaraDB RDS for MySQL instance
- Cloud Native Distributed Database PolarDB-X (formerly known as DRDS)
- User-created Oracle database

Objects for change tracking

The objects for change tracking include tables and databases. You can specify one or more tables from which you want to track data changes.

In change tracking, data changes include data manipulation language (DML) operations and data definition language (DDL) operations. When you configure a change tracking task, you can select the operation type.

Change tracking channel

A change tracking channel is the basic unit of change tracking and data consumption. To track data changes from an RDS instance, you must create a change tracking channel for the RDS instance in the DTS console. The change tracking channel pulls incremental data from the RDS instance in real time and locally stores the incremental data. You can use the DTS SDK to consume the incremental data from the change tracking channel. You can also create, manage, or delete change tracking channels in the DTS console.

A change tracking channel can be consumed by only one downstream SDK client. To track data changes from an RDS instance by using multiple downstream SDK clients, you must create an equivalent number of change tracking channels. The channels pull incremental data from the same RDS instance.

The following table describes the statuses of a change tracking channel.

Channel statuses

Channel status	Description	Available operation
Prechecking	The configuration of the change tracking channel is complete and a precheck is being performed.	Delete the change tracking channel
Not Started	The change tracking channel has passed the precheck but has not been started.	• Start the change tracking channel • Delete the change tracking channel
Performing Initial Change Tracking	The initial change tracking is in progress. This process takes about 1 minute.	Delete the change tracking channel
Normal	Incremental data is being pulled from the source RDS instance.	• View the demo code • View the tracked data • Delete the change tracking channel
Error	An error occurs when the change tracking channel pulls incremental data from the source RDS instance.	• View the demo code • Delete the change tracking channel

Advanced features

You can use the following advanced features that are provided for change tracking:

- Add or remove the objects for change tracking

  You can add or remove the required objects when a change tracking task is running.

- View the tracked data

  You can view the data that is tracked from the change tracking channel in the DTS console.

- Modify consumption checkpoints

  You can modify consumption checkpoints.

- Monitor change tracking channels

  DTS allows you to monitor the status of change tracking channels. If the threshold for consumption delay is reached, you will receive an alert. You can set the alert threshold based on the sensitivity of your businesses to consumption delays.

16.6. Scenarios

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

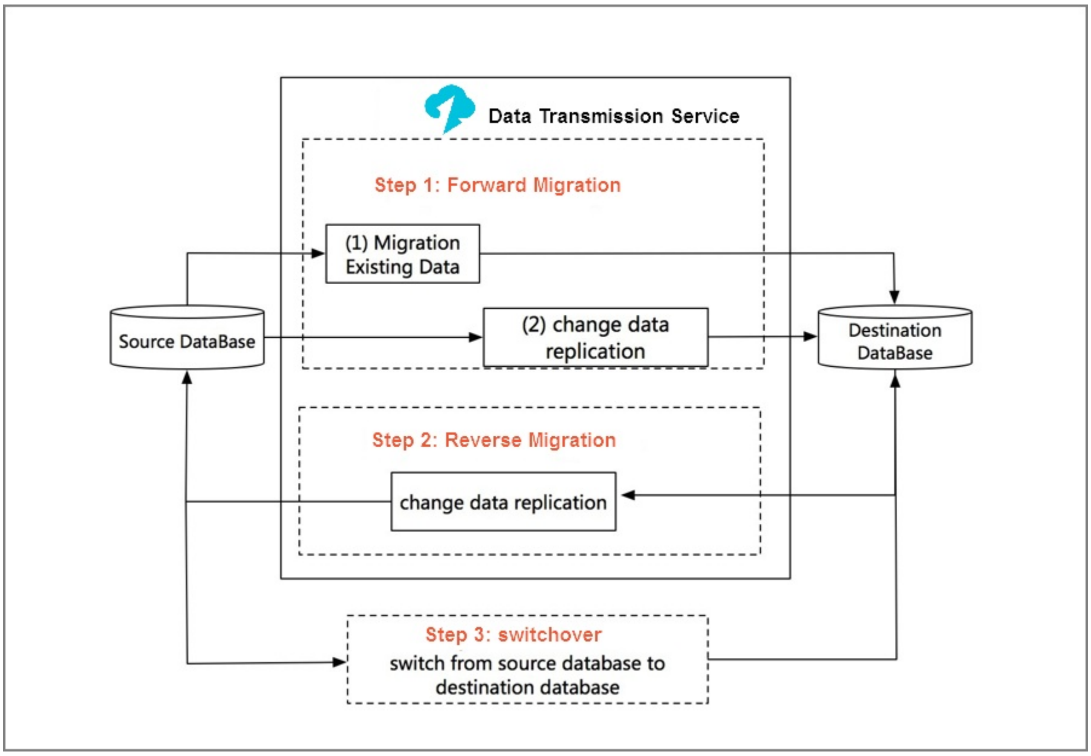
Migration with service downtime reduced to minutes

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

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

Interruption-free migration shows how interruption-free migration works.

Interruption-free migration


Data Transmission Service
Step 1: Forward Migration
(1) Migration Existing Data
Source DataBase
(2) change data replication
Destination DataBase
Step 2: Reverse Migration
change data replication
Step 3: switchover
switch from source database to destination database

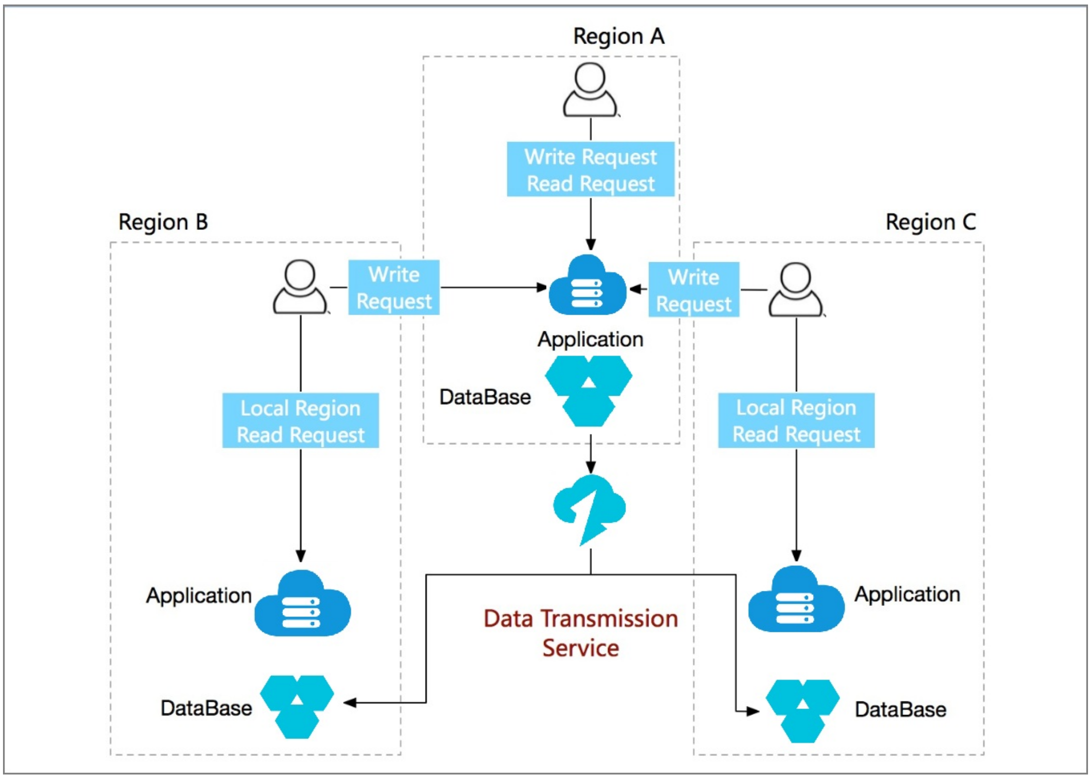
The interruption-free data migration process involves schema migration, full data migration, and incremental data migration. In the incremental data migration phase, data is synchronized between the source and destination instances in real time. You can validate the service in the destination database.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

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

Accelerated access to global services to empower cross-border businesses

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

Reduced cross-region access latency



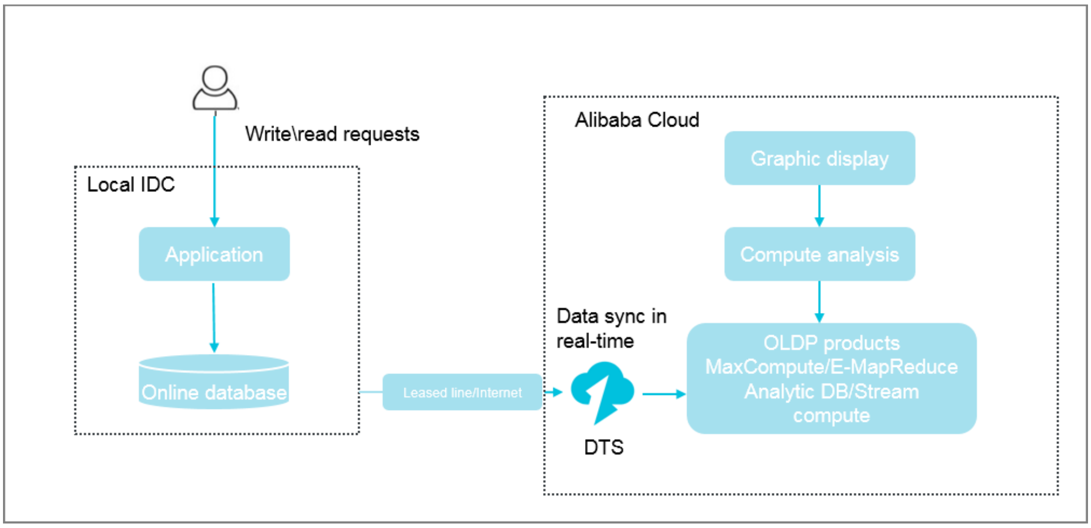
This architecture consists of one center and multiple units. Write requests of users in all regions are routed back to the center. DTS synchronizes data in the center to all units. Read requests of users in different regions can be routed to nearby units to avoid remote access and reduce access latency. In this way, access to global services is accelerated.

Custom cloud BI system built with more efficiency

User-created business intelligence (BI) systems cannot meet the increasing demand for real-time performance and are difficult to manipulate. With the Apsara Stack BI architecture, you can quickly build a BI system without affecting the current architecture. For this reason, more and more users choose to build BI systems that meet their own business requirements on Apsara Stack.

DTS can help you synchronize data stored in local databases to an Apsara Stack BI system (such as MaxCompute or StreamCompute) in real time. You can then perform subsequent data analysis with various compute engines while viewing the computing results in real time with a visualization tool. You can also synchronize those results back to the local IDC with a migration tool. Cloud BI architecture shows the implementation architecture.

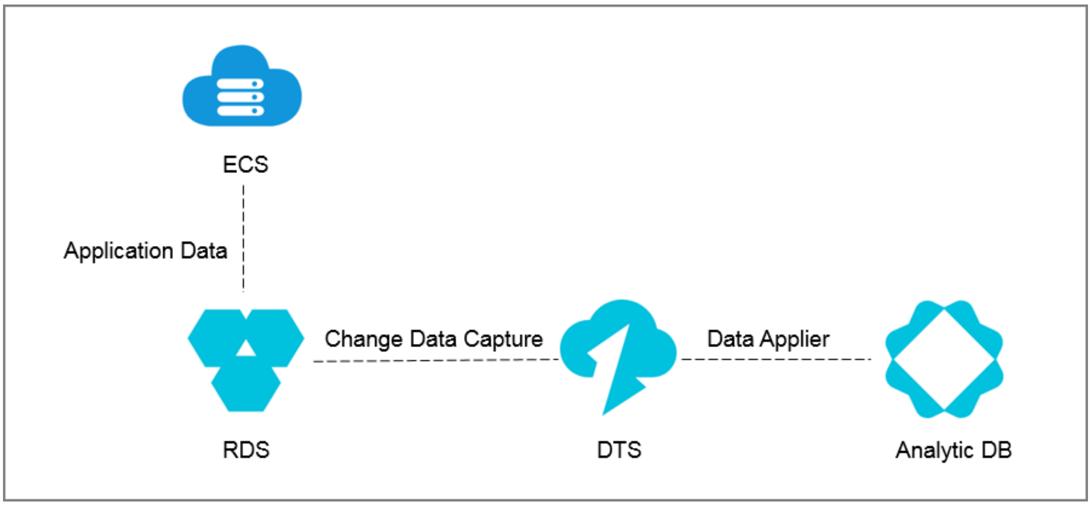
Cloud BI architecture



Real-time data analysis to rapidly respond to market conditions

Data analysis is essential in improving enterprise insights and user experience. Real-time data analysis enables enterprises to adjust marketing strategies more quickly and flexibly so that they can adapt to the rapidly changing marketing conditions and demands for higher user experience. To implement real-time data analysis without affecting online services, service data needs to be synchronized to the analysis system in real time. For this reason, acquiring service data in real time becomes essential. In DTS, the data subscription feature can help you acquire real-time incremental data without affecting online services and synchronize the data to the analysis system using the SDK for real-time data analysis, as shown in Real-time data analysis.

Real-time data analysis

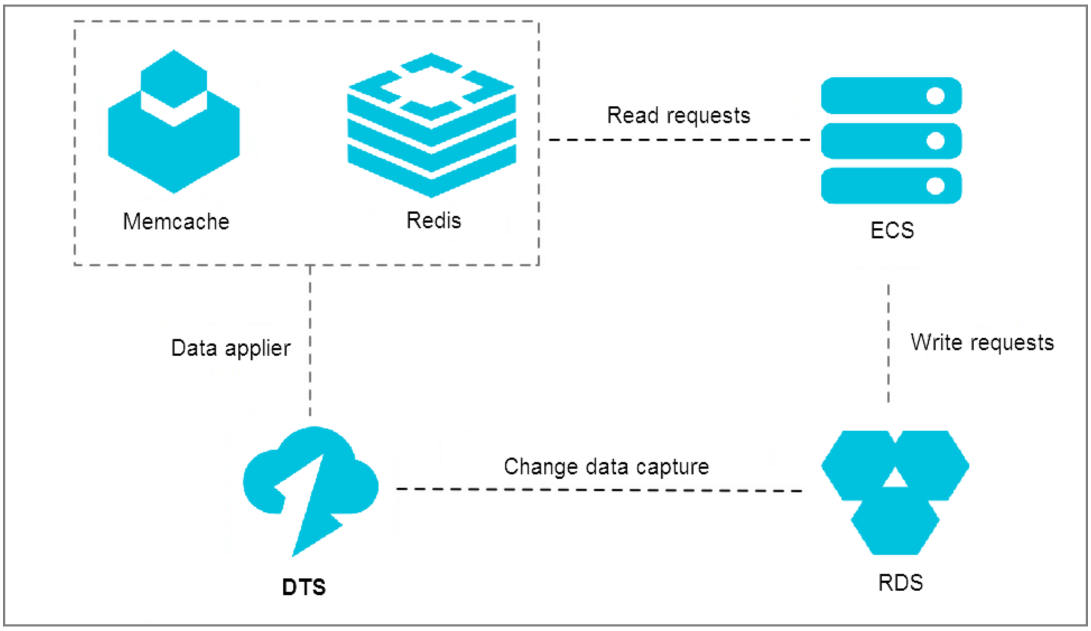


Lightweight cache update policies to make core services more simple and reliable

To accelerate service access and improve concurrent read performance, many enterprises introduce the caching layer to the service architecture. In this architecture, all the read requests are routed to the caching layer, and the memory reading mechanism greatly improves read performance. Cached data cannot persist. If caching ends abnormally, data in the cache memory is lost. To ensure data integrity, the updated service data is kept in a persistent storage medium, such as a database.

In this condition, the service data is inconsistent between the cache and the persistent databases. The data subscription feature can help asynchronously subscribe to the incremental data in those databases and update the cached data to implement lightweight cache update policies. Cache update policies shows the architecture of these policies.

Cache update policies



Cache update policies offer the following benefits:

- Quick update with low latency

  Cache invalidation is an asynchronous process, and the service returns data directly after the database update is complete. For this reason, you do not need to consider the cache invalidation process, and the entire update path is short with low latency.

- Simple and reliable applications

  The complex doublewrite logic is not required for the application. You only need to start the asynchronous thread to monitor the incremental data and update the cached data.

- Application updates without extra performance consumption

  Because data subscription acquires incremental data by parsing incremental logs in the database, the acquisition process does not damage the performance of services and databases.

Asynchronous service decoupling to make core services simpler and more reliable

Data subscription optimizes intensive coupling to asynchronous coupling by using real-time message notifications. This makes the core service logic simpler and more reliable. This application has been widely implemented in Alibaba. Tens of thousands of downstream services in the Taobao ordering system acquire real-time data updates through data subscription to trigger the business logic every day.

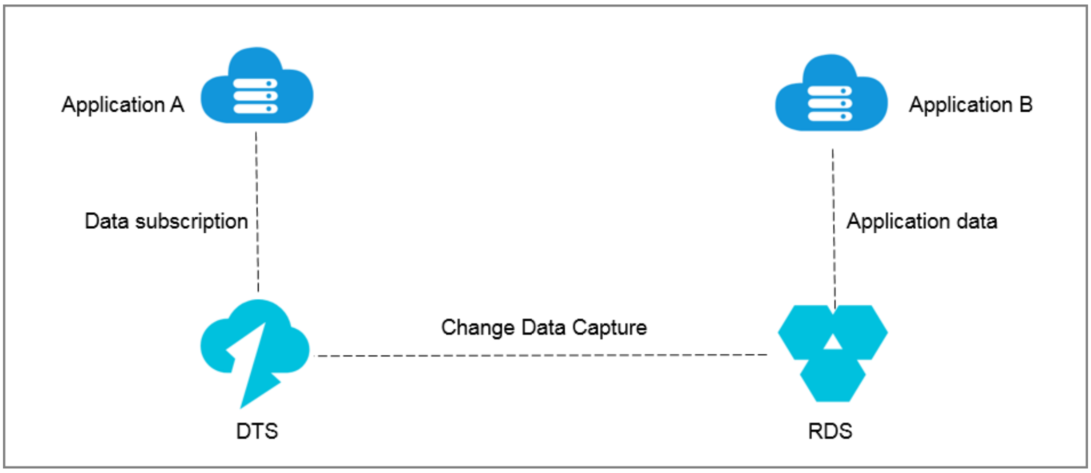
The following uses a simple example to describe the benefits of implementing data subscription in this scenario.

The e-commerce industry involves multiple services including the order management system, inventory management, and the shipping of goods. An ordering process with all of those services included is as follows: After a user places an order, downstream services including seller inventory notification and goods shipping are modified. When all logic modifications are complete, the order result is returned to the user. However, this ordering logic has the following issues:

- The lengthy ordering process results in poor user experience.
- The system is unstable and any downstream fault directly affects the availability of the ordering system.

To improve user experience of core applications, you can decouple the core applications and the dependent downstream services so that they can work asynchronously. In this way, the core applications become more stable and reliable. Asynchronous service decoupling shows how to adjust the logic.

Asynchronous service decoupling



The ordering system returns the order result directly after order placement. With DTS, the underlying layer acquires the updated data from the ordering system in real time. Then, the downstream service subscribes to the modified data using the SDK and triggers the service logic such as inventory and shipping. In this way, the ordering system becomes simpler and more reliable.

Horizontal scaling to improve read performance and quickly adapt to business growth

A single RDS instance may not be able to support a large number of read requests, which may affect the main service process. To elastically improve the read performance and reduce database workload, you can create read-only instances using the real-time synchronization feature of DTS. These read-only instances take on large amounts of the database reading workload and expand the throughput of applications.

16.7. Concepts

Precheck

Precheck is an essential stage before a migration task starts. It mainly checks the prerequisites that may affect a successful migration, such as the connectivity of the source and destination instances and the permissions of the migration accounts. If the precheck fails, you can fix the problems as instructed and run the precheck again.

Schema migration

Schema migration is a type of migration tasks. In database migration, it refers to migrating the schema syntax, including tables, views, triggers, stored procedures, stored functions, and synonyms. For migration between heterogeneous databases, data types are mapped during schema migration, and the schema syntax is adjusted according to the schema syntax of the source and destination instances.

Full data migration

Full data migration is a type of migration task. It refers to migrating all the data except the schema syntax from the source instance to the destination instance. If you select Full Data Migration only and leave Schema Migration unselected, new data generated in the source instance will not be migrated to the destination instance.

Incremental data migration

Incremental data migration is a type of migration tasks. It refers to synchronizing the new data written to the source instance to the destination instance during the migration. When creating a migration task, if you select both Full Data Migration and Incremental Data Migration, DTS will first perform a static snapshot on the source instance, migrate the snapshot data to the destination instance, and then synchronize the new data from the source instance to the destination instance during the migration. Incremental data migration is a process of synchronizing data between the source and destination instances in real time. This process does not automatically end. If you want to stop migrating data, you must manually disable the task in the console.

Initial synchronization

Initial synchronization refers to synchronizing the historical data of the objects to be synchronized to the destination instance before synchronizing the incremental data through the synchronization channel.

Initial synchronization includes initial schema synchronization and initial full data synchronization. Initial schema synchronization refers to synchronizing the required schema syntax in the initial stage. Initial full data synchronization refers to synchronizing the data of the objects for the first time.

Synchronization performance

Synchronization performance is measured based on the number of records that are synchronized to the destination instance per second. The measurement unit is records per second (RPS).

Synchronization delay

Synchronization delay refers to the duration between the timestamp when the latest data in the destination instance is starting to be synchronized from the source instance and the current timestamp of the source instance. It reflects the time difference between the data in the source and destination instances. If the synchronization delay is zero, data in the source instance is in sync with that in the destination instance.

Subscription channel ID

The subscription channel ID is a unique identifier of a subscription channel. After you purchase a subscription channel, DTS automatically generates a subscription channel ID. To consume the incremental data using the SDK, you must configure a correct subscription channel ID. You can find the ID that corresponds to each subscription channel in the subscription list of the DTS console.

Data update

In DTS, you can update data or its schema. A data update only modifies the data. The schema syntax is not changed. Operations including INSERT, UPDATE, and DELETE fall into this category.

Schema update

In DTS, you can update data or its schema. Schema update modifies the schema syntax. Operations including CREATE TABLE, ALTER TABLE, and DROP VIEW fall into this category . You can choose whether to subscribe to schema update when you create a subscription channel.

Data range

Data range refers to the range of timestamps of incremental data stored in the subscription channel. The timestamp of a piece of incremental data is the time when the incremental data is applied and written to the transaction log in the database instance. By default, only data generated on the most recent day is retained in the subscription channel. DTS regularly cleans the expired incremental data and updates the data range of the subscription channel.

Consumption checkpoint

The consumption checkpoint is the timestamp of the latest consumed incremental data that is subscribed using the downstream SDK. The SDK sends an ACK message to DTS for every piece of data that is consumed. The server updates and saves the consumption checkpoint corresponding to the SDK. When the SDK encounters an exception, the server restarts and automatically pushes the data at the latest consumption checkpoint.

17.Data Management (DMS)

17.1. What is DMS?

Data Management (DMS) is an integrated database solution that includes data, schema, and server management, access control, BI insights, data trend analysis, data tracking, and performance optimization.

17.2. Benefits

Extensive options for data sources

- ApsaraDB RDS for MySQL
- ApsaraDB RDS for SQL Server
- PolarDB and ApsaraDB RDS for PostgreSQL

Visualized data analysis

Visualized analysis of the numbers of read, inserted, deleted, and updated rows in business tables

Efficient R&D

- Schema comparison
- Smart SQL completion
- Convenient reuse of custom SQL statements and SQL templates
- Automatic restoration of work environments
- Export of dictionary files

17.3. Architecture

DMS consists of the business layer, scheduling layer, and connection layer. DMS processes real-time data access and schedules data-related background tasks for relational databases.

Business layer

- The business layer supports online GUI-based database operations and can be scaled to improve the general service capabilities of DMS.
- DMS supports stateless failover to ensure 24/7 availability.

Scheduling layer

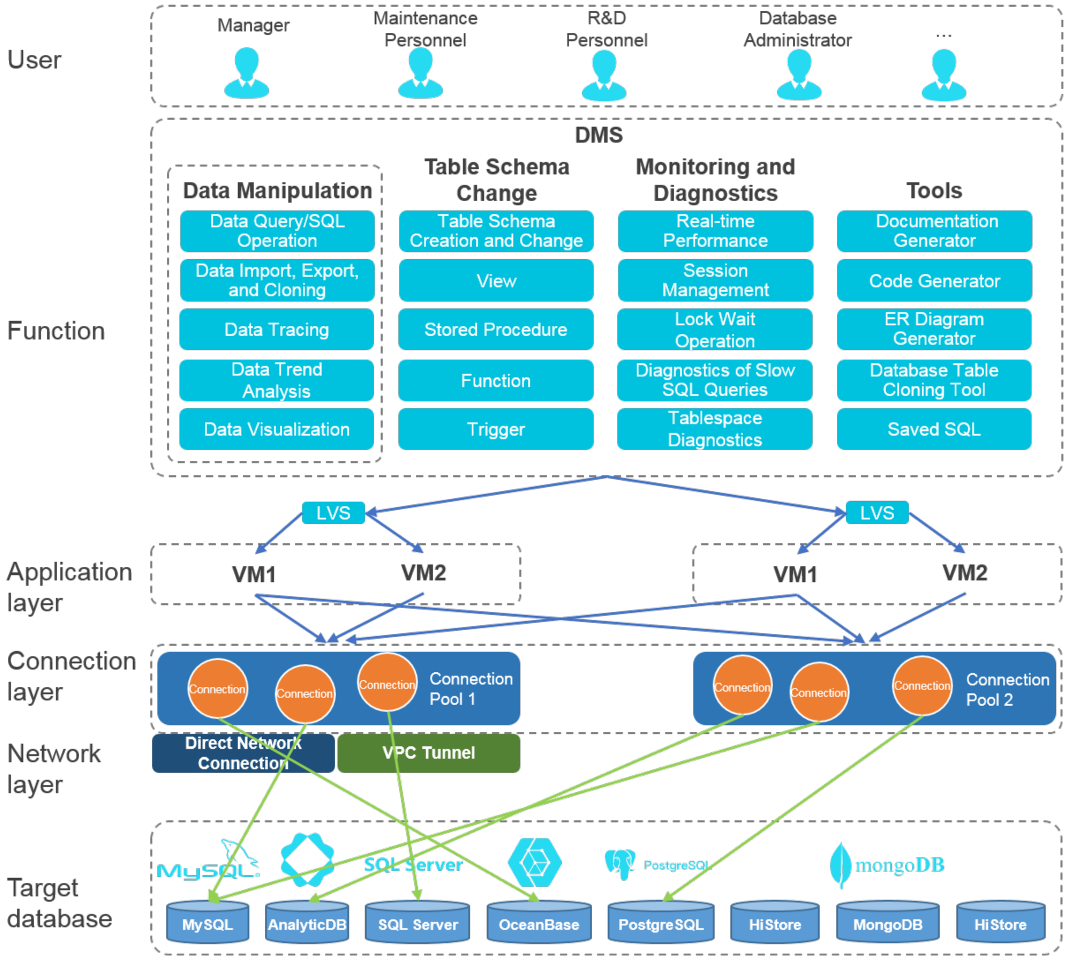
- The scheduling layer allows you to import and export tables and compare schemas. This layer schedules tasks by using the thread pool in the real-time scheduling or background periodic scheduling mode.
- Real-time scheduling allows you to schedule and run tasks on the frontend. After you submit a task, DMS automatically runs the task in the background. After the task is completed, you can download or view the execution result.
- Background periodic scheduling allows you to periodically obtain specified data such as data trends. DMS collects business data in the background for your reference and analysis based on scheduled tasks.

Connection layer

The connection layer is the core component for accessing data in DMS. It has the following characteristics:

- Processes requests from MySQL, SQL Server, and PostgreSQL databases.
- Supports session isolation and persistence. SQL windows opened in DMS are isolated from each other and the sessions in each SQL window are persistent to simulate the client experience.
- Controls the number of instance sessions to prevent a large number of connections from being established to a single instance.
- Provides different connection release policies for different features. This improves user experience and reduces the number of connections to the databases.

DMS system architecture



17.4. Features

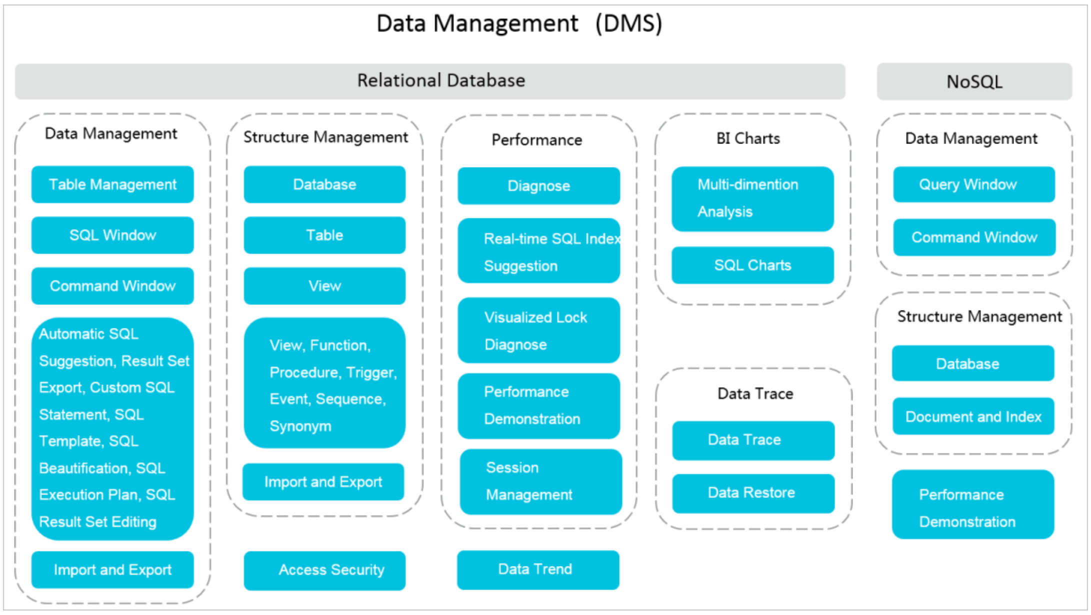
Relational database management

- Data management: includes functions such as SQL windows, SQL command lines, table data, intelligent SQL prompts, SQL formatting, custom SQL statements, SQL templates, SQL execution plans, and import and export operations.
- Structure management: includes functions such as table structure comparison, and management of objects (databases, tables, views, functions, storage procedures, triggers,events, series, and

synonyms).

Feature diagram

Feature diagram



17.5. Scenarios

17.5.1. Convenient data operations

Pain point

You need a lightweight product that features full functionality to create SQL statements, save frequently used SQL statements, and use these statements in your business.

Solution

- You can open a table in DMS and perform operations on table data as you would in an Excel worksheet. You can add, delete, change, query, and make statistical analysis of table data without understanding SQL.
- You can customize SQL statements, save frequently used SQL statements, and apply these SQL statements to databases or instances.

17.5.2. Prohibiting data export

Paint point

When cooperating with a partner, an enterprise manages data and its partner develops functions. The partner needs to have access to view the enterprise's data but cannot have the ability to export data to ensure data security.

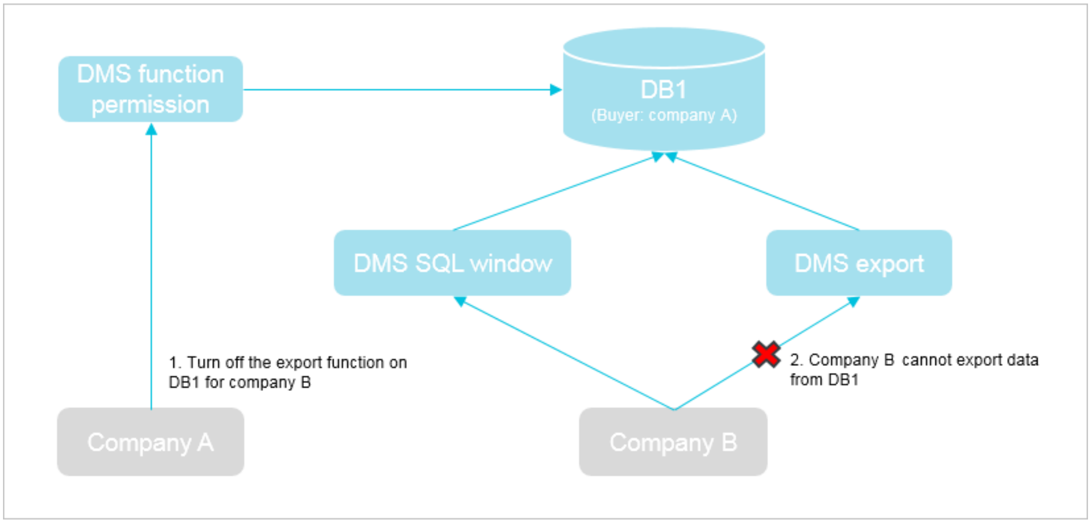
Solution

Enterprise users can log on to the DMS console to grant their partners access permissions on the corresponding database instances, disabling data exporting to protect their data.

Partners are permitted only to query and view data, eliminating the risk of data leakage.

DMS – Function-based authorization shows how to use the function-based authorization feature to prohibit partners from exporting data.

DMS – Function-based authorization


DMS function permission
DB1
(Buyer: company A)
DMS SQL window
DMS export
1. Turn off the export function on DB1 for company B
2. Company B cannot export data from DB1
Company A
Company B

17.5.3. SQL statement reuse

Pain point

SQL statements are used when you access a database. While simple queries are easy to use, rewriting SQL queries for complex data analysis or SQL queries that contain service logic is time-consuming. Even if you save these SQL queries to files, you have to maintain the files and you cannot use them without access to the files.

Solution

You can use the **My SQL** function provided to save frequently used SQL statements to DMS. As the SQL statements are not saved locally, they can be reused in any databases or instances.

17.6. Limits

Relational databases

Support for relational databases

Module	Function	MySQL	PostgreSQL
	Table data management	√	√
	SQL windows	√	√
	SQL command lines	√	√

Module	Function	MySQL	PostgreSQL
Data management	SQL templates	√	
	SQL formatting	√	√
	Custom SQL statements	√	
	Intelligent SQL prompts	√	
	SQL execution plans	√	√
Structure management	Database management	√	√
	Table management	√	√
	Management of objects such as indexes, views, stored processes, functions, triggers, and events	√	√
	Entity relationship diagram display	√	
	Data dictionaries	√	
Import and export	Basic import and export functions	√	√
	Export of large volumes of data	√	√

18.Server Load Balancer (SLB)

18.1. What is SLB?

This topic provides an overview of Server Load Balancer (SLB). SLB distributes inbound network traffic across multiple Elastic Compute Service (ECS) instances that act as backend servers based on forwarding rules. You can use SLB to improve the responsiveness and availability of your applications.

Overview

After you add ECS instances that reside in the same region to an SLB instance, SLB uses virtual IP addresses (VIPs) to virtualize these ECS instances into backend servers in a high-performance server pool that ensures high availability. Client requests are distributed to the ECS instances based on forwarding rules.

SLB checks the health status of the ECS instances and automatically removes unhealthy ones from the server pool to eliminate single points of failure (SPOFs). This enhances the resilience of your applications.

Components

SLB consists of three components:

- SLB instances

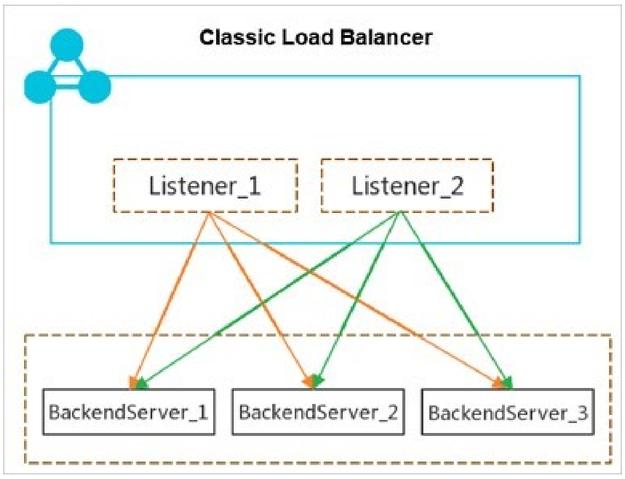
  An SLB instance is a key load-balancing component in SLB. It receives traffic and distributes traffic to backend servers. To get started with SLB, you must create an SLB instance and add at least one listener and two ECS instances to the SLB instance.

- Listeners

  A listener checks for connection requests from clients, forwards requests to backend servers, and performs health checks on backend servers.

- Backend servers

  ECS instances are used as backend servers in SLB to receive and process distributed requests. ECS instances can be added to the default server group of an SLB instance. You can also add multiple ECS servers to VServer groups or primary/secondary server groups after the corresponding groups are created.



Benefits

- High availability

  SLB features full redundancy that avoids SPOFs and supports zone-disaster recovery. You can use SLB with Apsara Stack DNS to achieve geo-disaster recovery with an availability of up to 99.95%.

  SLB can be scaled based on network traffic to protect your services from outages caused by fluctuating traffic flows.

- Strong scalability

  You can increase or decrease the number of backend servers to adjust the load balancing capacity for your applications.

- Low costs

  SLB can save 60% of load balancing costs compared with using traditional hardware solutions.

- Outstanding security

  You can use SLB with Apsara Stack Security to defend your applications against 5 Gbit/s distributed denial of service (DDoS) attacks.

- High concurrency

  An SLB cluster supports hundreds of millions of concurrent connections, and a single SLB instance supports tens of millions of concurrent connections.

18.2. High availability

This topic describes the high-availability architecture of CLB. You can use CLB in concert with DNS to implement geo-disaster recovery. CLB is designed to offer a multi-zone service availability of 99.99% and a single-zone service availability of 99.90%.

High availability of the CLB architecture

CLB instances are deployed in clusters to synchronize sessions and protect backend servers from SPOFs, improving redundancy and ensuring service stability. Layer-4 CLB uses the open-source Linux Virtual Server (LVS) and Keepalived software to balance loads, whereas Layer-7 CLB uses Tengine. Tengine, a web server project launched by Taobao, is based on NGINX and adds advanced features dedicated for high-traffic websites.

Requests from the Internet reach an LVS cluster along Equal-Cost Multi Path (ECMP) routes. In the LVS cluster, each machine uses multicast packets to synchronize sessions with the other machines. At the same time, the LVS cluster performs health checks on the Tengine cluster and removes unhealthy machines from the Tengine cluster to ensure the availability of Layer-7 CLB.

Best practice:

You can use session synchronization to prevent persistent connections from being affected by server failures within a cluster. However, for short-lived connections or if the session synchronization rule is not triggered by the connection (the three-way handshake is not completed), server failures in the cluster may still affect user requests. To prevent session interruptions caused by server failures within the cluster, you can add a retry mechanism to the service logic to reduce the impact on user access.

The high-availability solution with one CLB instance

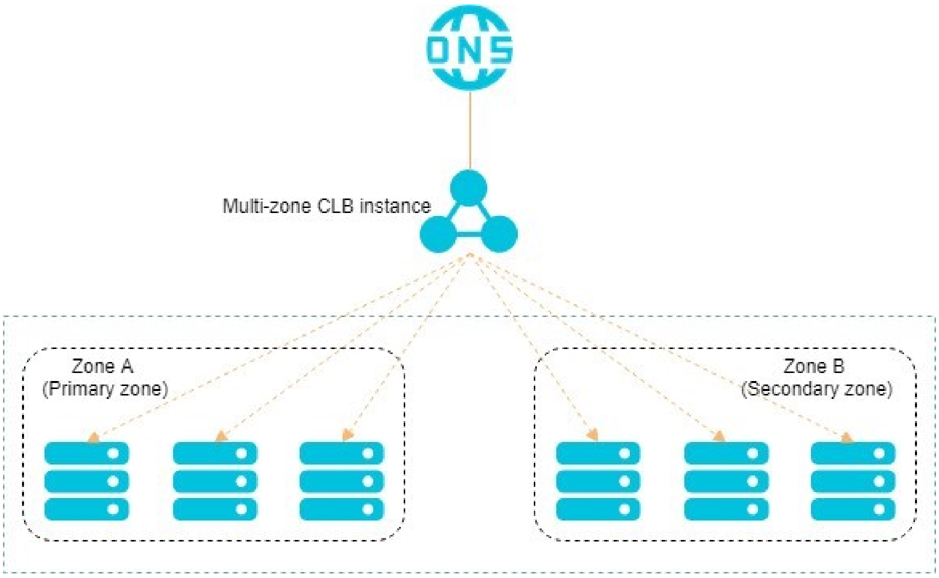
To provide more stable and reliable load balancing services, you can deploy CLB instances across multiple zones in most regions to achieve cross-data-center disaster recovery. Specifically, you can deploy a CLB instance in two zones within the same region whereby one zone acts as the primary zone and the other acts as the secondary zone. If the primary zone suffers an outage, a failover is triggered to redirect requests to the servers in the secondary zone within approximately 30 seconds. After the primary zone is restored, traffic will be automatically switched back to the servers in the primary zone.

> ⓘ **Note** Zone-disaster recovery is implemented between the primary and secondary zones. CLB implements failovers only when the whole CLB cluster within the primary zone is unavailable or fails, for example, due to power outage or optical cable failures. A failover will not be triggered when a single backend server fails.

Best practice:

1. We recommend that you create CLB instances in regions that support primary/secondary deployment for zone-disaster recovery.
2. You can choose the primary zone for your CLB instance based on the distribution of ECS instances. That is, select the zone where most of the ECS instances are located as the primary zone for minimized latency.

   However, we recommend that you do not deploy all ECS instances in the primary zone. When you develop a failover solution, you must deploy several ECS instances in the secondary zone to ensure that requests can still be distributed to backend servers in the secondary zone for processing when the primary zone experiences a downtime.

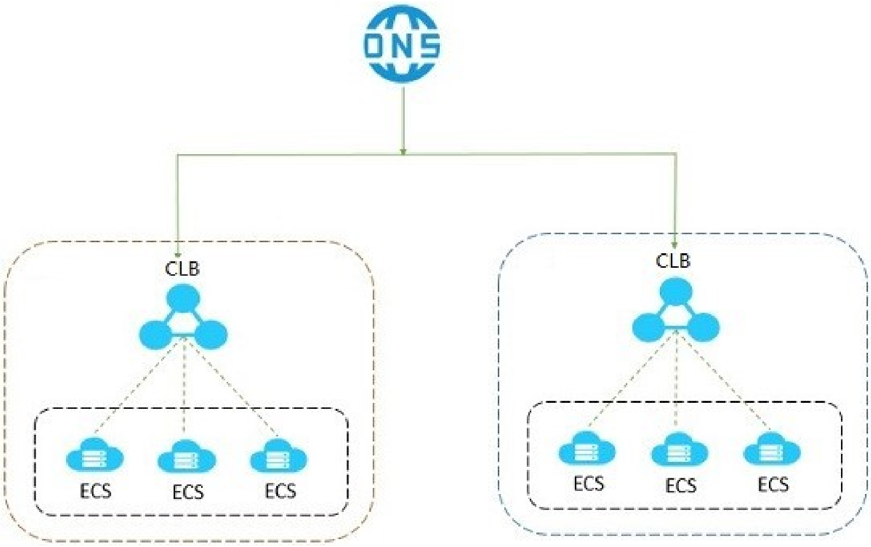


The high-availability solution with multiple CLB instances

In the context of one CLB instance, traffic distribution for your applications can still be compromised by network attacks or invalid CLB configurations, because the failover between the primary zone and the secondary zone is not triggered. As a result, the load-balancing performance is impacted. To avoid this situation, you can create multiple CLB instances to form a global load-balancing solution and achieve cross-region backup and disaster recovery. Also, you can use the instances with DNS to schedule requests so as to ensure service continuity.

Best practice:

You can deploy CLB instances and ECS instances in multiple zones within the same region or across different regions, and then use DNS to schedule requests.



The high-availability solution with backend ECS instances

With health check enabled, CLB verifies the availability of backend ECS instances (or backend servers), and thus improves the availability of frontend services by minimizing downtime that is caused by health issues of ECS instances.

After you enable the health check feature, when an ECS instance is detected unhealthy, CLB distributes new requests to other healthy ECS instances. CLB will only send requests to this backend ECS instance when it is restored and considered healthy. For more information, see *Health check overview* in the *CLB User Guide*.

Best practice:

Make sure health check is enabled and properly configured. For more information, see *Configure health check* in the *CLB User Guide*.

18.3. Architecture

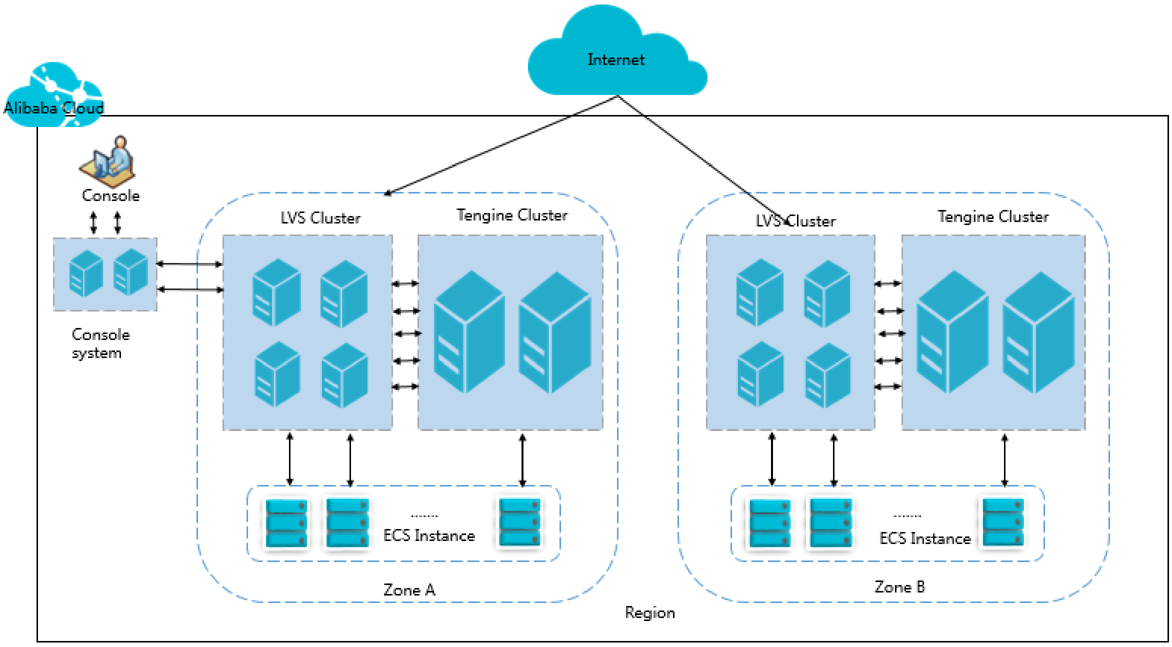
This topic describes the SLB architecture. SLB instances are deployed in clusters to synchronize sessions and protect backend servers from SPOFs, improving redundancy and ensuring service stability. SLB supports Layer-4 load balancing of Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) traffic, and Layer-7 load balancing of HTTP and HTTPS traffic.

SLB forwards client requests to backend servers by using SLB clusters and receives responses from backend servers over internal networks.

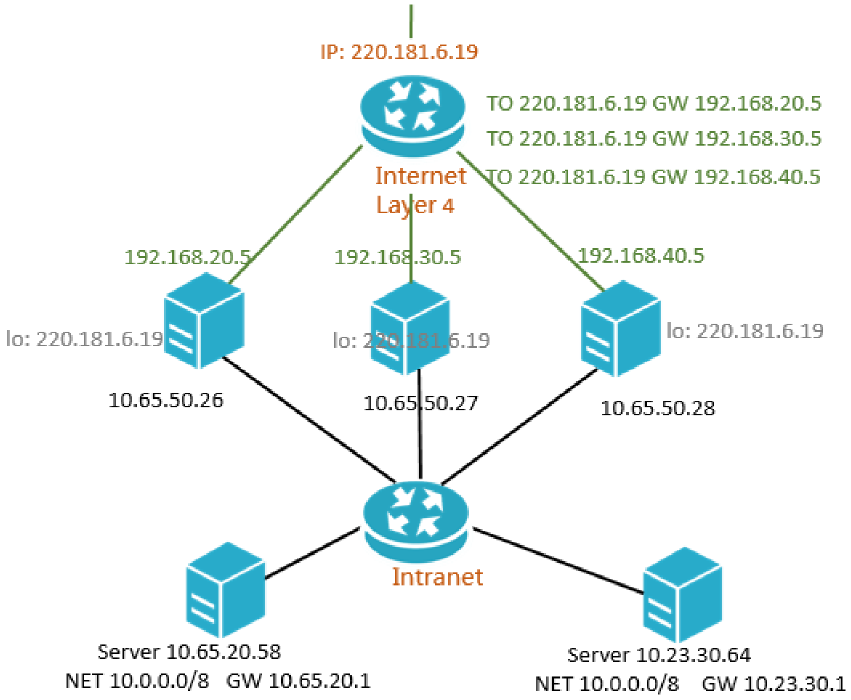
SLB design

Apsara Stack provides Layer-4 (TCP and UDP) and Layer-7 (HTTP and HTTPS) load balancing.

- Layer-4 SLB combines the open-source Linux Virtual Server (LVS) with Keepalived to balance loads, and implements customized optimizations to meet cloud computing requirements.
- Layer-7 SLB uses Tengine to balance loads. Tengine is a web server project launched by Taobao. Based on NGINX, Tengine has a wide range of advanced features optimized for high-traffic websites.

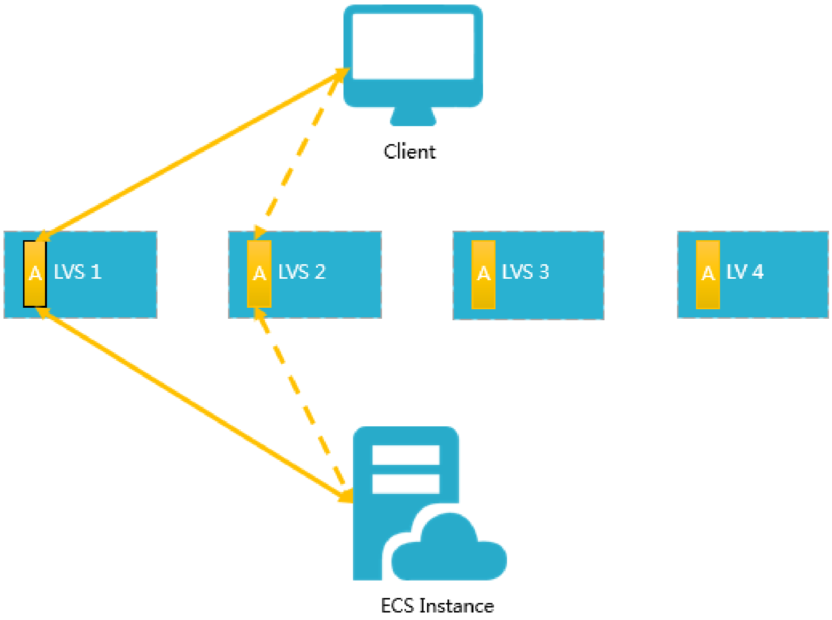


Layer-4 SLB runs in a cluster of LVS machines for higher availability, stability and scalability of load balancing in abnormal cases.



In an LVS cluster, each machine synchronizes sessions with other machines via multicast packets. As shown in the below figure, Session A is established on LVS1 and is synchronized to other LVS machines after the client transfers three data packets to the server. Solid lines indicate the current active connections, while dotted lines indicate that the session requests will be sent to other normally working machines if LVS1 fails or is being maintained. In this way, you can perform hot updates, machine maintenance, and cluster maintenance without affecting business applications.

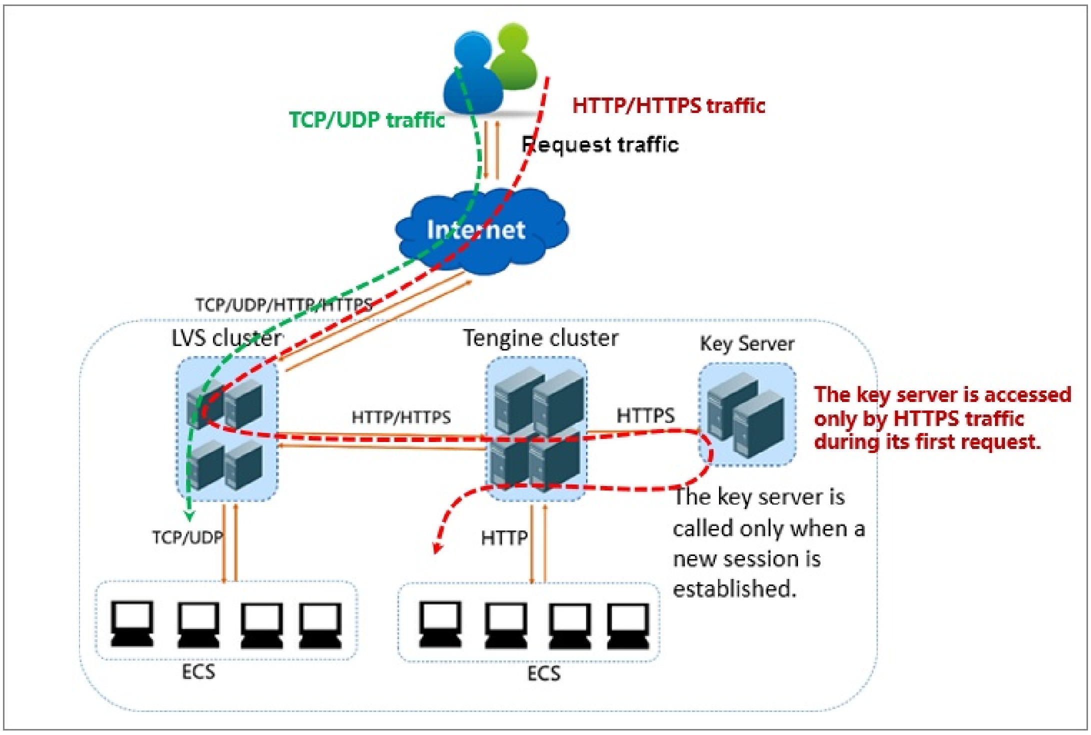
Note If a connection is not established (the three-way handshake is not completed), or if a connection has been established but session synchronization is not triggered during a hot upgrade, your service may be interrupted. In this case, the client needs to re-initiate the connection.



Inbound network traffic flow

SLB distributes incoming traffic according to the forwarding rules configured in the console or by using APIs. The following figure shows the inbound network traffic flow.

Inbound network traffic flow



1. For TCP, UDP, HTTP, and HTTPS protocols, the incoming traffic must be forwarded through the LVS cluster first.
2. Large amounts of access requests are evenly distributed among all servers in the LVS cluster. Servers synchronize sessions to guarantee high availability.
   - For layer-4 listeners (the frontend protocol is UDP or TCP), the node servers in the LVS cluster distribute requests directly to backend ECS instances according to the configured forwarding rules.
   - For layer-7 listeners that use the frontend protocol HTTP, the node servers in the LVS cluster first distribute requests to the Tengine cluster. Then, the node servers in the Tengine cluster distribute the requests to backend ECS instances according to the configured forwarding rules.
   - For layer-7 listeners that use the frontend protocol HTTPS, the request distribution is similar to the HTTP protocol. However, before distributing requests to backend ECS instances, the system calls the Key Server to validate certificates and decrypt data packets.

18.4. Features

This topic describes the key features of Apsara Stack SLB, including Layer-4 and Layer-7 load balancing, health check, and session persistence for high availability of backend servers.

Feature	Layer-4 load balancing	Layer-7 load balancing
Scheduling algorithms SLB supports round robin (RR), weighted round robin (WRR), and weighted least connections (WLC).	Yes	Yes
Health checks SLB uses health checks to inspect the availability of backend servers. If a backend server is detected unhealthy, it stops receiving new requests. In this case, SLB distributes traffic to other healthy backend servers.	Yes	Yes
Session persistence SLB provides session persistence that enables the redirection of requests from one backend server to another within the life of a session.	Yes	Yes
Access control SLB allows you to use whitelists to implement access control for your applications.	Yes	Yes
High availability SLB distributes inbound traffic to backend servers across zones. In addition, SLB allows you to implement primary/secondary deployment in most regions, where failovers can be automatically triggered if the primary zone suffers an outage.	Yes	Yes

Feature	Layer-4 load balancing	Layer-7 load balancing
Security You can use SLB with Apsara Stack Security to defend your applications against 5 Gbit/s DDoS attacks.	Yes	Yes
Public and internal load balancing SLB can be used to balance the traffic load from the Internet or within internal networks. You can create an internal SLB instance to balance traffic within a Virtual Private Cloud (VPC), or create an Internet-facing SLB instance to distribute inbound network traffic from the Internet.	Yes	Yes
Support for IPv6 Internet-facing SLB instances can forward requests from IPv6 clients.	Yes	Yes
Certificate management SLB provides a centralized certificate management solution for HTTPS listeners that allows certificate uploading and decryption on SLB. You do not need to upload certificates to backend servers, which reduces server CPU overhead.	No	Yes
Support for WebSocket Secure (WSS) and WebSocket (WS) The WS protocol is a new addition to the HTML5 specification. It provides bidirectional communication channels between clients and servers. You can use the WS protocol to minimize server resources, reduce bandwidth consumption, and achieve real-time communication.	No	Yes
Support for HTTP/2 HTTP/2 is the second version of the HTTP protocol. It features significant performance improvement and is backward compatible with HTTP/1.x.	No	Yes

18.5. Scenarios

Classic Load Balancer (CLB) can be used to improve the availability and reliability of applications with high access traffic.

Balance the loads of your applications

You can configure listening rules to distribute heavy traffic among ECS instances that are attached as backend servers to CLB instances. You can also use the session persistence feature to forward all of the requests from the same client to the same backend ECS instance to enhance access efficiency.

Scale your applications

You can extend the service capability of your applications by adding or removing backend ECS instances to suit your business needs. CLB can be used for both web servers and application servers.

Eliminate single points of failure (SPOFs)

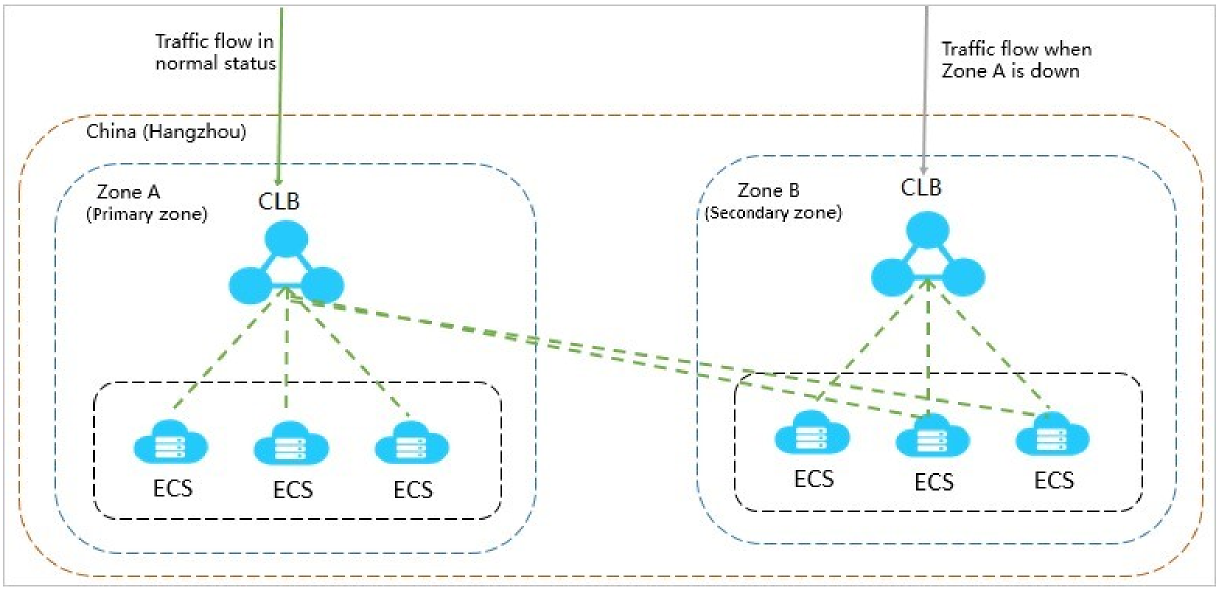
You can attach multiple ECS instances to a CLB instance. When an ECS instance malfunctions, CLB automatically isolates this ECS instance and distributes inbound requests to other healthy ECS instances, ensuring that your applications continue to run properly.

Implement zone-disaster recovery (multi-zone disaster recovery)

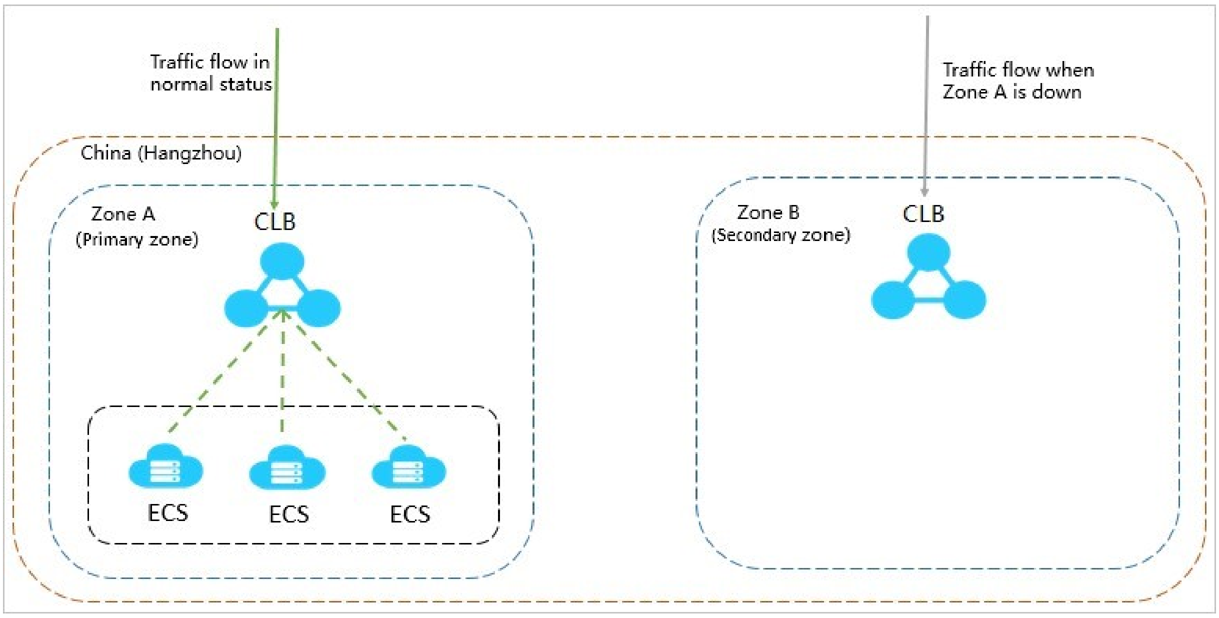
To provide more stable and reliable load balancing services, Apsara Stack allows you to deploy CLB instances across multiple zones in most regions for disaster recovery. Specifically, you can deploy a CLB instance in two zones within the same region. One zone is the primary zone, while the other zone is the secondary zone. If the primary zone fails or becomes unavailable, the CLB instance will fail over to the secondary zone in about 30 seconds. When the primary zone recovers, the CLB instance will automatically switch back to the primary zone.

We recommend that you create a CLB instance in a region that has multiple zones for zone-disaster recovery. We recommend that you plan the deployment of backend servers based on your business needs. In addition, we recommend that you add at least one backend server in each zone to achieve the highest load balancing efficiency.

As shown in the following figure, ECS instances in different zones are attached to a single CLB instance. In normal cases, the CLB instance distributes inbound traffic to ECS instances both in the primary zone (Zone A) and in the secondary zone (Zone B). If Zone A fails, the CLB instance distributes inbound traffic only to Zone B. This deployment mode helps avoid service interruptions caused by zone-level failure and reduce latency.

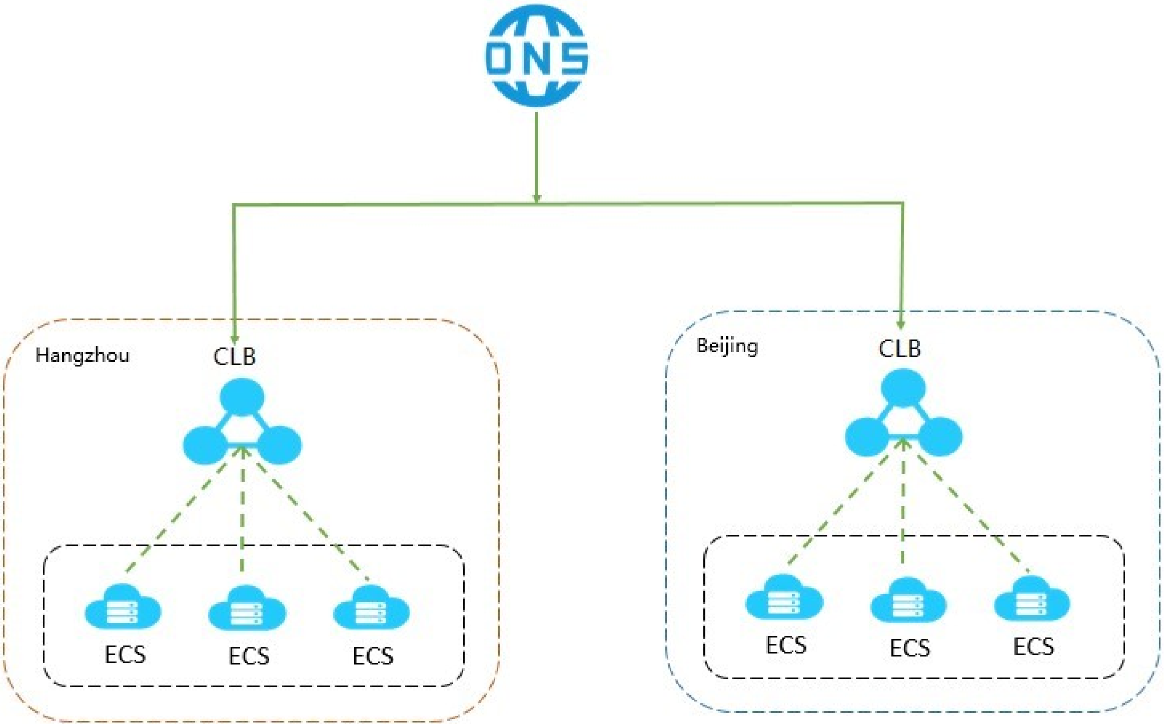


Assume that you deploy all ECS instances in the primary zone (Zone A) and no ECS instances in the secondary zone (Zone B) as shown in the following figure. If Zone A fails, your services will be interrupted because no ECS instances are available in Zone B. This deployment mode achieves low latency at the cost of high availability.



Geo-disaster recovery

You can deploy CLB instances in different regions and attach ECS instances of different zones within the same region to a CLB instance. You can use DNS to resolve domain names to service addresses of CLB instances in different regions for global load balancing purposes. When a region becomes unavailable, you can temporarily stop DNS resolution within that region without affecting user access.



18.6. Limits

This topic describes the limits for SLB.

Item	Limit
SLB instances	
Maximum number of listeners that can be added to an SLB instance	500
Certificates	
Maximum number of server certificates that can be uploaded in a region	1,000
Maximum number of client Certificate Authority (CA) certificates that can be uploaded in a region	1,000

18.7. Terms

This topic introduces the terms used in CLB.

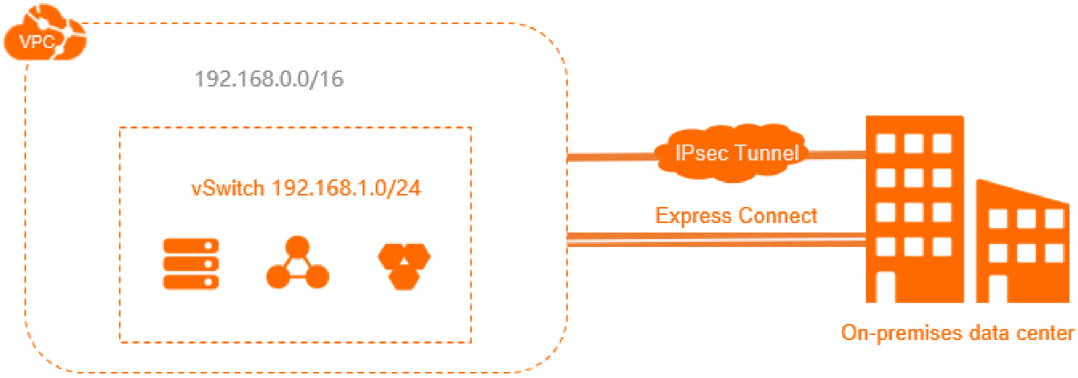
Term	Description
CLB	CLB distributes traffic across ECS instances. CLB provides Layer-4 and Layer-7 load balancing.
CLB instance	A load-balancing instance in CLB. To get started with CLB, you must create a CLB instance.
Endpoint	An IP address assigned to a CLB instance. The IP address can be either public or private, depending on the type of the CLB instance. You can resolve a domain name to a public IP address of a CLB instance to provide external services.
Listener	A listener distributes requests to backend servers. Each CLB instance must have at least one listener.
Backend server	A backend server is an ECS instance that receives client requests distributed by a CLB instance.
Default server group	A group of ECS instances that process distributed requests. If a listener is not configured with any VServer group or primary/secondary server group, the listener distributes traffic to the backend servers in the default server group.
VServer group	A group of ECS instances that process distributed requests. You can create multiple VServer groups for different listeners of a CLB instance to specify traffic distribution with specific listeners.
Primary/secondary server group	Each primary/secondary server group contains two ECS instances, where one acts as the primary server and the other acts as the secondary server. If the primary server is detected unhealthy, new requests are automatically distributed to the secondary server.

19.Virtual Private Cloud (VPC)

19.1. What is a VPC?

A virtual private cloud (VPC) is a private network dedicated for your use. You have full control over your VPC, which you can define and customize by specifying the Classless Inter-domain Routing (CIDR) block, configuring route tables, and creating gateways. You can launch Apsara Stack resources such as Elastic Compute Service (ECS) instances, ApsaraDB for RDS (RDS) instances, and Server Load Balancer (SLB) instances in your VPC.

Furthermore, you can connect your VPC to other VPCs or on-premises networks to create a custom network environment. In this way, you can smoothly migrate applications and extend on-premises data centers to the cloud.



Components

Each VPC consists of one VRouter, at least one private CIDR block, and one or more VSwitches.

- Private CIDR block

  When you create a VPC or a VSwitch, you must specify its private IP address range in the form of a CIDR block.

  You can use the standard private CIDR blocks listed in the following table and their subsets as CIDR blocks for your VPCs. For more information, see the Plan and design a VPC section in this *User Guide*.

CIDR block	Number of available private IP addresses (excluding those reserved by the system)
192.168.0.0/16	65,532
172.16.0.0/12	1,048,572
10.0.0.0/8	16,777,212

- VRouter

  A VRouter is a hub that connects all VSwitches in a VPC and serves as a gateway between the VPC and other networks. After a VPC is created, a VRouter is automatically created for the VPC. Each VRouter is associated with a route table.

  For more information, see the Route table overview section in this *User Guide*.

- VSwitch

  A VSwitch is a basic network component that connects different cloud resources in a VPC. After you create a VPC, you can create VSwitches to partition your VPC into multiple subnets. VSwitches within a VPC can communicate with each other over the private network. You can deploy your applications in VSwitches that belong to different zones to improve service availability.

  For more information, see the Create a VSwitch sectionin this *User Guide*.

19.2. Benefits

This topic describes the benefits of using VPCs.

High security

Each VPC has a unique tunnel ID, and each tunnel ID corresponds to a virtual network. Different VPCs are isolated by tunnel IDs:

- Similar to traditional networks, VPCs can also be divided into subnets. ECS instances in the same subnet use the same VSwitch to communicate with each other, while ECS instances in different subnets use VRouters to communicate with each other.
- VPCs are completely isolated from each other and can only be interconnected by mapping an EIP or a NAT IP address.
- ECS IP packets are encapsulated by using the tunneling technique. Therefore, information about the data link layer (layer-2 MAC address) of ECS does not go to the physical network. As a result, the layer-2 network between different ECS instances or between different VPCs is isolated.
- ECS instances in a VPC use security groups as firewalls to control traffic going to and from ECS instances. This is layer-3 isolation.

High flexibility

You can use security groups or whitelists to flexibly control traffic going to and from the cloud resources in a VPC.

Ease of use

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

High scalability

You can create multiple subnets in a VPC to deploy different services. Additionally, you can connect a VPC to other VPCs or on-premises data centers to expand your network.

19.3. Architecture

Based on the tunneling technique, VPCs isolate virtual networks. Each VPC has a unique tunnel ID, and each tunnel ID corresponds to only one VPC.

Background information

With the development of cloud computing, a variety of network virtualization techniques have been developed to meet the increasing demands for virtual networks with higher scalability, security, reliability, privacy, and connectivity.

Earlier solutions combined the virtual network with the physical network to form a flat network, for example, the large layer-2 network. However, with the increase of virtual network scale, problems such as ARP spoofing, broadcast storms, and host scanning are becoming more serious. To resolve these problems, various network isolation techniques are developed to completely isolate the physical network from the virtual network. One of these techniques can isolate users with a VLAN. However, a VLAN only supports up to 4,096 nodes, which are insufficient for the large number of users in the public cloud.

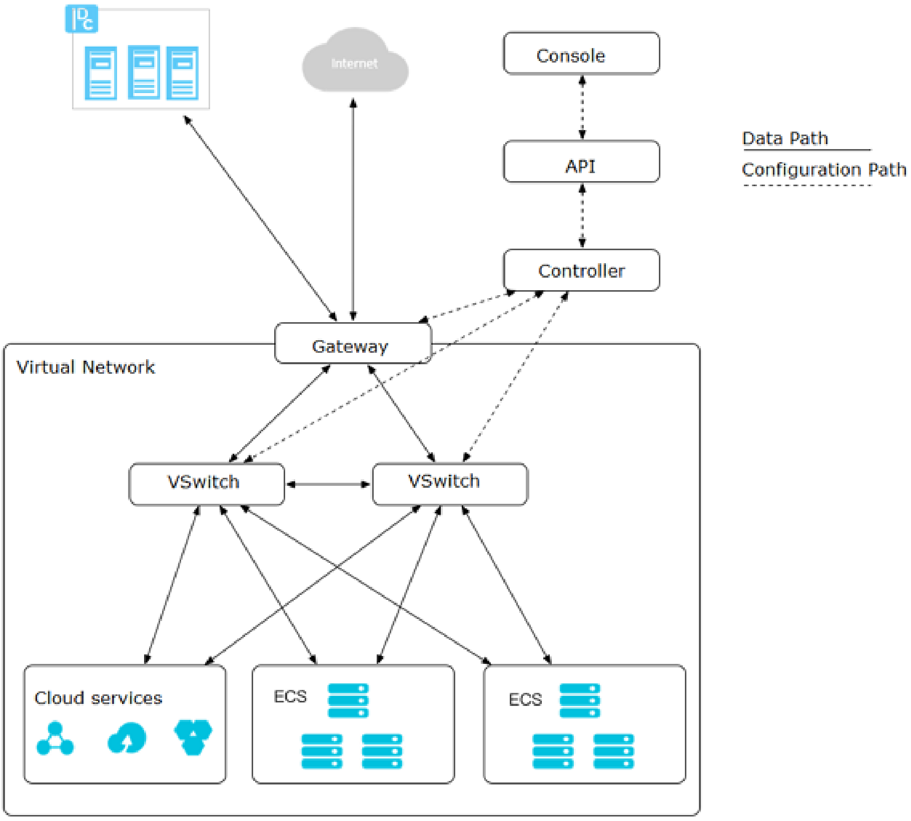
Principles

Based on the tunneling technique, VPCs isolate virtual networks. Each VPC has a unique tunnel ID, and each tunnel ID corresponds to only one VPC. A tunnel encapsulation carrying a unique tunnel ID is added to each data packet transmitted over the physical network between ECS instances in a VPC. In different VPCs, ECS instances with different tunnel IDs are located on two different routing planes. Therefore, these ECS instances cannot communicate with each other.

Based on the tunneling and Software Defined Network (SDN) techniques, Alibaba Cloud has developed VPCs that are integrated with gateways and VSwitches.

Logical architecture

As shown in the following figure, a VPC consists of a gateway, a controller, and one or more VSwitches. The VSwitches and gateway form a key data path. By using a protocol developed by Alibaba Cloud, the controller distributes the forwarding table to the gateway and VSwitches to provide a key configuration path. In the overall architecture, the configuration path and data path are separated from each other. The VSwitches are distributed nodes, the gateway and controller are deployed in clusters, and all links are equipped with disaster recovery. These features improve the availability of the VPC.



19.4. Features

A Virtual Private Cloud (VPC) is a private network logically isolated from other virtual networks. This topic describes the features of VPCs.

Custom private networks

You can customize VPCs. When you create VPCs or VSwitches, you can specify private CIDR blocks for them. Furthermore, you can divide a VPC into multiple subnets and deploy services in different subnets to improve service availability.

Custom routes

You can add custom routes to the route table of a VPC to forward traffic to the specified next hops. The route table uses the longest prefix match algorithm for traffic routing. If one destination address may match more than one entry in the route table, the algorithm is used to select the entry with the longest subnet mask because it is the most specific route.

Varied connection methods

A VPC provides you with varied connection methods. You can connect a VPC to the public network, an on-premises data center, or another VPC.

- Connect a VPC to the public network

  You can connect a VPC to the public network by associating an Elastic IP address with the VPC or configuring NAT Gateway, so that cloud services in the VPC can communicate with the public network.

- Connect a VPC to another VPC

  You can connect a VPC to another VPC by creating a pair of router interfaces to enable high speed and secure communication between the VPCs over the internal network.

- Connect a VPC to an on-premises data center

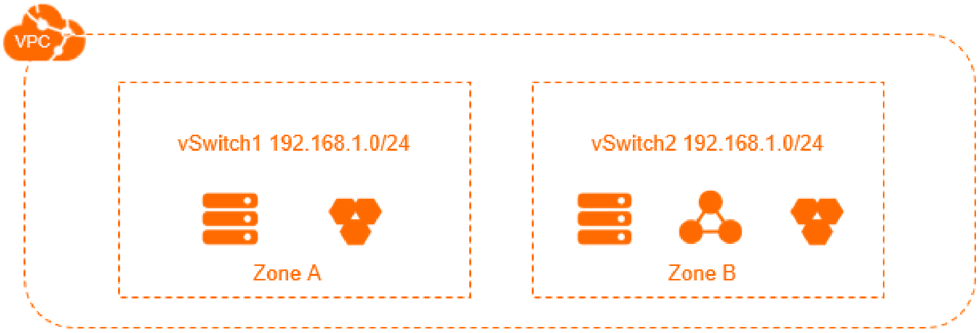
  You can connect a VPC to an on-premises data center by using a leased line to smoothly migrate local applications to the cloud.

19.5. Scenarios

This topic describes the scenarios in which VPCs are used to guarantee a high level of data security and service availability.

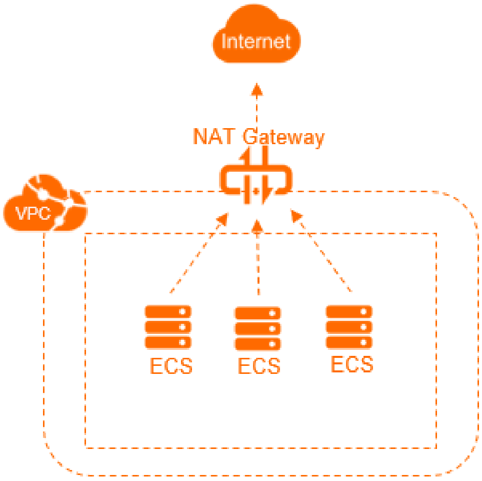
Host applications that provide external services

You can host applications that provide external services in a VPC and control access to these applications from the Internet by creating security group rules and access control whitelists. You can also isolate Internet-based mutual access between the application server and the database. For example, you can deploy the web server in a subnet that can access the Internet and deploy the application database in a subnet that cannot access the Internet.



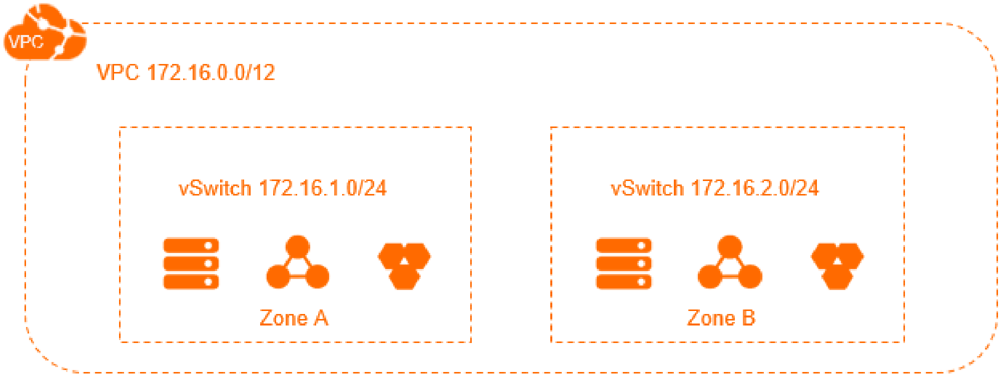
Host applications that require access to the Internet

You can host applications that require access to the Internet in a subnet of a VPC and route traffic through network address translation (NAT). After you configure SNAT rules, instances in the subnet can access the Internet without exposing their private IP addresses, which can be changed to public IP addresses any time to avoid external attacks.



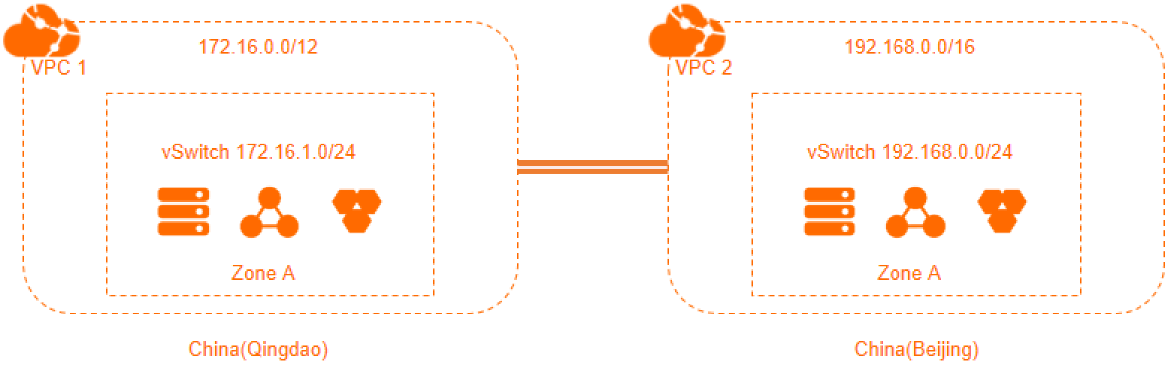
Implement disaster tolerance across zones

You can create one or multiple subnets in a VPC by creating VSwitches. VSwitches in a VPC can communicate with each other. You can deploy resources on VSwitches in different zones for disaster tolerance.



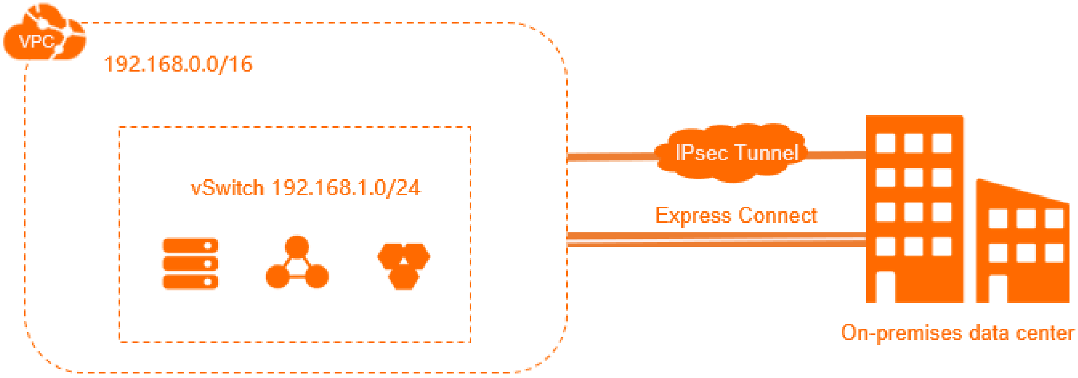
Isolate business systems

VPCs are logically isolated from each other. Therefore, you can create multiple VPCs to isolate multiple business systems, for example, isolate the production environment from the test environment. You can also create a peering connection between two VPCs if they need to communicate with each other.



Build a hybrid cloud

You can create a dedicated connection to connect your VPC to an on-premises data center to expand your local network. By doing so, you can seamlessly migrate your local applications to the cloud without changing the method of access to these applications.



19.6. Terms

This topic describes the key concepts for VPCs.

Term	Description
Virtual Private Cloud (VPC)	A VPC is a private network established on Alibaba Cloud. VPCs are logically isolated from each other. You can create and manage cloud resources in your VPC, such as Elastic Compute Service (ECS), Server Load Balancer (SLB), and ApsaraDB for RDS instances.
VSwitch	A VSwitch is a basic network device that connects different cloud resources in a VPC. When you create a cloud resource in a VPC, you must specify the VSwitch to which the cloud resource is connected.
VRouter	A VRouter is a hub that connects all VSwitches in a VPC and serves as a gateway that connects the VPC to other networks. A VRouter also forwards network traffic according to the route entries in the route table.
Route table	A route table is a list of route entries in a VRouter.
Route entry	Each item in a route table is a route entry. A route entry specifies the next hop address for the network traffic directed to a destination CIDR block. Route entries are divided into system route entries and custom route entries.

19.7. Limits

This topic describes the limits of Virtual Private Cloud (VPC). Before you use this service, we recommend that you understand these limits.

VPC

Item	Limit
The maximum number of VRouters in a VPC	1
The maximum number of route tables in a VPC	1
The maximum number of VSwitches in a VPC	24
The maximum number of route entries in a route table	48

VRouters and VSwitches

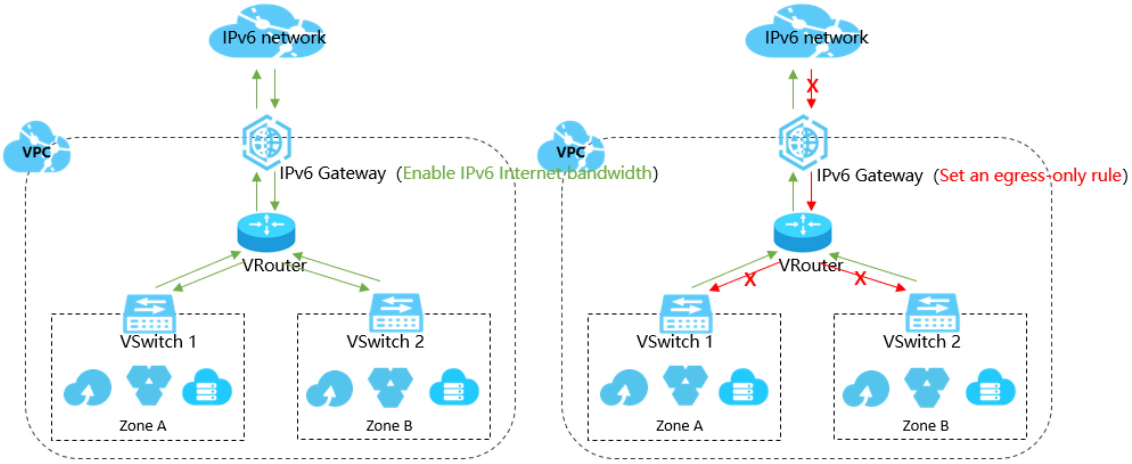
Item	Limit
VRouters	• Each VPC can have only one VRouter. • Each VRouter can only include one route table. • Dynamic routing protocols such as Border Gateway Protocol (BGP) and Open Shortest Path First (OSPF) are not supported.

Item	Limit
VSwitch	Layer-2 broadcasting and multicasting are not supported.

20.IPv6 Gateway

20.1. What is an IPv6 Gateway?

This topic provides an overview of the IPv6 Gateways of Virtual Private Cloud (VPC). An IPv6 Gateway functions as an IPv6 traffic gateway for a VPC. You can configure the IPv6 Internet bandwidth and egress-only rules to manage the inbound and outbound IPv6 traffic.



Functions

The functions of an IPv6 gateway are as follows:

- **IPv6 internal network communication**

  By default, an IPv6 address in a VPC is allocated with an Internet bandwidth of 0 Mbit/s and only supports communication over the internal network. Specifically, the cloud instances in a VPC can only access other IPv6 addresses in the same VPC through the IPv6 address. The resources cannot access the Internet with these IPv6 addresses or be accessed by IPv6 clients over the Internet.

- **IPv6 public network communication**

  You can purchase an Internet bandwidth for the IPv6 address for which you have applied. In this way, the resources in the VPC can access the Internet through the IPv6 address and be accessed by IPv6 clients over the Internet.

  You can set the Internet bandwidth to 0 Mbit/s at any time to deny the IPv6 address Internet access. After this configuration, the IPv6 address can only communicate over the internal network.

- **IPv6 public network communication with an egress-only rule**

  You can set an egress-only rule for an IPv6 Gateway. In this way, the IPv6 address can access the Internet, but IPv6 clients are denied access to your cloud resources in the VPC over the Internet.

  You can delete the egress-only rule at any time. After the rule is deleted, your resources in the VPC can access the Internet through the IPv6 address for which you have purchased Internet bandwidth, and IPv6 clients can access the resources in the VPC over the Internet.

The network access capability of IPv6 addresses is dependent on the settings of the network type, Internet bandwidth, and egress-only rule, as shown in the following table.

IPv6 network type	Enable IPv6 Internet bandwidth?	Set an egress-only rule?	IPv6 network access capability
Internal network	No	No	Internal network communication
Public network	Yes	No	Internal network communication Public network communication
		Yes	Internal network communication Public network communication when access is initiated by VPCs

Benefits

IPv6 Gateway provides the following benefits:

- **High availability**

  IPv6 Gateways provide cross-zone high availability and stable IPv6 Internet gateway services.

- **High performance**

  A single IPv6 Gateway provides a 10-gigabit level throughput.

- **Flexible management of public network communication**

  You can manage the Internet communication capability of an IPv6 Gateway by adjusting its Internet bandwidth and setting an egress-only rule.

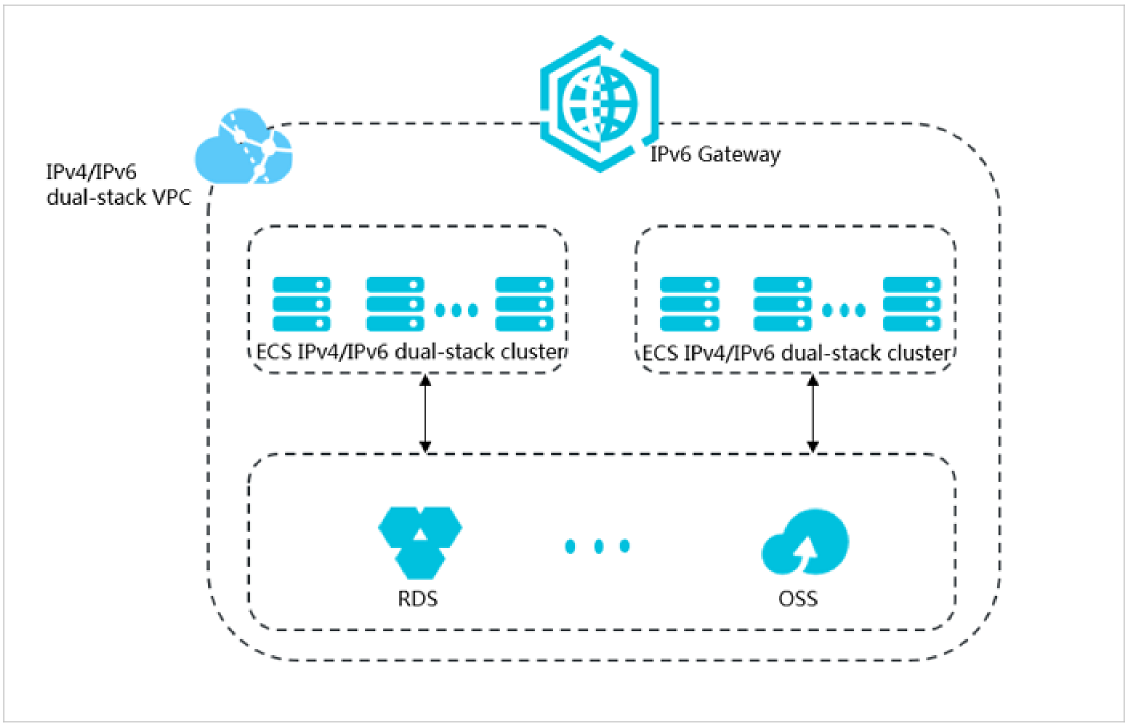
20.2. Scenarios

This topic describes three different scenarios where IPv6 Gateway can be established to provide IPv6 addressing services.

Scenario 1: Provide IPv6 support with an isolated IPv6 cloud environment

You can enable IPv6 for an existing VPC, which allows the VPC to support both IPv4 and IPv6 protocol stacks. You can also allocate IPv6 addresses to the ECS instances in the VPC. In this way, the ECS instances have both IPv4 and IPv6 addresses. Note that IPv6 addresses of ECS instances can only communicate with other IPv6 addresses over the intranet in the VPC by default.

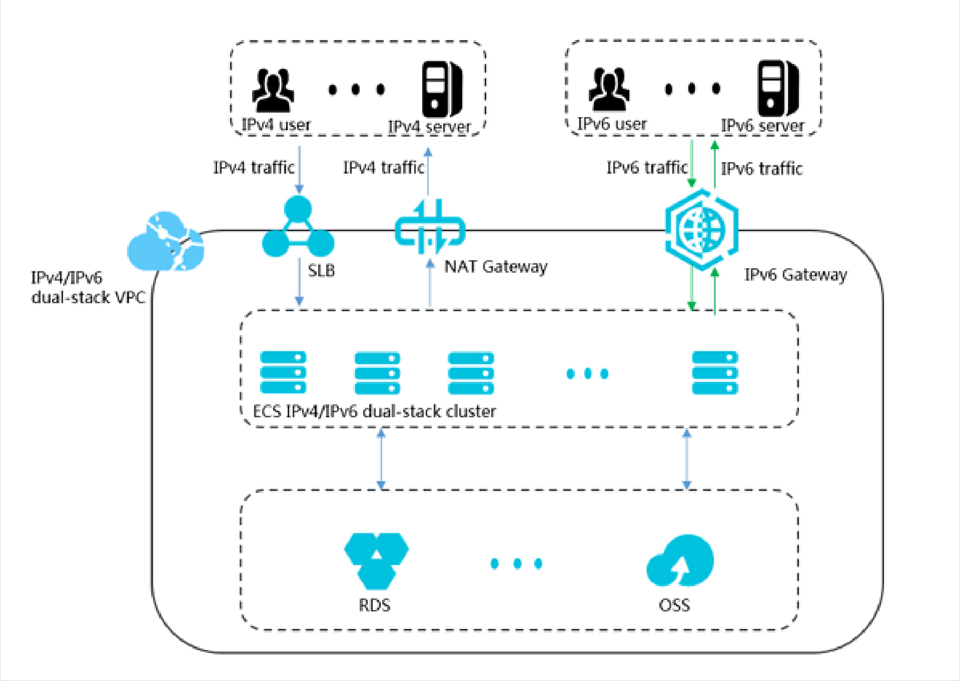
IPv4 and IPv6 dual-stack ECS clusters can connect to each other through IPv4 intranets or IPv6 networks. However, the ECS instances cannot use IPv6 addresses to access the Internet or be accessed by IPv6 clients over the Internet.


IPv4/IPv6
dual-stack VPC
IPv6 Gateway
ECS IPv4/IPv6 dual-stack cluster
ECS IPv4/IPv6 dual-stack cluster
RDS
OSS

Scenario 2: IPv6-based communication between cloud resources in your VPC and the Internet

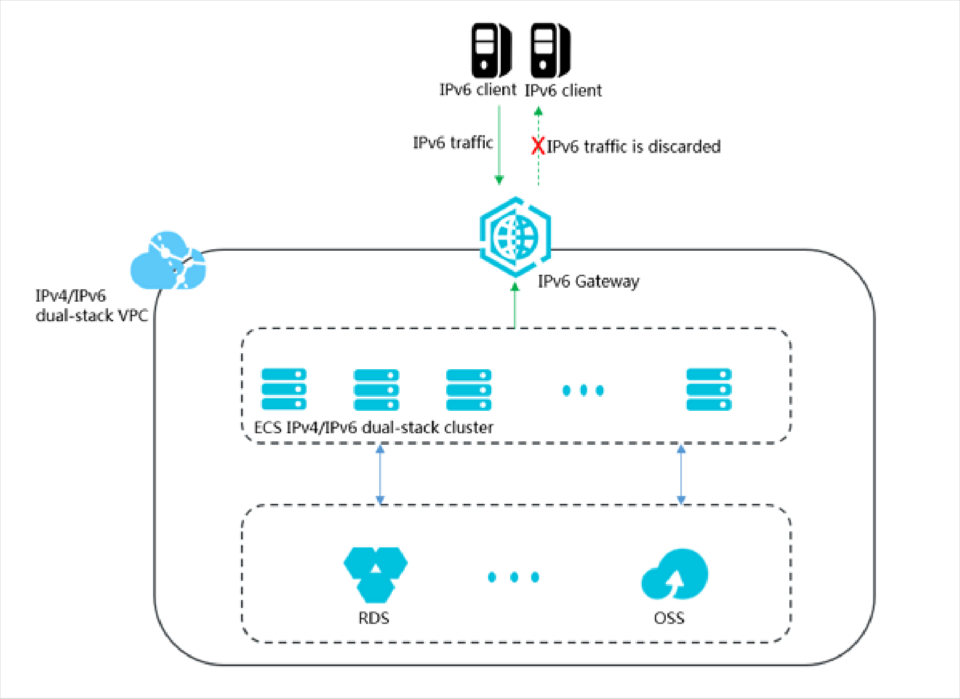
After you purchase an Internet bandwidth for your IPv6 addresses, the IPv6 addresses can access the Internet. Specifically, the IPv6 traffic between the cloud resources in the VPC and the IPv6 network passes through IPv6 Gateway, which functions as the transfer hub of IPv6 Internet traffic for dual-stack VPCs.

The IPv4 addresses of ECS clusters in the VPC communicate with IPv4 clients over the Internet through SLB and NAT Gateway, which serve as the inlet and outlet of IPv4 Internet traffic for dual-stack VPCs.



Scenario 3: Egress-only IPv6 traffic over the Internet

If you need your services to actively access IPv6 clients but do not want the IPv6 addresses of ECS instances to be accessed by external IPv6 clients, you can set an egress-only rule. In this way, the ECS instances you specified can access IPv6 networks by using IPv6 addresses. However, external IPv6 clients cannot access the specified ECS instances and access requests initiated by external IPv6 clients are discarded by IPv6 Gateway.



20.3. Terms

This topic describes the terms used in IPv6 Gateway.

Term	Description
IPv6 address	The IPv6 address allocated by the system to an instance in a VPC. An IPv6 address is made of 128 binary bits that are divided into eight 16-bit groups separated by colons (:). Each group is represented as a 4-digit hexadecimal number. The following is an example of an IPv6 address: 2001:xxx:0102::0304
IPv6 gateway	The Internet gateway for IPv6 traffic flowing in and out of a VPC. You can use an IPv6 gateway to control and manage the bandwidth used by IPv6 traffic. IPv6 gateways allow you to create egress-only rules to funnel egress traffic.
IPv6 Internet bandwidth	The Internet bandwidth of an IPv6 address that limits the bandwidth of Internet connectivity for the IPv6 address. You must purchase and add IPv6 Internet bandwidth to an IPv6 address before the IPv6 address can be used to communicate over the Internet.
Egress-only rule	A rule by which an IPv6 gateway implements egress control for IPv6 traffic. After you configure an egress-only rule for an IPv6 address, the IPv6 gateway allows outbound only communication to the Internet over IPv6 using the IPv6 address, and prevents the Internet from initiating IPv6 connections with the instance associated with the IPv6 address.
IPv6 CIDR block for VPC	A /61 IPv6 CIDR block automatically allocated to a VPC after IPv6 is enabled for the VPC.
IPv6 CIDR block for VSwitch	An IPv6 CIDR block allocated to a VSwitch. The default subnet mask for the IPv6 CIDR block of a VSwitch is /64. When you enable IPv6 for a VSwitch, you can specify the last eight bits of the IPv6 CIDR block of the VSwitch.

20.4. Limits

This topic describes the limits for IPv6 Gateway.

IPv6 gateways

- A VPC can be configured with only one IPv6 gateway.
- After IPv6 is enabled for a VPC, you cannot delete the IPv6 CIDR block of the VPC.
- After IPv6 is enabled for a VSwitch, you cannot delete the IPv6 CIDR block of the VSwitch.
- Before you delete an IPv6 gateway, make sure that the IPv6 Internet bandwidth and egress-only rules configured for all IPv6 addresses of the IPv6 gateway have been deleted.
- You can only create IPv6 routes with the destination of ::/0 and whose next hops are IPv6 gateway instances.

Egress-only rules

Before you create an egress-only rule, make sure that you have purchased and added IPv6 Internet bandwidth to the corresponding IPv6 address.

The following table lists the maximum number of egress-only rules that can be created for IPv6 gateways of different editions.

IPv6 gateway edition	Maximum number of egress-only rules per IPv6 gateway
Free Edition	0
Enterprise Edition	50
Enhanced Enterprise Edition	200

IPv6 Internet bandwidth

The following table lists the bandwidth limits for IPv6 traffic supported by different editions of IPv6 gateways.

IPv6 gateway edition	Maximum bandwidth per IPv6 address
Free Edition	2Gbps
Enterprise Edition	2Gbps
Enhanced Enterprise Edition	2Gbps

IPv6 Internet speeds

IPv6 Gateway implements traffic throttling with address-level and gateway-level limits:

- If the total amount of bandwidth allocated to the IPv6 addresses of an IPv6 gateway does not exceed the maximum forwarding bandwidth of the IPv6 gateway, the IPv6 communication speed of each instance within the VPC is limited based on the bandwidth limit specified for the associated IPv6 address.
- If the total amount of bandwidth allocated to the IPv6 addresses of an IPv6 gateway exceeds the maximum forwarding bandwidth of the IPv6 gateway, the Internet speed of each IPv6 address is subject to the maximum forwarding bandwidth of the IPv6 gateway.

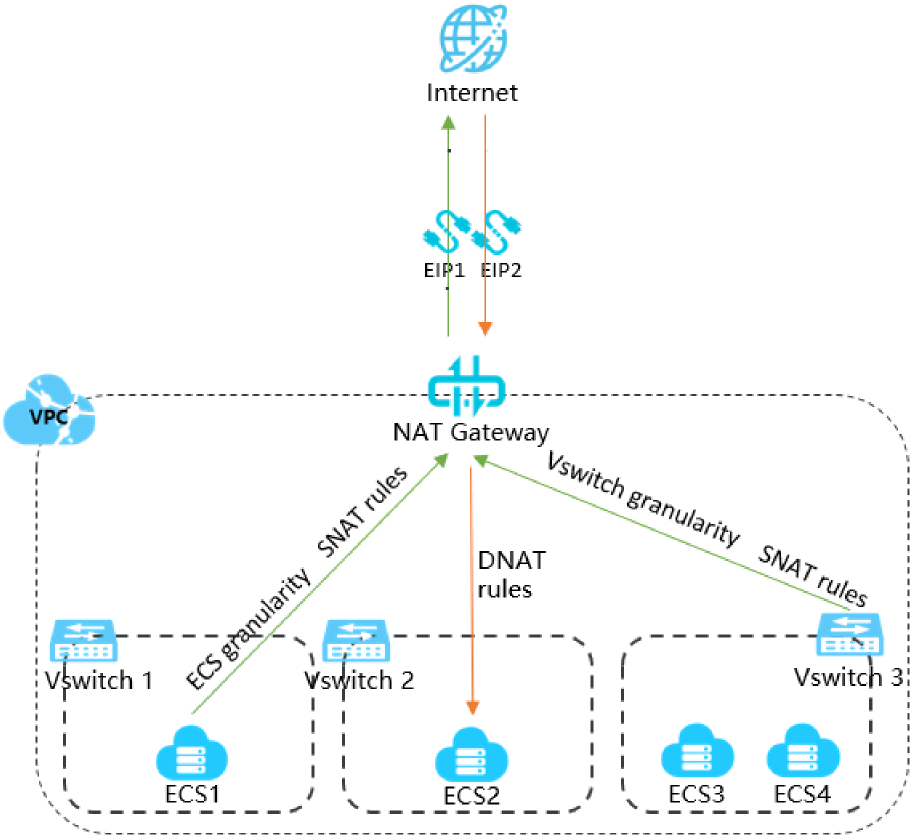
The following table lists the maximum forwarding bandwidth supported by different editions of IPv6 gateways.

IPv6 gateway edition	Maximum forwarding bandwidth per IPv6 gateway
Free Edition	10 Gbit/s
Enterprise Edition	20 Gbit/s
Enhanced Enterprise Edition	50 Gbit/s

21.NAT Gateway

21.1. What is NAT Gateway?

A NAT gateway is an enterprise-grade Internet gateway. NAT Gateway provides source network address translation (SNAT) and destination network address translation (DNAT) features, a maximum forwarding capacity of 10 Gbit/s, and support for cross-zone disaster recovery.



Features

NAT gateways must be associated with public IP addresses. After you create a NAT gateway, you can associate it with one or more elastic IP addresses (EIPs).

NAT Gateway supports SNAT and DNAT.

- SNAT allows Elastic Compute Service (ECS) instances that are deployed in a virtual private cloud (VPC) and not associated with public IP addresses to access the Internet.
- DNAT maps public IP addresses of a NAT gateway to ECS instances so that the ECS instances can be accessible from the Internet.

21.2. Features

NAT Gateway provides SNAT and DNAT features.

SNAT

SNAT allows ECS instances without public IP addresses in a VPC to access the Internet.

SNAT can also be used as a firewall to prevent unwanted access to backend servers. After you configure SNAT entries to allow backend servers to initiate connections with specific external terminals, only these external terminals will be able to access the backend servers.

DNAT

DNAT maps a public IP address of a NAT gateway to an ECS instance so that the ECS instance can be accessible from the Internet.

DNAT supports port mapping and IP mapping.

21.3. Benefits

NAT Gateway features easy configuration, high performance, high availability, and flexible specification adjustment.

Easy configuration

A NAT gateway is an enterprise-grade Internet gateway that controls the traffic flowing to and from a VPC and provides SNAT and DNAT. NAT gateways are reliable, flexible, and easy-to-use, saving you the trouble of building an Internet gateway yourself.

High performance

A NAT gateway is a virtual network device developed based on the software-defined networking (SDN) technology and the distributed gateway design of Alibaba Cloud. NAT Gateway supports a forwarding capability of up to 10 Gbit/s, meeting the requirements of large-scale Internet applications.

High availability

NAT Gateway supports cross-zone deployment, which guarantees high availability and ensures service continuity during a zone failure.

Flexible specification adjustment

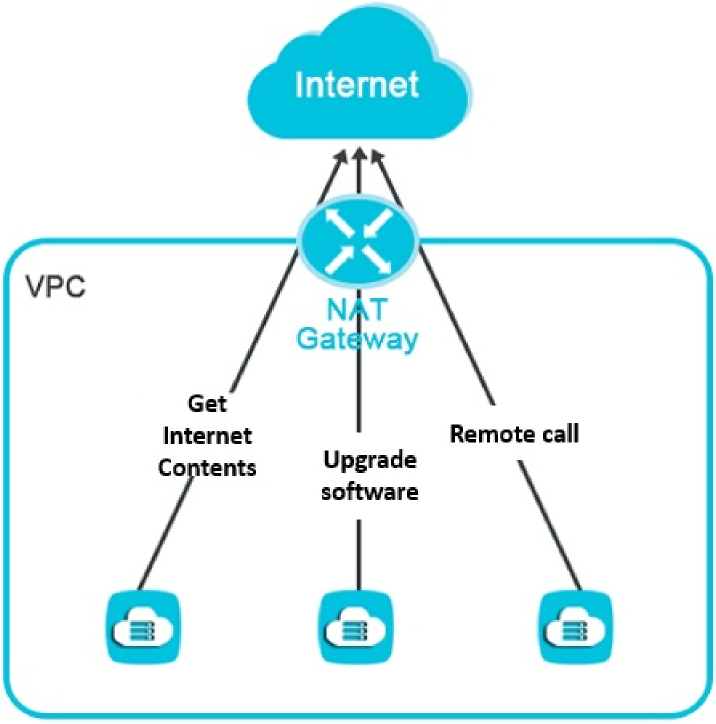
You can change the specification of your NAT gateway, or the number and specifications of the EIPs associated with the NAT gateway at any time to provide dynamic support for your services.

21.4. Scenarios

You can use NAT Gateway to allow access to and from the Internet for ECS instances in VPCs.

Scenario 1: Set up a high-availability SNAT gateway

You can use the SNAT feature of NAT Gateway to connect ECS instances to the Internet and prevent their IP addresses from being exposed to the Internet.



Scenario 2: Provide services accessible from the Internet

You can use the DNAT feature of NAT Gateway to allow applications on ECS instances to provide publicly accessible services through IP or port mapping.

21.5. Terms

The following table lists the terms used in NAT Gateway and their descriptions.

Term	Description
NAT gateways	A NAT gateway is an enterprise-grade Internet gateway that controls the traffic flowing to and from a VPC. NAT Gateway provides SNAT and DNAT features, a maximum forwarding capacity of 10 Gbit/s, and support for cross-zone disaster recovery.
DNAT tables	A DNAT table is a configuration table of a NAT gateway and is used to configure DNAT. DNAT allows you to map a public IP address of a NAT gateway to an ECS instance through IP or port mapping.
SNAT tables	An SNAT table is a configuration table of a NAT gateway and is used to configure SNAT. You can configure an SNAT entry for a VSwitch or for a specific ECS instance. • Configure an SNAT entry for a VSwitch: All ECS instances in the VSwitch use the specified public IP addresses to access the Internet. • Configure an SNAT entry for an ECS instance: The ECS instance uses the specified public IP addresses to access the Internet.

Term	Description
EIPs	An EIP is a public IP address that you can purchase and own independently. You can associate an EIP with an ECS instance deployed in a VPC, an internal SLB instance deployed in a VPC, or a secondary elastic network interface (ENI) attached to a VPC. You can also associate an EIP with a NAT gateway or a High-availability Virtual IP Address (HAVIP).

21.6. Limits

This topic describes the limits that apply to NAT Gateway and NAT Gateway associations with EIPs.

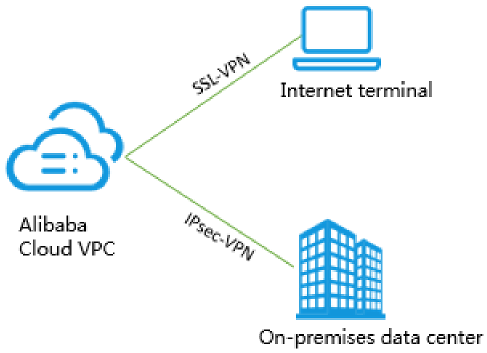
Item	Limit
The maximum number of NAT gateways that can be configured for a VPC	1
Using a public IP address for both SNAT and DNAT	Not supported
The maximum number of DNAT entries that can be configured for a NAT gateway	100
The maximum number of SNAT entries that can be configured for a NAT gateway	40
The maximum number of public IP addresses that can be specified in an SNAT entry	64
Creating a NAT gateway for a VPC that contains a custom route entry whose destination CIDR block is 0.0.0.0/0	Not supported **Note** You must delete the custom route entry with 0.0.0.0/0 as the destination CIDR block before you can create a NAT gateway for the VPC.
Limits on VSwitch bandwidth by the maximum bandwidth of the EIPs specified in the SNAT entry configured for the VSwitch	Yes
The maximum number of EIPs that can be associated with a NAT gateway	20
When multiple ECS instances without public IP addresses in a VPC access the same destination IP address and port on the Internet through a NAT gateway, the maximum number of concurrent connections supported by the NAT gateway is limited based on the number of EIPs specified in the corresponding SNAT entry. • If only one EIP is specified, the maximum number of concurrent connections supported by the NAT gateway is 55,000. • If multiple EIPs are specified, the maximum number of concurrent connections supported by the NAT gateway is n x 55,000 (n refers to the number of EIPs).	

22.VPN Gateway

22.1. What is VPN Gateway?

VPN Gateway is an Internet-based service that allows you to connect enterprise data centers, office networks, or Internet-facing terminals to Alibaba Cloud Virtual Private Cloud (VPC) networks through secure and reliable connections. VPN Gateway supports both IPsec-VPN connections and SSL-VPN connections.

Note The Alibaba Cloud VPN Gateway service complies with the local regulations and policies. VPN Gateway does not provide Internet access services.



Features

VPN Gateway supports the following features:

- IPsec-VPN

  Route-based IPsec-VPN allows you to route network traffic in multiple ways, and also facilitates the configuration and maintenance of VPN policies.

  You can use IPsec-VPN to connect an on-premises data center to a VPC network or connect two VPC networks. IPsec-VPN supports the IKEv1 and IKEv2 protocols. Any devices that support these two protocols can connect to Alibaba Cloud VPN Gateway, such as devices manufactured by Huawei, H3C, Hillstone, Sangfor, Cisco ASA, Juniper, SonicWall, Nokia, IBM, and Ixia.

- SSL-VPN

  SSL-VPN is implemented based on the OpenVPN framework. You can create an SSL-VPN connection to connect a remote client to applications and services deployed in a VPC network. After you deploy your applications or services, you only need to import the certificate to the client to initiate a connection.

Benefits

VPN Gateway offers the following benefits:

- High security: You can use the IKE and IPsec protocols to encrypt data for secure and reliable data transmission.
- High availability: VPN Gateway adopts the hot-standby architecture to achieve failover within a few seconds, session persistence, and zero service downtime.

- Cost-effectiveness: The encrypted Internet connections provided by VPN Gateway are more cost-effective than leased lines.
- Ease of use: VPN Gateway is a ready-to-use service. VPN gateways start to work immediately after they are deployed.

22.2. Scenarios

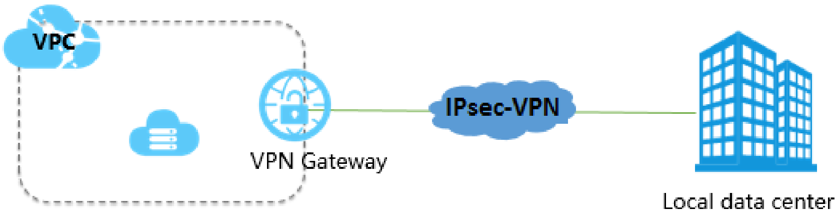
VPN Gateway is an Internet-based service that allows you to connect enterprise data centers, office networks, or Internet-facing terminals to Alibaba Cloud Virtual Private Cloud (VPC) networks through encrypted connections. VPN Gateway can be configured based on your requirements and applied in multiple scenarios.

Connect a VPC network to an on-premises data center

You can use IPsec-VPN to connect an on-premises data center to a VPC network and build a hybrid cloud.

Route-based IPsec-VPN allows you to route network traffic in multiple ways, and also facilitates the configuration and maintenance of VPN policies.

> ⓘ **Note** Before you create an IPsec-VPN connection between an on-premises data center and a VPC network, make sure that the IP address of the on-premises data center is different from that of the VPC network. In addition, you must set a static public IP address for your VPN gateway.

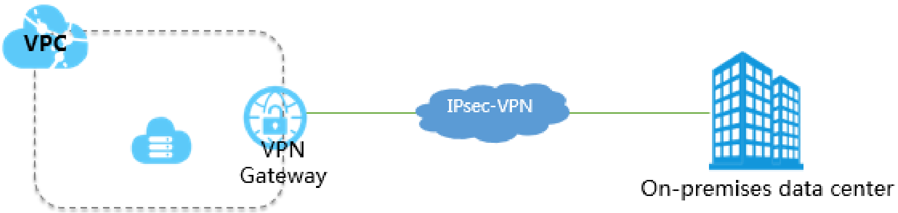


Connect two VPC networks

You can use IPsec-VPN to connect two VPC networks. This way, cloud resources can be shared across these VPC networks.

Route-based IPsec-VPN allows you to route network traffic in multiple ways, and also facilitates the configuration and maintenance of VPN policies.

> ⓘ **Note** The CIDR blocks of the two VPC networks must not overlap with each other.

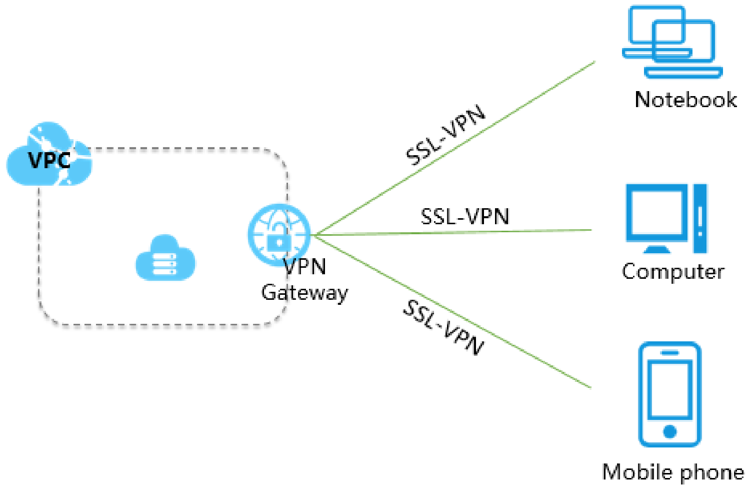


Connect a VPC network to a mobile client

If you require office automation, you can use SSL-VPN to connect a mobile client to a VPC network. This way, the mobile client can access cloud resources deployed in the VPC network anytime and anywhere.

You can initiate an SSL-VPN connection from clients that run Windows, Linux, macOS, iOS, and Android.

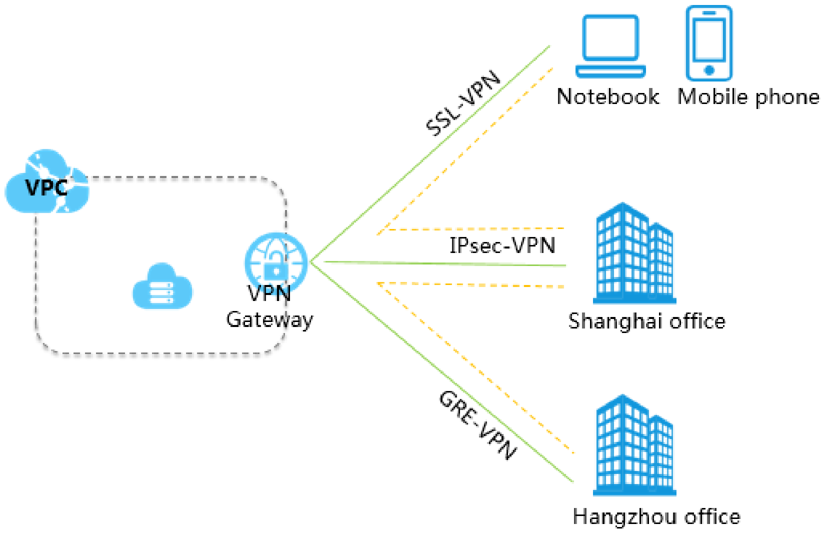
> Note The CIDR block assigned to the client must not overlap with that of the VSwitch in the VPC network.



Use IPsec-VPN and SSL-VPN together

You can use IPsec-VPN and SSL-VPN connections together to expand your network topology. After the connections are established, the client can access the applications deployed in the connected VPC network, and can also access the applications deployed in the connected office networks.

> Note The private CIDR blocks to be interfaced must not overlap with each other.



22.3. Limits

Before you use VPN Gateway, note the following limits.

VPN gateways

Item	Limit
Maximum number of VPN gateways for each account	30 **Note** The maximum number of VPN gateways that can be created is determined by the account, regardless of the region where the VPN gateways are deployed or the connected VPC networks. For example, for each account: • If there is only one VPC network in a region, you can create up to 30 VPN gateways for the VPC network in the region. • If there are multiple regions or VPC networks, you can create a total number of 30 VPN gateways for the VPC networks in all regions.
Maximum number of policy-based routes for a VPN gateway	20
Maximum number of destination-based routes that can be created for a VPN gateway	20

Customer gateways

Item	Limit
Maximum number of customer gateways that you can create in a region	100

IPsec-VPN connections

Item	Limit
Maximum number of IPsec-VPN connections for a VPN gateway	10
Maximum number of local CIDR blocks that can be added to an IPsec-VPN connection	5
Maximum number of remote CIDR blocks that can be added to an IPsec-VPN connection	5

SSL-VPN connections

Item	Limit
Maximum number of SSL client certificates that an account can reserve	50
Maximum number of SSL servers that can be associated with a VPN gateway	1
Maximum number of local CIDR blocks that can be added to an SSL server	5
Maximum number of remote CIDR blocks that can be added to an SSL server	1
Ports that are not supported by SSL servers	22, 2222, 22222, 9000, 9001, 9002, 7505, 80, 443, 53, 68, 123, 4510, 4560, 500, and 4500
The validity period of an SSL client certificate	Three years

23.Elastic IP Address

23.1. What is an EIP?

This topic provides an overview of Elastic IP Address. An elastic IP address (EIP) is a public IP address that you can purchase and hold as an independent resource. You can associate an EIP with an Elastic Compute Service (ECS) instance deployed in a virtual private cloud (VPC), an internal Server Load Balancer (SLB) instance deployed in a VPC, or a secondary elastic network interface (ENI) attached to a VPC. You can also associate an EIP with a NAT gateway, or a High-Availability Virtual IP Address (HAVIP).

An EIP is also a NAT IP address that is provisioned in a public-facing gateway of Alibaba Cloud and is mapped to the associated cloud resource with NAT. After an EIP is associated with a cloud resource, the cloud resource can connect to the Internet by using this EIP.

Benefits

EIPs have the following benefits:

- Independent purchase and possession

  You can purchase and hold an EIP as an independent resource. EIPs are not bundled with other computing or storage resources.

- Flexible association

  You can dissociate an EIP from a cloud resource and then release the EIP if the EIP is no longer needed.

- Configurable network capabilities

  You can adjust the peak bandwidth of an EIP at any time. The bandwidth changes take effect immediately.

23.2. Limits

This topic describes the limits that apply to EIP and EIP associations with other cloud resources.

General limits

The following general limits apply to EIPs:

- An EIP can be associated with only one cloud resource. In addition, the EIP must be in the available state. An association immediately takes effect when the configuration is complete.
- EIPs that are locked due to security reasons cannot be released, associated with, or disassociated from other cloud resources.

Limits on EIP associations with ECS instances

Note the following limits before you associate an EIP with an ECS instance:

- The ECS instance must be deployed in a VPC.
- The ECS instance and the EIP must reside in the same region.
- The ECS instance must be running or stopped.
- The ECS instance is not associated with a public IP address or another EIP.
- An ECS instance can be associated with only one EIP. If you want to associate multiple EIPs with an

ECS instance, you can associate the EIPs with one or more secondary ENIs and then associate the ENIs with the ECS instance.

Limits on EIP associations with NAT gateways

Note the following limits before you associate an EIP with a NAT gateway:

- The NAT gateway and the EIP must reside in the same region.
- A NAT gateway can be associated with a maximum of 20 EIPs.

Limits on EIP associations with SLB instances

Note the following limits before you associate an EIP with an SLB instance:

- The SLB instance must be deployed in a VPC.
- The SLB instance and the EIP must reside in the same region.
- An SLB instance can be associated with only one EIP.

Limits on EIP associations with HAVIPs

Note the following limits before you associate an EIP with an HAVIP:

- The HAVIP and the EIP must reside in the same region.
- The HAVIP must be available or allocated.
- An HAVIP can be associated with only one EIP.

Limits on EIP associations with secondary ENIs

Note the following limits before you associate an EIP with a secondary ENI:

- The secondary ENI must be attached to a VPC.
- The secondary ENI and the EIP must reside in the same region.
- You can associate an EIP with a secondary ENI in one of the following modes: NAT mode and cut-through mode.
    - In the NAT mode, the number of EIPs with which a secondary ENI can be associated depends on the number of private IP addresses of the secondary ENI.
    - In the cut-through mode, a secondary ENI can be associated with only one EIP.

24.Express Connect

24.1. What is Express Connect?

Express Connect allows you to establish private, flexible, stable, and secure communication channels between Virtual Private Clouds (VPCs) in different networking environments. You can use Express Connect to prevent data breaches and provide greater stability compared with Internet connections.

You can create a peering connection between your own VPCs, or with a VPC in another account. The VPCs can be in different regions.

Benefits

Express Connect provides the following benefits:

- High-speed interconnection

  Powered by the network virtualization technology, Express Connect allows networks to connect and exchange traffic at high speeds within internal networks without carrying traffic across the Internet. The impact of distance on network performance is minimized to ensure low-latency and high-bandwidth communication.

- Stability and reliability

  Built on the state-of-the-art infrastructure of Apsara Stack Cloud, Express Connect guarantees stable and reliable communication between networks.

- Security

  Express Connect implements cross-network communication at the network virtualization layer, where data is transmitted over separate and private channels within the infrastructure of Alibaba Cloud, mitigating the risks of data breaches.

- Buy-as-you-need service

  Express Connect delivers connectivity with a wide range of bandwidth options. You can choose based on the needs of your business to get the best value for your purchase.

Differences between Express Connect and Internet connections

A VPC is a logically isolated network. The traffic between VPCs or between VPCs and data centers flows over different networks.

Compared with Internet connections, Express Connect offers more enhanced connection performance and security for communication between VPCs. The following table summarizes their differences.

Item	Connect VPCs over the Internet	Connect VPCs with Express Connect
Communication performance and availability	Prone to high latency and packet loss	Guaranteed by the state-of-the-art infrastructure of Alibaba Cloud

Item	Connect VPCs over the Internet	Connect VPCs with Express Connect
Expense	High costs for Internet bandwidth or data transfer	Low bandwidth costs for cross-region communication Free for communication between VPCs within the same region
Security	Sensitive to data breaches	Guaranteed by communication isolation based on the network virtualization technology of Alibaba Cloud

24.2. Benefits

A VPC is a logically isolated network. The traffic between VPCs flows over different networks.

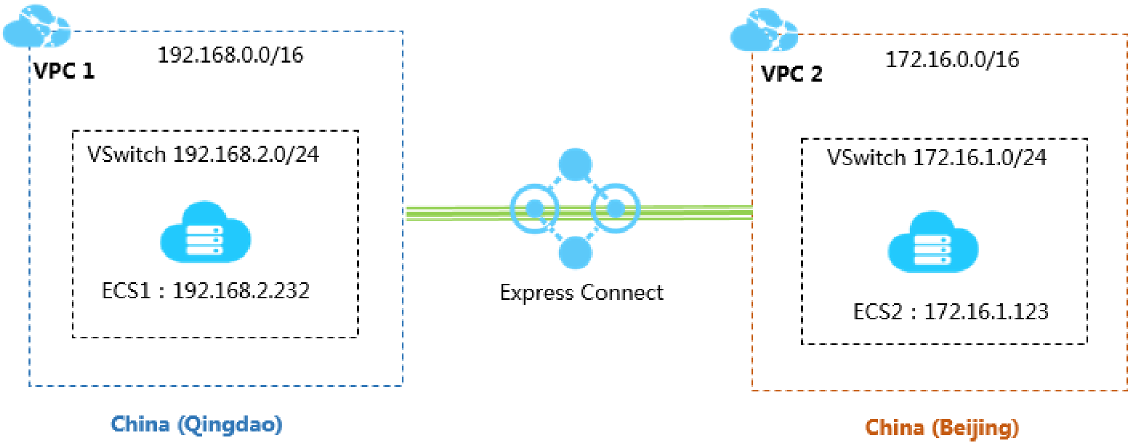
Compared with Internet connections, Express Connect offers more enhanced connection performance and security for communication between VPCs. The following table summarizes their differences.

Item	Connect VPCs over the Internet	Connect VPCs with Express Connect
Communication performance and availability	Prone to high latency and packet loss	Guaranteed by the state-of-the-art infrastructure of Alibaba Cloud
Expense	High costs for Internet bandwidth or data transfer	Low bandwidth costs for cross-region communication Free for communication between VPCs within the same region
Security	Sensitive to data breaches	Guaranteed by communication isolation based on the network virtualization technology of Alibaba Cloud

24.3. Architecture

Based on the Layer-3 overlay scheme of Software-Defined Networking (SDN) and the network virtualization technology, Apsara Stack Cloud isolates the ports accessed by physical connections and abstracts the ports into virtual border routers (VBRs). Before data is transferred to VPCs, data packets are encapsulated in switches and then go through tunnel encapsulation when they travel over physical connections to the VRouters of VPCs.

The following figure shows the architecture of using Express Connect to connect two VPCs.



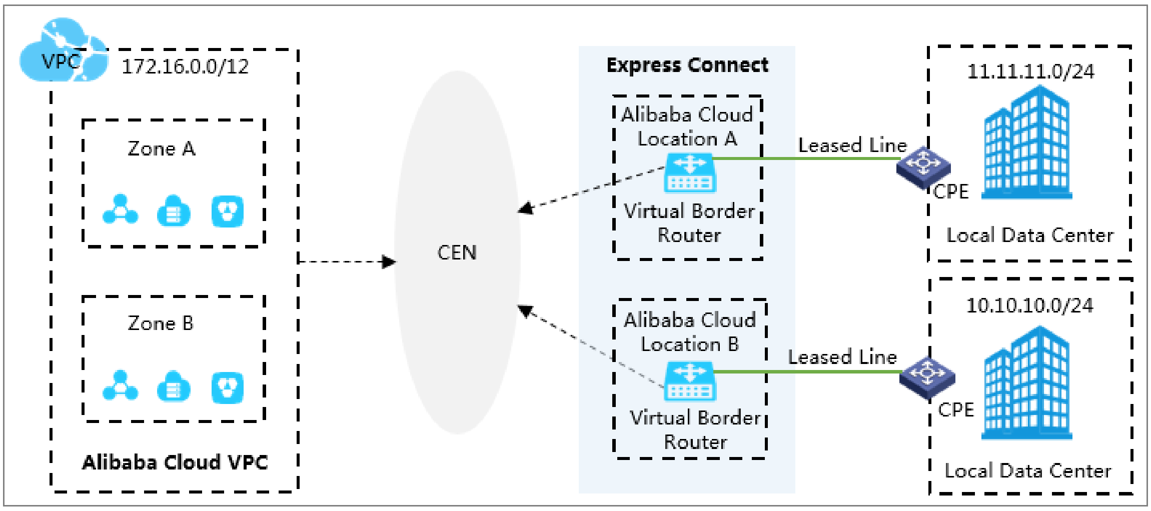
24.4. Scenarios

Express Connect enables reliable, secure, and high-speed communication between data centers and VPCs. You can use Express Connect in the following scenarios to facilitate communication in different network architectures.

Scenario 1: Implement disaster recovery for large and medium-sized enterprises

You can establish multiple physical connections between different access points and a VPC to provide high availability for your services. This prevents service outages caused by severed fiber-optic cables, device faults, or access point malfunctions.

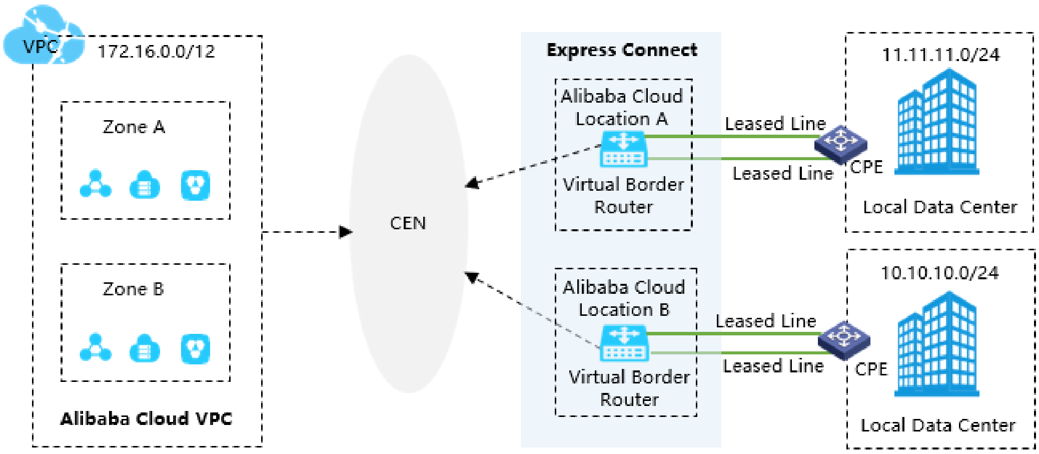
You can use a physical connection interface or use a shared pre-deployed leased line of a partner in this scenario.



Scenario 2: Build highly-scalable and high-availability architectures for large enterprises

When you deploy services on the cloud to provide better support for fast-growing workloads together with on-premises solutions, you can use the equal-cost multi-path routing (ECMP) strategy of Express Connect to offer substantial increases in bandwidth that handles terabytes of data. This prevents service outages caused by device faults, leased line failures, or access point malfunctions.

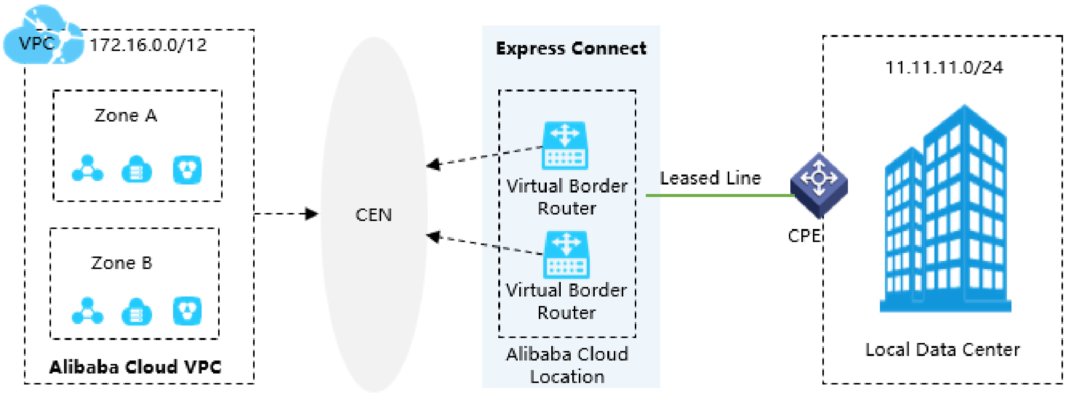
You can only use a physical connection interface in this scenario.



Scenario 3: Establish a basic network architecture for non-critical services

For services that do not require high scalability and availability, such as running tests in a testing environment on the cloud, we recommend that you use Express Connect to directly connect your data center to Alibaba Cloud. This ensures security and reliability of communication between your data center and Alibaba Cloud.

You can use a physical connection interface or use a shared pre-deployed leased line of ad partner in this scenario.



24.5. Terms

This topic describes the terms used in Express Connect.

The following table lists the terms and their descriptions.

Term	Description

Term	Description
Express Connect	An Apsara Stack Cloud service that helps you build private network communication channels between VPCs or between VPCs and data centers. It provides better flexibility for your network topology and enhances the performance and security of cross-network communication.
Virtual Private Cloud (VPC)	A custom private network created on Apsara Stack Cloud. VPCs are logically isolated from each other. You can create and manage your cloud instances, such as Elastic Compute Service (ECS) instances, ApsaraDB instances, and Server Load Balancer (SLB) instances in your VPC.
Virtual Border Router (VBR)	A router between the customer-premises equipment (CPE) and a VPC. It serves as a bridge for data exchange between your data center and the VPC.
VRouter	A hub that connects all VSwitches in a VPC and serves as a gateway to connect the VPC with other networks.
Cloud Enterprise Network (CEN)	A network that helps you build private network communication channels between VPCs or between VPCs and data centers. It features automatic route distribution and learning, faster convergence, and improved quality and security of cross-network communication. CEN allows you to connect all your network resources to build a network with enterprise-level communication capabilities. By using CEN, you do not need to configure routes for VPCs, which simplifies network operation and maintenance.
Physical connection interface	A physical interface that is used by a leased line to access an Apsara Stack Cloud access point.
Access point	A geographical location where a physical connection is connected to Apsara Stack Cloud. Two access devices are deployed in each access point. Each region has one or more access points to enable communication between local data centers and VPCs.
Border Gateway Protocol (BGP)	A routing protocol designed to exchange routing and reachability information among gateways, Internet, or autonomous systems.
Equal-cost multi-path (ECMP) link aggregation	A mechanism that is used to increase the bandwidth of physical connections by using ECMP routing, or in other words, by connecting multiple leased lines to the same device and associating multiple physical connections with the same VBR.
Exclusive physical connection	A physical connection that uses a dedicated physical port and requires a dedicated leased line of a third-party service provider to connect a data center to an Apsara Stack Cloud access point.

25.Log Service

25.1. What is Log Service?

Log Service (SLS) is a one-stop logging service developed by Alibaba Cloud that is widely used by Alibaba Group in big data scenarios. You can use Log Service to collect, query, and consume log data without the need to invest in in-house data collection and processing systems. This enables you to focus on your business, improving business efficiency and helping your business to expand.

Log Service provides the following features:

- Log collection: Log Service allows you to collect events, binary logs, and text logs in real time through multiple methods, such as Logtail and JavaScript.
- Query and analysis: Log Service allows you to query and analyze the collected log data and view analysis results on charts and dashboards.
- Status alert: Log Service can automatically run query statements at regular intervals after you create an alert task. If the query results meet the conditions of the alert task, Log Service sends an alert to the specified recipients in real time.
- Real-time consumption: Log Service provides real-time consumption interfaces through which log consumers can consume log data.

25.2. Benefits

This topic describes the benefits of Log Service.

Fully managed service

- Log Service is easy to access and use.
- LogHub provides all features of Kafka, outputs monitoring and alert data, and supports auto scaling (by PBs per day).
- LogSearch/Analytics allows you to query log data, view log data on dashboards, and configure alerts.
- Log Service provides more than 30 access methods to seamlessly connect with open-source software, such as Storm and Spark Streaming.

Comprehensive ecosystem

- LogHub supports more than 30 types of log data sources, such as embedded devices, web pages, servers, and programs. LogHub can connect with consumption systems, such as Storm and Spark Streaming.
- LogSearch/Analytics provides complete query and analysis syntax and supports connection with Grafana based on the SQL-92 and JDBC protocols.

Real-time response

- LogHub: Data can be consumed immediately after being written to Log Service. Logtail acts as an agent to collect and deliver data in real time.
- LogSearch/Analytics: Data can be queried and analyzed immediately after being written to Log Service. In scenarios where multiple query and analysis conditions are specified, data can be queried and analyzed in seconds.

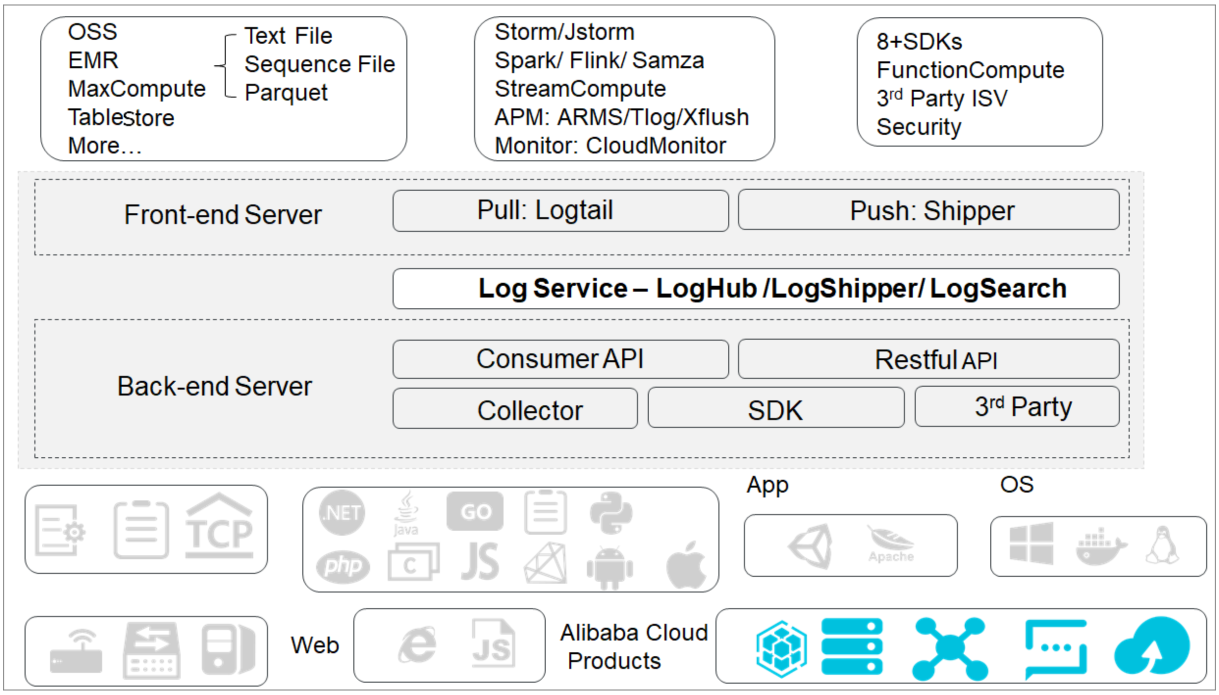
Scalability

Log Service provides auto scaling capabilities for petabytes of data.

25.3. Architecture

This topic describes the architecture of Log Service.

The following figure shows the architecture of Log Service.



Logtail

Logtail is an agent that is used to collect logs. It has the following features:

- Non-intrusive file-based log collection
  - Logtail only reads log files.
  - Log collection is non-intrusive.
- High security and reliability
  - Logtail can rotate logs without data loss.
  - Logtail supports local caching.
  - Logtail re-attempts to collect logs if network exceptions occur.
- Ease management
  - Logtail can be accessed by using a web client.
  - Logtail can be configured in the console.
- Comprehensive self-protection
  - Logtail monitors CPU and memory usage of its processes in real time.
  - Logtail allows you to set an upper limit on the resource usage of its processes.

Frontend servers

Frontend servers are built on LVS and NGINX to provide the following features:

- Data transfer over HTTP and REST protocols
- Horizontal scaling
    - The processing capabilities can be increased when traffic increases.
    - Frontend servers can be added.
- High throughput and low latency
    - Asynchronous processing: If an exception occurs when a single request is sent, other requests are not affected.
    - LZ4 compression: This compression method increases the processing capabilities of individual servers and reduces network bandwidth consumption.

Backend servers

The backend service is a distributed process that is deployed on multiple servers. The service can store, index, and query logs in a Logstore in real time. The backend service has the following features:

- High data security
    - Each log is saved to three copies and stored on different servers.
    - If the disk is damaged or the server fails, data is automatically copied and restored.
- Stable services
    - If the process does not respond or the server fails, data in Logstores is automatically migrated.
    - Automatic load balancing ensures that traffic is evenly distributed among different servers.
    - Strict resource quota limits prevent unexpected operations of a single user from affecting other users.
- Horizontal scaling
    - A shard is the basic unit for horizontal scaling.
    - You can add shards to increase the throughput based on your business requirements.

25.4. Features

25.4.1. Core features

This topic describes the two core features of Log Service: real-time log collection and consumption (LogHub) and real-time log search and analysis (Log Search/Analytics).

LogHub

LogHub allows you to collect logs without data loss by using various methods. These methods include clients, websites, protocols, SDKs, and API operations (for mobile apps and games). You can also consume logs by using SDKs, Storm Spout, and Spark Client. LogHub supports real-time log collection and consumption in multiple formats. You can use LogHub to streamline the collection and consumption of logs across multiple devices and sources.

Features:

- LogHub collects real-time log data from Elastic Compute Service (ECS) instances, containers, mobile terminals, open-source software, and JavaScript. The log data can be metrics, events, binary logs, and text logs.

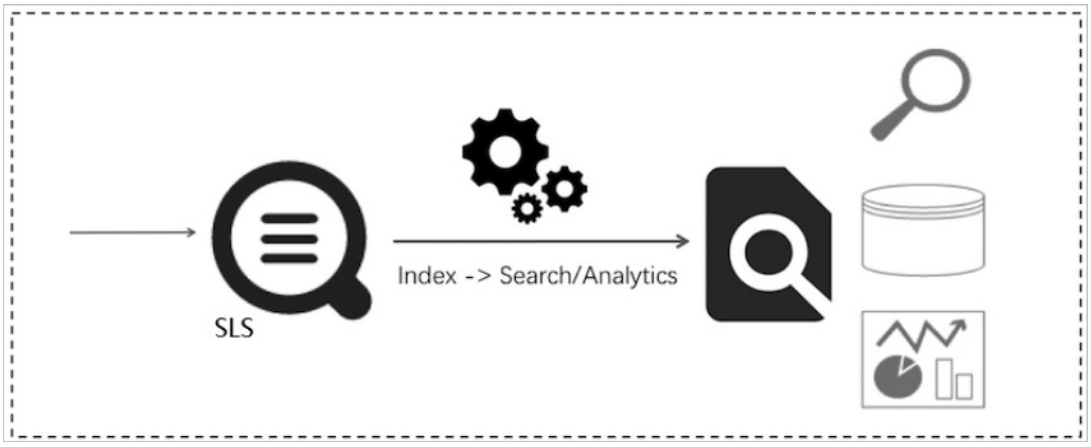
- A real-time consumption interface is provided to connect Log Service to real-time computing engines and services.

Log Search/Analytics

You can use Log Service to index, query, and analyze collected log data in real time. Log Service can generate dynamic reports based on the query and analysis results. It can also generate visualized reports of log data in multiple scenarios.

- Query: You can query data by using keywords, fuzzy match, and contextual query. You can also query data within a specified range.
- Statistics: You can use various query methods, such as SQL aggregate functions.
- Visualization: You can view log analysis on dashboards and export the results as reports.
- Interconnection: You can connect Log Service to Grafana based on the Java Database Connectivity (JDBC) and SQL-92 protocols.

Log search/analytics



25.4.2. Other features

25.4.2.1. Logs

Logs are records of changes that occur in a system. The operations on specific objects and the relevant results are recorded in logs in chronological order.

Logs in Log Service

Log Service supports different types of logs, such as log files, events, binary logs, and metrics. Each log file consists of one or more log entries. Each log entry describes a single system event and is the smallest unit of data that can be processed in Log Service.

Log Service uses a semi-structured data model to define logs. This model consists of the topic, time, content, and source fields.

Log Service has different format requirements on different log fields, as described in the following table.

Field	Description	Format

Field	Description	Format
Topic	The user-defined field in a log. This field can be used to mark a group of logs. For example, access logs can be marked based on sites.	The value of this field can be a string up to 128 bytes in length, including an empty string. The default value of this field is an empty string.
Time	The time when a log was generated. This field is a reserved field. The value of this field is directly extracted from the time in the log.	This value is a Unix timestamp representing the number of seconds that have elapsed since the epoch time January 1, 1970, 00:00:00 UTC.
Content	The specific content of a log. The content consists of one or more content items. Each item is a key-value pair.	The key is a UTF-8 encoded string up to 128 bytes in length. It can contain letters, digits, and underscores (_). The key cannot start with a digit and cannot contain the following keywords: _time , source , topic__ , partition time _ , extract others , and __extract_others__ . The value can be a string up to 1024 × 1024 bytes in length.
Source	The source of a log. For example, the source can be the IP address of the server where the log was generated.	The value of this field can be a string up to 128 bytes in length. The default value of this field is an empty string.

Logs are used in various formats in different scenarios. The following example shows how to map a raw NGINX access log to the log data model of Log Service. In this example, the IP address of your NGINX server is `10.249.201.117`. A raw log generated on this server is as follows:

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

The following table describes how to map this raw log to the log data model of Log Service.

Field	Field value	Description
Topic	""	An empty string is used.
Time	1330589527	The exact time when the log was generated is used. The value indicates the number of seconds that have elapsed since the epoch time January 1, 1970, 00:00:00 UTC. The value is converted from the timestamp in the raw log.

Field	Field value	Description
Content	Key-value pair	The specific content of the log is used.
Source	"10.249.201.117"	The IP address of the server is used as the log source.

You can decide how to extract the content of a raw log to create key-value pairs. The following table lists some key-value pairs.

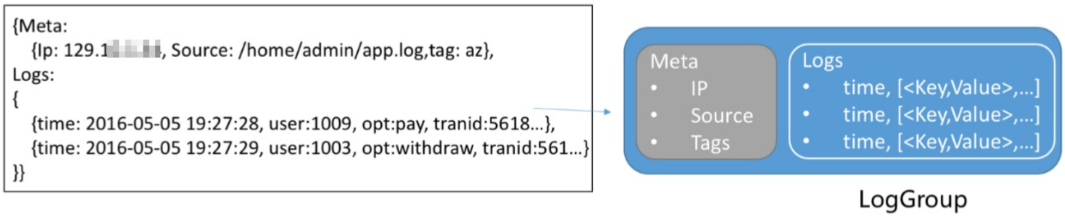
Key	Value
ip	10.1.168.193
method	GET
status	200
length	5
ref_url	N/A
browser	Mozilla/5.0 (X11; Linux i686 on x86_64; rv:10.0.2) Gecko/20100101 Firefox/10.0.2

Log groups

A log group is a collection of logs and is the basic unit for read and write operations.

The maximum capacity of a log group is 4,096 logs or 10 MB of logs.

Log group



```
{Meta:
  {Ip: 129.1[redacted], Source: /home/admin/app.log,tag: az},
Logs:
{
  {time: 2016-05-05 19:27:28, user:1009, opt:pay, tranid:5618...},
  {time: 2016-05-05 19:27:29, user:1003, opt:withdraw, tranid:561...}
}}
```

25.4.2.2. Project

This topic introduces the concept of project. It also describes the features of a project.

A project in Log Service is the resource management unit that is used to separate and manage different resources. You can use a project to manage all the logs and related log sources of an application. You can also use a project to manage Logstores and Logtail configuration files that are used to collect logs. A project serves as an endpoint that allows authorized access to the resources of Log Service.

Projects provide the following features:

- Projects allow you to manage different Logstores. Logs that you collect and store in Log Service are generated from different projects, services, and environments. You can specify different projects for these logs to facilitate data consumption, exporting, and indexing. You can also grant permissions on these projects to different users.

- Projects serve as endpoints that allow authorized access to the resources of Log Service. Log Service allocates an exclusive endpoint to each project. You can use the endpoint of a project to read, write, and manage log data.

25.4.2.3. Logstore

This topic introduces the concept of Logstore. It also describes the features of a Logstore.

A Logstore in Log Service is the unit that is used to collect, store, and query data. Each Logstore belongs to only one project. However, you can create multiple Logstores for a single project. In most cases, a Logstore is created for each type of log in an application. For example, you have a gaming application called big-game, and three types of logs are stored on the server: operation_log, application_log, and access_log. You can create a project named big-game, and then create three Logstores in this project to collect, store, and query these logs.

You must specify a Logstore when you write or query logs. If you transfer log data to MaxCompute for offline analysis, the logs in each Logstore are transferred to a MaxCompute Table.

Logstores provide the following features:

- Logs can be written to Logstores in real time.
- Logs stored in Logstores can be consumed in real time.
- Indexes can be created for Logstores to support real-time log queries.

25.4.2.4. Shard

This topic describes the range, read/write capacities, and status of the shards of a Logstore.

Log data is stored on a shard of a Logstore where read/write operations are performed. A Logstore consists of multiple shards. Each shard has an MD5 hash key range. Each range is left-closed and right-open and does not overlap with the ranges of other shards. The entire range of MD5 hash key values consists of all these different ranges.

Range

You must specify the number of shards when you create a Logstore. The entire range of MD5 hash key values is evenly divided by Log Service based on the specified number of shards. The MD5 hash key ranges of the shards must be within the following range: [00000000000000000000000000000000, ffffffffffffffffffffffffffffffff).

The range of each shard is a left-closed and right-open interval. The range consists of the following keys:

- BeginKey: indicates the start of a shard. This value is included in the range.
- EndKey: indicates the end of a shard. This value is excluded from the range.

You can use the hash key values of the ranges to write logs to specified shards. You can also use the values to identify shards that you want to split or merge. When you read data from a shard, you must specify the shard. When you write data to a shard, you can use the load balancing method or specify a hash key. If you use the load balancing method, each data packet is randomly written to an available shard. If you specify a hash key, data is written to the shard whose range includes the value of the specified hash key.

For example, a Logstore has four shards and the MD5 hash value range of the Logstore is [00, FF). The following table describes the ranges of the shards.

Shard ID	MD5 hash key range
Shard0	[00,40)
Shard1	[40,80)
Shard2	[80,C0)
Shard3	[C0,FF)

If you set the MD5 hash key value to 5F when you write logs in the hash key mode, the log data is written to Shard1 because the range of Shard1 contains the MD5 hash key value 5F. If you set the MD5 hash key value to 8C, the log data is written to Shard2 because the range of Shard2 contains the MD5 hash key value 8C.

Read/write capacities

Each shard has limits on read/write capacities. We recommend that you adjust the number of shards based on the actual data traffic. If the data traffic exceeds the read/write capacities of a shard, you can split the shard into two shards to increase the total read/write capacities. If the data traffic is much less than the read/write capacities of a shard, you can merge the shard with the adjacent shard to reduce the total read/write capacities and save costs.

Note

- If a 403 or 500 error code is repeatedly returned when you use API operations to write logs to a Logstore, you can check the Log Service monitoring metrics. You can then determine whether to increase the number of shards.
- If the data traffic exceeds the read/write capacities of your shards, Log Service attempts to provide the best services possible. However, service quality cannot be guaranteed.

The status of the shards

A shard can be in one of the following states:

- readwrite: The shard supports read/write operations.
- readonly: The shard supports only read operations.

When you create a shard, the shard is in the readwrite state. After you split or merge shards, the original shards are in the readonly state and the new shards are in the readwrite state. The status of a shard does not affect its read performance. You can write data to a shard in the readwrite state, but you cannot write data to a shard in the readonly state.

When you split a shard, you must specify the ID of a shard that is in the readwrite state and an MD5 hash key. The value of the MD5 hash key must be greater than the value of the BeginKey and less than the value of the EndKey of a shard. After a shard is split, two shards are added. The status of the original shard changes from readwrite to readonly. Data can still be read from the shard, but cannot be written to the shard. The two new shards are in the readwrite state. They are placed next to the original shard. The total MD5 hash key ranges of the two shards cover the range of the original shard.

When you merge shards, you must specify a shard that is in the readwrite state. In addition, the shard cannot be the last shard that is in the readwrite state. After you specify a shard that is in the readwrite state, Log Service finds the adjacent shard and merges these two shards into a single shard. After you merge the shards, the status of the original shards change from readwrite to readonly. Data can still be read from the shards, but cannot be written to the shards. A shard in the readwrite state is generated. The MD5 hash key range of the shard covers the total range of the original shards.

25.4.2.5. Log topic

This topic introduces the concept of log topic.

A log topic is used to classify logs in a Logstore. You can specify a topic for logs that are written to Log Service. You can also specify a topic when you query logs. For example, you can use your user ID as the log topic when you write logs to Log Service. In this way, you can view only your own logs based on the log topic when you query the logs. If you do not need to classify logs in a Logstore, use the same topic for all logs.

> ? **Note** The default log topic for log data writes and queries is an empty string. If you do not specify a log topic, you can use the default topic to write or query logs.

25.5. Scenarios

This topic describes the application scenarios of Log Service. The scenarios include data collection, real-time computing, data warehousing, offline analysis, product operation and analysis, and O&M.

Data collection and consumption

You can use the LogHub feature of Log Service to collect large amounts of log data in real time. The log data can be metrics, events, binary logs, and text logs.

Benefits:

- Ease of use: More than 30 real-time data collection methods are provided. This simplifies the construction of log O&M platform and reduces your O&M workloads.
- Elastic scalability: Log Service supports automatic scaling based on traffic and business requirements. This simplifies the handling of traffic spikes that result from business growth.

ETL and stream processing

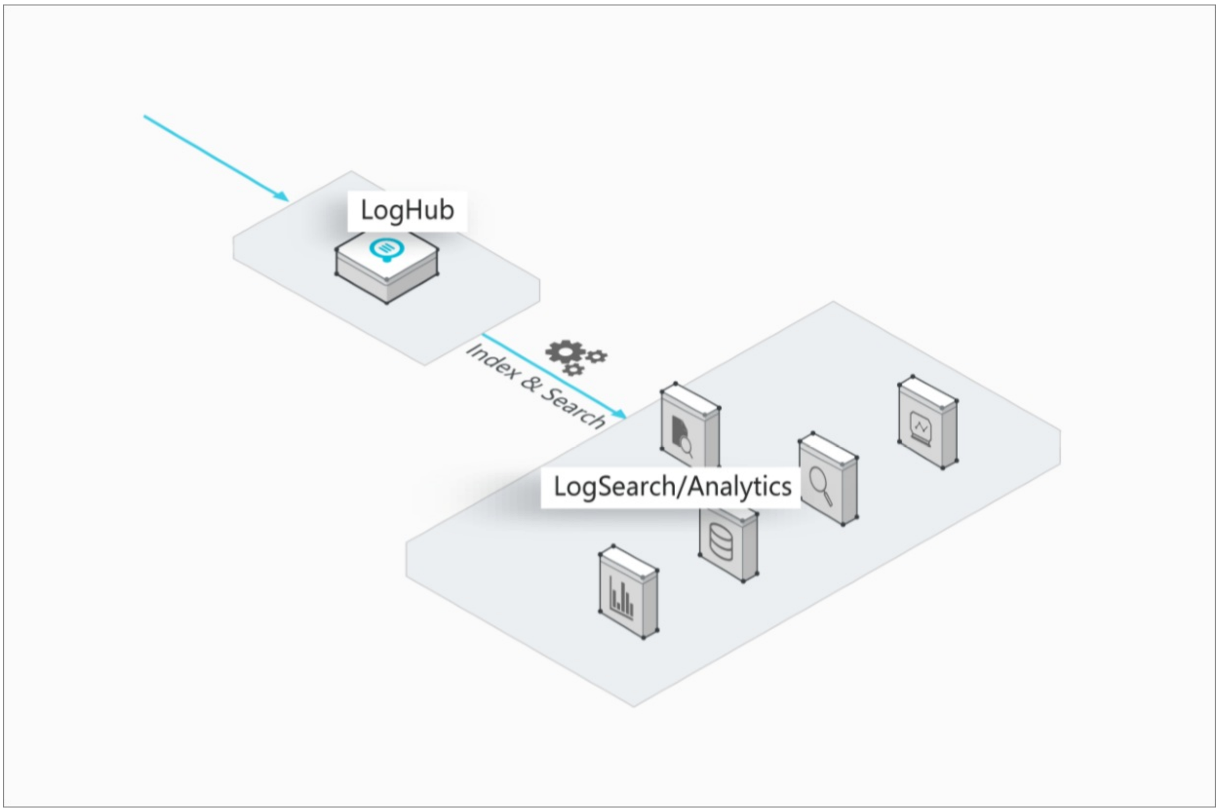
You can connect LogHub to multiple real-time computing engines and services. LogHub can monitor the processing progress and generate alerts based on the monitoring results. You can also use SDKs or API operations to consume logs.

- Ease of use: LogHub provides SDKs in multiple programming languages and frameworks. You can connect LogHub to various stream computing engines.
- Comprehensive features: LogHub can monitor large amounts of data and generate alerts based on the monitoring results.
- Elastic scalability: LogHub supports scaling to handle petabyte of data with zero latency.

Real-time log search and analysis

The log search/analytics feature allows you to index LogHub data in real time and query data by using keywords, fuzzy match, contextual query, and SQL aggregate functions. You can also query data within a specified range.

- Real-time query: Data can be immediately queried after it is written to Log Service.
- High efficiency and low costs: You can index petabytes of data each day. Costs are 85% lower compared with user-defined systems.
- Strong analysis capability: Multiple query methods and SQL aggregate functions are supported. Log Service can also generate visualized reports and alerts based on log data.



25.6. Limits

This topic describes the limits of Log Service.

Limits of resources

Resource	Limit	Note
Project	You can create a maximum of 10 projects in an organization.	If you need to increase the quota, submit a ticket.
Logstore	You can create a maximum of 100 Logstores in a project.	If you need to increase the quota, submit a ticket.

Resource	Limit	Note
Shard	● You can create a maximum of 10 shards in a Logstore. However, you can **split** the shards to increase the number of shards. ● You can create a maximum of 100 shards in a project.	If you need to increase the quota, submit a ticket.
Dashboard	● You can create a maximum of five dashboards for a project. ● Each dashboard can contain a maximum of 10 charts.	If you need to increase the quota, submit a ticket.
Saved search	You can create a maximum of 10 saved searches for a project.	If you need to increase the quota, submit a ticket.
Logtail configuration file	You can create a maximum of 100 Logtail configuration files in a project.	If you need to increase the quota, submit a ticket.
Consumer group	You can create a maximum of 10 consumer groups in a project.	If you need to increase the quota, submit a ticket.
Machine group	You can create a maximum of 100 machine groups in a project.	If you need to increase the quota, submit a ticket.
Log retention time	Logs that are collected to the server can be kept for a maximum of 365 days.	If you need to increase the quota, submit a ticket.

25.7. Terms

This topic introduces the basic concepts of Log Service.

Log

Logs are abstract records of changes within a system. The records are ordered by time. The records contain information about operations on specific objects and results of the operations. Log data is stored in different forms such as log files, log events, binary logs, and metric data. Each log file consists of one or more log entries. A log entry is the basic unit of data that can be processed in Log Service. Each log entry describes a single system event.

Log group

A log group is a collection of logs. The groups are the basic units that are used for read and write operations.

Log topic

A log topic is used to classify logs in a Logstore. Topics can be specified when logs are written to Log Service. Topics also serve as filters when logs are queried.

Project

A project in Log Service is the resource management unit that is used to separate and manage different resources. You can use a project to manage all the logs and related log sources of an application. You can also use a project to manage Logstores and Logtail configuration files. A project serves as an endpoint to access the resources of Log Service.

Logstore

A Logstore is the unit used in Log Service for log data collection, storage, and query. Each Logstore belongs to only one project. However, you can create multiple Logstores for a single project.

Shard

A Logstore consists of multiple shards. Each shard has an MD5 hash key range. The range is left-closed and right-open and does not overlap with the ranges of other shards. The entire range of MD5 hash key values consists of all these different ranges.

26.Apsara Stack Security

26.1. What is Apsara Stack Security?

Apsara Stack Security is a solution that provides Apsara Stack assets with a full suite of security features, such as network, server, application, data, and security management.

Background information

Traditional security solutions for IT services detect attacks on network perimeters. These solutions use hardware products such as firewalls and intrusion prevention systems (IPSs) to protect networks against attacks.

With the development of cloud computing, an increasing number of enterprises and organizations use cloud computing services instead of traditional IT services. Cloud computing features low costs, on-demand flexible configuration, and high resource utilization. Cloud computing environments do not have definite network perimeters. As a result, traditional security solutions cannot effectively safeguard cloud assets.

With the powerful data analysis capabilities and professional security operations team of Alibaba Cloud, Apsara Stack Security provides integrated security protection services for networks, applications, and servers.

Complete security solution

Apsara Stack Security consists of Apsara Stack Security Standard Edition and optional security services to provide a comprehensive security solution.

Security domain	Service name	Description
Security management	Threat Detection Service (TDS)	Monitors traffic and overall security status to audit and centrally manage security.
Server security	Server Guard	Protects ECS instances against intrusions and malicious code.
Application security	Web Application Firewall (WAF)	Protects web applications against attacks and ensures that mobile and PC users can securely access web applications over the Internet.
Network security	Anti-DDoS	Ensures the availability of network links and improves business continuity.
	Cloud Firewall	Allows you to centrally manage access control policies for traffic transferred within your business system (east-west) and between the Internet and your business system (north-south).
Data security	Sensitive Data Discovery and Protection (SDDP)	Prevents data leaks and helps your business system meet compliance requirements.
Security O&M service	On-premises security service	Helps you establish and optimize your cloud security system to protect your business system against attacks by using security features of Apsara Stack Security and other Apsara Stack services.

Security domain	Service name	Description

26.2. Benefits

As a pioneer of cloud security, Apsara Stack Security has received a variety of authoritative certifications. Through mature security systems and advanced security technologies, Apsara Stack Security can fully protect the security of the Apsara Stack environment.

Pioneer of cloud security

The Apsara Stack Security team has accumulated a wealth of security experience by protecting all internal business systems of Alibaba Group since 2005. Since its release in 2011, Apsara Stack Security has become a pioneer in providing comprehensive protection for cloud security.

Apsara Stack Security protects more than 40% of all websites in China. It prevents more than 50% of all distributed denial of service (DDoS) attacks and blocks up to 3.6 billion attacks every day. It has fixed over 6.13 million vulnerabilities over the last year.

Mature systems and advanced technologies

Apsara Stack Security is a service born from ten years of protection experience. After a decade of experience in providing security operations services for the internal business systems of Alibaba Group, Alibaba has obtained considerable security research achievements, security data, and security operations methods, and has built a professional cloud security team. Apsara Stack Security brings together the rich experience of these experts to develop the sophisticated systems that provide enhanced security for cloud computing platforms. This service can protect the cloud platform, cloud network environments, and cloud business systems of Apsara Stack users.

Comparison with traditional security products

Feature	Traditional security product	Apsara Stack Security
Comprehensive industry-leading security capabilities among Internet enterprises	A traditional security service provider only has limited products and features and cannot provide a comprehensive security protection system.	Alibaba has accumulated a large number of intelligence sources through years of attack prevention experience. This has allowed it to detect common Internet attacks including zero-day exploits, and provide comprehensive security capabilities.
Early risk detection	Traditional security service providers cannot detect risks due to a lack of complete monitoring systems.	Apsara Stack Security can detect and respond to critical vulnerabilities and security events quickly to prevent security issues.

Feature	Traditional security product	Apsara Stack Security
Security big data modeling analysis	Traditional security service providers cannot detect threats through signature scanning. The traditional log analysis feature only provides data collection and reporting. It does not provide data modeling analysis.	Big data modeling analysis enables Apsara Stack Security to detect threats in the entire network and display the security data. More than 30 algorithmic models are used to analyze the historical data, network data, and server data. This enables security situation awareness.
Scalability and decoupling with hardware	Traditional security products are developed based on the existing hardware devices. Security product software relies on the virtual machines created on virtualization platforms.	• Hardware and software decoupling: All modules are developed based on the cloud computing architecture and the common x86 hardware platform, and therefore do not rely on specific hardware. • Scalability: You can increase the amount of hardware for higher performance without the need to change the network architecture.
Collaboration between the network and servers	Traditional security service providers increase security features by adding devices. The devices can only collect device logs and status data and display the data on the management platform. They cannot collaborate to provide more features.	Apsara Stack Security provides complete Internet protection to guarantee the security of networks, applications, and servers. The security modules interact with each other to form a comprehensive protection system that blocks attacks effectively.
Compatibility with all data center environments and decoupling with specific cloud platforms	Most traditional security products are provided as hardware appliances. This makes the product incompatible with the cloud platforms based on Software Defined Network (SDN) technology.	Based on the interactions between servers and the operating system, Apsara Stack Security detects threats at the network perimeter through data analysis. This enables the service compatibility with all data center environments by avoiding the complex network topology inside the data centers.

26.3. Architecture

Apsara Stack Security consists of Apsara Stack Security Standard Edition and optional security services.

Apsara Stack Security Standard Edition

- Threat Detection Service

This module collects network traffic and server information, and detects possible vulnerability exploits, intrusions, and virus attacks through machine learning and data modeling. It also provides you with up-to-date information about ongoing attacks to help you monitor the security status of your businesses.

- Network Traffic Monitoring System

  This module is deployed on the network perimeter of Apsara Stack. It allows you to inspect and analyze each inbound or outbound packet of an Apsara Stack network through traffic mirroring. The analysis results are used by other Apsara Stack Security modules.

- Server Guard

  This module safeguards ECS instances by providing security features such as vulnerability management, baseline check, intrusion detection, and asset management. To do this, the module performs operations such as log monitoring, file analysis, and signature scanning.

- Web Application Firewall

  This module protects web applications against common web attacks reported by Open Web Application Security Project (OWASP), such as Structured Query Language (SQL) injections, cross-site scripting (XSS), exploitation of vulnerabilities in web server plug-ins, Trojan uploads, and unauthorized access. It also blocks a large number of malicious visits to avoid data leaks and ensure both the security and availability of your websites.

Apsara Stack Security Standard Edition also provides on-premises security services. These services help you better use the features of Apsara Stack products such as Apsara Stack Security to secure your applications.

On-premises security services include pre-release security assessment, access control policy management, Apsara Stack Security configuration, periodic security check, routine security inspection, and urgent event handling. These services cover the entire lifecycle of your businesses in Apsara Stack and help you create a security operations system. This system enhances the security of your application systems and ensures both the security and stability of your businesses.

Optional security services

You can also choose the following service modules to enhance your system security.

- DDoS Traffic Scrubbing

  This module detects and blocks distributed denial of service (DDoS) attacks.

- Cloud Firewall

  This module sorts and isolates businesses based on visualized business data to implement access control over east-west traffic in Apsara Stack.

- Sensitive Data Discovery and Protection

  This module uses Alibaba Cloud's big data analytics capabilities and artificial intelligence (AI) technologies to detect and classify sensitive data based on your business requirements. It masks sensitive data both in transit and at rest, monitors dataflows, and detects abnormal activities. This module provides visible, controllable, and industry-compliant security protection for your sensitive data by means of precise detection and analysis.

26.4. Features

26.4.1. Apsara Stack Security Standard Edition

Apsara Stack Security is developed based on the Apsara Stack environment and adopts a cloud security architecture that enables in-depth defense and multi-module collaboration. Unlike traditional software and hardware security products, Apsara Stack Security provides comprehensive and integrated cloud security protection at multiple layers, such as the network layer, application layer, and server layer.

The following table describes the features of Apsara Stack Security Standard Edition.

Features of Apsara Stack Security Standard Edition

Module	Feature	Description
Network Traffic Monitoring System	Traffic collection and analysis	Uses a bypass in traffic mirroring mode to collect inbound and outbound traffic that passes through the interconnection switch (ISW), and generates a traffic diagram.
	Malicious server identification	Detects attacks launched by internal servers and identifies the internal servers.
	Abnormal traffic detection	Uses a bypass in traffic mirroring mode to detect abnormal traffic that has exceeded a specific threshold.
	Web application protection	Uses a bypass to block common attacks on web applications at the network layer based on default web attack detection rules.
Server Guard	Baseline check	Performs security baseline checks for Elastic Compute Service (ECS) instances. The check items include accounts, weak passwords, and at-risk configuration items. The baseline checks ensure that the ECS instances comply with the security standards for enterprise servers.
	Vulnerability management	• Scans ECS instances for software vulnerabilities, and provides suggestions on vulnerability fixes. • Provides quick fixes for critical vulnerabilities in applications and operating systems on your ECS instances.
	Webshell detection and removal	Accurately detects and removes webshells based on specified rules, and allows you to manually quarantine webshells.
	Brute-force attack blocking	Detects and blocks brute-force attacks in real time.
	Unusual logon alerting	Detects unusual logons based on the approved logon settings, and generates alerts.
	Suspicious server detection	Detects suspicious activities such as reverse shells, Java processes running CMD commands, and unusual file downloads with Bash.
	Asset fingerprints	Collects up-to-date information about the servers, such as ports, accounts, processes, and applications, to perform event tracking.

Module	Feature	Description
	Log retrieval	Centrally manages server logs of processes, networks, and system logons. This helps you quickly locate the cause of an issue from the logs.
	Detection overview	Provides the following modules: • Detection overview: provides statistics on protection for the last 24 hours and the last 30 days. • Access status monitor: displays the top 100 access requests in real time. • Export detection report: allows you to export daily reports, weekly reports, and scheduled task reports. • Attack detection statistics: provides statistics on attack detection.
	Protection logs	Provides the following modules: • Attack detection logs: provides attack detection logs. The log list displays the processing results, attacked addresses, attack types, attacker IP addresses, and attack time. You can view log details about each attack. • HTTP flood detection logs: provides HTTP flood protection logs. The log list displays logs for matched HTTP flood protection rules, including the request URL, the name of the matched rule, and the match time. You can filter logs based on the event generation time and the name of the HTTP flood protection rule. • System operations log: provides system operations logs, including usernames, operations, and IP addresses. • Access log: provides access logs, including the access address, destination IP address, source IP address, request method, and response code.

Module	Feature	Description
WAF	Protection configuration	Provides the following modules: • Protection site management: allows you to create, delete, modify, enable, and disable function forwarding proxies of a protected site. • Customized rules: allows you to create, delete, enable, and disable custom rules. This implements fine-grained HTTP access control for websites. • Website protection policies: ○ Supports decoding methods, such as URL decoding, JSON parsing, Base64 decoding, hexadecimal conversion, backslash unescape, XML parsing, PHP deserialization, and UTF-7 decoding. ○ Detects SQL injections, cross-site scripting (XSS), intelligence, cross-site request forgery (CSRF), server-side request forgery (SSRF), Hypertext Preprocessor (PHP) deserialization, Java deserialization, Active Server Pages (ASP) code injections, file inclusion attacks, file upload attacks, PHP code injections, command injections, crawlers, and server responses. ○ Provides five built-in protection templates, including the template with default protection policies, monitoring mode template, anti-DDoS template, template for financial customers, and template for Internet customers. WAF allows you to customize the decoding algorithms in the templates, enable or disable each attack detection module separately, and configure the detection granularity. WAF also allows you to specify the Block Status Code parameter. ○ Allows you to enable HTTP response detection and configure the length of the response body in detection rules. ○ Allows you to configure the length of the request body in detection rules. ○ Allows you to enable or disable detection timeout settings. • HTTP flood protection: allows you to configure access frequency control rules for domain names and URLs. This restricts the access frequency of IP addresses or sessions that meet the criteria, or blocks these IP addresses or sessions. Restricts the access frequency of known IP addresses or sessions or blocks these IP addresses or sessions. Supports the HTTP flood protection whitelist function. HTTP flood protection rules are not applicable to IP addresses or sessions in a whitelist.
	System management	Displays the workload, network, and detection statuses of a node. You can configure syslog to send logs and also configure the service- and system-related alert thresholds.

Module	Feature	Description
Threat Detection Service	Overview	Provides a comprehensive security overview with statistics on security score, asset status, unhandled alerts, and handled alerts.
	Visualization	Displays the security data on the dashboard, including assets, vulnerabilities, baselines, attack sources, and attack distribution.
	Security alerts	Allows you to view and handle security events, including suspicious process, webshells, unusual logons, sensitive file tampering, malicious processes, suspicious network connections, and web application threat detection.
	Attack analysis	Protects against common attacks on web applications and brute-force attacks.
	Cloud service check	Checks the security configurations of cloud services from the aspects of network access control and data security. It supports periodic checks that run automatically and manual checks. You can verify the check results or configure whitelist policies for the check results.
	Application whitelists	Allows you to add servers to the whitelist based on intelligent learning and identifies programs as trusted, suspicious, or malicious based on the whitelist. Unauthorized processes will be terminated.
	Assets	• Server: displays the security statuses for servers. You can view the numbers of all servers, risky servers, unprotected servers, inactive servers, and new servers. • Cloud product: provides security status information for cloud services and supports SLB and NAT.
	Security reports	Allows you to query reports. For example, you can retrieve historical reports by report name.
Security Operations Center (SOC)	Dashboard	Allows security administrators to view the overall statistics and perform operations.
	Security monitoring	Allows you to view the security events of all users and the platform.
	Asset management	Allows you to view the security status of user assets and platform assets.
	Log analysis	Analyzes logs from multiple data sources, detects unexpected alerts, and improves alert detection of Apsara Stack.
	Report management	Allows security administrators to quickly export reports.

Module	Feature	Description
	System configurations	Allows you to configure system features such as alerts, upgrades, global policies, and accounts.

26.4.2. On-premises security operations services

Apsara Stack Security Standard Edition provides on-premises security operations services that guarantee the security of your business systems.

The following table describes the on-premises security operations services that Apsara Stack Security provides.

On-premises security operations services

Category	Service	Description
Security operations	Asset research	Periodically researches your business systems in the cloud under your authorization and develops a business list containing information such as the business system name, Elastic Compute Service (ECS) information, Relational Database Service (RDS) information, IP address, domain name, and owner.
	New system assessment	• Detects system and application vulnerabilities in a new business system by using both automation tools and manual operations before you migrate the system to the cloud. • Provides suggestions and verification on vulnerability fixes.
	Periodic security assessment	• Periodically uses automation tools to detect system vulnerabilities, application vulnerabilities, and security risks in running business systems. • Provides suggestions on handling detected risks, including but not limited to security policy settings, patch updates, and application vulnerability handling.
	Access control management	Provides inspection and guidance on applying access control policies when a new business system is migrated to the cloud.
	Access control routine inspection	Periodically checks for access control risks of your business systems.
	Security risk routine inspection	Monitors and inspects security events in Apsara Stack Security, informs you of the verified events, and provides suggestions on event handling.
Apsara Stack Security maintenance	Rule update	Periodically updates the rules repository of Apsara Stack Security.
	Service integration	• Provides support for integrating Apsara Stack Security with your business systems. • Helps you customize and optimize security policies.

Category	Service	Description
Security event response	Event alerts	Synchronizes security events information from Alibaba Cloud, and helps you remove the risks.
	Event handling	Handles urgent events such as attacker intrusions.

26.4.3. Optional security services

The following table describes the optional security services that Apsara Stack Security provides.

Optional security services

Service	Feature	Description
DDoS Traffic Scrubbing	Traffic scrubbing against distributed denial of service (DDoS) attacks	Detects and prevents attacks such as SYN flood, ACK flood, ICMP flood, UDP flood, NTP flood, DNS flood, and HTTP flood.
	DDoS attack display	Allows you to search for DDoS attack events by IP address, status, and event information.
	DDoS traffic analysis	Allows you to monitor and analyze the traffic of a DDoS attack, and view the attack traffic protocol and the top 10 IP addresses that have launched the most attacks.
Cloud Firewall	Access control	● Centrally controls east-west traffic between servers and north-south traffic between the Internet and the servers. ● Controls both inbound and outbound traffic. ● Controls outbound traffic of external connections based on domain names. ● Analyzes external connections to help you detect exceptions on ECS instances.
	Real-time traffic monitoring	● Monitors external connections. ● Analyzes traffic from the Internet to your ECS instances. ● Presents asset information and access relationships in a visualized way to help you detect abnormal traffic.

Service	Feature	Description
Cloud Firewall	Real-time defense	• Intelligently detects and blocks intrusions in real time. Analyzes the network traffic blocked by Cloud Firewall and IPS. • Synchronizes all malicious IP addresses detected across Alibaba Cloud and defends against potential threats, such as malicious visitors, scanners, and command-and-control servers. • Integrates the best practices of intrusion prevention policies on Alibaba Cloud to ensure high accuracy in threat detection. • Supports installation-free virtual patches for business systems. Protects against common vulnerabilities and high-risk zero-day and N-day vulnerabilities.
	Behavior backtracking	• Provides event logs to show threats or intrusions detected and blocked by IPS in real time. • Provides logs to record traffic that flows through Cloud Firewall. When a threat occurs, you can view traffic logs to analyze traffic, identify its source, and check configured access control policies. • Provides system operations logs to record all configurations and operations in Cloud Firewall.
Sensitive Data Discovery and Protection	Security situation overview	Allows you to view the overall security status of sensitive data.
	Detection and processing of suspicious activities	Detects suspicious activities related to sensitive data and allows you to confirm or exclude the activities after manual verification.
	Sensitive data detection	Detects sensitive data in services such as MaxCompute, Tablestore, Object Storage Service (OSS), AnalyticDB, and ApsaraDB for RDS.
	Static data masking	Uses data masking algorithms to mask sensitive data at rest.
	Intelligent audit	Creates audit rules to intelligently audit services such as OSS, MaxCompute, and ApsaraDB for RDS.
	Data permission management	Displays departments and users hierarchically, displays users and accounts by type, and allows you to query and manage detailed permissions of accounts.
	Dataflow monitoring	Allows you to view the dataflow details of DataHub and Cloud Data Pipeline (CDP).
	Rule configuration	Allows you to configure detection rules, risk levels, and abnormal output rules to detect sensitive data.

Service	Feature	Description
	Access authorization	Supports department-based authorization and protects the data assets of authorized departments.

26.5. Restrictions

None

26.6. Terms

DDoS attacks

An attacker combines multiple computers by using the client-server model to form an attack platform and initiates a large number of valid requests to one or more targets from this platform to cause network failures. Distributed denial of service (DDoS) attacks are much stronger than common denial of service (DoS) attacks.

SQL injections

An attacker makes the server run malicious Structured Query Language (SQL) commands by inserting these commands in Web tables or inserting malicious strings in URL requests.

Traffic scrubbing

The traffic scrubbing service monitors the inbound traffic of a data center in real time and detects unusual traffic that may be from DDoS attacks and other attacks. This service scrubs the unusual traffic without affecting businesses.

Brute-force attacks

Brute-force attacks work by iterating through all possible combinations that can make up a password.

Webshells

A webshell is a script written in languages such as Active Server Pages (ASP) and Hypertext Preprocessor (PHP). Attackers can run a webshell on a Web server to perform risky operations. This enables attackers to obtain sensitive information or control the server through server penetration or privilege escalation.

Server intrusion detection

By analyzing server logs, Apsara Stack Security can detect attacks, such as system password cracking and logons from unusual IP addresses, and generate real-time alerts.

27.Key Management Service (KMS)

27.1. What is KMS?

Key Management Service (KMS) is a one-stop service platform for key management and data encryption. KMS provides simple, reliable, secure, and standard-compliant capabilities to encrypt and protect data. KMS greatly reduces your costs of purchase, operations and maintenance (O&M), and research and development (R&D) on cryptographic infrastructure and data encryption services. This helps you focus on the business development.

KMS provides the following features:

- Encryption key hosting

  KMS supports encryption key hosting. An encryption key hosted on KMS is called a customer master key (CMK). You can manage the lifecycle of a CMK by enabling or disabling the CMK.

- BYOK

  KMS supports Bring Your Own Key (BYOK). You can import your own keys to KMS to encrypt data on the cloud. This facilitates key management. You can import the following types of keys to KMS:

  - Keys in your on-premises key management infrastructure (KMI)
  - Keys in user-managed hardware security modules (HSMs) of Data Encryption Service

  > **Note** Keys imported to managed HSMs in KMS cannot be exported by using any method because secure key exchange algorithms are used in KMS. Operators or third parties are not allowed to check the plaintext of keys.

- Automatic rotation of encryption keys

  A CMK in KMS can have multiple key versions. Each version represents an independently generated key and does not have any relation with other versions. KMS automatically rotates encryption keys. This helps you implement the best security practices and comply with audit requirements. For more information, see the Overview and Automatic key rotation topics of *Key rotation* in *User Guide*.

- Fully managed HSMs

  KMS provides fully managed HSMs. You can host keys in HSMs. Cryptographic operations are implemented in HSMs to protect key security.

  > **Note** To use this feature, you must purchase an HSM and the KMS license of the Advanced edition.

- Simple cryptographic API operations

  - KMS provides cryptographic API operations that are simpler than those for traditional cryptographic modules or cryptographic software libraries.
  - Encryption keys in KMS support authenticated encryption with associated data (AEAD) and deliver additional authenticated data (AAD) to protect data integrity. For more information, see the EncryptionContext topic of *Use symmetric keys* in *User Guide*.

- CMK aliases

KMS allows you to create CMK aliases, which facilitate CMK usage. For more information, see the Use aliases topic in *User Guide*. For example, you can use CMK aliases to manually rotate CMKs in specific scenarios.

- Resource tags

  KMS supports resource tags, which facilitate key resource management.

27.2. Benefits

KMS has advantages over traditional KMI, such as multi-service integration and ease of use.

Multi-service integration

KMS is integrated with multiple Alibaba Cloud services such as ECS, ApsaraDB for RDS, and OSS. You can use CMKs in KMS to encrypt and control the data stored in these services and maintain control over the cloud computing and storage environments.

Ease of use

- Easy encryption

  KMS simplifies abstract cryptographic concepts and provides cryptographic API operations that allow you to encrypt and decrypt data. For applications that require a key hierarchy, KMS provides convenient envelope encryption that implements a key hierarchy. For example, KMS generates a data key (DK) and uses a CMK as a key encryption key (KEK) to protect the DK. For more information, see the Envelope encryption topic of *Features* in *Technical White Paper*.

- Centralized key hosting

  KMS provides centralized key hosting and control.

  You can import keys from KMI or the HSMs of Data Encryption Service to KMS. No matter whether a key is imported from an external source or created in KMS, the confidential information or sensitive data in the key is used to encrypt data of an Alibaba Cloud service.

- BYOK

  KMS supports BYOK. You can import your own keys to KMS to encrypt data on the cloud. This facilitates key management. You can import the following types of keys to KMS:

  - Keys in your on-premises KMI
  - Keys in user-managed HSMs of Data Encryption Service

  > **Note** Keys imported to managed HSMs in KMS cannot be exported by using any method because secure key exchange algorithms are used in KMS. Operators or third parties are not allowed to check the plaintext of keys.

- Custom key rotation policies

  KMS supports automatic rotation of symmetric encryption keys based on security policies. You only need to configure a custom rotation cycle for a CMK. KMS automatically generates new CMK versions. A CMK can have multiple key versions. Each version can be used to decrypt corresponding ciphertext data. The latest key version (called the primary version) is an active encryption key and is used to encrypt current data. For more information, see the Automatic key rotation topic of *Key rotation* in *User Guide*.

27.3. Scenarios

This topic describes the common scenarios in which you can use KMS.

Typical scenarios

Role	Demand	Scenario
Application developer	Protect the security of sensitive data in applications.	Sensitive data is used in applications. Application developers need to use encryption keys to encrypt sensitive data and use KMS to protect the encryption keys.

Encryption methods

- Envelope encryption

  Your CMKs are stored in KMS. You only need to deploy enveloped data keys (EDKs). You can call the Decrypt API operation to decrypt the EDKs and use the returned plaintext DKs to encrypt or decrypt your local business data.

  For more information about envelope encryption, see the Envelope encryption topic of *Features* in *Technical White Paper*.

- Direct encryption

  You can call the Encrypt or Decrypt API operation to directly encrypt or decrypt sensitive data by using CMKs.

- Server-side encryption

  If you use Alibaba Cloud services to store data, you can use the server-side encryption feature of these services to encrypt and protect data in a simple and effective way. For example, you can use the server-side encryption feature of OSS to protect buckets that store sensitive data or use transparent data encryption (TDE) to protect tables that store sensitive data.

27.4. Terms

This topic introduces the terms used in KMS.

Note For more information about API operations mentioned in this topic, see *API reference* in *Developer Guide*.

Key Management Service (KMS)

KMS provides features such as key hosting and cryptographic operations. KMS implements security practices such as key rotation and can be integrated with other Alibaba Cloud services to encrypt the user data managed by these services. KMS frees you from the need to manually maintain the security, integrity, and availability of your keys. You only need to focus on data encryption, data decryption, and digital signature generation and verification based on your business requirements.

customer master key (CMK)

A CMK is used to encrypt data keys (DKs) and generate enveloped data keys (EDKs). It can also be used to encrypt a small volume of data. You can call the CreateKey API operation of KMS to create a CMK.

envelope encryption

To encrypt business data, you can call the GenerateDataKey or GenerateDataKeyWithoutPlaintext operation to generate a symmetric key and encrypt the symmetric key by using a CMK. An EDK is returned. The EDK is secure even if it is stored and transferred over unsecured communication channels. If you want to use the symmetric key, you only need to call the Decrypt operation to decrypt the EDK.

For more information about envelope encryption, see the Envelope encryption topic of *Features* in *Technical White Paper*.

data key (DK)

A DK refers to a plaintext key used to encrypt data.

> **Note** You can call the GenerateDataKey operation to generate a DK and encrypt the DK by using a CMK. The plaintext and ciphertext of the DK are returned. In a general sense, DK refers to the plaintext and EDK refers to the ciphertext.

enveloped data key (EDK)

An EDK refers to a ciphertext DK generated by using envelope encryption.

> **Note** If the plaintext of a DK is not needed, you can call the GenerateDataKeyWithoutPlaintext operation to return only the ciphertext of the DK.

hardware security module (HSM)

An HSM is a hardware device that performs cryptographic operations and securely generates and stores keys. The Managed HSM feature provided by KMS meets both the testing and validation requirements of regulatory agencies and provides you with high security for your keys managed in KMS.

EncryptionContext

EncryptionContext refers to the encapsulation of authenticated encryption with associated data (AEAD) in KMS. KMS uses imported encryption context as the additional authenticated data (AAD) of the symmetric encryption algorithm for cryptographic operations. In this way, KMS provides additional integrity and authenticity for encrypted data. For more information, see the EncryptionContext topic of *Use symmetric keys* in *User Guide*.

28.Apsara Stack DNS

28.1. What is Apsara Stack DNS?

Apsara Stack DNS is a service that runs on Apsara Stack to resolve domain names over internal networks, such as VPCs, data centers, and the classic network. You can configure rules to map domain names to IP addresses. Apsara Stack DNS then distributes domain name requests from clients to cloud resources, user-created business applications, business systems on your internal networks, or the business resources of Internet service providers (ISPs).

Apsara Stack DNS provides the DNS resolution and Global Server Load Balancer (GSLB) services in VPCs, data centers, and the classic network. You can perform the following operations by using Apsara Stack DNS in these internal networks:

- Access other ECS instances deployed in the same VPC.
- Access other cloud service instances on Apsara Stack.
- Access enterprise business systems.
- Access services over the Internet.
- Use the GSLB service to implement multiple-active solutions and disaster recovery, such as local active-active, local multi-active, remote active-active, active geo-redundancy, and geo-disaster recovery.
- Connect to Apsara Stack DNS with your own DNS servers over a leased line.

28.2. Edition comparison

This topic describes the differences between Apsara Stack DNS editions.

Category	Feature	DNS Lightweight Basic Edition	DNS Basic Edition	DNS Standard Edition	Internal GTM Standard Edition
Internal DNS resolution management	Global basic DNS resolution	Supported	Supported	Supported	Supported
	Global load balancing (weight)	Supported	Supported	Supported	Supported
	Global domain name forwarding	Supported	Supported	Supported	N/A
	Global default forwarding	Supported	Supported	Supported	N/A

Category	Feature	DNS Lightweight Basic Edition	DNS Basic Edition	DNS Standard Edition	Internal GTM Standard Edition
	Internet recursive resolution	Not supported by default. (However, it can be supported if you use the NAT Gateway solution.)	Supported	Supported	N/A
PrivateZone (tenant isolation)	Tenant-based basic DNS resolution	Not supported	Not supported	Supported	N/A
	Tenant load balancing (weight)	Not supported	Not supported	Supported	N/A
	Tenant domain name forwarding	Not supported	Not supported	Supported	N/A
	Tenant-based default forwarding	Not supported	Not supported	Supported	N/A
	VPC binding and unbinding	Not supported	Not supported	Supported	N/A
Internal Global Traffic Manager	Scheduling instance management	N/A	N/A	N/A	Supported
	Address pool management	N/A	N/A	N/A	Supported
	Scheduling domain management	N/A	N/A	N/A	Supported
	Data synchronization management of clusters (multi-cloud)	N/A	N/A	N/A	Supported
	Global data synchronization (multi-cloud)	N/A	N/A	N/A	Supported

Category	Feature	DNS Lightweight Basic Edition	DNS Basic Edition	DNS Standard Edition	Internal GTM Standard Edition
Others	Independent physical machine deployment	Not required	Required	Required	Required. It must be deployed on the physical machines of the DNS Basic Edition or DNS Standard Edition.
	Web UI console	Supported	Supported	Supported	Supported
	Upgrade to DNS Basic Edition	Supported (New physical machines must be deployed.)	N/A	N/A	N/A
	Upgrade to DNS Standard Edition	Supported (New physical machines must be deployed.)	Supported	N/A	N/A

28.3. Benefits

Enterprise domain name management

Apsara Stack DNS provides management and resolution services for your domain names. It supports the following features:

- Performs forward and reverse DNS resolutions for domain names of cloud service instances, such as ECS instances.
- Performs forward and reverse DNS resolutions for your internal domain names.
- Allows you to add, modify, and delete DNS records of the following types: A, AAAA, CNAME, NS, MX, TXT, SRV, and PTR.
- Allows you to add multiple A, AAAA, or PTR records at a time. DNS servers randomly respond to all DNS queries through round robin to achieve load balancing.

Flexible integration with data centers

Apsara Stack DNS can forward enterprise domain names and provide the following services for you to flexibly build your network and cascade DNS servers with user-created DNS servers:

- Global default forwarding
- Forwarding queries for specific domain names

Internet access from enterprise servers

Apsara Stack DNS supports recursive resolution for Internet domain names, which allows your servers to access the Internet.

Tenant isolation

Apsara Stack DNS allows you to manage private zones in VPCs, resolve internal domain names, and isolate DNS records by tenant.

- You can add, delete, modify, and query private authoritative zones. You can also bind and unbind private authoritative zones to and from VPCs.
- You can add, delete, modify, and query private forwarding zones. You can also bind and unbind private forwarding zones to and from VPCs.

GSLB

Global Server Load Balancer (GSLB) provides the following features on internal networks:

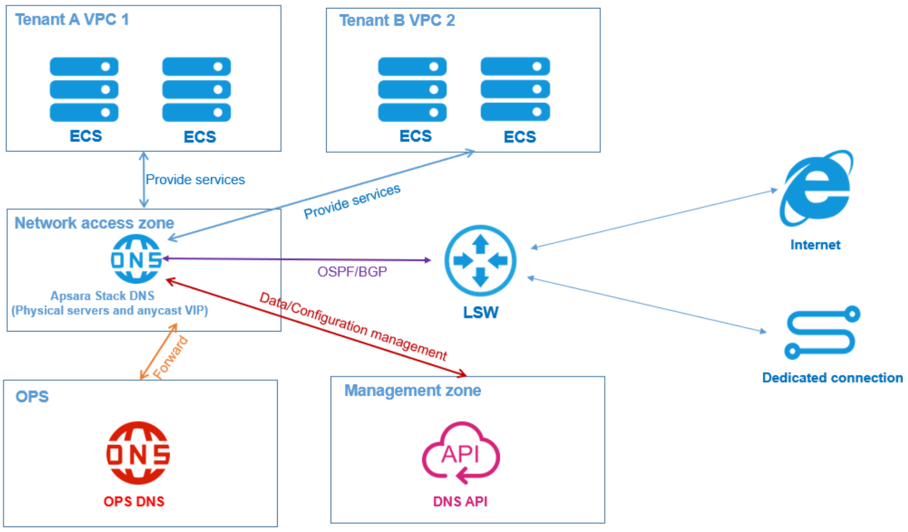
- Allows you to add multiple A, AAAA, or CNAME records at a time. DNS servers respond to DNS queries based on the weight of each record type to achieve load balancing.
- Synchronizes configuration data for resolution among multiple clusters for which GSLB is activated. This feature is supported in multi-cloud scenarios.
- Supports address pool management to centrally manage enterprise applications by application service cluster.
- Supports custom global scheduling domains. You can centrally manage and code global scheduling instances based on your naming conventions.

Centralized management console

You can access DNS and other cloud services in the Apsara Stack Cloud Management (ASCM) console with one account. This implements web-based data and service management, which simplifies operations.

28.4. Architecture

Architecture of Apsara Stack DNS



Architecture of Apsara Stack DNS (DNS Basic Edition and DNS Standard Edition)

- Uses two independent physical machines that are deployed in the network access zone to improve service availability. Apsara Stack DNS in this architecture can be scaled in or out.
- Issues anycast virtual IP address (VIP) routing requests over the LAN switch (LSW) by using Open Shortest Path First (OSPF) or Border Gateway Protocol (BGP). Anycast VIPs provide DNS services for VPCs and the classic network of tenants. The outbound IP address configured on the DNS servers can be used to forward requests to the OPS DNS server, Internet, or a dedicated enterprise network based on forwarding and recursive rules.
- Manages data and configurations by using APIs in the management zone.
- Allows you to create and query domain names on a web UI, forwards requests for cloud service domain names to the OPS DNS server, performs recursive DNS queries for Internet domain names, allows you to add, modify, delete, and query authoritative domain names and forwarding domain names of private zones, and binds and unbinds a private zone to and from a VPC.

Architecture of Apsara Stack DNS (DNS Lightweight Edition)

- Supports the deployment with two physical machines on the OPS3 or OPS4 base, which eliminates the need to apply for an independent physical machine. The two physical machines achieve high availability. Apsara Stack DNS in this architecture cannot be scaled in or out.
- Issues anycast VIP routing requests over the LSW by using OSPF or BGP. Anycast VIPs provide DNS services for VPCs and the classic network of tenants. The outbound IP address configured on the DNS servers can be used to forward requests to the OPS DNS server, Internet, or a dedicated enterprise network based on forwarding and recursive rules.
- Manages data and configurations by using APIs in the management zone.
- Allows you to create and query domain names on a web UI, forwards requests for cloud service domain names to the OPS DNS server, and performs recursive DNS queries for Internet domain names.

Architecture of Apsara Stack DNS (internal GTM Standard Edition)

- Depends on the deployment of DNS Basic Edition or DNS Standard Edition. Apsara Stack DNS of the internal GTM Standard Edition is deployed on the two physical machines of DNS Basic Edition or DNS Standard Edition in the network access zone. Apsara Stack DNS in this architecture can be scaled in or out.
- Issues anycast VIP routing requests over the LSW by using OSPF or BGP. Anycast VIPs provide DNS services for VPCs and the classic network of tenants.
- Manages data and configurations by using APIs in the management zone.
- Allows you to manage domain names on a web UI, allows you to add, modify, delete, and query address pools, access policies, and scheduling instances. You can also create and delete Global Traffic Manager (GTM) synchronization clusters.

28.5. Features

1 Internal DNS resolution management

Internal DNS resolution management allows you to manage global internal domain names, global forwarding configurations, and global recursive resolution configurations that you have created in Apsara Stack. Changes to these configurations take effect on all VPCs and the classic network.

This feature provides the same global DNS resolution service to all servers in VPCs. DNS servers use anycast IP addresses within a region. This way, seamless service failover and failback can be achieved in a specific region where data centers support disaster recovery. Note: If you do not need to upgrade Apsara Stack DNS to the Standard Edition, you can configure DNS server addresses as global anycast IP addresses to implement seamless service failover and failback over the entire network if data centers support disaster recovery.

1.1 Global internal domain names

Allows you to register, search, and delete global internal domain names and add descriptions for these domain names. You can also add, delete, and modify DNS records. The following DNS record types are supported: A, AAAA, CNAME, MX, PTR, TXT, SRV, NAPTR, CAA, and NS.

Allows you to add multiple DNS records of the A, AAAA, and PTR types on one host. By default, the resolution results include all the matching records. Records can be randomly rotated for load balancing.

Allows you to add multiple DNS records of the A, AAAA, and CNAME types on one host. DNS servers respond to DNS queries based on the weight of each record type to achieve load balancing.

1.2 Global forwarding configurations

Forwards domain name requests to another DNS server for resolution.

Supports global default forwarding, which forwards requests of domain names that do not have forwarding configurations to another DNS server for resolution.

Apsara Stack DNS forwards requests with or without recursion.

- Forward All Requests (without Recursion): Only the specified DNS server is used to resolve domain names. If the resolution fails or the request times out, a message is returned to the DNS client to indicate that the query failed.
- Forward All Requests (with Recursion): The specified DNS server is preferentially used to resolve domain names. If the resolution fails, the local DNS server is used.

1.3 Global recursive configurations

Supports recursive resolution for Internet domain names, which enables your servers to access the Internet.

Allows you to enable, disable, or modify the global default forwarding configurations.

2 PrivateZone (DNS Standard Edition only)

The PrivateZone feature allows you to create tenant-specific domain names in VPCs. You can bind and unbind the domain names to and from VPCs as needed to isolate tenants. Changes to these configurations take effect only in the VPCs to which the domain names are bound.

This feature provides personalized DNS resolution service to servers in the VPCs to which the domain names are bound. DNS servers use anycast IP addresses within a region. This way, seamless service failover and failback can be achieved in a specific region where data centers support disaster recovery.

2.1 Tenant internal domain names

Allows you to register, search, and delete tenant internal domain names and add descriptions for these domain names. You can also add, delete, and modify DNS records. The following DNS record types are supported: A, AAAA, CNAME, MX, PTR, TXT, SRV, NAPTR, CAA, and NS.

Allows you to add multiple DNS records of the A, AAAA, and PTR types on one host. By default, the resolution results include all the matching records. Records can be randomly rotated for load balancing.

Allows you to add multiple DNS records of the A, AAAA, and CNAME types on one host. DNS servers respond to DNS queries based on the weight of each record type to achieve load balancing.

Allows you to bind and unbind a domain name to and from a VPC.

2.2 Tenant forwarding configurations

Forwards domain name requests to another DNS server for resolution.

Supports global default forwarding, which forwards requests of domain names that do not have forwarding configurations to another DNS server for resolution.

Apsara Stack DNS can forward requests with or without recursion.

- Forward All Requests (without Recursion): Only the specified DNS server is used to resolve domain names. If the resolution fails or the request times out, a message is returned to the DNS client to indicate that the query failed.
- Forward All Requests (with Recursion): The specified DNS server is preferentially used to resolve domain names. If the resolution fails, the local DNS server is used.

Allows you to bind and unbind a domain name to and from a VPC.

3 Internal Global Traffic Manager (internal GTM Standard Edition only)

Internal Global Traffic Manager (GTM) provides multi-cloud disaster recovery for your domain names. You can connect your domain names to an internal GTM instance to manage traffic loads between Apsara Stack systems.

Internal GTM supports internal Global Server Load Balancer (GSLB). This feature intelligently allocates IP addresses for DNS queries from request sources based on configured scheduling policies. It also supports multi-cloud, hybrid deployment and configuration data synchronization between cloud networks.

3.1 Scheduling instance management

Allows you to manage scheduling instances. Each scheduling instance corresponds to an application instance.

Allows you to manage address pools. Each address pool corresponds to a service cluster of an application instance.

Allows you to manage scheduling domains and set the scheduling domains to which scheduling instances belong. You can centrally manage and code global scheduling instances based on your own naming conventions.

3.2 Data synchronization management

Allows you to manage global data synchronization links. You can create data synchronization links, manage data synchronization configurations, and view data synchronization information of multiple internal GTM services. The information includes local system information, information of cluster nodes on which data synchronization relationship has been established, and primary and secondary relationships.

Allows you to manage the messages for changes to data synchronization links, which helps you confirm request messages for primary nodes to actively add secondary nodes.

28.6. Scenarios

Apsara Stack DNS is a key network service that controls data traffic for Apsara Stack. It resolves domain names, balances server loads, and connects Apsara Stack with data centers and Alibaba Cloud public cloud. Apsara Stack DNS provides a complete suite of solutions to deploy a cloud environment, achieve high availability and disaster recovery for data centers, and balance server loads to secure your IT services.

Apsara Stack DNS provides four categories of solutions in the following eleven scenarios for enterprise users:

(1) Basic DNS resolution

Scenario 1: Access cloud resource instances over a VPC

You can use Apsara Stack DNS to access ApsaraDB for RDS, Server Load Balancer (SLB), and Object Storage Service (OSS) instances from ECS or Docker instances over a VPC.

Scenario 2: Access ECS or Docker instances by using hostnames over a VPC

You can use Apsara Stack DNS to create hostnames for your ECS or Docker instances, and access and manage them by using hostnames over a VPC.

Scenario 3: Access internal service domain names over a VPC

You can use Apsara Stack DNS to develop your own PaaS or SaaS services on Apsara Stack and access the PaaS or SaaS services by using domain names over a VPC.

Scenario 4: Schedule internal service requests on Apsara Stack by using the round robin algorithm

If the SaaS service in scenario 3 is deployed in multiple data centers or regions, you can use Apsara Stack DNS to access the service and evenly distribute requests to different nodes in a VPC.

Scenario 5: Access Internet services over a VPC or the classic network

You can use Apsara Stack DNS to access Internet services from a VPC or the classic network.

Scenario 6: Establish connections between Apsara Stack and other networks

You can use Apsara Stack DNS to establish connections between Apsara Stack and other networks, such as internal networks, Alibaba Cloud public cloud, and other external networks by using domain names.

(2) Tenant isolation (DNS Standard Edition)

Scenario 7: Isolate tenant resources on Apsara Stack

You can use Apsara Stack DNS to isolate tenant resources on Apsara Stack so that the internal DNS resolution data and the default forwarding configurations of each tenant are invisible to other tenants. You can configure your private DNS resolution data to complete business addressing and scheduling.

Scenario 8: Establish connections among global resources on Apsara Stack

If you need to allow all tenants to share global resources and configurations on Apsara Stack, system administrators can configure global DNS resolution data and configurations to complete business addressing and scheduling.

Scenario 9: Perform VPC-based intelligent scheduling on Apsara Stack

You can use Apsara Stack DNS to provide different DNS records for different VPCs of a tenant based on the same host record.

(3) Global scheduling (internal GTM Standard Edition)

Scenario 10: Schedule traffic loads of internal network services on Apsara Stack based on weights

If the PaaS or SaaS service in scenario 3 is deployed in multiple data centers or regions, you can use Apsara Stack DNS to access the service and distribute requests to different nodes in a VPC based on the weights of the nodes of backend service clusters.

(4) Enterprise disaster recovery (internal GTM Standard Edition)

Scenario 11: Schedule traffic loads of internal network services on Apsara Stack to achieve disaster recovery

If the PaaS or SaaS service in scenario 3 is deployed in multiple data centers or regions, you can use Apsara Stack DNS to access the service and switch access traffic of internal services from backend service cluster A (primary data center) to backend service cluster B (secondary data center) in a disaster recovery scenario in a VPC.

28.7. Limits

This topic describes limits on Apsara Stack DNS clusters.

Basic Edition

Cluster	Module	Server type	Minimum configuration	Quantity
Access cluster	Resolution	Q46S1.2C	16-core CPU, 96 GB of memory, 600 GB of hard disk space, two GE ports, and four 10GE ports. The downgraded Q46 model is used, and the network must support the Q46S1.2C server type.	2

Cluster	Module	Server type	Minimum configuration	Quantity
Management cluster	Management	Container	4-core CPU, 8 GB of memory, 60 GB of hard disk space, and network connections.	2

Standard Edition

Cluster	Module	Server type	Minimum configuration	Quantity
Access cluster	Resolution	Q46S1.2C	40-core CPU (two Intel Xeon Silver 4114 CPUs), 192 GB of memory, 1,200 GB of hard disk space, two GE ports, and four 10GE ports.	2
Management cluster	Management	Container	4-core CPU, 8 GB of memory, 60 GB of hard disk space, and network connections.	2

Lightweight Basic Edition

Cluster	Module	Server type	Minimum configuration	Quantity
Access cluster	Resolution	OPS3 or OPS4	8-core CPU, 16 GB of memory, 480 GB of hard disk space, and two GE ports.	2
Management cluster	Management	Container	4-core CPU, 8 GB of memory, 60 GB of hard disk space, and network connections.	2

> ? **Note** Tenant-related features are not supported in this edition. You must plan two anycast IP addresses for DNS resolution of ECS instances. By default, the resolution module of this edition provides DNS resolution for ECS instances in new deployment scenarios since Apsara Stack DNS V3.11. You must plan a source IP address for forwarding DNS queries.

28.8. Terms

DNS

Domain Name System (DNS) is a distributed database that is used for TCP/IP applications. It translates domain names into IP addresses, and selects paths for emails.

Domain name resolution

Domain name resolution maps domain names to IP addresses by using the DNS system. It includes both authoritative DNS and recursive DNS.

Recursive DNS

Recursive DNS queries domain names cached on the local DNS server or sends a request to the authoritative DNS system to obtain the corresponding IP addresses. You can use recursive DNS to resolve Internet domain names.

Authoritative DNS

Authoritative DNS resolves the names of root domains, top-level domains, and various other domains.

Authoritative domain names

Authoritative domain names are domain names resolved by the local DNS server. You can configure and manage the domain name resolution data on the local DNS server.

DNS forwarding

DNS forwarding uses two DNS servers to provide DNS resolution services. The local DNS server is used to configure and manage the domain name resolution data. The other DNS server is used to resolve domain names.

Default forwarding

If DNS queries for authoritative domain names are not resolved by the local DNS server, they are forwarded to another DNS server for resolution.

29.API Gateway

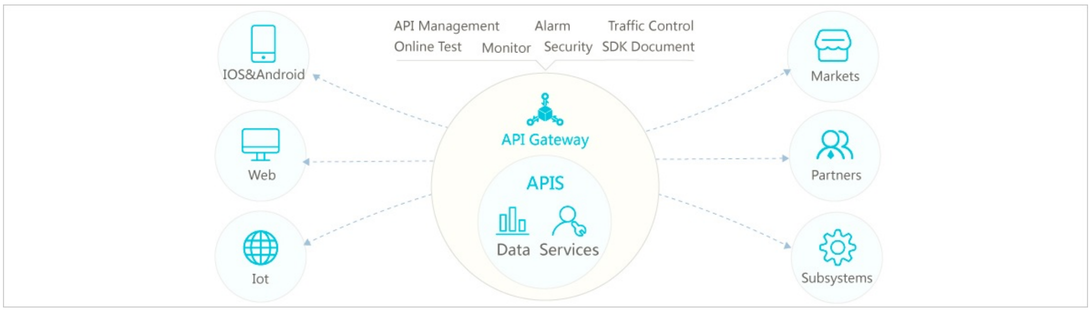
29.1. What is API Gateway?

API Gateway provides a comprehensive suite of API hosting services that help you share capabilities, services, and data with partners in the form of APIs.

- API Gateway provides multiple security mechanisms to secure APIs and reduce the risks arising from open APIs. These mechanisms include protection against replay attacks, request encryption, identity authentication, permission management, and throttling.
- API Gateway provides API lifecycle management that allows you to define, publish, and unpublish APIs. This improves API management and iteration efficiency.

API Gateway allows enterprises to reuse and share their capabilities with each other so that they can focus on their core business.

API Gateway



29.2. Features

API lifecycle management

- Manages APIs throughout their entire lifecycle, including publishing, testing, and unpublishing APIs.
- Supports maintenance features such as routine management, version management, and quick rollback.

Comprehensive security protection

- Supports multiple authentication methods and HMAC (SHA-1 and SHA-256) algorithms.
- Supports HTTPS and SSL encryption.
- Provides multiple security mechanisms to prevent injections, replay attacks, and tampering.

Flexible access control

- API Gateway implements access control on apps that are used to make API requests.
- Only authorized apps can call APIs.
- API providers can authorize apps to call APIs.

Various plug-in features

- Allows you to use various plug-ins to expand the features of APIs in a pluggable manner.
- Provides various plug-ins, including throttling, IP address-based access control, backend signature,

JWT authentication, cross-origin resource sharing (CORS), caching, backend routing, access control, circuit breaker, and error code mapping.

Request verification

Supports the verification of parameter types and values (range, enumerated value, and regular expression). Requests with invalid parameter types or values are denied by API Gateway. This reduces backend resources wasted on invalid requests and significantly reduces backend service processing costs.

Data conversion

Enables you to configure mapping rules to translate frontend and backend data.

- Supports data conversion for frontend requests.

Automated tools

- Automatically generates API documentation.
- Provides SDK samples in multiple programming languages.
- Provides graphical debugging tools for quick testing and deployment.

Monitoring and alerting

- Provides a graphical real-time API monitoring panel that displays information such as the number of API calls, response time, and error rate.
- Allows you to configure alerts and to track the operating status of each API in real time.
- Supports API-based full log query by using Log Service.

29.3. Benefits

Easy maintenance

After you create APIs in API Gateway, API Gateway performs all the other API management functions. This significantly reduces routine maintenance costs.

High request processing performance

API Gateway uses a distributed deployment and auto-scaling model to respond to a large number of API call requests at very low latency. It provides secure and efficient gateway functions for your backend services.

Security and stability

Your services are open to API Gateway only over an internal network. API Gateway also provides enhanced permission management functions and precise request throttling functions. It makes your services secure, stable, and controllable.

Comprehensive functionality

API Gateway supports API lifecycle management and provides graphical debugging tools and various plug-ins. It can also automatically generate SDKs of different programming languages. This meets your requirements on API usage.

29.4. Terms

Before you use API Gateway, familiarize yourself with the following basic concepts.

app

An app defines the identity of an API caller. To call an API, you must create an app first.

AppKey and AppSecret

Each app has an AppKey and AppSecret pair. This pair is encrypted and attached to a request as the signature.

encrypted signature

An encrypted signature is attached to each API request and is authenticated by API Gateway.

authorization

The API service provider grants an app the permission to call an API. Only authorized apps can call specified APIs.

API lifecycle

The API service provider manages an API in stages, including the creation, test, publish, unpublish, and version change stages.

API definition

An API definition is a set of rules defined by the API service provider during the creation of an API. The API definition specifies the backend service, request format, receive format, and return format.

parameter mapping

Parameter mapping is configured by the API service provider. It is used when the parameters in a request are inconsistent with those of the API backend service.

parameter verification

Parameter verification is performed based on a set of rules defined by the API service provider. API Gateway filters out invalid requests based on these rules.

constant parameter

API users do not need to enter the constant parameters. The constant parameters are always received by the backend service.

system parameter

API Gateway includes system parameters such as CaClientIP (request IP address) in the requests sent to your backend service.

API group

An API group is a group of APIs that are managed by the API service provider as a whole. Before you create an API, you must create an API group.

second-level domain

A second-level domain is a domain name that you bind to an API group when you create the group. This domain name is used to test API calling.

independent domain

An independent domain is a domain name that you bind to an API group when you open an API in the group. You must access the independent domain to call an API.

plugin

A plugin is a pluggable extension used to implement API functions by performing binding and unbinding operations.

signature key

A signature key is created by the API service provider and bound to an API. The signature information is added to each request sent from API Gateway to the backend service. The backend service checks the signature information to ensure security.

throttling policy

The API service provider can configure a throttling policy to limit the maximum number of requests for an API, and the maximum number of API requests that can be initiated by a user or an app. The throttling granularity can be day, hour, or minute.

30.Enterprise Distributed Application Service (EDAS)

30.1. What is EDAS?

Enterprise Distributed Application Service (EDAS) is a Platform as a Service (PaaS) platform for application hosting and microservice management, providing full-stack solutions such as application development, deployment, monitoring, and O&M. It supports microservice runtime environments such as Dubbo and Spring Cloud, helping you easily migrate applications to the cloud.

Diverse application hosting environments

You can select instance-exclusive Elastic Compute Service (ECS) clusters, Container Service Kubernetes clusters, and user-created Kubernetes clusters based on your application systems and resource needs.

Abundant microservice frameworks

You can develop applications and services in the native Dubbo, native Spring Cloud, and High-speed Service Framework (HSF) frameworks, and host the developed applications and services to EDAS.

- You can host Dubbo and Spring Cloud applications to EDAS by adding dependencies and modifying a few configurations. You have access to the features of EDAS, such as enterprise-level application hosting, service governance, monitoring and alerting, and application diagnosis, without having to build ZooKeeper, Eureka, and Consul. This lowers the costs of deployment and O&M.
- HSF is the distributed remote procedure call (RPC) framework that is widely used within Alibaba Group. It interconnects different service systems and decouples inter-system implementation dependencies. HSF unifies the service publishing and call methods for distributed applications to help you conveniently and quickly develop distributed applications. HSF provides or uses common functional modules, and frees developers from various complex technical details involved in distributed architectures, such as remote communication, serialization, performance loss, and the implementation of synchronous and asynchronous calls.

Comprehensive application management

You can perform lifecycle management, service governance, and microservice management for your applications in the EDAS console.

- Application lifecycle management

  EDAS provides application lifecycle management, allowing you to deploy, scale out, scale in, stop, and delete applications. Applications of all sizes can be managed in the EDAS console.

- Service governance

  EDAS integrates a wide variety of service governance components, such as auto scaling, throttling and degradation, and health check, to deal with unexpected traffic spikes and crashes caused by dependencies. This greatly improves platform stability.

- Microservice governance

  EDAS provides features such as service topology, service statistics, and trace query, to help you manage every component and service in a distributed system.

Comprehensive monitoring and diagnosis

You can monitor the status of resources and services in applications in the EDAS console to promptly identify problems and quickly locate their causes by using the logging and diagnosis components.

- Application monitoring

  EDAS monitors the health status of application resources at the IaaS layer in real time, helping you quickly locate problems.

- Application diagnosis

  EDAS provides the container-based application diagnosis feature. Based on the provided data, this feature allows you to identify application runtime errors, such as errors in garbage collection (GC), class loading, connectors, memory allocated for objects, thread hotspots, Druid database connection pools, and Commons Pool.

30.2. Benefits

EDAS supports more than 99% large-scale application systems within the Alibaba Group, including all the key online systems involving memberships, transactions, products, stores, logistics, and customer reviews. It also delivers enhanced stability and reliability.

Reliability

- EDAS is a core product that has been used and tested within the Alibaba Group for nearly 10 years.
- It ensures the stable operation of all of Alibaba's key applications.
- It has supported Alibaba through Double 11 Shopping Festival.
- Its complete authentication system ensures that every single service call is made securely and reliably.

Comprehensiveness

- EDAS is a PaaS platform that supports application lifecycle management.
- The complete service governance solution provides an effective way to manage distributed services.
- The comprehensive application diagnosis system helps you easily identify the root causes of problems.
- Online load testing and capacity planning offer you easy access to online operation performance metrics and real-time operation capabilities.
- Auto scaling helps you deal with unexpected traffic spikes.

Thoroughness

- EDAS provides in-depth, global metrics reporting.
- It performs all-around monitoring for comprehensive troubleshooting.
- It analyzes every single distributed call through tracing.
- It identifies every possible bottleneck of the system with dependency analysis.

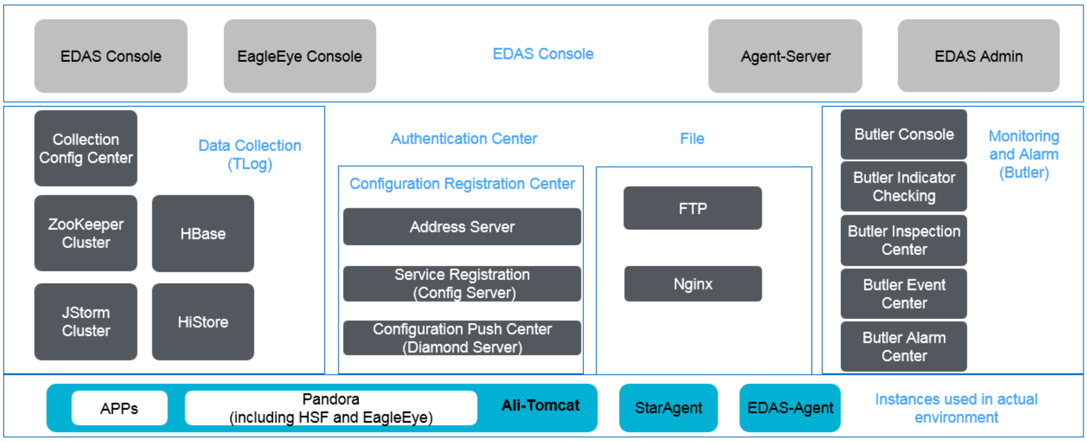
Openness

- Multiple Internet middleware products are now open-source.
- First-class Apache projects are openly shared and enjoy an excellent reputation in the industry.
- EDAS comes with no bundles and its functions can be easily replaced with open-source software.

30.3. Architecture

EDAS consists of the console, data collection system, configuration registry, and authentication center. EDAS architecture shows the EDAS architecture.

EDAS architecture



- EDAS console

  It is a GUI where you can directly use EDAS system functions. In the console, you can implement resource management, application lifecycle management, O&M control, service governance, three-dimensional monitoring, and digital operations.

- Data collection system

  It collects trace logs and the runtime statuses of EDAS clusters and all customer application instances, and summarizes, computes, and stores data in real time.

- Configuration registry

  It is a central server used to publish and subscribe to HSF services (RPC framework) and push distributed configurations.

- Authentication center

  It controls permissions for user data to ensure data security.

- O&M system

  It is a major tool of EDAS for daily monitoring and alarms of all EDAS components.

- Command channel system

  It is a control center that remotely sends commands to application instances.

- File system

  It stores WAR packages and required components, such as JDK and Ali-Tomcat, uploaded by users.

30.4. Features

As a core product of the Alibaba distributed service architecture, EDAS provides a wide variety of features ranging from application lifecycle management to O&M control.

30.4.1. Application hosting

You can deploy applications in ECS clusters and Container Service Kubernetes clusters. You can isolate environments by using namespaces.

Currently, different types of clusters impose limits on application frameworks and application packaging.

Application	Optional cluster	Packaging mode
Spring Cloud, Dubbo, and HSF	ECS cluster	WAR and JAR
	Container Service Kubernetes cluster	WAR, JAR, and image

You can host applications on EDAS in the console or by using tools.

30.4.2. Application lifecycle management

After applications are deployed, you can perform other application lifecycle management operations in the EDAS console.

Lifecycle management allows you to create, deploy, scale out, scale in, stop, and delete applications. Lifecycle management operations vary depending on the types of deployed clusters.

30.4.3. Service governance

EDAS integrates a wide variety of service governance components to address unexpected traffic spikes and the crashes caused by dependencies, improving platform stability.

- Auto scaling: This feature perceives the status of each instance in an application and implements dynamic scale-out and scale-in accordingly. This ensures the quality of service (QoS) and improves application availability.
- Throttling and degradation: This feature solves slow system responses or crashes caused by high pressure on the backend core services. This feature is generally used in high-traffic scenarios, such as flash sales, shopping sprees, major promotions, and empty-box scam protection.
- Health check: This feature periodically checks containers and applications and reports the results to the console. This keeps you informed of the general application runtime status in the cluster environment and helps you locate and troubleshoot problems.
- Canary release: Canary release is divided into single-application canary release and end-to-end canary release. It ensures smooth transition from earlier to later application versions.

30.4.4. Application development

Applications that are developed based on native Spring Cloud, native Dubbo, and HSF can be hosted to EDAS.

- You can host Spring Cloud applications in EDAS by adding dependencies and modifying a few configurations. You have access to the features of EDAS, such as enterprise-level application hosting, service governance, monitoring and alerting, and application diagnosis, without having to build Eureka and Consul. This lowers the costs of deployment and O&M.
- You can host Dubbo applications in EDAS simply by adding dependencies and modifying a few configurations. You have access to the features of EDAS, such as enterprise-level application hosting, service governance, monitoring and alerting, and application diagnosis, without having to build

ZooKeeper and Redis. This lowers the costs of deployment and O&M.

- HSF is the distributed RPC service framework widely used in the Alibaba Group. It interconnects different service systems and decouples inter-system implementation dependencies. It unifies the service publishing and call methods for distributed applications to help you develop distributed applications conveniently and quickly. It provides or uses common function modules and frees developers from various complex technical details related to distributed architectures.

30.4.5. Microservice management

EDAS provides the service query and inter-service trace query functions to help you manage every component and service in a distributed system.

- Service topology: A topology intuitively presents the calling relationships between services and relevant performance data.
- Service query: You can view the HSF, Spring Cloud, and Service Mesh services of applications in a specific namespace of a region.
- Service statistics: You can view the runtime statuses of all the services of all the applications within the current tenant over the past 24 hours, including the number of service calls, time consumption, and call errors. These statistics allow you to easily compare all services in the system.
- Trace query: By setting filter criteria, you can accurately locate the services with poor performance and errors.
- Trace details: Based on the trace query results, you can view the trace details of slow services and services with errors and reorganize their dependencies. This information allows you to identify frequent failures, performance bottlenecks, strong dependencies, and other problems. You can also evaluate service capacities based on call ratios and peak QPS.

30.4.6. Configuration management

EDAS integrates Application Configuration Management (ACM). In EDAS, you can centrally manage and push application configurations through ACM. In addition, you can isolate and synchronize configurations between different environments by namespace.

Configuration management allows you to create configurations, view push status, query push trajectories, and query and roll back versions.

30.4.7. Application monitoring

EDAS monitors its hosted applications through infrastructure monitoring, service monitoring, logs, and notifications and alarms.

- Application monitoring: EDAS monitors the health status of application resources at the IaaS layer in real time, helping you locate problems quickly. You can also activate Application Real-Time Monitoring Service (ARMS) for advanced monitoring.
- Logs: You can view the application runtime logs of an instance without logging on to the instance. You can check logs during troubleshooting.
- Real-time logs (applicable to applications deployed in Kubernetes clusters): You can check real-time logs to troubleshoot pod-related problems.
- Notifications and alarms: When some resources are overused, the EDAS system sends text messages or emails to contacts, instructing them to promptly troubleshoot online problems.

30.4.8. Application diagnosis

HSF applications are deployed and run in EDAS containers. EDAS provides container-based diagnosis that lets you diagnose application runtime errors based on the provided data.

- GC diagnosis: The GC diagnosis and memory diagnosis modules are provided.
  - GC diagnosis: This module monitors certain performance metrics of the selected application instance for the occurrence of GC and analyzes the GC status of the current instance based on the selected time range. These metrics help you determine whether an application instance is healthy. For example, it checks whether the application has a memory leak or large objects.
  - Memory diagnosis: This module provides statistics on the heap memory and non-heap memory of the JVM process of the Tomcat container where the application instance is located.
- Class loading: This component provides real-time loading information for JAR packages. When the JAR package of an application has a version conflict, you can use this function to easily locate the path to which the JAR package is loaded. This simplifies troubleshooting for such problems.
- Connector: A Tomcat connector is <Connector /> in the XML configuration of Ali-Tomcat. The information pulled from the <Connector /> line can be thought of as the configuration of the connector. This view displays the runtime status of the corresponding connector over the past 10 minutes.
- Memory allocated for objects: After you select the system class, Java primitive object class, and class loading, the system displays the number of objects, occupied space, and usage of the total system memory in a pie chart and a list.
- Method tracing: This component adopts the JVM bytecode enhancement technique to record the consumed time and sequence during the entire call process of the selected method. This allows you to check the execution sequence while execution is in progress. This helps you quickly fix application runtime errors.
- Thread hotspot: This provides the thread snapshot retrieval and call statistics analysis functions.
  - Retrieve thread snapshots

    Similar to the jstack command, the thread hotspot function obtains the stack frames of all the current threads from the target instance, and then filters out identified idle threads, such as HSF, Tomcat, and GC threads. To avoid excessive overhead, it only returns the data of 30 of the remaining threads by default.
  - Analyze call statistics

    The thread hotspot function collects statistics on and analyzes the method calls in an application within a certain period of time and displays the call methods and call relationships, namely, call stacks. The final result is displayed in two views, including the tree graph and flame graph. In addition, your service methods are automatically highlighted so that you can quickly locate the call sources of the service methods that consume the most time.
- Druid database connection pool monitoring: For an application whose data connection pool uses the Druid database, EDAS monitors the data connection pool and SQL execution.
- Commons Pool: When an application or application class library uses Commons Pool 2 (v2.0) (for example, the Jedis and Commons DBCP2 connection pools on a Redis client), the EDAS Commons Pool monitoring component monitors the configuration and usage of these pools.

30.4.9. Component center

Based on a distributed microservice system, the EDAS component center focuses on service integration and helps build a more open ecosystem for PaaS platforms. Related functions need to be implemented by using the corresponding components.

Microservice components

- Cloud Service Bus (CSB): In the EDAS console, you can create an exclusive CSB instance that allows you to expose applications in the target environment or introduce applications to the target environment for management . In addition, it allows you to expose EDAS applications in a VPC so that you can perform testing and joint debugging on them over the Internet in your own development environment.
- ARMS: This is an application performance monitoring product from Alibaba Cloud. ARMS helps quickly and conveniently build application monitoring capabilities with response speeds measured in seconds for enterprises.
- SchedulerX: This is a distributed task scheduling product. It provides an accurate, highly reliable, and highly available timed (Cron expression-based) task scheduling service with response speeds measured in seconds. It supports distributed task execution models, such as grid tasks, in which massive volumes of subtasks are evenly distributed to all worker nodes (SchedulerX clients) for execution.

Application diagnosis components

Currently, EDAS provides the following application diagnosis modules: method tracing, logging, performance analysis, Druid database connection pool monitoring, and Apache Commons Pool monitoring. All these five modules provide online diagnosis services for applications.

30.4.10. System management

EDAS provides account, role, and permission management functions for system administration, allowing you to manage and control permissions.

- Primary account/RAM user system: This system allows you to build primary account and RAM user relationships on the EDAS platform based on your enterprise's organization at the department, team, and project levels. ECS instances are organized based on these primary account and RAM user relationships so that you can easily allocate resources.
- Role and permission control: Application lifecycle management generally involves development, O&M, instance resources, and other roles. Different roles are permitted to perform different application management operations. EDAS provides a role and permission control mechanism that allows you to define roles for and assign permissions to different accounts.
- Service authentication: This feature ensures the reliability and security of each distributed call. Strict authentication is implemented in every phase, from service registration and subscription to service calling.

30.5. Scenarios

Publish and manage applications

Application publishing and management can be complicated in complex cloud environments. For locally developed applications, you need to deploy each of them to an instance and log on to each instance to publish and deploy them. You also need to restart and scale out the applications as your business keeps growing. The increasing number of instances creates a major challenge for the maintenance personnel.

For this scenario, EDAS provides a visual application publishing and management platform that allows you to easily perform application lifecycle management in a web console regardless of the cluster size.

Build a distributed system

After you transform a centralized system into a distributed system, it is always a challenge to ensure reliable service calls between systems in the distributed architecture. For example, you have to nail down a lot of technical details in network communication and serialization protocol design.

EDAS provides a high-performance RPC framework. It systematically considers the technical details, such as distributed service discovery, service routing, service calling, and service security between applications, allowing you to build highly available distributed systems.

Analyze the system runtime status by digital means

After applications are developed and deployed in the production environment, you need to monitor the application runtime statuses, including their CPU usage, instance load, memory usage, and network traffic. However, such infrastructure monitoring cannot meet all service needs. For example, you may not be able to locate the bottleneck when system operating performance degrades or identify the specific call error when you open a page.

To address these challenges, EDAS provides a series of digital operation components, allowing you to precisely monitor and track every single component or service in the distributed system and quickly pinpoint the bottleneck.

30.6. Limits

Currently, EDAS imposes limits on programming languages and package sizes used in application development.

Item	Description
Programming language	Currently, only Java applications can be published in EDAS.
Package size	Currently, only application deployment packages no larger than 500 MB can be published in EDAS.

30.7. Terms

Ali-Tomcat

Ali-Tomcat is a container that EDAS depends on to run services. It integrates service publishing, subscription, call tracing, and other core functions. Applications must be published to Ali-Tomcat in both development and runtime environments.

Dubbo

Dubbo is a distributed service framework that provides high-performance and transparent remote procedure calls (RPC). It is the core framework of Alibaba's SOA service governance solution, providing support for over 3 billion access requests for more than 2,000 services every day. It is widely used by various member sites of the Alibaba Group.

Cluster

A cluster is a collection of cloud resources that are required to run applications. ECS instances need to be added to a cluster. If you do not select one, the ECS instances are added to the default cluster.

Namespace

A namespace is an isolated resource environment that is established in a region. It contains one or more clusters. Different namespaces are logically isolated from each other by nature. Clusters need to be created in a namespace. If you do not select one, the clusters are created in the default namespace.

EDAS Agent

EDAS Agent is a daemon of EDAS that is installed on ECS instances. It is responsible for the communication between an EDAS service cluster and the applications deployed on the ECS instances in the cluster. EDAS Agent is used for application management, status reporting, and information retrieval. It also serves as the communication channel between the EDAS console and your applications.

RPC

The EDAS RPC service provides support for the Dubbo framework. An application that is developed by using the Dubbo framework and deployed with a WAR package can be seamlessly published and managed in EDAS and use the service governance and data operation functions of EDAS.

Application lifecycle

Applications are the basic management units in EDAS. A single application generally contains multiple instances. EDAS EDAS provides a comprehensive application lifecycle management mechanism, covering the entire process from application publishing to operation, including application creation, deployment, startup, rollback, scale-out, scale-in, stop, and deletion.

Application instance quota

The application instance quota sets the maximum number of instances for all applications held by a primary account and its RAM users.

31.MaxCompute

31.1. What is MaxCompute?

MaxCompute is a highly efficient, highly available, and low-cost EB-level computing service for big data. It is independently developed by Alibaba Cloud. This service is used within Alibaba Group to process exabytes of data each day. MaxCompute is a distributed system designed for big data processing. As one of the core services in the Alibaba Cloud computing solution, MaxCompute is used to store and compute structured data.

MaxCompute is designed to support multiple tenants and provide features, such as data security and horizontal scaling. The service provides centralized programming interfaces for various data processing tasks of different users based on an abstract job processing framework.

MaxCompute is used to store and compute large amounts of structured data. It provides various data warehousing solutions as well as big data analytics and modeling services. MaxCompute is designed to make the analysis and processing of large amounts of data easier. You can analyze big data without deep knowledge about distributed computing.

MaxCompute has the following features:

- Uses a distributed architecture that can be scaled as needed.
- Provides an automatic storage and fault tolerance mechanism to ensure high data reliability.
- Allows all computing tasks to run in sandboxes to ensure high data security.
- Uses RESTful APIs to provide services.
- Supports both uploads and downloads of high-concurrency, high-throughput data.
- Supports two service models: the offline computing model and the machine learning model.
- Supports data processing methods based on programming models such as SQL, MapReduce, Graph, and MPI.
- Supports multiple tenants, which allows multiple users to collaborate on data analytics.
- Provides user permission management based on access control lists (ACLs) and policies, which allows you to configure flexible data access control policies to prevent unauthorized access to data.
- Supports Spark on MaxCompute for enhanced applications.
- Supports Elasticsearch on MaxCompute for enhanced applications.
- Supports the access to and processing of unstructured data.
- Supports the deployment of multiple clusters in a single region.
- Supports multi-region deployment.
- Uses the column store method, and supports Key Management Service (KMS) to encrypt data files.
- Stores audit logs and dumps them to a specified server directory for long-term storage and management.

31.2. Integration with other Alibaba Cloud services

MaxCompute has been integrated with several other Alibaba Cloud services to quickly implement a variety of business scenarios.

MaxCompute and DataWorks

DataWorks uses MaxCompute as the core computing and storage engine to provide offline processing and analysis of large amounts of data. DataWorks offers fully hosted services for visual workflow development, scheduling, and O&M.

MaxCompute works with DataWorks to provide complete ETL and data warehouse management capabilities, as well as classic distributed computing models such as SQL, MapReduce, and Graph. These models enable you to process large amounts of data while reducing business costs and maintaining data security.

MaxCompute and Data Integration

You can use Data Integration to load data from different sources such as MySQL databases into MaxCompute, and export data from MaxCompute to various business databases.

Data Integration has been integrated into DataWorks and is configured and run as a data synchronization task. You can directly configure MaxCompute data sources in DataWorks, and then configure tasks to read or write data from or to MaxCompute tables. The entire process is completed on a single platform.

MaxCompute and Machine Learning Platform for AI

Machine Learning Platform for AI (PAI) is a machine learning algorithm platform based on MaxCompute. PAI provides an end-to-end machine learning platform for data processing, model training, service deployment, and prediction without data migration. After creating a MaxCompute project and activating PAI, you can use the algorithm components of the machine learning platform to perform operations such as model training on MaxCompute data.

MaxCompute and Quick BI

After processing data in MaxCompute, you can add the project as a Quick BI data source. Then, you can create reports based on MaxCompute table data in the Quick BI console for visual data analysis.

MaxCompute and AnalyticDB for MySQL

AnalyticDB for MySQL is a cloud computing service designed for online analytical processing (OLAP). It can process huge amounts of data in a highly concurrent and real-time manner. AnalyticDB for MySQL can be used in combination with MaxCompute to implement big data-driven business systems. You can use MaxCompute to compute and mine data offline and generate high-quality data. Then, you can import the data to AnalyticDB for MySQL for business systems to perform analysis.

You can use one of the following methods to import data from MaxCompute to AnalyticDB for MySQL:

- Use the import and export feature of DMS for AnalyticDB for MySQL.
- Use DataWorks to configure a data synchronization task and configure MaxCompute Reader and AnalyticDB Writer.

MaxCompute and Recommendation Engine

Recommendation Engine (RecEng) is a recommendation service framework established in the Alibaba Cloud computing environment. The recommendation service is typically composed of three parts: log collection, recommendation computing, and product connection. The input and output of offline recommendation computing are both MaxCompute tables.

On the Resources page of the RecEng console, you can add a MaxCompute project as a RecEng computing resource in the same way as you add a cloud computing resource.

MaxCompute and Table Store

Table Store is a distributed NoSQL data storage service built on the Apsara distributed operating system of Alibaba Cloud. MaxCompute 2.0 allows you to directly access and process table data in Table Store through external tables.

MaxCompute and Object Storage Service

Object Storage Service (OSS) is a secure, cost-efficient, and highly reliable cloud storage service that can store large amounts of data. MaxCompute 2.0 allows you to directly access and process table data in OSS through external tables.

MaxCompute and OpenSearch

Alibaba Cloud OpenSearch is a large-scale distributed search engine platform developed by Alibaba Cloud. After data is processed by MaxCompute, you can access MaxCompute data by adding data sources on the OpenSearch platform.

MaxCompute and Mobile Analytics

Mobile Analytics is a product launched by Alibaba Cloud to collect and analyze mobile application usage data, providing developers with an end-to-end digital operations service. When the basic analysis reports that come with Mobile Analytics cannot meet the personalized needs of app developers, app developers can synchronize data to MaxCompute with a single click. They can further process and analyze their data based on their business requirements.

MaxCompute and Log Service

Log Service allows you to quickly complete data collection, consumption, delivery, query, and analysis. After collecting log data, you need more personalized analysis and mining. You can synchronize Log Service data to MaxCompute through Data Integration in DataWorks, and then use MaxCompute to perform personalized and in-depth data analysis and mining on log data.

MaxCompute and RAM

Resource Access Management (RAM) is a service provided by Alibaba Cloud to manage user identities and resource access permissions.

Character sets supported by other Alibaba Cloud services

Service	Supported character set
Table Store	UTF-8
PAI	UTF-8
OSS	UTF-8
Quick BI	UTF-8
DataWorks	UTF-8, GBK, CP936, and ISO-8859 are supported when data is uploaded in DataStudio. However, all data is encoded in UTF-8 in DataWorks. UTF-8 and GBK are supported when data is downloaded.

31.3. Benefits

Excellent big data cloud service and real data sharing platform in China

- MaxCompute can be used for data warehousing, mining, analytics, and sharing.
- Alibaba Group implements this unified data processing platform in several of its own services, such as Aliloan, Data Cube, DMP (Alimama), and Yu'e Bao.

Support for large numbers of clusters, users, and concurrent jobs

- A single cluster can contain more than 10,000 servers and maintain 80% linear scalability.
- A single MaxCompute service supports more than 1 million servers in multiple clusters without limits. However, linear scalability is slightly affected. It also supports the local multi-data center mode.
- A single MaxCompute service supports more than 10,000 users, more than 1,000 projects, and more than 100 departments of multiple tenants.
- A single MaxCompute service supports more than 1 million jobs (daily submitted jobs on average) and more than 20,000 concurrent jobs.

Big data computing at your fingertips

You do not need to worry about the storage difficulties and prolonged computing time caused by the increase of the data volume. MaxCompute automatically expands the storage and computing capabilities of clusters based on the volume of data that needs to be processed. This allows you to focus on data analytics and mining to maximize your data value.

Out-of-the-box service

You do not need to worry about cluster creation, configuration, and O&M. Only a few simple steps are required to upload data, analyze data, and obtain analysis results in MaxCompute.

Secure and reliable data storage

The multi-level data storage and access control mechanisms are used to protect user data against loss, leak, and interception. These mechanisms include multi-replica technology, read/write request authentication, and application and system sandboxes.

Reliable management nodes

Multi-node cluster architecture is used. The management nodes of each component feature high availability. Faults that occur on O&M management nodes do not affect normal business operations.

Powerful fault tolerance

MaxCompute supports automatic fault tolerance for the failures of server hard disks in a cluster and supports hot swapping of hard disks. In the event of a hard disk failure, services can be restored within two minutes.

Comprehensive storage space management

MaxCompute allows you to query information about both the storage capacity and usage of distributed file systems. It supports data lifecycle management. MaxCompute also allows you to store data in different locations based on the data value or tag. For example, you can write temporary files to SSDs to accelerate I/O operations. This facilitates efficient use of cluster data. MaxCompute also supports the self-optimizing Zstandard compression algorithm, which achieves the optimal compression ratio.

Comprehensive data backup

- MaxCompute allows you to perform full or incremental data backup and restore data from storage media.
- MaxCompute allows you to back up data for clusters in different data centers. This meets the requirements of mutual data backups among multiple data centers. You can use Apsara Bigdata Manager (ABM) to manage the backup process in a visualized manner.

Secure and reliable access control

- MaxCompute allows you to manage data access permissions, including logon permissions, table creation permissions, read/write permissions, and whitelist-related permissions.
- MaxCompute allows you to use the Apsara Stack Cloud Management (ASCM) console to manage administrative permissions, including administrator classification.
- MaxCompute allows you to use the ASCM console to manage user permissions in a centralized manner. You can manage the access control features of all components in the system. You can also block common users from querying access control details and simplify access control for administrators. This improves the usability and user experience of access control.

Multi-tenancy for multi-user collaboration

By configuring different data access policies, you can enable multiple data analysts in an organization to work together and make data accessible to users with permissions granted by the organization. This ensures data security and maximizes productivity.

- **Isolation**: You can submit the tasks of multiple tenants (projects) to different queues for concurrent running. Resources are isolated among tenants.
- **Permission**: You can manage different tenants in a centralized manner and perform dynamic configuration, management, isolation, and usage statistics of tenant resources. The management of multi-level tenants is supported.
- **Scheduling**: MaxCompute supports multi-tenant scheduling for multiple clusters and multiple resource pools.

Multi-region deployment

- You can specify compute clusters to efficiently use computing resources.
- Data exchanges between clusters are completed within MaxCompute, and data replication and synchronization between clusters are managed based on configured policies. Therefore, cross-region data processing is no longer involved, which significantly reduces the waiting time for data processing.

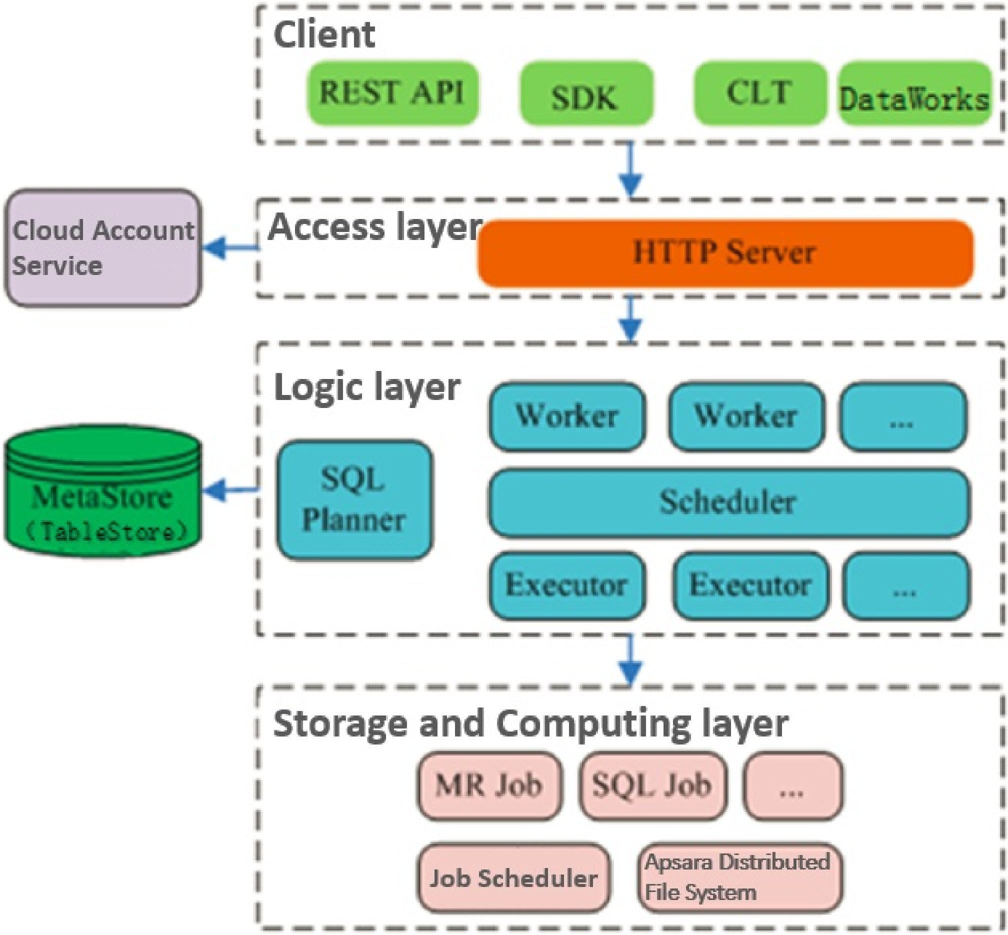
Multi-device support

You can use CPUs, hard disks, memory, and network interface controllers with different specifications in a single-component cluster without an effect on cluster running performance. This ensures maximum compatibility with existing devices.

31.4. Architecture

MaxCompute architecture shows the MaxCompute architecture.

MaxCompute architecture



The MaxCompute service is divided into four parts: **client**, **access layer**, **logic layer**, and **storage and computing layer**. Each layer can be scaled out.

The following methods can be used to implement the functions of a MaxCompute client:

- **API**: RESTful APIs are used to provide offline data processing services.
- **SDK**: RESTful APIs are encapsulated in SDKs. SDKs are currently available in programming languages such as Java.
- **Command line tool (CLT)**: This client-side tool runs on Windows and Linux. CLT allows you to submit commands to manage projects and use DDL and DML.
- **DataWorks**: DataWorks provides upper-layer visual ETL and BI tools that allow you to synchronize data, schedule tasks, and create reports.

The access layer of MaxCompute supports HTTP, HTTPS, load balancing, user authentication, and service-level access control.

The logic layer is at the core of MaxCompute. It supports project and object management, command parsing and execution logic, and data object access control and authorization. The logic layer is divided into control and compute clusters. The control cluster manages projects and objects, parses queries and commands, and authorizes access to data objects. The compute cluster executes tasks. Both control and compute clusters can be scaled out as required. The control cluster is comprised of three different roles: Worker, Scheduler, and Executor. These roles are described as follows:

- **The Worker role processes all RESTful requests** and manages projects, resources, and jobs. Workers forward jobs that need to launch Fuxi tasks (such as SQL, MapReduce, and Graph jobs) to the Scheduler for further processing.
- **The Scheduler role schedules instances**, splits instances into multiple tasks, sorts tasks that are pending for submission, and queries resource usage from FuxiMaster in the compute cluster for throttling. If there are no idle slots in Job Scheduler, the Scheduler stops processing task requests from Executors.
- **The Executor role is responsible for launching SQL and MapReduce tasks.** Executors submit Fuxi tasks to FuxiMaster in the compute cluster and monitor the operating status of these tasks.

In summary, when you submit a job request, the Web server at the access layer queries the IP addresses of registered Workers and sends API requests to randomly selected Workers. The Workers then send these requests to the Scheduler for scheduling and throttling. Executors actively poll the Scheduler queue. If the necessary resources are available, the Executors start executing tasks and return the task execution status to the Scheduler.

The storage and computing layer of MaxCompute is a core component of the proprietary cloud computing platform developed by Alibaba Cloud. The architecture diagram illustrates only major modules.

31.5. Features

31.5.1. Tunnel

31.5.1.1. Terms

Tunnel is the data tunnel service provided by MaxCompute. You can use Tunnel to import data from various heterogeneous data sources into MaxCompute or export data from MaxCompute. As the unified channel for MaxCompute data transmission, Tunnel provides stable and high-throughput services.

Tunnel provides RESTful APIs and Java SDKs to facilitate programming. You can upload and download only table data (excluding view data) through Tunnel.

31.5.1.2. Tunnel features

- The channel through which data flows in to and out of MaxCompute
- Highly concurrent upload and download
- Horizontal expansion of service capabilities
- Tools based on MaxCompute Tunnel, such as TT, CDP, Flume, and Fluentd
- Support for reads and writes of tables (excluding views)
- Support for data writes in append mode
- Concurrency capabilities to improve total throughput
- Support for data upload only when target partitions exist

- Real-time upload mode

31.5.1.3. Data upload and download through Tunnel

Tunnel commands

```
odps@ > tunnel upload log.txt test_project.test_table/p1="b1",p2="b2 “;
```

```
odps@ > tunnel download test_project.test_table/p1="b1",p2="b2" log.txt;
```

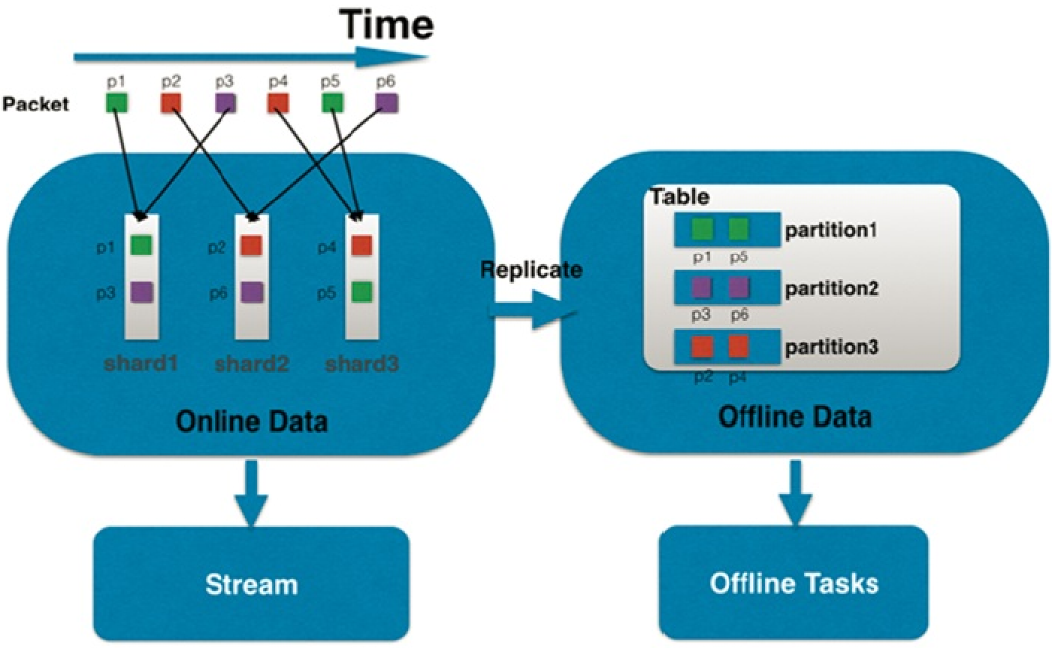
Notes

- Tunnel is a CLT based on the Tunnel SDK and can be used to upload local text files to MaxCompute or download data tables to your local device.
- You must create table partitions before using Tunnel.
- DataX, CDP, and TT provide enhanced Tunnel-based tools, which are used to exchange data between MaxCompute and relational databases.
- You can import log data by using the Flume and Fluentd tools.
- In some scenarios, you can develop custom tools based on Tunnel.

Real-time upload

- Upload in small batches
- High QPS performance
- Latency within milliseconds
- Subscription available

Real-time upload



31.5.2. SQL

31.5.2.1. Terms

The syntax of MaxCompute SQL is similar to SQL. It can be considered as a subset of standard SQL. However, MaxCompute SQL is not equivalent to a database, because it does not possess many characteristics that a database has, such as transactions, primary key constraints, and indexes. The maximum SQL statement size currently allowed in MaxCompute is 2 MB.

MaxCompute SQL offline computing is applicable to scenarios that have a large amount of data (measured in TBs) and that do not have high real-time processing requirements. It takes a relatively long time to prepare and submit each job. Therefore, MaxCompute SQL is not optimal for services that need to process thousands of transactions per second. MaxCompute SQL online computing is applicable to scenarios that require near-real-time processing.

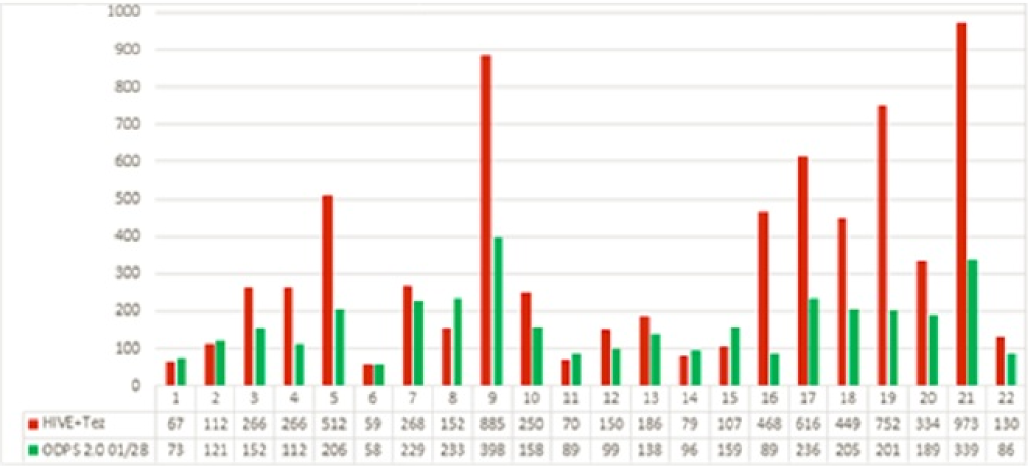
31.5.2.2. SQL characteristics

- It is suitable for processing large volumes of data (TBs or PBs).
- It has relatively high latency. The runtime of each SQL statement ranges from dozens of seconds to several hours.
- Its syntax is similar to that for Hive HQL. It is extended based on standard SQL syntax.
- It does not involve transactions or primary keys.
- It does not support UPDATE and DELETE operations.

31.5.2.3. Comparison with open source products

- TPC-H 1 TB data benchmark: Compared with Hive (Apache Hive-1.2.1-bin + Tez-UI-0.7.0 with CBO), MaxCompute has a 95.6% improvement in performance.

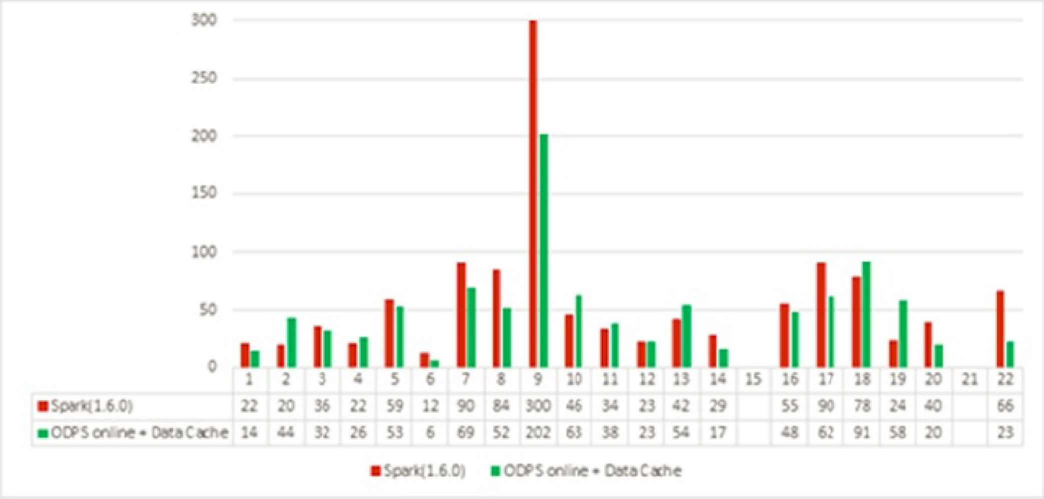
MaxCompute 2.0 VS Hive



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
🟥 HIVE+Tez	67	112	266	266	512	59	268	152	885	250	70	150	186	79	107	468	616	449	752	334	973	130
🟩 ODPS 2.0 01/28	73	121	152	112	206	58	229	233	398	158	89	99	138	96	159	89	236	205	201	189	339	86

- TPC-H 450 GB data benchmark: Compared with Spark SQL V1.6.0 (the latest release), MaxCompute has a 17.8% improvement in performance.

MaxCompute 2.0 VS Spark SQL



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
🟥 Spark(1.6.0)	22	20	36	22	59	12	90	84	300	46	34	23	42	29		55	90	78	24	40		66
🟩 ODPS online + Data Cache	14	44	32	26	53	6	69	52	202	63	38	23	54	17		48	62	91	58	20		23

31.5.3. MapReduce

31.5.3.1. Terms

MapReduce is a programming model, which is basically equivalent to Hadoop MapReduce. The model is used for parallel MaxCompute operations on large-scale data sets (measured in TBs).

MaxCompute provides a MapReduce programming interface. You can use Java APIs, which is provided by MapReduce, to write MapReduce programs for processing data in MaxCompute.

> **Note** All data in MaxCompute is stored as tables. The inputs and outputs of MaxCompute MapReduce can only be tables. Custom output formats are not supported, and no interface, such as a file system, is provided.

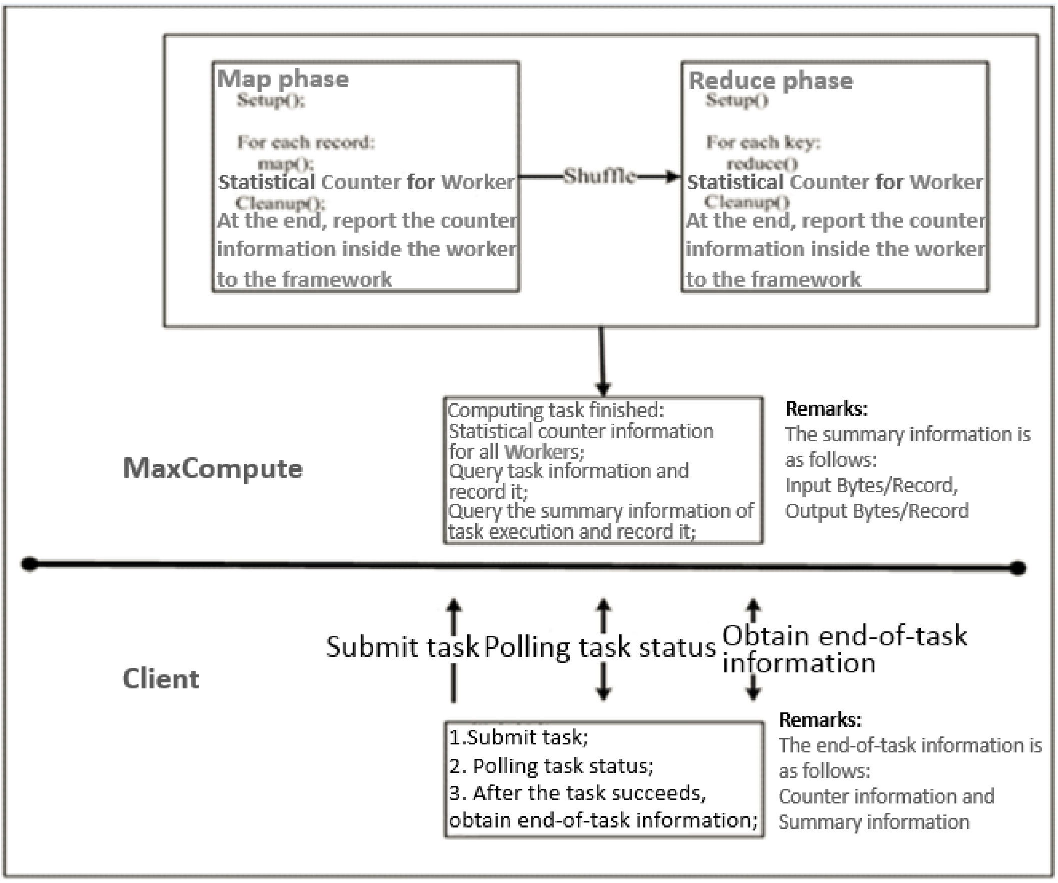
31.5.3.2. MapReduce characteristics

- It only supports the input and output of MaxCompute built-in data types.
- It supports the input and output of multiple tables to different partitions.
- It reads resources.
- It does not support using views as data inputs.
- It supports MapReduce programming only in the JDK 1.8 environment.
- It provides a limited sandbox security environment.

31.5.3.3. MaxCompute MapReduce process

The following figure shows the MapReduce process in MaxCompute:

MapReduce process


Map phase
Setup():
For each record:
map();
Statistical Counter for Worker
Cleanup():
At the end, report the counter information inside the worker to the framework
Shuffle
Reduce phase
Setup()
For each key:
reduce()
Statistical Counter for Worker
Cleanup()
At the end, report the counter information inside the worker to the framework
MaxCompute
Computing task finished:
Statistical counter information for all Workers;
Query task information and record it;
Query the summary information of task execution and record it;
Remarks:
The summary information is as follows:
Input Bytes/Record,
Output Bytes/Record
Client
Submit task
Polling task status
Obtain end-of-task information
1.Submit task;
2. Polling task status;
3. After the task succeeds, obtain end-of-task information;
Remarks:
The end-of-task information is as follows:
Counter information and
Summary information

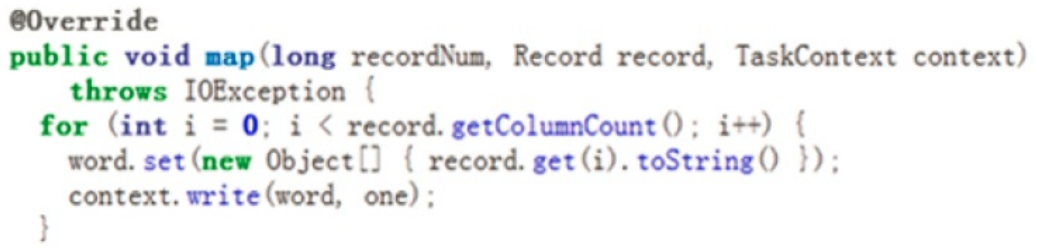
31.5.3.4. Hadoop MapReduce VS MaxCompute MapReduce

The following table describes the comparison between Hadoop MapReduce and MaxCompute MapReduce.

Mapper/Reducer

Mapper/Reducer	
Hadoop MapReduce	MaxCompute MapReduce
Map (InKey key, InputValue value, OutputCollector<OutKey, OutValue> output, Reporter reporter)	Map (long key, Record record, TaskContext context)
Reduce (InKey key, Iterator<InValue> values, OutputCollector<OutKey, OutValue> output, Reporter reporter)	Reduce (lRecord key, Iterator<Record> values, TaskContext context)

MapReduce



```
@Override
public void map(long recordNum, Record record, TaskContext context)
    throws IOException {
  for (int i = 0; i < record.getColumnCount(); i++) {
    word.set(new Object[] { record.get(i).toString() });
    context.write(word, one);
  }
```

31.5.4. Graph

31.5.4.1. Terms

Graph is the computing framework of MaxCompute designed for iterative graph processing. It provides programming interfaces similar to Pregel, allowing you to use Java SDKs to develop efficient machine learning and data mining algorithms.

Graph jobs use graphs to build models. This process outputs a result after performing iterative graph editing and evolution.

31.5.4.2. Graph characteristics

- It is a graphic computing programming model (similar to Google Pregel).
- It loads data to the memory, which is superior in multiple iteration scenarios.
- It can be used to develop machine learning algorithms.
- It can support 10 billion vertices and 150 billion edges.
- Its typical applications include:
  - PageRank
  - K-means clustering
  - Level 1 and level 2 relationships and shortest path
- Graph jobs process graph data.
- The original data is stored in tables. The user-defined Graph Loader loads data in the table as vertexes and edges.
- It supports iterative computing.

31.5.4.3. Graph relational network models

A relational network engine provides a variety of business-oriented relational network models. It helps you quickly implement relational data mining at finer granularities.

Community discovery

- Input to the engine: relational data.
- Engine output: IDs and community IDs.
- Computing logic: locates N communities with the optimal global network connection. The communities are close enough internally, and sparse enough in between.

Semi-supervised category

- Input to the engine: problematic IDs.
- Engine output: potentially problematic IDs and weights.
- Computing logic: uses existing problematic IDs (of one or more categories) to determine potential problematic IDs of the same or multiple categories and corresponding weights based on the entire network connection relationships.

Isolated point detection

- Input to the engine: relational data.
- Engine output: isolated points and weights.
- Computing logic: determines whether there are relatively isolated nodes using the connection relationships in a relational network, and generates the result.

Key point mining

- Input to the engine: relational data.
- Engine output: key point IDs and categories.
- Computing logic: calculates the key type nodes in a computing network using the connection relationships (such as centrality, influence, and betweenness centrality) in a relational network.

Level N relationships

- Input to the engine: relational data.
- Engine output: retrievable relational networks.
- Computing logic: manages multi-dimensional relationships using the connection relationships in the relational network, and creates indexes to facilitate the query for specific associations of an ID.

31.5.5. Unstructured data processing (integrated computing scenarios)

Alibaba Cloud introduced the MaxCompute-based unstructured data processing framework so that MaxCompute SQL commands can directly process external user data, such as unstructured data from OSS. You are no longer required to first import data into MaxCompute tables.

You can run a simple DDL statement to create an external table in MaxCompute and associate MaxCompute tables with external data sources. This table can then act as an interface between MaxCompute and external data sources. The external table can be accessed in the same way as a MaxCompute table, and computed by MaxCompute SQL.

MaxCompute allows you to process the following data sources by creating external tables:

- Internal data sources: OSS, Table Store, AnalyticDB, ApsaraDB for RDS, HDFS (Alibaba Cloud), and TDDL.
- External data sources: HDFS (Open Source), ApsaraDB for MongoDB, and Hbase.

31.5.6. Unstructured data processing in MaxCompute

MaxCompute has the following problems when processing unstructured data: MaxCompute stores data as volumes and must export generated unstructured data to an external system for processing.

To alleviate these problems, MaxCompute uses external tables to enable connections between MaxCompute and various data types. MaxCompute uses external tables to read and write data volumes as well as process unstructured data from external sources such as OSS.

31.5.7. Enhanced features

31.5.7.1. Spark on MaxCompute

31.5.7.1.1. Terms

Spark on MaxCompute is a solution developed by Alibaba Cloud to enable seamless use of Spark on the MaxCompute platform, extending the functions of MaxCompute.

Spark on MaxCompute provides a native Spark user experience with its native Spark components and APIs. It allows access to MaxCompute data sources and better security for multi-tenant scenarios. It also offers a management platform enabling Spark jobs to share resources, storage, and user systems with MaxCompute jobs. This guarantees high performance and low costs. Spark can work with MaxCompute to create better and more efficient data processing solutions. Spark Community applications can run seamlessly in **Spark on MaxCompute**.

Spark on MaxCompute has an independent data development node in DataWorks and supports data development in DataWorks.

31.5.7.1.2. Features of Spark on MaxCompute

Processing of data from MaxCompute and unstructured data sources

- Processes MaxCompute tables through APIs based on Scala, Python, Java, and R programming languages.
- Processes MaxCompute tables through components such as Spark SQL, Spark MLlib, GraphX, and Spark Streaming.
- Can process unstructured data from Alibaba Cloud OSS.

User-friendly experience and management functions

- Supports job submission in a way similar to Spark on YARN. **Spark on MaxCompute** is compatible with YARN and HDFS APIs.
- Supports components including Spark SQL, Spark MLlib, GraphX, and Spark Streaming.
- Can work with SQL and Graph components of MaxCompute to form optimized solutions.
- Can connect to the native Spark UI.
- Allows you to directly use the powerful management functions of MaxCompute.
- Supports not only Spark Community but also tools such as client, Livy, and Hue.

Scalability

Spark and MaxCompute share cluster resources. Spark resources can be scaled from large-scale MaxCompute clusters.

31.5.7.1.3. Spark features

The following table describes **Spark on MaxCompute** features.

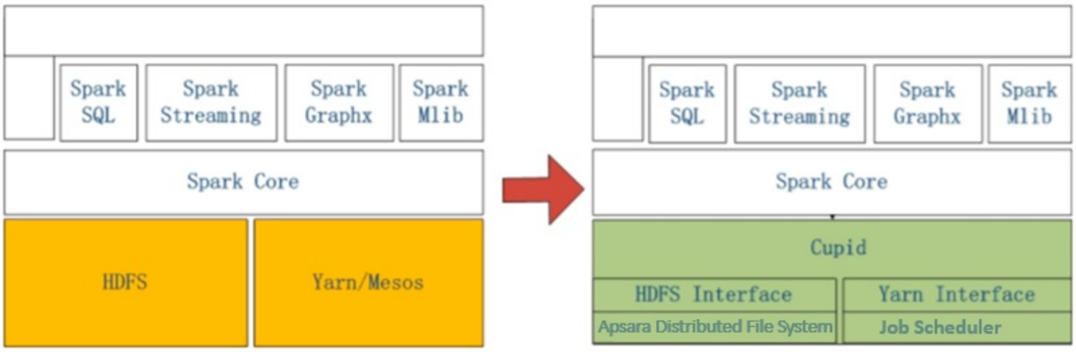
Features

Type	Feature	Description
Distributed cluster	Cluster deployment Cluster monitoring	Provide an O&M platform to monitor clusters and nodes.
Data processing component	Support for components such as Spark SQL, Spark MLlib, GraphX, and Spark Streaming	Provide native Spark components.
Job management	Centralized resource management, life cycle management, and authentication	The features are available through compatible YARN APIs.
Data sources	Unstructured data Table data sources in MaxCompute	Provide data processing capabilities of SQL and MapReduce on MaxCompute.
Security management	User identification, data authentication, and multi-tenant job isolation	Harden Spark security through authentication and sandboxes.

31.5.7.1.4. Spark architecture

The following figure shows the architectural comparison between **Spark on MaxCompute** and **native Spark**.

Architectural comparison between Spark on MaxCompute and native Spark



Note On the left is the native Spark architecture and on the right is the **Spark on MaxCompute** architecture.

As shown in the figure, Spark on MaxCompute has the computing capabilities of native Spark and the functions related to management, O&M, scheduling, security, and data interconnection. The management function of Spark is implemented by starting a Cupid Task instance of MaxCompute. The resource application function is realized through layer-1 YARN APIs provided by MaxCompute. The security function is offered through the sandbox mechanism of MaxCompute. The processing of and interconnection between data and metadata are also made available. The module details are described as follows:

- The MaxCompute control cluster starts a Spark driver by using the Cupid Task instance. The Spark driver uses YARN APIs to apply for resources from FuxiMaster, the central resource manager.
- The MaxCompute control cluster manages user quota consumed by running Spark instances, life cycles of Spark instances, and permissions on accessible data sources.
- The MaxCompute computing cluster starts a Spark driver and Executor as parent and child processes and executes Spark code in the sandbox of MaxCompute, ensuring security in multi-tenant scenarios.
- MaxCompute allows you to use the native Spark UI through its Proxy Server and manage job information through its management components.

31.5.7.1.5. Benefits of Spark on MaxCompute

Support for the complete Spark ecosystem

Provides consistent user experience with that of open source Spark.

Full integration with MaxCompute

Implements centralized management of resources, data, and security features for both Spark and MaxCompute.

Combination of Spark and the Apsara system

Combines the flexibility and ease of use of Spark with the high availability, scalability, and stability of the Apsara system.

Support for multi-tenancy

Reduces costs by centrally scheduling resources in large-scale clusters and ensuring high performance of physical machines.

Support for cross-cluster scheduling

Maximizes the efficiency of cluster resources by effectively allocating clusters and scheduling resources across clusters.

Support for real-time scaling of Spark resources

Scales resources in Spark Community in real time to better utilize resources and avoid waste.

> **Note** Real-time Spark resource scaling is not enabled on all MaxCompute clusters. To use this function, contact the MaxCompute team.

31.5.7.2. Elasticsearch on MaxCompute

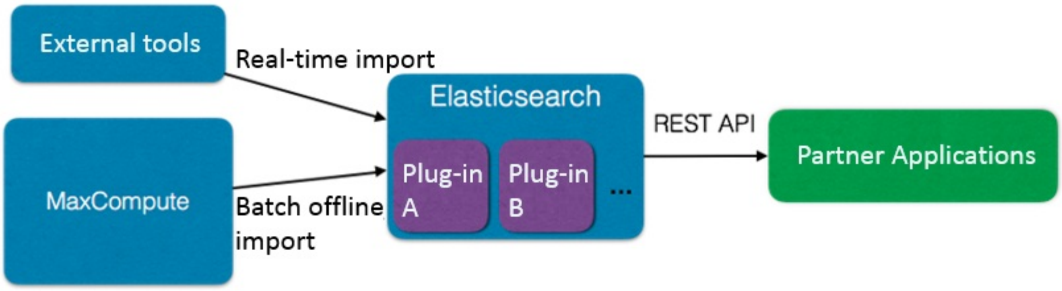
31.5.7.2.1. Overview

Elasticsearch on MaxCompute is an enterprise-class full-text retrieval system developed by Alibaba Cloud to retrieve large volumes of data with near-real-time search performance.

Elasticsearch on MaxCompute provides elastic full-text retrieval and supports native Elasticsearch APIs. You can import data from heterogeneous data sources and perform O&M for clusters and services. The centralized scheduling and management capabilities of MaxCompute allow Elasticsearch to provide more efficient core services for data retrieval at large volumes. **Elasticsearch on MaxCompute** can also work with plug-ins available from the Elasticsearch open source community to enhance retrieval functions.

Elasticsearch on MaxCompute allows you to use tools to import data from external sources in real time. You can also import offline data from MaxCompute. After the imported data is indexed, Elasticsearch on MaxCompute provides retrieval services through RESTful APIs. The following figure shows its usage.

Elasticsearch on MaxCompute usage



31.5.7.2.2. Features of Elasticsearch on MaxCompute

Distributed cluster architecture

- Improves retrieval and reliability of data with a distributed architecture.
- Supports elastic scaling.
- Supports dynamic scaling.
- Supports service-level O&M and monitoring.

Robust full-text retrieval

- Performs full-text retrieval at the word, phrase, sentence, and section levels.
- Available in languages such as Chinese and English.
- Provides precise word segmentation with 100% recall for Chinese information retrieval.
- Supports complex searching methods, such as Boolean retrieval, proximity search, and fuzzy search.
- Sorts search results by relevance, field, and custom weight, and allows for secondary sorting.
- Performs statistical classification and analysis of search results.
- Allows real-time indexing and retrieval, so that inserted data can be retrieved immediately.
- Allows an index to be used multiple times after it is created.
- Allows you to modify the index structure in real time or rebuild the index to re-distribute data.

Support for multiple data sources

- Imports data from native Elasticsearch interfaces.
- Provides data import tools for MaxCompute.
- Supports full and incremental update.

Reliability

- Stores data in multiple copies, preventing user data from being lost during the downtime of machines.
- Implements a high availability architecture and comprehensive failover for nodes and services.
- Provides comprehensive O&M and monitoring functions.
- Authenticates access to protect data from malicious operations and ensure security.

31.5.7.2.3. Elasticsearch features

Elasticsearch on MaxCompute features are described as follows:

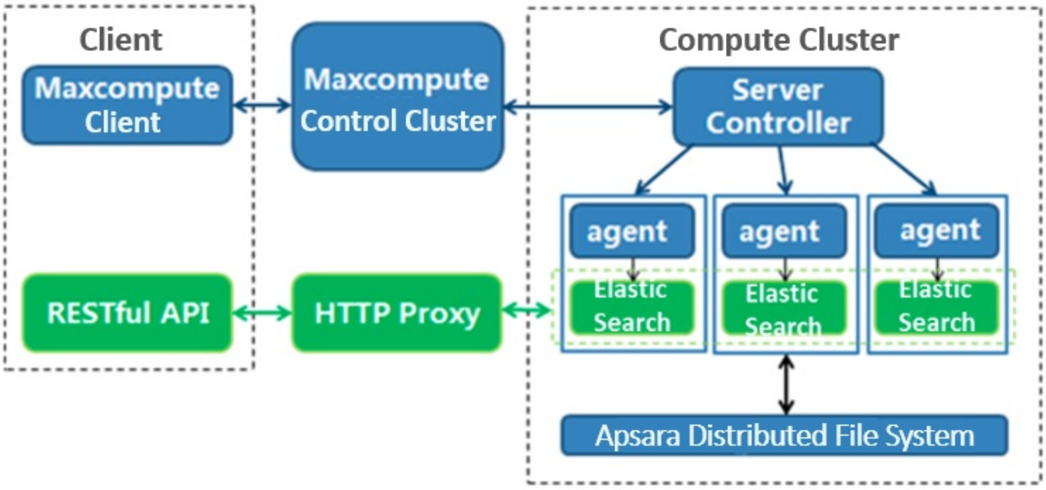
Features

Type	Feature	Description
Distributed cluster	Cluster deployment Cluster monitoring	Provide an O&M platform to monitor clusters, nodes, and indexes.
Retrieval management	Index configuration management Structure definition and index rebuilding	Provide a retrieval management platform and support configuration.
Full-text retrieval	Retrieval Sorting Statistical analysis	The features are provided through RESTful APIs.
Data collection	Elasticsearch data import APIs MaxCompute data import tools Full and incremental collection	Support a variety of interfaces to collect native data. Provide integrated tools to import MaxCompute data.
Service authentication	Service-level user authentication	Allow you to configure user authentication in a centralized manner.

31.5.7.2.4. Elasticsearch architecture

Elasticsearch on MaxCompute provides core search engine services, management platforms for O&M and indexes, MaxCompute management system, and MaxCompute data import tools. It can work with universal data import interfaces and data retrieval SDKs of Elasticsearch, enabling you to retrieve applications and perform full-text retrieval of large volumes of data. The following figure shows the overall architecture.

Overall architecture



An Elasticsearch cluster corresponds to a MaxCompute Server Task instance in MaxCompute. You can quickly and flexibly deploy, operate, and expand Elasticsearch clusters on a MaxCompute client. In the overall architecture,

- The MaxCompute control cluster starts Server Controller and forwards control requests from a client.
- Server Controller is the core component for Elasticsearch cluster management. It applies for resources, starts each Elasticsearch node, and responds to the control requests that a client forwards through the control cluster. It also returns the running status of Elasticsearch clusters or adjusts the clusters.
- An agent starts Elasticsearch node processes, monitors node running status, handles failover events, and executes tasks distributed by Server Controller.
- **Elasticsearch on MaxCompute** stores its data in Apsara Distributed File System. Once a node is started successfully, Elasticsearch on MaxCompute can provide services through HTTP Proxy and allow users to use **its functions** through RESTful APIs.

31.5.7.2.5. Benefits

Integration of big data computing and data retrieval for resource sharing

Elasticsearch on MaxCompute can access and import MaxCompute data to an Elasticsearch cluster to perform a full-text search. This facilitates centralized data management and usage.

Centralized management of computing and storage resources

You do not need to worry about the storage problems and prolonged computing tasks caused by the increase of the data volume. Elasticsearch on MaxCompute supports automatic scaling of your cluster storage and retrieval capacities based on the volume of your data. This way, you can focus on data analytics and mining to maximize your data value.

Provision of services such as Elasticsearch cluster deployment and O&M

You do not need to worry about cluster creation, configuration, and O&M. Only a few simple steps are required to upload data, analyze data, and obtain analysis results in the offline analysis service.

Secure and reliable data storage

Elasticsearch on MaxCompute uses the multi-replica technology to store user data at multiple layers. This prevents the loss, leak, and interception of data.

Open service interfaces

Elasticsearch on MaxCompute provides Elasticsearch SDKs that are native and open. This allows you to import, index, and retrieve data by using **Elasticsearch on MaxCompute**.

Multi-tenancy for multi-user collaboration

- **Isolation**: Elasticsearch on MaxCompute supports the cross-cluster search feature and allows you to submit tasks to different clusters for execution. Resources among clusters are isolated. Elasticsearch on MaxCompute does not support single-cluster multi-tenancy.
- **Permissions**: You can manage clusters in a centralized manner to implement the configuration, management, isolation, and usage statistics of cluster resources. In addition, it supports multi-level permissions and multi-level tenant management based on Apsara Stack.
- **Scheduling**: Elasticsearch on MaxCompute supports multi-tenant scheduling for multiple clusters and multiple resource pools.

31.5.8. Multi-region deployment

31.5.8.1. Multi-region deployment of MaxCompute

MaxCompute can be deployed across regions. Control clusters are deployed in a unified manner and are used to configure resources and manage computing tasks. Compute clusters are separately deployed in each region to create projects and distribute computing tasks.

The multi-region deployment of MaxCompute has the following features:

- One MaxCompute service can manage multiple clusters in different regions.
- Data exchanges between clusters are completed within MaxCompute, and data replication and synchronization between clusters are managed based on configured policies.
- Metadata is stored in a centralized manner. Therefore, the infrastructure requirements, such as the network connections of different data centers, are relatively high.
- A unified account system is required.
- The development systems for big data applications, such as DataWorks, are used for all clusters in all regions.
- MaxCompute must run in multi-cluster mode to support multi-region deployment.

> **Note** Take note the following conditions and limits on changes to the cluster mode:
>
> - The network bandwidth must be sufficient to support multi-region data synchronization and link redundancy.
> - Control clusters in the central region have a high latency for basic services such as Alibaba Cloud DNS and Tablestore. Therefore, we recommend that you deploy basic services in the same data center to ensure that network latency remains within 5 ms.
> - The network latency between control clusters in the central region and compute clusters in other regions is within 20 ms.
> - Clocks must be synchronized between clusters in different regions and between machines in the same cluster.
> - The network bandwidth must be sufficient to support data replication among clusters.
> - Alibaba Cloud DNS is required.
> - Machines in different clusters can communicate with each other, and the clusters have similar network infrastructure (1-Gigabit or 10-Gigabit).

- The O&M and upgrades for multi-region deployment are different from those for single-cluster deployment. Multi-region deployment requires higher on-site O&M capabilities.
- MaxCompute supports cross-region multi-cluster (sub data centers) distributed computing. It uses the global job scheduling feature of the primary data center to balance the resource usage among clusters. It schedules jobs to the most appropriate cluster based on cluster information, such as the default settings, historical analysis, data distribution, and cluster load. Then, it executes the jobs and generates the query results. MaxCompute supports history- and cost-based optimization policies of SQL queries.

31.6. Scenarios

31.6.1. Scenario 1: Migrate data to the cloud cost-effectively and quickly

Usage scenario: The customer is a data and information service provider focusing on the new energy power sector. The customer's target is to build a cloud platform for Internet big data application services of the new energy industry.

Results: The customer's entire business system has been migrated to the cloud within three months. The data processing time is decreased to less than one third when compared with the customer-built system. Cloud data security is ensured through multiple security mechanisms.

Customer benefits:

- **More focus on its core business**: The entire business system is migrated to the cloud within three months, which enables the customer to use a variety of cloud resources to improve the business.
- **Low investment and O&M costs**: The cloud platform helps to significantly lower the costs of infrastructure construction, O&M personnel, and R&D when compared with a customer-built big data platform.
- **Security and stability**: Alibaba Cloud's comprehensive service and stable performance guarantee data security on the cloud.

31.6.2. Scenario 2: Improve development efficiency and reduce storage and computing costs

Usage scenario: Massive log analysis services for weather query and advertising business are provided to meet the business needs of an emerging mobile Internet company aiming for an excellent weather service provider.

Results: After the Internet company's log analysis business is migrated to MaxCompute, the development efficiency is improved by more than five times, the storage and computing costs are reduced by 70%, and 2 TB of log data is processed and analyzed every day. This more efficiently empowers its personalized marketing strategies.

Customer benefits:

- **Improved work efficiency**: All log data is analyzed by using SQL, and the work efficiency is increased by more than 5 times.
- **Improved storage usage**: The overall storage and computing cost is reduced by 70%, and the performance and stability are also improved.
- **Personalized service**: Machine learning algorithms on MaxCompute are used to perform in-depth data mining and provide personalized services for users.
- **Easy use of big data**: MaxCompute provides plugins for a variety of open-source software to easily migrate data to the cloud.

31.6.3. Scenario 3: Use mass data to achieve precision marketing for millions of users

Usage scenario: To meet the business needs of a community-oriented vertical e-commerce app that focuses on the manicure industry, you can use MaxCompute to build a big data platform for the app. It is mainly used in four aspects: business monitoring, business analysis, precision marketing, and recommendation.

Results: This e-commerce app uses the big data platform built based on MaxCompute to achieve precision marketing for millions of users through the computing capability of MaxCompute, making e-commerce business more agile, intelligent, and insightful. The platform can quickly respond to the data and analysis needs of new business.

Customer benefits:

- **Improved business insights**: Through the computing capabilities of MaxCompute, precision marketing for millions of users is achieved.
- **Data-driven business**: The platform improves the business data analysis capability and effectively monitors business data to better empower businesses.
- **Fast response to business needs**: The MaxCompute ecosystem can quickly respond to changing business data analysis needs.

31.6.4. Scenario 4: Achieve precision marketing with big data

Usage scenario: MaxCompute is used to meet the business needs of an Internet company that focuses on precision marketing and advertising technologies and services. A core big data-based precision marketing platform will be built for the company.

Results: Based on MaxCompute, the company builds a core big data-based precision marketing platform. All log data is stored in MaxCompute, and offline scheduling and analysis are performed through DataWorks.

Customer benefits:

- **Efficient and low-cost analysis of massive data**: Statistical analysis of massive data can reduce expenditures by half to meet the same business needs, effectively saving costs and helping startup enterprises grow rapidly.
- **Real-time data query and analysis**: MaxCompute helps the enterprise establish technical advantages, overcoming the technical bottleneck of massive data processing and analysis, and real-time query and analysis. MaxCompute collects, analyzes, and stores more than 2 billion visitor activities every day. At the same time, it performs millisecond-level queries in hundreds of millions of log tables based on user requirements.
- **Machine learning platform with low entry barrier**: As for a precision marketing and advertising provider, the quality of algorithm models is directly linked to its final revenue. Therefore, selecting the ease-of-use MaxCompute machine learning platform with low entry barrier can get twice the result with half the effort.

31.7. Limits

None.

31.8. Terms

project

The basic unit of operation in MaxCompute. A MaxCompute project is similar to a database or schema in a traditional database. MaxCompute projects set boundaries for isolation and access control between different users. A user can have permissions on multiple projects.

> ? **Note** After being authorized, a user can access objects within a project, such as tables, resources, functions, and instances, from other projects.

table

The data storage unit in MaxCompute. A table is a two-dimensional data structure composed of rows and columns. Each row represents a record, and each column represents a field of the same data type. One record can contain one or more columns. The column names and data types comprise the schema of a table.

> ? **Note** There are two types of MaxCompute tables: external tables and internal tables.

partitioned table

A logical structure used to divide a large table into smaller pieces called partitions. You can specify a partition when creating a table. Specifically, several fields in the table can be specified as partition columns. If you specify the name of a partition you want to access, the system only reads data from the specified partition instead of scanning the entire table, thus reducing costs and improving efficiency.

lifecycle

The validity period of a MaxCompute table or partition. The lifecycle of a MaxCompute table or partition is measured from the last update time. If the table or partition has not undergone any changes within a specified amount of time, MaxCompute will automatically recycle it. This amount of time is specified by the lifecycle.

- Lifecycle unit: days, positive integers only.
- When a lifecycle is specified for a non-partitioned table, the lifecycle is counted from the last time the table data was modified (LastDataModifiedTime). If the table has not been modified before the end of the lifecycle, MaxCompute will automatically recycle the table in a manner similar to the DROP TABLE operation.
- When a lifecycle is specified for a partitioned table, you can decide whether a partition should be recycled based on the LastDataModifiedTime value of the partition. Unlike non-partitioned tables, a partitioned table will not be deleted even if its last partition has been recycled.

  > **Note** Lifecycle scanning is performed at a scheduled time each day, and entire partitions are scanned. If a partition has not undergone any changes within its lifecycle, MaxCompute will automatically recycle it. Assume that the lifecycle of a partitioned table is one day and that the partition data was last modified at 15:00 on the 17th. If the table is scanned before 15:00 on the 18th, the aforementioned partition will not be recycled. During the lifecycle scanning scheduled on the 19th, if the last modification time of the partition exceeds the lifecycle period, the partition will be recycled.

- You can configure a lifecycle for tables, but not for partitions. You can specify a lifecycle when creating a table.
- If no lifecycle is specified, the table or partition cannot be automatically recycled by MaxCompute.

data type

A property of a field that defines the kinds of data the field can store. Columns in MaxCompute tables must be of one of the following data types: TINYINT, SMALLINT, INT, BIGINT, STRING, FLOAT, BOOLEAN, DOUBLE, DATETIME, DECIMAL, VARCHAR, BINARY, TIMESTAMP, ARRAY, MAP, and STRUCT.

resource

A unique concept in MaxCompute. To accomplish tasks by using user-defined functions (UDFs) or MapReduce features in MaxCompute, you must use resources.

> **Note** Resource types in MaxCompute include file, MaxCompute table, JAR (compiled JAR package), and archive. Compressed files are identified by the extensions of resource names. Supported file types include .zip, .tgz, .tar.gz, .tar, and .jar.

function

A piece of code that operates as a single logical unit. MaxCompute provides SQL computing capabilities. In MaxCompute SQL, you can use built-in functions for computing and calculation. When the built-in functions are not sufficient to meet your requirements, you can use the Java programming interface provided by MaxCompute to develop UDFs.

> ? **Note** UDFs can be further divided into scalar-valued functions, user-defined aggregate functions (UDAFs), and user-defined table functions (UDTFs).

task

The basic computing unit of MaxCompute. Computing jobs such as those involving SQL and MapReduce functions are completed by using tasks.

task instance

A snapshot of a task taken at a specified time. In MaxCompute, some tasks are converted into instances when being executed and subsequently exist as MaxCompute instances.

resource quota

A per-process limit on the use of system resources. There are two types of quotas: storage and computing. MaxCompute allows you to set an upper limit of storage for a project. When the storage space occupied approaches the upper limit, MaxCompute triggers an alert. The computing quota limits the use of memory and CPU resources. The memory usage and CPU utilization of running processes in a project cannot exceed the specified upper limit.

ACID semantics

This topic describes the ACID semantics of MaxCompute for concurrent jobs. ACID is an acronym that stands for Atomicity, Consistency, Isolation, Durability.

Terms

- Operation: a single job submitted in MaxCompute.
- Data object: an object that contains data, such as a non-partitioned table or partition.
- INTO jobs: such as INSERT INTO and DYNAMIC INSERT INTO.
- OVERWRITE jobs: such as INSERT OVERWRITE and DYNAMIC INSERT OVERWRITE.
- Data upload with Tunnel: an INTO or OVERWRITE job.

Description of ACID semantics

- Atomicity: An operation is either fully completed or not executed at all.
- Consistency: The integrity of data objects is not compromised during the entire period of an operation.
- Isolation: An operation is completed independent of other concurrent operations.
- Durability: After an operation is complete, data is available in its current state even in the event of a system failure.

Description of ACID semantics in MaxCompute

- Atomicity
  - At any time, MaxCompute ensures that only one job succeeds in the case of a conflict, and all other conflicting jobs fail.

  - The atomicity of the CREATE, OVERWRITE, and DROP operations on a single table or partition can be guaranteed.
  - Atomicity is not supported in cross-table operations such as MULTI-INSERT.
  - In extreme cases, the following operations may not be atomic:
    - The DYNAMIC INSERT OVERWRITE operation is performed on more than 10,000 partitions.
    - INTO operations fail because the data cleansing fails during transaction rollback. This does not cause raw data loss.

- Consistency
  - OVERWRITE jobs ensure consistency.
  - If an OVERWRITE job fails due to a conflict, data from the failed job may remain.

- Isolation
  - Non-INTO operations ensure that read operations are committed.
  - INTO operations can be performed in scenarios where read operations are not committed.

- Durability

  MaxCompute ensures data durability.

32.DataWorks

32.1. What is DataWorks?

DataWorks is an end-to-end big data platform based on compute engines such as MaxCompute and E-MapReduce. It integrates all processes from data collection to data display and from data analysis to application running. DataWorks provides various features to help you complete the entire research and development (R&D) process in a quick and effective manner. The entire R&D process involves data integration, data development, data governance, data service provisioning, data quality control, and data security assurance.

DataWorks is an all-in-one solution for collecting, presenting, and analyzing data, and driving application development. It not only supports offline processing, analysis, and mining of large amounts of data, but also integrates core data-related technologies such as data development, data integration, production and operations and maintenance (O&M), real-time analysis, asset management, data quality control, data security assurance, and data sharing. In addition, it provides the DataService Studio and Machine Learning Platform for Artificial Intelligence (PAI) services.

In 2018, Forrester, a globally recognized market research company, named Alibaba Cloud DataWorks and MaxCompute as a world-leading cloud-based data warehouse solution. This solution is by far the only solution from a Chinese company to receive such an acknowledgment. Building on the success of the previous version, DataWorks V2.0 incorporates several new additions, such as workflows and script templates. DataWorks V2.0 supports dual workspaces for development, isolates the development environment from the production environment, adopts standard development processes, and uses a specific mechanism to reduce errors in code.

32.2. Benefits

This topic describes the benefits of DataWorks.

- Powerful computing capabilities

  DataWorks integrates with compute engines that can process large amounts of data.

  - DataWorks supports join operations for trillions of data records, millions of concurrent jobs, and petabytes (PB) of I/O throughput per day.
  - The offline scheduling system can run millions of concurrent jobs. You can configure rules and alerts to monitor the running statuses of nodes in real time.
  - DataWorks provides efficient and easy-to-use SQL and MapReduce engines, and supports most standard SQL syntax.
  - MaxCompute protects user data from loss, breach, or theft by using multi-layer data storage and access security mechanisms, including triplicate backups, read/write request authentication, application sandboxes, and system sandboxes.

- End-to-end platform

  DataWorks provides the graphical user interface (GUI) and allows multiple users to collaborate on a workspace.

  - DataWorks integrates all processes from data integration, processing, management, and monitoring to output.
  - You can create and edit workflows in a visual manner by using the workflow designer.

  - DataWorks provides a collaborative development environment. You can create and assign roles for varying nodes, such as development, online scheduling, maintenance, and data permission management, without locally processing data and nodes.

- Integration of heterogeneous data stores

  DataWorks supports batch synchronization of data among heterogeneous data stores at custom intervals in minutes, days, hours, weeks, or months. More than 400 pairs of heterogeneous data stores are supported.

- Web-based software

  DataWorks is an out-of-the-box service. You can use it on the Internet or an internal network without the need for installation and deployment.

- Multitenancy

  Data is isolated among different tenants. Each tenant controls permissions, processes data, allocates resources, and manages members in a unified and independent manner.

- Intelligent monitoring and alerting

  By setting monitoring thresholds, you can control the entire process of all nodes as well as monitor the running status of each node.

- Easy-to-use SQL editor

  The SQL editor supports automatic code and metadata completion, code formatting and folding, and pre-compilation. It offers two editor themes. These features ensure a good user experience.

- Comprehensive data quality monitoring

  DataWorks allows you to control the quality of data in heterogeneous data stores, offline data, and real-time data. You can check data quality, configure alert notifications, and manage connections.

- Convenient API development and management

  The DataService Studio service of DataWorks interacts with API Gateway. This makes it easy for you to develop and publish APIs for data sharing.

- Secure data sharing

  DataWorks enables you to de-identify sensitive data before you share it with other tenants, which ensures the security of your big data assets and maximizes their value.

32.3. Architecture

DataWorks is an end-to-end big data platform launched by Alibaba Group, which supports big data processing, management, analysis, mining, sharing, and transmission. It releases you from cluster deployment and management. DataWorks adopts MaxCompute (formerly known as ODPS) as the compute engine to process large volumes of data.

DataWorks is developed based on MaxCompute. DataWorks provides a management console and supports functions such as data processing, management, analysis, and mining.

32.3.1. Service architecture

This topic describes the service architecture of DataWorks.

DataWorks provides the following services:

- Data Integration: supports integration of large amounts of data from heterogeneous data stores to a big data platform.
- DataStudio: supports data warehouse design and whole extract, transform, load (ETL) procedure design.
- Operation Center: supports management and monitoring of online ETL nodes, and supports monitoring of large amounts of nodes and instances based on business baselines.
- DataAnalysis: supports ad hoc queries and data analysis.
- Data Asset Management: supports features such as metadata management and provides data maps, data lineages, and data asset dashboards.
- Data Quality: supports data quality check, monitoring, verification, and grading.
- Data Protection: supports permission management, data management based on security levels, data de-identification, and data auditing.
- DataService Studio: supports data sharing and transmission by using APIs.

32.3.2. System architecture

This topic describes the system architecture of DataWorks.

DataWorks is an end-to-end big data platform that enables you to process data by using services such as Data Integration, DataStudio, Data Asset Management, and DataService Studio. It serves as a basis for upper-layer applications, which satisfies all user requirements.

32.3.3. Security architecture

This topic describes the security architecture of DataWorks.

The security architecture of DataWorks features error proofing, basic security, and optional security tools.

- Error proofing ensures proper running of DataWorks during coding, deployment, and configuration.
- Basic security ensures the security of data for DataWorks by using features such as resource isolation among tenants, user identity verification, authentication, and log auditing.
- Optional security tools in DataWorks allow you to customize security policies for the protection and management of your system and data.

32.3.4. Multitenancy

DataWorks adopts multitenancy.

- Storage and computing resources are scalable. Tenants can apply for resource quotas as needed.
- Tenants are isolated and can manage only their own data, permissions, accounts, and roles. This ensures data security.

32.4. Services

32.4.1. Data Integration

This topic provides an overview of the Data Integration service of DataWorks and describes the features that the service provides.

Overview

Data Integration is a stable and efficient data synchronization service provided by Alibaba Cloud. It allows you to add data stores to and remove them from DataWorks. Data Integration is designed to transmit and synchronize data fast and stably between various heterogeneous data stores in complex network environments.

Data Integration can monitor and read data directly from your database. It provides you with an overview of all data stores. Supported data stores include but are not limited to the following types: relational databases, NoSQL databases, big data stores, and FTP servers. Data Integration also supports data synchronization between heterogeneous data stores in complex network environments. Supported synchronization methods include batch synchronization, full synchronization, and incremental synchronization. Data can be synchronized at an interval of minutes, hours, days, weeks, or months.

Various data stores

Data Integration supports the following data stores:

- Metadata

  Data Integration can collect metadata from more than 20 types of common data stores, such as MySQL, SQL Server, and Oracle databases and MaxCompute projects. It generates a clear view of all data assets from the collected metadata and allows you to take inventory of data assets and synchronize data.

- Relational databases

  Data Integration allows you to perform read/write operations on relational databases such as MySQL, SQL Server, Oracle, Distributed Relational Database Service (DRDS), PostgreSQL, IBM Db2, and ApsaraDB RDS for PPAS.

- NoSQL databases

  Data Integration allows you to perform read/write operations on NoSQL databases such as HBase, MongoDB, and Tablestore.

- MPP databases

  Data Integration allows you to perform read/write operations on massively parallel processing (MPP) databases such as AnalyticDB for MySQL and AnalyticDB for PostgreSQL.

- Big data stores

  Data Integration allows you to perform read/write operations on MaxCompute projects and Hadoop Distributed File System (HDFS). It also allows you to write data to AnalyticDB databases.

- Unstructured data stores

  Data Integration allows you to perform read/write operations on Object Storage Service (OSS) and FTP servers.

  > **Note** Data Integration supports data exchanges between more than 400 pairs of data stores.

Inbound data control

Data Integration supports conversion between various data types. It accurately identifies, filters, collects, and displays dirty data to facilitate inbound data control. It provides you with statistics such as data volume, data throughput, and job duration. It can also detect dirty data for each job.

Fast transmission speed

Data Integration makes full use of the network interface card (NIC) on each server and adopts a distributed architecture. It can transmit gigabytes or terabytes of data within a short period of time.

Accurate throttling

Data Integration implements accurate throttling on channels, record streams, and byte streams. It also supports fault tolerance and allows you to rerun specific or all threads, processes, and jobs.

Synchronization agents

Data Integration provides synchronization agents for you to connect to data store servers and collect data.

Cross-network transmission

Data Integration supports data transmission in complex network environments. For example, it can transmit data across local private networks or virtual private clouds (VPCs).

Note Transmission of large amounts of data over a long distance is accelerated by specific protocols, which ensures high stability and efficiency.

32.4.2. DataStudio

32.4.2.1. Overview

DataStudio is an integrated development environment (IDE) that allows you to develop ETL and data mining algorithms, and build data warehouses in DataWorks.

Before using DataStudio, you need to add data stores by using Data Integration. Then, you can use DataStudio to process the data retrieved from the data stores.

32.4.2.2. Workflows

This topic describes workflows in DataStudio.

Overview

In DataStudio, you can organize various data development nodes in a workflow. DataStudio provides you with a directed acyclic graph (DAG) for nodes in each workflow. It also provides professional tools and supports administrative operations for workflows, which promotes intelligent development and management.

A workflow can contain the following types of nodes: ODPS SQL, ODPS MR, shell, machine learning, data synchronization, PyODPS, SQL component, and zero-load node. You can configure dependencies between nodes within the same workflow or across workflows. You can also schedule a whole workflow or specific nodes.

Manage nodes

You can organize the following types of nodes in a workflow: ODPS SQL, ODPS MR, shell, machine learning, data synchronization, PyODPS, SQL component, and zero-load node. Nodes can be scheduled based on node dependencies or schedules. Each node can depend on other nodes in the current workflow or nodes from other workflows.

Configure a node

After you double-click a node in the left-side navigation pane or in a DAG, the configuration tab of the node appears. Then, you can configure the node. For example, you can write SQL statements for an ODPS SQL node or configure data synchronization rules for a batch sync node. You can also click the tabs in the right-side navigation pane to view the version information or modify settings such as the scheduling properties and lineage of the node.

View the versions of a node

You can view the versions of a node, for example, an ODPS SQL node, an ODPS MR node, or a shell node. If required, you can roll back a node to an earlier version.

Deploy a node

In workspaces in the standard mode, you can deploy nodes that have passed tests to the production environment.

32.4.2.3. Solutions

In a DataWorks workspace, you can group multiple workflows in a solution.

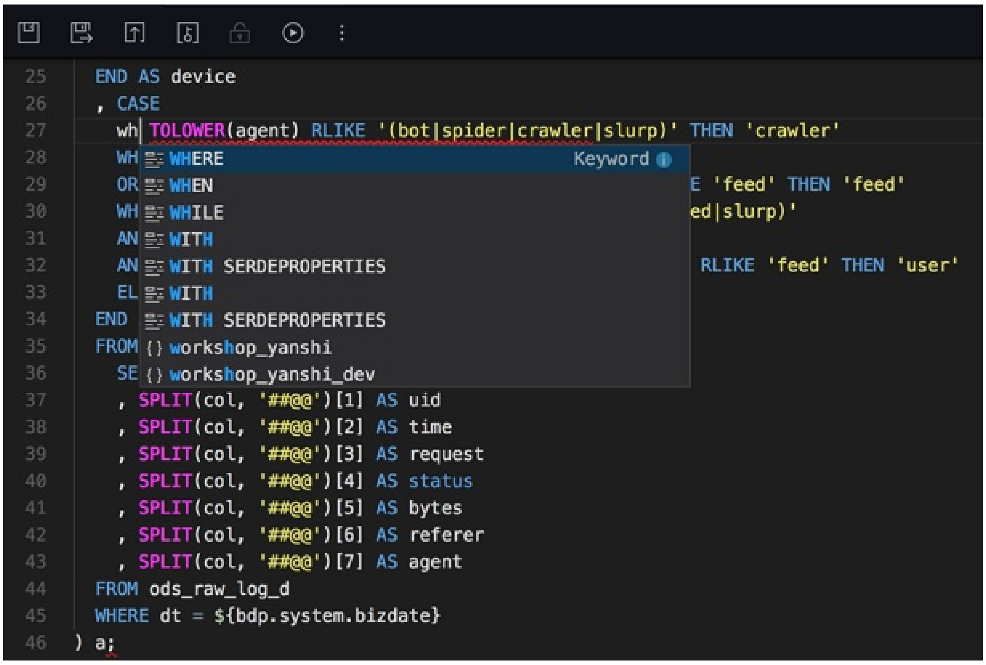
You can add one or more workflows to one solution so that you can manage them as a whole. In addition, a workflow can be added to multiple solutions, allowing you to assess your business based on solutions.

32.4.2.4. Code editor

DataStudio provides a code editor. You can configure ODPS SQL and ODPS MR nodes, upload files as resources, register user-defined functions (UDFs), and write shell scripts in the code editor.

Configure ODPS SQL nodes

The web-based code editor allows you to write SQL statements. It supports a variety of features such as automatic SQL statement completion, code formatting and highlighting, and debugging.



```
  END AS device
  , CASE
    wh| TOLOWER(agent) RLIKE '(bot|spider|crawler|slurp)' THEN 'crawler'
    WH WHERE                                   Keyword
    OR WHEN                                     E 'feed' THEN 'feed'
    WH WHILE                                    ed|slurp)'
    AN WITH
    AN WITH SERDEPROPERTIES                      RLIKE 'feed' THEN 'user'
    EL WITH
  END  WITH SERDEPROPERTIES
  FROM {} workshop_yanshi
    SE {} workshop_yanshi_dev
    , SPLIT(col, '##@@')[1] AS uid
    , SPLIT(col, '##@@')[2] AS time
    , SPLIT(col, '##@@')[3] AS request
    , SPLIT(col, '##@@')[4] AS status
    , SPLIT(col, '##@@')[5] AS bytes
    , SPLIT(col, '##@@')[6] AS referer
    , SPLIT(col, '##@@')[7] AS agent
  FROM ods_raw_log_d
  WHERE dt = ${bdp.system.bizdate}
) a;
```

Configure ODPS MR nodes

When you configure an ODPS MR node in the code editor, you can upload a Java Archive (JAR) file that contains MapReduce code as a JAR resource and then reference the file in the node.

Upload files as resources

DataWorks supports the following types of resources:

- JAR: You can upload JAR files as file resources. Then, UDFs or ODPS MR nodes can reference the resources.
- Python: You can upload Python files as Python resources. Then, UDFs can reference the resources.
- File: You can upload user-defined files such as shell scripts, XML configuration files, or TXT configuration files as file resources.
- Archive: You can upload compressed files as archive resources. The following file formats are supported: .zip, .tgz, .tar.gz, .tar, and .jar. DataWorks automatically identifies the format of an uploaded file based on the file name extension.

Register UDFs

You can register Java or Python UDFs in the code editor. Before you register UDFs, you must upload JAR or Python files as resources. Then, the UDFs can reference the resources.

Write shell scripts

You can use the code editor to write and debug shell scripts online.

32.4.2.5. Code repository and team collaboration

DataWorks allows multiple users to simultaneously work on the same workspace, which improves development efficiency.

DataWorks adopts a lock mechanism that allows you to lock workflows and nodes. This ensures that each workflow or node is edited by only one user at the same time. To edit a node that is locked by another user, you can force unlock the node and then lock the node yourself. This operation is called steal lock. After you steal the lock of a node, the system sends a notification to the user who locked the node previously.

In addition, DataWorks records each committed version of your node and workflow. You can compare two versions of a node and roll back a node to an earlier version.

32.4.3. Administration

32.4.3.1. Overview

Operation Center is a centralized data operations and management platform for data developers and administration experts. You can control and monitor the running of nodes and instances, and set node priorities in Operation Center.

Due to the volume, diversity, and complexity of data used in DataWorks, it is necessary to use a scheduling system that supports high concurrency, multiple cycles, and various data processing procedures.

Operation Center allows you to trace all the nodes that are committed to the scheduling system, view alerts when nodes do not run as scheduled or fail, and view daily reports of node statistics.

32.4.3.2. Overview page

The Overview page displays running statistics of nodes and instances.

You can view the following statistics on the Overview tab: the trend of node instances run today and in past days, rankings of nodes sorted by duration, by number of errors, and by number of overtime node instances within 30 days, and the distribution of nodes by status and by type.

32.4.3.3. Node O&M pages

You can view a node in a directed acyclic graph (DAG), which allows you to perform operations and maintenance (O&M) in a visual manner.

- You can rerun, stop, or suspend nodes, set the status of nodes to successful, and configure alerts to monitor the running status of nodes.
- You can view each node in a node list or the DAG. The DAG clearly shows the relationships between nodes.
- You can view the running status of auto triggered nodes, test nodes, and manually triggered nodes.
- You can view the operational logs, code, and property settings of nodes.

32.4.3.4. Intelligent Monitor service

Intelligent Monitor is a system that monitors and analyzes nodes in DataWorks.

Intelligent Monitor monitors the running status of nodes and sends alerts based on the intervals, notification methods, and recipients specified in alert triggers. When the alerting condition is met, Intelligent Monitor automatically selects the most appropriate alerting time, notification methods, and recipients.

Intelligent Monitor provides you with the following benefits:

- Improves your efficiency on configuring monitoring rules.
- Prevents invalid alerts.
- Automatically covers all important nodes.

Intelligent Monitor provides comprehensive monitoring and alerting logic. You only need to provide the names of important nodes in your business. Then, Intelligent Monitor automatically monitors the entire process of your nodes and creates standard alert triggers for them. Intelligent Monitor also allows you to customize the monitoring feature. You can define alert triggers based on your business requirements.

32.4.4. DataAnalysis

DataAnalysis provides two core features: ad hoc query and private table management. It expedites the analysis process by using the data collection tools of MaxCompute in the near real-time mode.

Benefits

By default, the near real-time mode is used.

You can run the `set ODPS.service.mode=[all|off|limited]` command to change the configuration.

The near real-time mode has the following advantages over the standard mode:

- In the near real-time mode, DataAnalysis preallocates thread pools based on the job size. The near real-time mode eliminates the need for Job Scheduler to plan jobs and reduces the preparation time to run jobs.
- In the near real-time mode, DataAnalysis shuffles data from Mappers to Reducers, without transmitting the data to Apsara Distributed File System.

Keynotes

- The near real-time mode is used if you set the ODPS.service.mode parameter to all. However, if MaxCompute resources are insufficient to run SQL nodes, DataAnalysis switches to the standard mode in which Job Scheduler is responsible for resource allocation. For example, Time Analysis switches to the standard mode if insufficient workers are available for creating instances.
- The scheduling process in the near real-time mode is still complex, but is much more time-saving than the scheduling process in the standard mode.
- If you set the ODPS.service.mode parameter to all, DataAnalysis preferentially uses the near real-time mode. DataAnalysis uses the standard mode if system resources are insufficient, or if known issues or unknown exceptions occur in the near real-time mode.

32.4.5. DataService Studio

DataService Studio aims to build a data service bus to help enterprises manage private and public APIs in a unified manner.

DataService Studio allows you to create APIs based on data tables. You can also register existing APIs to DataService Studio for unified management. DataService Studio and API Gateway are interconnected. This allows you to publish APIs to API Gateway with ease. DataService Studio, together with API Gateway, provides a secure, stable, cost-effective, and easy-to-use API development and management service. DataService Studio adopts a serverless architecture. This allows you to focus on the query logic of the API without worrying about the infrastructure, such as compute resources. DataService Studio supports automatic scaling for compute resources, which significantly reduces your operations and maintenance (O&M) costs.

DataService Studio serves the government as a secure, flexible, and reliable platform for data sharing across departments and networks within the government. It also enables the government to share data with the public.

Create an API operation

DataService Studio allows you to create APIs based on tables in relational databases, NoSQL databases such as Tablestore, and analytical databases such as AnalyticDB. You can create an API in the codeless UI within a few minutes without the need to write code. You can call an API immediately after it is created. DataService Studio also allows you to create an API in the code editor. You can write SQL statements to customize the query logic of the API. In the code editor, you can specify multi-table join queries, complex query criteria, and aggregate functions.

Register an API

You can register existing RESTful APIs to DataService Studio to manage them together with the APIs that are created in DataService Studio based on tables. Four request methods and three data formats are supported. The four request methods are GET, POST, PUT, and DELETE. The three data formats are tables, JSON, and XML.

API Gateway

API Gateway provides API lifecycle management services, including API publishing, management, maintenance, and monetization. API Gateway helps you integrate microservices, separate the frontend from the backend, and integrate systems at low costs and low risks in an easy and quick manner. API Gateway enables you to share features and data with partners and developers. Being integrated with API Gateway, DataService Studio allows you to publish APIs to API Gateway conveniently. Both APIs that you create based on data tables and APIs that you register to DataService Studio can be published to API Gateway for management, for example, for authorization, authentication, throttling, and billing.

32.4.6. Workspace Management

Workspace Management enables administrators to manage their organizations and workspaces.

Workspaces are organizational units for code, member, role, and permission management in DataWorks. Workspaces are isolated from each other. You can view and modify code in a workspace only if you are a member of the workspace and have been granted the required permissions.

> **Note** A user can be a member of multiple workspaces at the same time. The user's permissions in each workspace vary based on the role assigned to the user.

Workspace Management provides the Organizations, Workspaces, Members, and Authorizations pages for managing organizations, workspaces, members, and permissions, respectively.

Organizations

The Organizations page displays the account, AccessKey ID, and AccessKey secret of the owner of the current organization. On this page, you can manage all members in the organization.

Workspaces

On the Workspaces page, you can create, modify, activate, and disable workspaces.

Members

The **Members** page displays information, such as the name, logon username, and roles, about each member of the current workspace. On this page, you can perform the following operations:

- Search for workspace members in the fuzzy match mode and remove the target members from the current workspace.
- Search for users in the fuzzy match mode and add the target users as members of the current workspace.

> ? **Note** When you add a user as a member of a workspace, you must assign at least one role to the user.

Only workspace administrators can add members to and remove members from workspaces.

> ? **Note** After a user is removed from a workspace, all permissions that have been granted to the user within the workspace are revoked.

Authorizations

On the Authorizations page, you can manage roles and specific permissions for all users.

The following table describes the permissions of each role in DataWorks.

Role	Permissions
Administrator	An administrator can manage the basic properties, data stores, compute engine configurations, and members of the workspace. The administrator can also assign the administrator, developer, administration expert, deployment expert, and visitor roles to other members of the workspace.
Developer	A developer can create workflows, script files, resources, and user-defined functions (UDFs). The developer can also create and delete tables, and create deployment tasks. However, the developer cannot perform deploy operations.
Administration expert	An administration expert can perform deploy and administrative operations, but does not have the permissions of a developer. The administration permissions of an administration expert are assigned by an administrator.
Deployment expert	A deployment expert can perform all operations that an administration expert can, except administrative operations.
Visitor	A visitor can only view data, but cannot edit workflows or code in workspaces.

32.4.7. Data Asset Management

Data Asset Management is a tenant-level feature. To use this feature, you must first obtain required permissions on the Project Management page.

This feature allows you to manage your data assets, such as tables and APIs, in your business system and DataWorks. Before you use this feature, you must use Data Integration to synchronize data and then use DataStudio to process the data.

32.4.8. Data Protection

32.4.8.1. Overview

Data Protection is a data security management platform. It can be used to identify data assets, detect sensitive data, classify data, de-identify data, monitor data access behavior, report alerts, and audit risks.

Data Protection provides security management services for MaxCompute.

Data Protection provides the following features:

- Sensitive data detection

  Data Protection automatically detects an enterprise's sensitive data based on self-training models and algorithms, and displays statistics on data types, volume, and visitors. It also recognizes custom data types.

- Custom data classification

  Data Protection allows you to classify data and create custom levels for better data management.

- Flexible data de-identification

  Data Protection provides diverse and configurable methods for dynamic data de-identification.

- Monitoring and auditing of risky user behavior

  Data Protection uses various association analysis algorithms to detect risky user behavior. It also provides alerts and supports visualized auditing for detected risks.

32.4.8.2. Terms

This topic describes the terms that are used in Data Protection, for example, organization, workspace, and data de-identification.

Organization

An organization refers to all system settings and resources owned by a single tenant in DataWorks. The system settings and resources include account configurations, permission settings, and custom applications.

Workspace

Workspaces are organizational units in DataWorks. Similar to databases in a relational database management system (RDBMS), workspaces isolate resources among different users and offer boundaries for access control. Tables, resources, user-defined functions (UDFs), and nodes are isolated among different workspaces.

Regular expression

A regular expression is a sequence of characters that define a search pattern. You can use regular expressions to detect sensitive data.

> **Note** A regular expression consists of metacharacters and literal characters such as letters from a to z.

Data classification

Data is classified based on value, sensitivity, related risks, legal and regulatory requirements, and the potential impact of data breaches.

Sensitive data detection

Data Protection detects sensitive data on the user side based on user-defined rules.

Data de-identification

Data Protection de-identifies sensitive data based on user-defined rules.

MaxCompute

MaxCompute is a data processing platform developed by Alibaba Cloud for large-scale data warehousing. Being able to store and compute mass structured data, MaxCompute provides support for various data warehouse solutions as well as big data analysis and modeling.

32.4.8.3. Management

You can configure sensitive data detection rules on the Data Definition page as a security expert.

After you configure sensitive data detection rules, you can go to the **Data Recognition Rules** page or the **Manipulations and Queries** or **Export** tab of the Data Activities page to perform relevant operations.

32.4.8.4. Data recognition

On the next day after you configure data recognition rules, you can view the recognized data in the Overview, Level, and Fields Recognized sections on the Data Recognition page.

You can filter recognized data by project, rule name, rule type, and risk level.

32.4.8.5. Data Activities

This topic describes data activities.

Data activities include data manipulations and queries, and data export.

- Data manipulations and queries include successful create, insert, and select operations that are performed on data.
- Data export refers to the operation of exporting data from MaxCompute.

Manipulations and queries

On the next day after you configure sensitive data detection rules as a security expert, you can view data manipulations and queries on the Manipulations and Queries tab of the Data Activities page. The Manipulations and Queries tab displays information about data access activities, including the overview, trend, and records. You can filter the information by project, user, rule name, rule type, and risk level based on your query requirements.

Export

On the next day after you configure data recognition rules as a security expert, you can view data export activities on the Export tab of the Data Activities page. The Export tab displays information about data export from MaxCompute, including the overview, top N accounts that have exported the most data, and data export details. You can filter the information by rule name, rule type, and greater than condition based on your query requirements.

32.4.8.6. Data masking

On the Data Masking page, you can create, modify, delete, and test data masking rules.

You can configure data masking rules for each data recognition rule, and configure a whitelist to include recognized sensitive data that does not require data masking.

32.4.8.7. Levels

On the Levels page, you can configure the security levels of rules if the existing configuration cannot meet your needs.

You can create levels, delete levels, and adjust the priority of levels and rules on the **Levels** page.

32.4.8.8. Manual check

On the Manual Check page, you can manually modify recognition results if any sensitive data is recognized incorrectly. You can delete data that is incorrectly recognized, change the type of recognized data, and process multiple data records at a time.

32.4.8.9. Data risks

In Data Protection, data activities are audited manually or based on the risk identification rules and AI-based identification rules configured on the Risk Rules page. The Data Risks page lists data activities that are audited as risky. You can also comment audit results as required.

32.4.8.10. Risk Rules

The Risk Rules page allows you to configure risk identification rules.

You can configure risk identification rules or enable AI-based identification rules to identify risks in users' daily access to your data. The Data Risks page lists the data activities where risks are identified. You can check these data activities and mark them as secure or risky. On the Data Activities page, you can click an activity to view the risk rule that the activity hits.

32.4.8.11. Data Auditing

The Data Auditing page provides an overview and the trend of the total number data risks, number of data risks that have been handled, and number of data risks that have not been handled. This page also provides risk analysis from multiple dimensions.

You can view the data in the **Total Risks**, **Risks Handled**, **Risks Not Handled**, **Trend**, and **Risk Analysis by Dimension** sections.

32.5. Scenarios

32.5.1. Cloud-based data warehouse

Enterprises can use DataWorks in Apsara Stack to build large data warehouses.

DataWorks provides superior data processing capabilities:

- Mass storage: supports petabyte- and exabyte-level data warehouses and scalable storage.
- Data integration: supports data synchronization and integration across heterogeneous data stores to eliminate data siloes.
- Data analytics: supports MaxCompute-based big data analytics, programming frameworks such as SQL and MapReduce, and a visualized workflow designer.
- Data management: supports unified metadata management and permission-based data access control.
- Batch scheduling: supports real-time node monitoring and error alerts, periodic node execution, and processing for millions of nodes per day.

32.5.2. Business intelligence

This topic describes how to create reports by using DataWorks.

You can analyze the following items based on the network logs of your website:

- Page views, unique visitors, and device types such as Android devices, iPads, iPhones, and PCs. You can also create a daily report based on these statistics.
- Locations of visitors.

The following log entry is used as an example:

```
xx.xxx.xx.xxx - - [12/Feb/2014:03:15:52 +0800] "GET /articles/4914.html HTTP/1.1" 200 37666
"http://xxx.cn/articles/6043.html" "Mozilla/5.0 (Windows NT 6.2; WOW64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/xx.x.xxxx.xxx Safari/537.36" -
```

1. Create a destination table named ods_log_tracker in MaxCompute, and then import data to the table.
2. Configure dependencies among the tables to be analyzed.
3. Create a shell node, a sync node, and an ODPS SQL node.
4. Configure the created nodes.

32.5.3. Data-driven management

- Innovative business: Data mining, data modeling, and real-time decision making can be implemented based on big data analytics results provided by DataWorks.

- Small and medium enterprises: With DataWorks, data can be quickly analyzed and put into commercial use, which helps enterprises to generate operational strategies.

32.6. Limits

None.

32.7. Basic concepts

This topic introduces the basic concepts in DataWorks, including workspace, workflow, solution, SQL script template, node, instance, commit operation, script, resource, function, and output name.

Workspace

Workspaces are basic units for managing nodes, members, roles, and permissions in DataWorks. A workspace administrator can add members to the workspace and assign the workspace administrator, developer, administration expert, deployment expert, security expert, or visitor role to each member. In this way, workspace members with different roles can collaborate with each other.

Note We recommend that you create workspaces to isolate resources by department or business unit.

You can bind multiple compute engines such as MaxCompute, E-MapReduce, and Realtime Compute to a single workspace. Then, you can configure and schedule nodes in the workspace.

Workflow

Workflows are abstracted from business to help you manage and develop code based on business demands and improve the efficiency of node management.

Note A workflow can be used in multiple solutions.

Workflows help you manage and develop code based on business demands. A workflow has the following features:

- Allows you to organize nodes by type.
- Supports a hierarchical directory structure. We recommend that you create a maximum of four levels of sub-directories for a workflow.
- Allows you to view and optimize the workflow from the business perspective.
- Allows you to deploy and manage the workflow as a whole.
- Allows you to view the workflow on a dashboard to develop code with improved efficiency.

Solution

A solution contains one or more workflows.

Solutions have the following benefits:

- A solution can contain multiple workflows.
- A workflow can be used in multiple solutions.
- Workspace members can collaboratively develop and manage solutions in a workspace.

SQL script template

SQL script templates are general logic chunks that are abstracted from SQL scripts. They can be reused to improve the efficiency of code development.

Each SQL script template involves one or more source tables. You can filter source table data, join source tables, and aggregate source tables to generate a result table based on your business requirements. An SQL script template includes multiple input and output parameters.

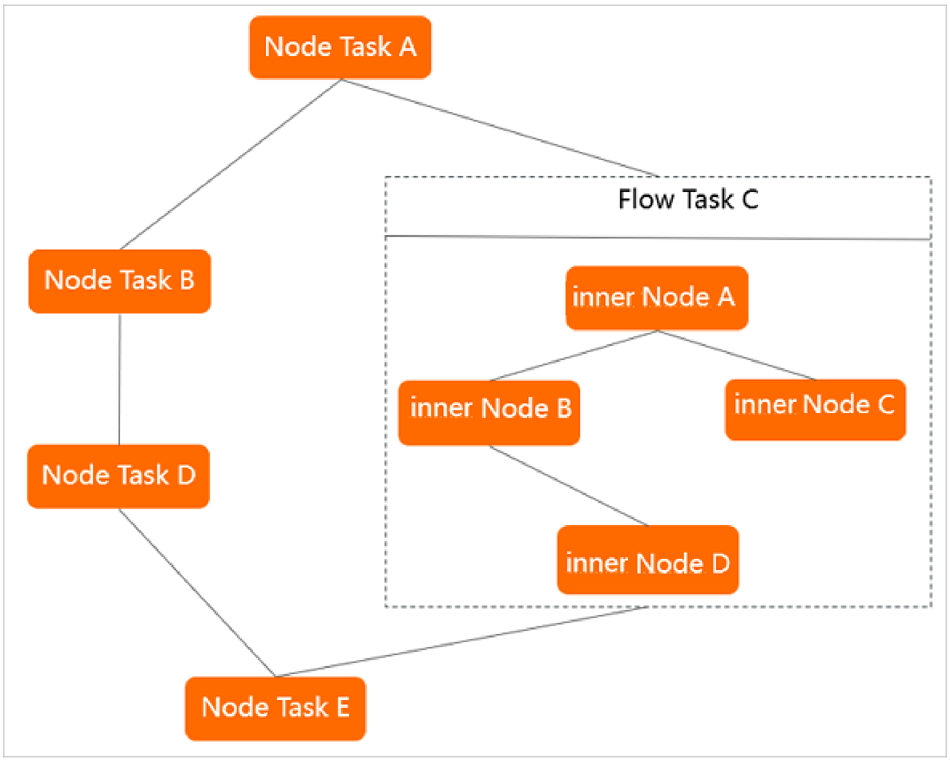
Node

Each type of node is used to perform a specific data operation. For example:

- A sync node is used to synchronize data from ApsaraDB for RDS to MaxCompute.
- An ODPS SQL node is used to convert data by running SQL statements that are supported by MaxCompute.

Each node has zero or more input tables or datasets and generates one or more output tables or datasets.

Nodes are classified into node tasks, flow tasks, and inner nodes.



Type	Description
Node task	A node task is used to perform a data operation. You can configure dependencies between a node task and other node tasks or flow tasks to form a directed acyclic graph (DAG).

Type	Description
Flow task	A flow task contains a group of inner nodes that process a workflow. We recommend that you create less than 10 flow tasks. Inner nodes in a flow task cannot be depended upon by other flow tasks or node tasks. You can configure dependencies between a flow task and other flow tasks or node tasks to form a DAG. **Note** In DataWorks V2.0 and later, you can find the flow tasks that are created in DataWorks V1.0 but cannot create flow tasks. Instead, you can create workflows to perform similar operations.
Inner node	An inner node is a node within a flow task. Its features are basically the same as those of a node task. You can configure dependencies between inner nodes in a flow task by performing drag-and-drop operations. However, you cannot configure a recurrence for inner nodes because they follow the recurrence configuration of the flow task.

Instance

An instance is a snapshot of a node at a specific time point. An instance is generated every time a node is run as scheduled by the scheduling system or manually triggered. An instance contains information such as the time point at which the node is run, the running status of the node, and operational logs.

Assume that Node 1 is configured to run at 02:00 every day. The scheduling system automatically generates an instance of Node 1 at 23:30 every day. At 02:00 the next day, if the scheduling system verifies that all the ancestor instances are run, the system automatically runs the instance of Node 1.

Note You can query the instance information on the **Cycle Instance** page of **Operation Center**.

Commit

You can commit nodes and workflows from the development environment to the scheduling system. The scheduling system runs the code in the committed nodes and workflows as configured.

Note The scheduling system runs nodes and workflows only after you commit them.

Script

A script stores code for data analysis. The code in a script can be used only for data query and analysis. It cannot be committed to the scheduling system for scheduling.

Resource and function

Resources and functions are concepts in MaxCompute. You can manage resources and functions in the DataWorks console. Note that you cannot query resources or functions in DataWorks if they are uploaded by using other services such as MaxCompute.

Output name

Under an Apsara Stack tenant account, each node has an output name that is used to connect to its descendant nodes.

When you configure dependencies for a node, you must use its output name instead of its node name or node ID. After you configure the dependencies, the output name of the node serves as the input name of its descendant nodes.

? **Note** Each output name distinguishes a node from other nodes under the same Apsara Stack tenant account. By default, an output name is in the following format: Workspace name.Randomly generated nine-digit number_out. You can customize the output name for a node. Note that the output name of each node must be unique under an Apsara Stack tenant account.

33.Realtime Compute

33.1. What is Realtime Compute?

Alibaba Cloud Realtime Compute is an advanced stream processing platform that provides real-time computations over data streams.

Background

We are seeing an increasing demand for high timeliness and operability of information, which requires software systems to process more data in less time. In traditional models for big data processing, online transaction processing (OLTP) and offline data analysis are separately performed at different times. These models cannot satisfy the growing demand for real-time big data processing.

Realtime Compute comes from the strict demand for the timeliness of data processing. The business value of data decreases as time passes by. Therefore, data must be computed and processed as soon as possible after it is generated. The traditional models for big data processing follow the scheduled processing mode, that is, accumulating and processing data with hours or even days as the computing cycle. This processing mode cannot satisfy the growing demand for computing data streams. Batch (or offline) processing is inapplicable to delay-sensitive scenarios such as real-time big data analytics, risk control and alerting, real-time prediction, and financial transactions. Realtime Compute enables real-time computing over data streams. With Realtime Compute, you can achieve a short data processing delay, easily implement real-time computational logic, and greatly reduce computing costs. This helps you meet the business needs for real-time processing of big data.

Streaming data

Broadly speaking, big data can be viewed as a series of discrete events. These discrete events form event streams or data streams along a timeline. Unlike traditional offline data, streaming data is continuously generated by thousands of data sources. Streaming data is usually sent in the form of data records. Compared with offline data, streaming data is on a smaller scale. Streaming data is generated from endless event streams, including:

- Log files
- Online shopping data
- In-game player activity information
- Social network information
- Financial transaction information
- Geospatial service information
- Telemetry data from devices or instruments

Features

Realtime Compute has the following features:

- Real-time and unbounded data streams

Realtime Compute can compute directly on a real-time, streaming data source. Realtime Compute subscribes to and consumes streaming data in order of time. Data streams are continuously and permanently collected into the Realtime Compute system as long as data is constantly generated. For example, in scenarios where Realtime Compute processes data streams from website visit logs, the log data streams continuously enter the Realtime Compute system before the website is shut down. Therefore, the data in the Realtime Compute system is in real time and unbounded.

- Continuous and efficient computing

  Realtime Compute is an *event-driven* system where unbounded event or data streams continuously trigger real-time computations. Once new streaming data enters Realtime Compute, Realtime Compute immediately initiates and performs a computing job. In this regard, the real-time computing of Realtime Compute is an ongoing process that never stops.

- Real-time integration of streaming data

  Once a real-time computing job is triggered by streaming data, the computing result is directly written to sinks. For example, you can directly write the computed report data to an RDS system for report display. Realtime Compute can continuously write the computing result of streaming data to sinks, in the same way as data is written to streaming data sources.

33.2. End-to-end real-time computing

Unlike offline or batch computing, end-to-end real-time computing of Alibaba Cloud runs real-time computations over data streams, including real-time data collection, computing, and integration. The real-time computational logic of Realtime Compute ensures a short processing delay.

1. Data collection

   You can use data collection tools to collect and send streaming data in real time to a publish-subscribe system for big data analysis. This publish-subscribe system continuously produces events for Realtime Compute in the downstream to trigger stream processing jobs.

2. Stream processing

   Data streams continuously enter Realtime Compute for real-time computing. At least one data stream must enter the Realtime Compute system to trigger a real-time computing job. Each batch of incoming data records initiates a stream processing procedure in Realtime Compute. The computing results for each batch of data records are then instantly provided.

3. Data integration

   Realtime Compute allows you to write the result data of stream processing to sinks, such as tables of data stores and message delivery systems. You can also integrate Realtime Compute with the alerting system that is connected to your business applications. This enables you to easily receive alerts if the specified business rules for alerting are satisfied. Unlike batch computing products such as MaxCompute and open source Apache Hadoop, Realtime Compute inherently comes with data integration modules that allow you to write result data to sinks.

4. Data consumption

   After the result data of stream processing is written to sinks, the data consumption phase is decoupled from real-time computing. You can use data stores, data transmission systems, or alerting systems to access the result data, send and receive the result data, or send alerts, respectively.

33.3. Differences between real-time computing and batch computing

33.3.1. Overview

Compared with batch computing, real-time computing has made groundbreaking progress in the field of big data computing. This section describes the differences between batch computing and real-time computing from two aspects: users and products.

ⓘ **Note** For more detailed theoretical analysis, see Wikipedia: Stream processing.

33.3.2. Batch computing

Batch computing models have been used for most traditional data computing and analysis services. In batch computing models, extract-transform-load (ETL) or online transaction processing (OLTP) systems are used to load data into data stores. The loaded data is then used for online data services, such as ad-hoc queries and dashboard services, based on SQL statements. You can also use SQL statements to obtain results from the analysis.

Batch computing models are widely accepted along with the evolution of relational databases in diversified industries. However, in the era of big data, with the increasing number of human activities being converted to information and then data, more and more data requires real-time and stream processing. The current processing models are facing great challenges in real-time processing.

A typical batch computing model is described as follows:

1. An ETL or OLTP system is used to build data stores and provides raw data for computing and analysis. The batch computing model where users load the data and the batch computing system optimizes queries on the loaded data using multiple methods, such as creating indexes, based on its storage and computing capabilities. In batch computing models, data must be loaded into the batch computing system. Newly arriving data records are collected into a batch and the entire batch is then processed after all data in the batch is loaded.
2. A user or system initiates a computing job, such as a MaxCompute SQL job or Hive SQL job, and submits requests to the ETL or OLTP system. The batch computing system then schedules computing nodes to perform computations on large amounts of data. This may take several minutes or even hours. The mechanism of batch computing determines that the data to be processed is the accumulated historical data. As a result, the data processing may not be in real time. In batch computing, you can change computational logic using SQL at any time to meet your needs. You can also perform ad-hoc queries instantly after changing the logic.
3. The computing results are returned in the form of data sets when a computing job is completed. If the size of result data is excessively large, the result data is stored in the batch computing system. In this scenario, you can integrate the batch computing system with another system to view the result data. Large amounts of result data lead to a lengthy process of data integration. The process may take several minutes or even hours.

Batch computing jobs are initiated by users or systems and are processed with a long delay. The batch computing procedure is described as follows:

1. You load data into the data processing system.
2. You submit computing jobs. In this phase, you can change computing jobs to meet your business

needs, and publish the changed jobs.

3. The batch computing system returns the computing results.

33.3.3. Real-time computing

Unlike batch computing, real-time computing runs real-time computations over data streams and allows for a low processing delay. The differences between real-time computing and batch computing are described as follows:

1. Data integration. For real-time computing, data integration tools are used to send streaming data in real time to streaming data stores such as DataHub. For batch computing, large amounts of data are accumulated and then processed. In contrast, streaming data is sent in micro batches in real time, which ensures a short delay for data integration.

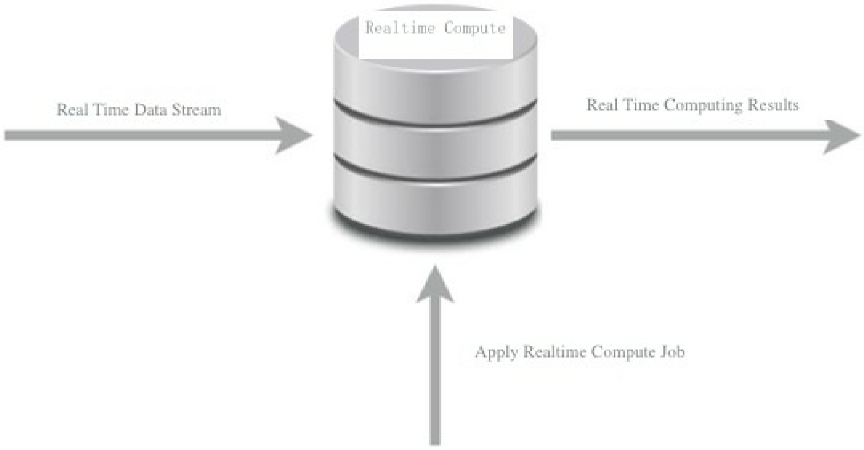
   The streaming data is continuously written to data stores in real time. You do not need to preload data for processing. Realtime Compute does not store real-time data that is continuously processed. The real-time data is discarded instantly after it has been processed.

2. Data computing. For batch computing, data is processed only after large amounts of data have been accumulated. In contrast, a real-time computing job is resident in the system and waits to be triggered by events once it is started. Each incoming micro batch of streaming data records initiates a real-time computing job. The computing results are instantly provided by Realtime Compute. Realtime Compute also divides large batches of data records into smaller batches for incremental computing. This effectively shortens the processing delay.

   For real-time computing, you must predefine the computational logic in Realtime Compute. You cannot change the computational logic when real-time computing jobs are running. If you terminate a running job and publish the job after changing the computational logic, the streaming data that has been processed before the change cannot be processed again.

3. Writing result data to target systems. For batch computing, result data can be written to online systems by batch only after all accumulated data has been processed. In contrast, real-time computing allows for writing result data to online and offline systems instantly after each micro batch of data records has been processed. This allows you to view the computing results in real time.

   Realtime computing



Realtime Compute runs real-time computations over data streams, which are continuously generated from data sources, based on an event-driven mechanism. Realtime Compute allows you to process data streams with a short delay. The real-time computing procedure is described as follows:

1. You publish real-time computing jobs.
2. Streaming data triggers real-time computing jobs.
3. Realtime Compute constantly returns the computing results.

33.3.4. Comparison between real-time computing and batch computing

Comparison between real-time computing and batch computing shows the differences between real-time computing and batch computing.

Comparison between real-time computing and batch computing

Item	Batch computing	Real-time computing
Data integration	You load data into the data processing system.	Data is loaded and processed in real time.
Computational logic	The computational logic can be changed, and data can be reprocessed.	After the computational logic is changed, data cannot be reprocessed. This is because streaming data is processed in real time.
Data scope	You can query and process all or most of the data in the data set.	You can query and process the latest data record or the data within the tumbling window.
Data size	It processes large batches of data.	It processes individual records or micro batches consisting of a few records.
Performance	It achieves a processing delay of several minutes or hours.	It achieves a processing delay of several seconds and even milliseconds.
Analysis	You can perform complex data analysis.	You can perform simple analysis, such as simple response functions, aggregates, and rolling metrics.

Realtime Compute uses a simple computing model. Real-time computing of Realtime Compute makes significant improvements to batch computing in most scenarios of big data computing. In particular, in scenarios where event streams need to be processed with an extremely low processing delay, real-time computing is a valuable service for big data computing.

33.4. Benefits

Realtime Compute provides competitive advantages in stream processing, which allows you to easily handle the demand for real-time big data analysis. Realtime Compute offers the following benefits:

Powerful real-time computing functions

Realtime Compute simplifies the development process by integrating a wide range of functions. These functions are described as follows:

- A powerful engine is used. This engine offers the following advantages:
  - Provides the standard Flink SQL that enables automatic data recovery from failures. This ensures accurate data processing when failures occur.
  - Supports multiple types of built-in functions, such as text functions, date and time functions, and statistics functions.
  - Enables an accurate control over computing resources. This ensures complete isolation of each tenant's jobs.

- The key performance metrics of Realtime Compute are three to four times higher than those of Apache Flink. For example, in Realtime Compute, the data processing delay is reduced to seconds or even to sub-second level. The throughput of a job reaches millions of data records per second. A cluster can contain thousands of nodes.
- Realtime Compute integrates cloud-based data stores such as MaxCompute, DataHub, Log Service, ApsaraDB for RDS, Table Store, and AnalyticDB for MySQL. With Realtime Compute, you can read data from and write data to these systems with the least efforts in data integration.

Managed real-time computing services

Unlike open source or user-developed stream processing services, Realtime Compute is a fully managed stream processing engine. You can query streaming data without deploying or managing any infrastructure. With Realtime Compute, you can use streaming data processing services with a few clicks. Realtime Compute integrates services such as development, administration, monitoring, and alerting. This allows you to use cost-effective streaming data services for trial and migrate your data for deployment.

Realtime Compute also enables complete isolation between tenants. This isolation and protection extends from the top application layer to the underlying infrastructure layer. This helps to ensure the security and privacy of your data.

Excellent user experience during development

Realtime Compute provides a standard SQL engine: Flink SQL. It also provides many built-in functions, such as the text functions, date and time functions, and statistics functions. The application of these functions greatly simplifies and accelerates the Flink-based development. With Flink SQL, even users with limited development knowledge, such as business intelligence (BI) analysts and marketers, can easily perform real-time analysis and processing of big data.

Realtime Compute provides an end-to-end solution for stream processing, including development, administration, monitoring, and alerting. On the Realtime Compute development platform, only three steps are required to publish a job.

Low costs in labors and compute clusters

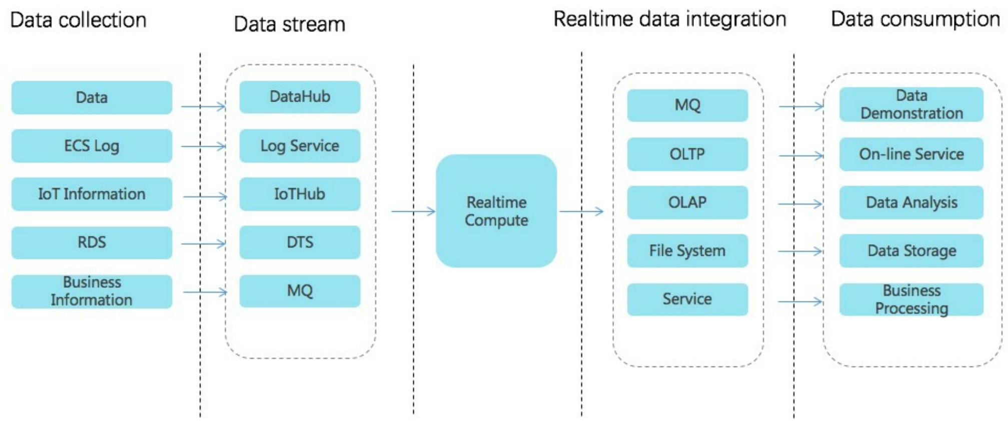
We have made many improvements to the SQL execution engine, allowing you to create jobs more cost-effectively than to create Flink jobs. Realtime Compute is more cost-effective than open source stream frameworks in both development and production costs. To create an Apache Storm job with complex computational logic, you have to incur high costs and devote a lot of effort, such as writing enormous lines of Java code, debugging, testing, performance tuning, publishing, and long-term administration of open source software applications like Apache Storm and Zookeeper. Realtime Compute allows you to offload the heavy lifting of handling these issues, which helps you focus on your business strategies and rapidly achieve market goals.

33.5. Product architecture

33.5.1. Business process

We recommend that you have a general knowledge about the stream processing architecture of StreamCompute before using this service. This helps you create effective plans for the design of stream processing systems. Architecture shows the stream processing architecture of StreamCompute.

Architecture



- Data collection

  You can use data collection tools to collect and send streaming data in real time to a publish-subscribe system for big data analysis. This publish-subscribe system continuously produces events for Realtime Compute in the downstream to trigger real-time computing jobs. The big data ecosystem of Alibaba Cloud offers a wide range of publish-subscribe systems to process streaming data in diversified scenarios. Realtime Compute integrates many of these systems, as shown in the preceding figure. This allows you to easily integrate multiple streaming data stores. To enable compatibility between the computing model of Realtime Compute and that of certain data stores, another data store may be required for data processing. Realtime Compute is seamlessly connected to the following data stores:

  - DataHub

    DataHub allows you to upload data into its system using a wide range of tools and interfaces. For example, you can easily upload logs, binary log files, and IoT streaming data into the DataHub system. DataHub also integrates open source business software applications. For more information about the data collection tools of DataHub, see DataHub documentation.

  - Log Service

    Log Service is a one-stop logging service that has been developed by Alibaba Group based on years of experience in addressing challenges involving large amounts of big data experienced by Alibaba Group. Log Service allows you to quickly collect, transfer, query, consume, and analyze log data.

  - IoT Hub

    IoT Hub is a service that enables developers of IoT applications to implement two-way communications between devices (such as sensors, final control elements, embedded devices, and smart home appliances) and the cloud by creating secure data channels.

    You can use the IoT Hub rule engine to easily send IoT data to DataHub, and use Realtime Compute and MaxCompute to process and perform computations on data.

  - Data Transmission Service (DTS)

    DTS supports data transmission between structured data stores represented by databases. DTS is a data exchange service that streamlines data migration, data synchronization, and data subscription. You can use the data transmission function of DTS to easily parse binary log files such as RDS logs and send data to DataHub. Realtime Compute and MaxCompute allow you to run computations over the data.

  - MQ

    Message Queue (MQ) is a key service that provides messaging capabilities, such as message publishing and subscription, message tracing, scheduled, and delayed messages, resource statistics, monitoring, and alerting. MQ offers a complete set of enterprise-level messaging functions powered by high-availability (HA) distributed systems and clusters.

- Realtime computing

  Data streams continuously enter Realtime Compute for real-time computing. At least one data stream must enter the Realtime Compute system to trigger a real-time computing job. In complex business scenarios, Realtime Compute allows you to perform association queries for static data from data stores and streaming data. For example, you can perform JOIN operations on DataHub and RDS tables based on the primary key of streaming data. You can then perform association queries on DataHub streaming data and RDS static data. Realtime Compute also enables you to associate multiple data streams. With Flink SQL, you can easily handle large amounts of data and complex business scenarios, such as those experienced by Alibaba Group.

- Realtime data integration

  To minimize the data processing delay and simplify data transmission links, Realtime Compute directly writes the result data of real-time computing to data sinks. Realtime Compute allows for a larger Alibaba Cloud ecosystem by integrating the following systems: online transaction processing (OLTP) systems such as ApsaraDB for RDS, NoSQL database services such as Table Store, online analytical processing (OLAP) systems such as AnalyticDB, message queue systems such as DataHub and RocketMQ, and mass storage systems such as Object Storage Service (OSS) and MaxCompute.

- Data consumption

  After the result data of real-time computing is written to the sinks, you can consume the data using custom applications.

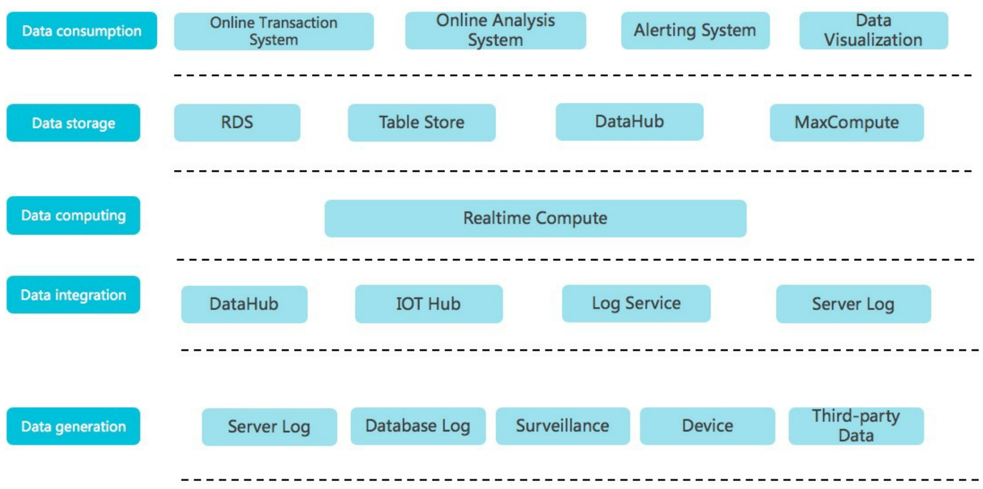
  - You can use data stores to access the result data.
  - You can use data transfer systems to send and receive the result data.
  - You can use alerting systems to send alerts.

33.5.2. Business architecture

Realtime Compute is a lightweight SQL-enabled streaming engine for real-time processing and analysis of data streams.

Business architecture

Business architecture



- Data generation

  In this phase, streaming data is generated from sources such as server logs, database logs, sensors, and third-party systems. The generated streaming data moves on to the next phase for data integration to drive real-time computing.

- Data integration

  In this phase, the streaming data is integrated. You can subscribe to and publish the integrated streaming data. The following Alibaba Cloud products can be used in this phase: DataHub for big data computing, IoT Hub for connecting IoT devices, and Log Service for integrating ECS logs.

- Data computing

  In this phase, the streaming data, which has been subscribed to in the data integration phase, acts as inputs to drive real-time computing in Realtime Compute.

- Data store

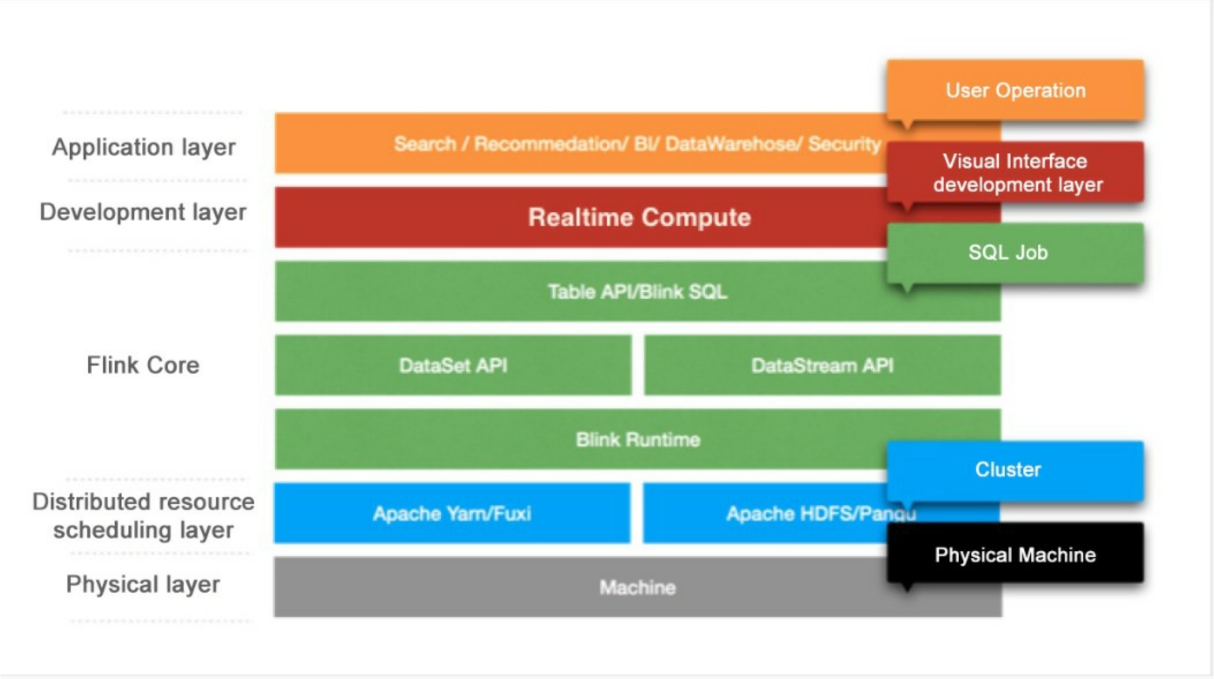
  Realtime Compute does not provide built-in data stores. Instead, it writes computing results to external data stores, such as relational databases, NoSQL databases, and online analytical processing (OLAP) systems.

- Data consumption

  Realtime Compute supports multiple data store types, which allows you to consume data in various ways. For example, data stores for message queues can be used to report alerts, and relational databases can be used to provide online support.

33.5.3. Technical architecture

Realtime Compute is a real-time data analysis platform for incremental computing. This platform provides statements that are similar to SQL statements and uses the MapReduceMerge (MRM) computing model for incremental computing. Realtime Compute offers a failover mechanism to ensure data accuracy when errors occur.



The Realtime Compute architecture consists of the following five layers.

- Application layer

  This layer allows you to create SQL files and publish jobs for real-time data processing based on a development platform. With a well-designed monitoring and alerting system, you would be notified of a processing delay for each job in a timely manner. You can also use systems like Flink UI to view the running information of published jobs and analyze performance bottlenecks. This allows you to quickly and effectively improve job performance.

- Development layer

  This layer parses Flink SQL and generates logical and physical execution plans. The execution plans are then conceptualized as executable directed acyclic graphs (DAGs). Based on these DAGs, directed graphs that consist of various models are obtained. Directed graphs are used to implement specific business logic. A model usually contains the following three modules:

  - Map: Operations such as data filtering, distribution (GROUP), and join (MAPJOIN) are performed.
  - Reduce: Realtime Compute processes streaming data by batch, and each batch contains multiple data records.
  - Merge: You can update the state by merging the computing results of the batch, which are produced from the Reduce module, with the previous state. Checkpoints are created after N (configurable) batches have been processed. In this way, the state is stored persistently in a data store, such as Tair and Apache HBase.

- Flink Core

  This layer provides a wide range of computing models, Table API, and Flink SQL. You can use DataStream API and DataSet API at the lower sublayer. At the bottom sublayer is Flink Runtime, which schedules resources to ensure that jobs can run properly.

- Distributed resource scheduling layer

  Realtime Compute clusters run based on the Gallardo scheduling system. This system ensures that Realtime Compute runs effectively and fault tolerance is provided for recovery.

- Physical layer

  This layer provides powerful hardware devices for clusters.

33.6. Features

Realtime Compute has the following features:

- **Data collection and storage**

  The premise of running a big data analysis system is that data has been collected into the system. To make full use of your existing streaming data store, Realtime Compute supports integration with multiple upstream streaming data stores, such as DataHub, Log Service, IoT Hub, Table Store, and MQ. You can use streaming data in existing data stores without operations of data collection and data integration.

  You can register data stores on the Realtime Compute development platform. This enables you to leverage the advantages of the one-stop Realtime Compute development platform. Realtime Compute provides the UI for managing different data stores, such as ApsaraDB for RDS, AnalyticDB, and Table Store. Realtime Compute allows you to manage cloud-based data stores in one stop.

- **Data development**

  - Realtime Compute provides a fully managed online development platform that integrates a wide range of SQL coding assistance features, such as Flink SQL syntax checking, intelligent code completion, and syntax highlighting.

    - **Syntax checking**

      On the Development page of Realtime Compute, the revised script is automatically saved. When the script is saved, an SQL syntax check is automatically performed. If a syntax error is detected, the Development page shows the row and column where the error is located, and the cause of the error.

    - **Intelligent code completion**

      When you enter SQL statements on the Development page of Realtime Compute, auto-completion prompts about keywords, built-in functions, and SQL statements are automatically displayed.

    - **Syntax highlighting**

      Flink SQL keywords are highlighted in different colors to differentiate data structures.

  - The Realtime Compute development platform allows you to manage different versions of SQL code.

    Realtime Compute provides key features that help you complete development tasks, such as coding assistance and code version management. On the data development platform, you can manage SQL code versions. Each time you commit code, the system generates a code version, which can be used for version tracking, modification, and rollback.

  - The Realtime Compute development platform allows you to register data stores on its **Development** page for effective data store management, such as data preview and auto DDL generation.
    - Data preview

      The Development page of Realtime Compute allows you to preview the data of multiple data store types. Data preview helps you efficiently analyze upstream and downstream data, identify key business logic, and complete development tasks.

    - Auto DDL generation

      In most cases, the DDL statements for data stores are manually translated into the DDL statements for real-time computing. Therefore, the DDL generation process includes a large number of repetitive tasks. Realtime Compute provides an auto DDL generation feature. This feature simplifies the way that you edit SQL statements for stream processing jobs, reduces the possibility of encountering errors when you manually enter SQL statements, and also improves efficiency.

  - Realtime Compute allows you to implement real-time data cleansing, statistics, and analysis using standard SQL. Realtime Compute also supports common aggregation functions, and association queries for streaming data and static data.
  - The Realtime Compute development platform provides a simulated running environment where you can customize uploaded data, simulate operations, and check output results.

- **Data operation**

  Realtime Compute allows you to manage stream processing jobs on the following tabs under the Administration page: Overview, Curve Charts, FailOver, CheckPoints, JobManager, TaskExecutor, Data Lineage, and Properties and Parameters.

- **Performance tuning**
  - Improve performance by automatic configuration

    The automatic configuration function of Realtime Compute helps you address performance issues, such as a low throughput of jobs and data piling up in the upstream.

  - Improve performance by manual configuration

    You can manually configure resources to improve job performance using one of the following methods:

    - Optimize resource configuration. You can modify the resources to improve performance by reconfiguring parameters, such as parallelism, core, and heap_memory.
    - Improve performance based on job parameter settings. You can specify the job parameters such as miniBatch to improve performance.
    - Improve upstream and downstream data stores based on parameter settings. You can specify related parameters to optimize the upstream and downstream data stores for a job.

- **Monitoring and alerting**

  This allows you to collect the performance metrics of cloud resources or other custom performance metrics, view service availability, and specify alerts based on the performance metrics. In this way, you can easily view the cloud resource usage and running information of jobs. You can also receive and respond to alerts in a timely manner to ensure that applications can run properly. With Realtime Compute, you can specify alerts for the following performance metrics:

  - Processing delay

  - Input RPS
  - Output RPS
  - Failover rate

33.7. Product positioning

Realtime Compute offers Flink SQL to support standard SQL semantics and help you easily implement the computational logic of stream processing. Realtime Compute also provides full-featured UDFs for some authorized users, helping you customize business-specific data processing logic in scenarios where SQL code functions cannot meet your business needs. In the field of streaming data analysis, you can directly use Flink SQL and UDFs to enable most of the streaming data analysis and processing logic. Realtime Compute focuses on the analysis, statistics, and processing of streaming data. It is less applicable to non-SQL businesses, such as complex iterative data processing and complex rule engine alerts.

Realtime Compute is applicable to the following scenarios:

- Collects the data about page views (PVs) and unique visitors (UVs) in real time.
- Collects the data about the average traffic flow at a traffic checkpoint every 5 minutes.
- Collects and displays the pressure data of hydroelectric dams.
- Reports alerts for financial thefts in online payment services based on fixed rules.

Realtime Compute is inapplicable to the following scenarios for now:

- Replacing Oracle stored procedures with Realtime Compute: Realtime Compute cannot implement all the functions of Oracle stored procedures, because they are designed to handle issues in different fields.
- Seamlessly migrating Spark jobs to Realtime Compute: Currently, you cannot seamlessly migrate Spark jobs to Realtime Compute. However, you can change the stream processing of Apache Spark and migrate this part to Realtime Compute. This eliminates various Apache Spark administration tasks and Spark-based development costs.
- Complex rule engines for alerting: Realtime Compute cannot handle scenarios where multiple complex alerting rules are specified for each data record, and the rules continue to change when the system is running. Specific rule engines need to be used to resolve these issues.

Realtime Compute provides a full set of development tools for streaming data analysis, statistics, and processing based on UDFs and Flink SQL. It allows you to devote the least efforts in developing the underlying code and simply write SQL statements to analyze streaming data. This makes Realtime Compute a good choice for users such as data warehouse developers and data analysts.

33.8. Scenarios

33.8.1. Overview

Realtime Compute uses Flink SQL to provide solutions for streaming data analysis.

- Real-time extract-transform-load (ETL)

  Realtime Compute allows you to cleanse, aggregate, and sort streaming data in real time by leveraging the advantages of multiple data channels and flexible data processing capabilities of SQL. Realtime Compute serves as an effective supplement and optimization of offline data warehouses and provides a computing channel for real-time data transmission.

- Real-time reports

Realtime Compute allows you to collect and process streaming data, monitor performance metrics of the business, and view corresponding reports in real time. This enables real-time data administration.

- Monitoring and alerting

  Realtime Compute allows you to monitor systems and analyze user behavior in real time, which helps to identify faults and risks in real time.

- Online systems

  Realtime Compute allows you to run real-time computations over data streams and view performance metrics in real time. You can shift strategies for online systems in a timely fashion. Realtime Compute can be widely used in various content delivery and intelligent mobile push scenarios.

33.8.2. Management of e-commerce activities

Realtime Compute has evolved into a reliable stream processing platform from Alibaba Group's big data architecture in the e-commerce industry. Realtime Compute is suitable for analyzing various streaming data and providing report support in the e-commerce industry. The e-commerce industry needs to process streaming data in real time in the following scenarios:

- Real-time analysis of user behavior, for example, display of transaction data and user data on big screens. In traditional batch processing models, large amounts of data are processed inefficiently with a long delay. The size of the result data may be excessively large, which poses considerable challenges for online systems that are used for displaying the result data. This may compromise the stability of the online systems.
- Real-time monitoring of users, services, and systems. For example, marketers and engineers can have knowledge of the transactions on the platform over a specified period by viewing the corresponding curve chart. If abnormal fluctuations occur, such as a sharp decrease in transactions, alerts must be instantly triggered and sent to users. This helps users effectively respond to abnormal fluctuations and reduces negative impacts on the business.
- Real-time monitoring of major promotional events. For example, marketers need to monitor the metrics of promotional events in real time, such as the Double 11 Shopping Festival created by Alibaba Group and 618 mid-year shopping festival started by JD.com, Inc. This helps marketers effectively decide whether to change strategies.

Integrating with Alibaba Cloud computing and storage systems, Realtime Compute allows you to meet your custom needs for streaming data analysis. Realtime Compute not only satisfies diverse business needs but also simplifies the business development process by using Flink SQL.

33.8.3. Multidimensional analysis of data from IoT sensors

Background

With the economic tidal wave of globalization sweeping over the world, industrial manufacturers are facing increasingly fierce competition. To increase competitiveness, manufacturers in the automotive, aviation, high-tech, food and beverage, textile, and pharmaceutical industries must innovate and replace the existing infrastructure. These industries have to address many challenges during the innovation process. For example, the existing traditional devices and systems have been used for decades, which results in high maintenance costs. However, replacing these systems and devices may slow down the production process and compromise the product quality.

These industries face two additional challenges, which are high security risks and the urgent need for complex process automation. The manufacturing industry has prepared to replace the existing traditional devices and systems. In this industry, high reliability and availability systems are needed to ensure the safety and stability of real-time operations. A manufacturing process involves a wide range of components, such as robotic arms, assembly lines, and packaging machines. This requires remote applications that can seamlessly integrate each stage of the manufacturing process, including the deployment, update, and end-of-life management of devices. The remote applications also need to handle failover issues.

Another requirement for these next-generation systems and applications is that they be able to capture and analyze the large amounts of data generated by devices, and respond appropriately in a timely manner. To increase competitiveness and accelerate development, manufacturers need to optimize and upgrade their existing systems and devices. The application of Realtime Compute and Alibaba Cloud IoT solutions allows you to analyze device running information, detect faults, and predict yield rates in real time. This topic describes a use case as an example. In this use case, a manufacturer uses Realtime Compute to analyze the large amounts of data collected from sensors in real time. Realtime Compute is also used to cleanse and aggregate data in real time, write data to an online analytical processing (OLAP) system in real time, and monitor the key metrics of devices in real time.

Scenario description

In this use case, the manufacturer has more than 1,000 devices from multiple factories in many cities. Each device is equipped with 10 types of sensors. These sensors send the collected data every 5 seconds to Log Service. The data collected from each sensor follows the format described in the following table.

s_id	s_value	s_ts
The ID of the sensor.	The current value from the sensor.	The time when the data was sent.

The sensors are distributed across devices from multiple factories. The manufacturer creates an RDS dimension table to display the distribution of sensors across devices and factories.

s_id	s_type	device_id	factory_id
The ID of the sensor.	The type of the sensor.	The ID of the device.	The ID of the factory.

The information included in this dimension table is stored in the RDS system. The manufacturer needs to organize the data from sensors based on this dimension table, and sort the data by device. To meet this need, Realtime Compute provides a summary table where the data sent from sensors is logically aggregated by device every minute.

ts	device_id	factory_id	device_temp	device_pres
The time when the data was sent.	The ID of the device.	The ID of the factory.	The temperature of the device.	The pressure of the device.

Assume that there are only two types of sensors in this use case: temperature and pressure. The computational logic is described as follows:

1. Realtime Compute identifies the devices whose temperatures are higher than 80°C and triggers

alerts at the downstream nodes. In this use case, Realtime Compute sends the data of the identified devices to MQ. MQ then triggers alerts that the manufacturer has specified in the downstream alerting system.

2. Realtime Compute writes the data to an OLAP system. In this use case, the manufacturer uses HybridDB for MySQL. To integrate with HybridDB for MySQL, the manufacturer has developed a set of business intelligence (BI) applications for multidimensional data display.

FAQ

- How can I aggregate data into a summary table?

  In most cases, each sensor only collects the IoT data of one dimension. This poses challenges for subsequent data processing and analysis. To create a summary table, Realtime Compute aggregates data based on windows and organizes data by dimension.

- Why is MQ used to trigger alerts?

  Realtime Compute allows you to write data to any type of storage system. We recommend that you use message storage systems like MQ for sending alerts and notifications. This is because the application of these systems helps to prevent the errors encountered by user-defined alerting systems. These errors may cause failures to report certain alerts and notifications.

Code description

Send the data uploaded from sensors to Log Service. The data format of a row is shown as follows:

```
{
  "sid": "t_xxsfdsad",
  "s_value": "85.5",
  "s_ts": "1515228763"
}
```

Define a Log Service source table s_sensor_data.

```
CREATE TABLE s_sensor_data (
  s_id    VARCHAR,
  s_value VARCHAR,
  s_ts    VARCHAR,
  ts      AS CAST(FROM_UNIXTIME(CAST(s_ts AS BIGINT)) AS TIMESTAMP),
   WATERMARK FOR ts AS withOffset(ts, 10000)
) WITH (
  TYPE='sls',
  endPoint ='http://cn-hangzhou-corp.sls.aliyuncs.com',
   accessId ='******',
   accessKey ='******',
   project ='ali-cloud-streamtest',
   logStore ='stream-test',
);
```

Create an RDS dimension table d_sensor_device_data. This dimension table stores the mappings between sensors and devices.

```
CREATE TABLE d_sensor_device_data (
  s_id   VARCHAR,
  s_type   VARCHAR,
  device_id BIGINT,
  factory_id BIGINT,
  PRIMARY KEY(s_id)
) WITH (
  TYPE='RDS',
   url='',
   tableName='test4',
   userName='test',
   password='*******'
);
```

Create an MQ result table r_monitor_data. This table specifies the logic for triggering alerts.

```
CREATE TABLE r_monitor_data (
  ts   VARCHAR,
  device_id   BIGINT,
  factory_id   BIGINT,
  device_TEMP   DOUBLE,
  device_PRES DOUBLE
) WITH (
  TYPE='MQ'
);
```

Create a HybridDB for MySQL result table r_device_data.

```
CREATE TABLE r_device_data (
  ts   VARCHAR,
  device_id BIGINT,
  factory_id BIGINT,
  device_temp   DOUBLE,
  device_pres DOUBLE,
  PRIMARY KEY(ts, device_id)
) WITH (
  TYPE='HybridDB'
);
```

Aggregate the data collected from sensors by minute and create a summary table based on the aggregated data. To clearly view the code structure and facilitate subsequent administration, we create views in this use case.

```
// Create a view to obtain the device and factory mapping each sensor.
CREATE VIEW v_sensor_device_data
AS
SELECT
  s.ts,
  s.s_id,
  s.s_value,
  s.s_type,
  s.device_id,
  s.factory_id
FROM
  s_sensor_data s
JOIN
  d_sensor_device_data d
ON
  s.s_id = d.s_id;
// Aggregate the data collected from sensors.
CREATE VIEW v_device_data
AS
SELECT
  // Specify the start time of a tumbling window as the time for the record.
  CAST(TUMBLE_START(v.ts, INTERVAL '1' MINUTE) AS VARCHAR) as ts,
  v.device_id,
  v.factory_id,
  CAST(SUM(IF(v.s_type = 'TEMP', v.s_value, 0)) AS DOUBLE)/CAST(SUM(IF(v.s_type = 'TEMP', 1, 0)) AS DOUBLE) device_temp, // Compute the average temperature by minute.
  CAST(SUM(IF(v.s_type = 'PRES', v.s_value, 0)) AS DOUBLE)/CAST(SUM(IF(v.s_type = 'PRES', 1, 0)) AS DOUBLE) device_pres // Compute the average pressure by minute.
FROM
  v_sensor_device_data v
GROUP BY
  TUMBLE(v.ts, INTERVAL '1' MINUTE), v.device_id, v.factory_id;
```

In the preceding core computational logic, the average temperature and pressure by minute are computed as the output. Tumbling windows are used in this use case. A new window is started every minute, and a new set of data is generated every minute. The generated data is then filtered and written to the MQ result table and HybridDB result table.

```
// Identify the sensors whose temperatures are higher than 80°C and write the data to the MQ result table to trigger alerts.
INSERT INTO r_monitor_data
SELECT
  ts,
  device_id,
  factory_id,
  device_temp,
  device_pres
FROM
  v_device_data
WHERE
  device_temp > 80.0;
// Write the result data to the HybridDB for MySQL result table for analysis.
INSERT INTO r_device_data
SELECT
  ts,
  device_id,
  factory_id,
  device_temp,
  device_pres
FROM
  v_device_data;
```

33.8.4. Big screen service for the Tmall Double 11 Shopping Festival

The annual Tmall Double 11 Shopping Festival has become the largest sales event for online shopping in the world. A large number of netizens demonstrate a strong desire to purchase products during the sales event each year. One of the key highlights of this event has been the increase in the overall turnover that is displayed on the Tmall big screen in real time. The real-time display of turnover on the big screen is a result of our senior engineers' hard work over several months. The big screen service excels in key performance metrics. For example, the end-to-end delay has been reduced within 5 seconds, from placing orders on the Tmall platform, to data collection, processing, verification, and to displaying the sales data on the big screen. As for the processing capability, hundreds of thousands of orders can be processed during the peak hours around 00:00 on November 11. Additionally, to ensure fault tolerance, multiple channels have been used to back up data.

Realtime Compute provides key support for the big screen service. The stream processing of the big screen service was previously based on the open source Apache Storm. The Storm-based development process took around one month. The application of Flink SQL shortened the development process of the big screen service to one week. The underlying layer of Realtime Compute removes the Apache Storm modules that are designed for execution optimization and troubleshooting. This enables higher efficiency and faster processing for Realtime Compute jobs.

- Online shopping rush

  During the Double 11 Shopping Festival, an enormous number of netizens join the online shopping rush on the Tmall platform. During the peak hours when "seckilling" activities occur, such as 00:00 on November 11, hundreds of thousands of sales orders need to be processed in real time. The word "seckilling" vividly describes fighting among buyers, which means that a buyer wins or loses all in a matter of seconds.

- Real-time data collection

  The data collection system collects and sends the logs of database changes to the DataHub system. With the application of Data Transmission Service (DTS), the data from online transaction processing databases can be written to DataHub tables within seconds at the peak hours around 00:00 on November 11.

- Real-time data computing

  Realtime Compute subscribes to the DataHub streaming data, continuously analyzes the streaming data, and calculates the total turnover up to the current time. In Realtime Compute, a cluster can contain up to thousands of nodes. The throughput of a job reaches millions of data records per second, fully meeting the system requirements of processing hundreds of thousands of transactions per second in Tmall. Realtime Compute subscribes to data and writes the result data to an online RDS system in real time.

- Frontend data visualization

  We also provide advanced data visualization components for the Tmall Double 11 Shopping Festival. These components allow you to view the total turnover on a dashboard, and the distribution of global transaction activities across the world in real time. To achieve astounding visual effects for the big screen, the frontend server performs periodic polling operations on the RDS system, and advanced web frontend applications are used.

33.8.5. Mobile data analysis

Realtime Compute allows you to analyze the data of mobile apps in real time. With Realtime Compute, you can analyze performance metrics of mobile apps, such as crash detection and distribution, and distribution of app versions. Mobile Analytics is a product provided by Alibaba Group to analyze the data of mobile apps. This product allows you to analyze user behavior and logs from multiple dimensions. It also helps mobile developers implement fine-grained operations based on big data analysis, improve product quality and customer experience, and enhance customer stickiness. The underlying big data computing of Mobile Analytics is implemented based on big data products of Alibaba Cloud, such as Realtime Compute and MaxCompute. Mobile Analytics uses Realtime Compute as the underlying engine for streaming data analysis. This allows Mobile Analytics to offer a wide range of real-time data analysis and reporting services for mobile apps.

- Data collection

  To collect data, developers can include the software development kit (SDK) provided by Mobile Analytics into an app installation package. This SDK offers data collection components based on mobile operating systems. These components collect and send the data about mobile phones and user behavior to the backend of Mobile Analytics for analysis.

- Data reporting

  The backend of Mobile Analytics offers a data reporting system, which allows you to collect the data reported by mobile phones using the specified SDK. The data reporting system preliminarily removes dirty data, and sends the processed data to DataHub.

  > **Note** In the future, DataHub provides an SDK for mobile phones to directly report data. The removal of dirty data is performed in Realtime Compute instead of Mobile Analytics, reducing the host costs of Mobile Analytics.

- Stream processing

Realtime Compute continuously subscribes to the DataHub streaming data. It also continuously reads and runs computations over the data about the performance metrics of mobile apps. Realtime Compute then writes the result data of stream processing during each period to an online RDS or Table Store system.

- Data display

  Mobile Analytics provides a complete set of performance metrics that allow you to quickly view the running information and usage of mobile apps. For example, you can quickly know user locations, visited pages, browsing duration, end devices and network environments, and slow responses or crashes. With Mobile Analytics, you can also analyze crashes by device, and view the details of crashes. The data display is based on the result data that is obtained in the stream processing phase.

33.9. Restrictions

None

33.10. Terms

Project

In Realtime Compute, a project is a basic unit for managing clusters, jobs, resources, and users. Project administrators can create projects, or add users to other existing projects. Realtime Compute projects can be collaboratively managed by Apsara Stack tenant accounts and RAM users.

Job

Similar to a MaxCompute or Hadoop job, a Realtime Compute job implements the computational logic of stream processing. A job is a basic unit for stream processing.

CU

In Realtime Compute, a compute unit (CU) defines the minimum capabilities of stream processing for a job with the specified CPU cores, memory, and input/output capacities. A Realtime Compute job can use one or more CUs.

Currently, `a CU is assigned with one CPU core and 4 GB memory` .

Flink SQL

Unlike most open source stream processing systems that provide basic APIs, Realtime Compute offers Flink SQL that includes standard SQL semantics and advanced semantics for stream processing. Flink SQL is designed to satisfy diversified business needs, and it allows developers to perform stream processing by using standard SQL. With Realtime Compute, even users with limited technological skills, such as data analysts, can quickly and easily process and analyze streaming data.

UDF

Realtime Compute allows you to use user-defined functions (UDFs) that are similar to Apache Hive UDFs. We recommend that you use UDFs to implement your custom computational logic. UDFs are a supplement to Flink SQL that can be used for standard stream processing. Currently, Realtime Compute only supports Java UDFs.

Resource

Currently, Realtime Compute only supports Java UDFs. A JAR file uploaded by a user is defined as a resource.

Data collection

During a typical data collection process, data is collected from sources and ingested into a big data processing engine. The data collection process of Realtime Compute focuses on the phases where data is collected from the source and then transferred into a data bus.

Data store

Realtime Compute is a lightweight computing engine without built-in data stores. Data sources and sinks of Realtime Compute are based on external data stores. For example, you can use RDS to store result tables for Realtime Compute.

Data development

During the data development process, you edit Flink SQL statements to create a Realtime Compute job. Realtime Compute offers an online integrated development environment (IDE) where you can edit SQL statements and debug data before publishing a Realtime Compute job.

Data administration

The data administration page of the Realtime Compute platform allows for online management of jobs. Realtime Compute helps you easily and effectively manage stream processing jobs.

34.Machine Learning Platform for AI

34.1. What is machine learning?

Machine learning is a process of using statistical algorithms to learn large amounts of historical data and generate an empirical model to provide business strategies.

Apsara Stack Machine Learning Platform for AI is a set of data mining, modeling, and prediction tools. It is developed based on MaxCompute (also known as ODPS). Machine Learning Platform for AI supports the following functions:

- Provides an all-in-one algorithm service covering algorithm development, sharing, model training, deployment, and monitoring.
- Allows you to complete the entire procedure of an experiment either through the GUI or by running PAI commands. This function is typically intended for data mining personnel, analysts, algorithm developers, and data explorers.
- In Apsara Stack, Machine Learning Platform for AI runs on MaxCompute. Machine Learning Platform for AI allows you to call algorithms to decouple the applications and compute engines after you have deployed algorithm packages in MaxCompute clusters.
- Provides various algorithms and reliable technical support, providing more options to resolve service issues. In the Data Technology (DT) era, you can use Machine Learning Platform for AI to implement data-driven services.

Machine Learning Platform for AI can be applied in the following scenarios:

- Marketing: commodity recommendations, user profiling, and precise advertising.
- Finance: loan delivery prediction, financial risk control, stock trend prediction, and gold price prediction.
- Social network sites (SNSs): microblog leader analysis and social relationship chain analysis.
- Text: news classification, keyword extraction, text summarization, and text analysis.
- Unstructured data processing: image classification and image text extraction through OCR.
- Other prediction cases: rainfall forecast and football match result prediction.

Machine learning can be divided into three types:

- Supervised learning: Each sample has an expected value. You can create a model and map input feature vectors to target values. Typical examples of this learning mode include regression and classification.
- Unsupervised learning: No samples have a target value. This learning mode is used to discover potential regular patterns from data. Typical examples of this learning mode include simple clustering.
- Reinforcement learning: This learning mode is complex. A system constantly interacts with the external environment to obtain external feedback and determines its own behavior to achieve a long-term optimization of targets. Typical examples of this learning mode include AlphaGo and driverless vehicles.

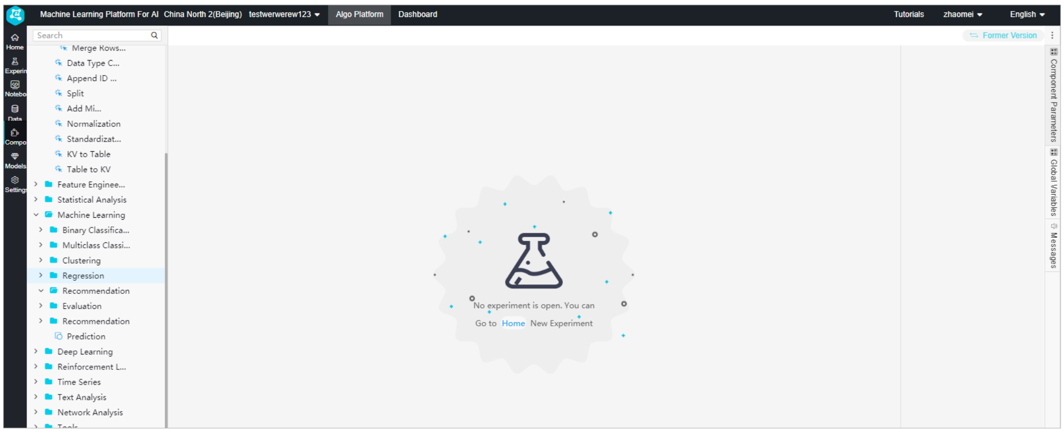
34.2. Benefits

Alibaba Cloud Machine Learning Platform for AI has the following benefits:

All-in-one visual user interface

- Machine Learning Platform for AI provides a Web interface for you to mine data by dragging and dropping components without programming, like piling up blocks, as shown in User interface.

User interface



No experiment is open. You can

Go to Home New Experiment

- Machine Learning Platform for AI provides the data model visualization function. It allows you to use charts to view data analysis results and algorithm evaluation.
- Machine Learning Platform for AI provides an all-in-one solution for data processing, model training, prediction, evaluation, model deployment, service building, and task scheduling.
- In addition to the Web interface, Machine Learning Platform for AI also provides command line tools to easily integrate algorithms into your projects.

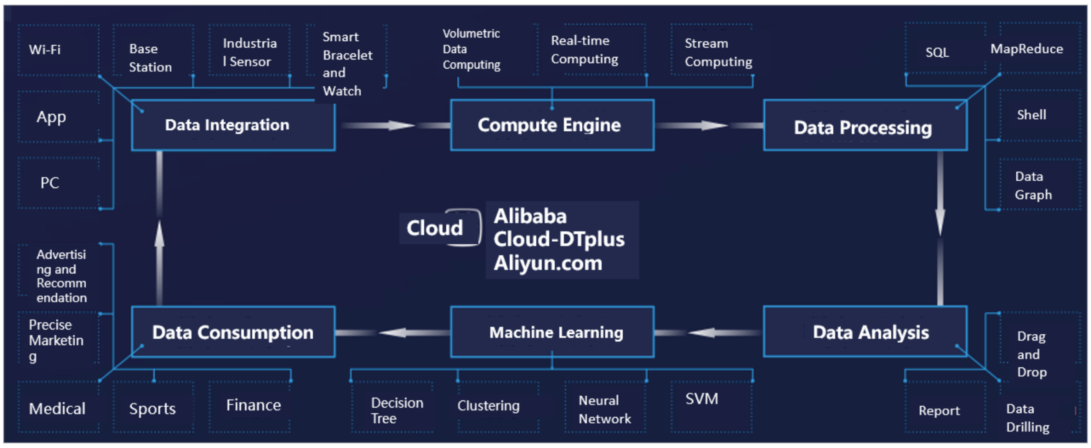
Multiple high-performance machine learning algorithms

- Machine Learning Platform for AI provides nearly 100 machine learning algorithms that can be applied to multiple business scenarios, such as data preprocessing, clustering, regression, text analysis, and feature processing algorithms.
- Compared with traditional software, Machine Learning Platform for AI adopts the latest and optimal algorithms in the machine learning industry to improve the computing capability and accuracy.
- Machine Learning Platform for AI supports deep learning and GPU job scheduling. Machine Learning Platform for AI integrates and completely optimizes the TensorFlow framework. You can get started with TensorFlow for model training.
- Machine Learning Platform for AI provides open-source algorithms that are developed based on years of experience of Alibaba Cloud in big data mining and utilization. This greatly shortens the data modeling, model deployment, and model utilization period.

Full compatibility with Alibaba Cloud services

- Apsara Stack has established a big data ecosystem, such as Machine Learning Platform for AI. All services are ready for use after you activate them.
- Machine Learning Platform for AI runs on MaxCompute and is integrated with DTplus DataWorks to help data mining, parent and child node data collection, experiment scheduling, and data utilization, as shown in Alibaba Cloud DTplus services.
- Based on the MPI, PS, graph algorithms, and MapReduce computing frameworks and distributed algorithms, Machine Learning Platform for AI easily handles a large amount of data.

Alibaba Cloud DTplus services



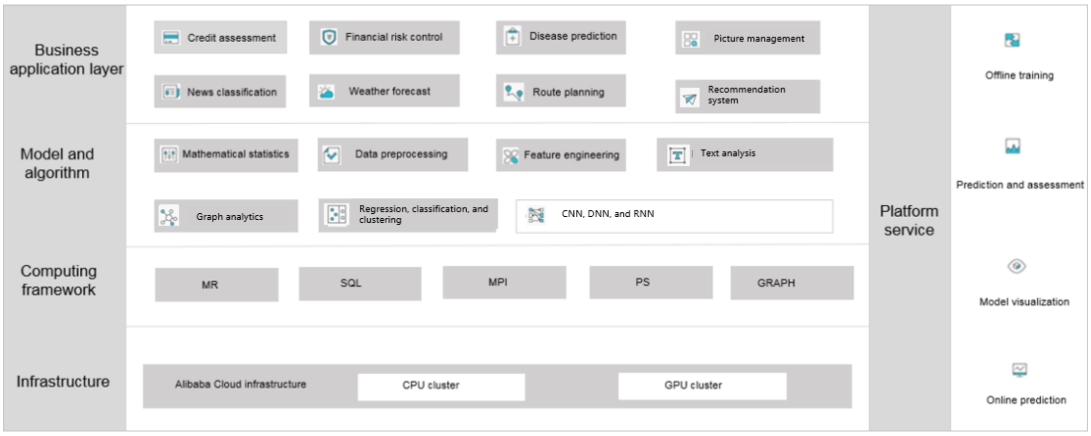
High-quality technical support

Machine Learning Platform for AI is supported by Alibaba algorithm scientists and Apsara Stack technical support. If you have any issues, submit a ticket through the ticket system or contact Apsara Stack technical personnel.

34.3. Architecture

Basic architecture of Machine Learning Platform for AI shows the basic architecture of Machine Learning Platform for AI.

Basic architecture of Machine Learning Platform for AI



The architecture is composed of the following layers:

- Infrastructure layer: provides the cluster resources for computing. You can choose CPU or GPU computing clusters based on the algorithm type.
  - The CPU cluster runs machine learning algorithms and provide computing resources such as CPU and memory resources. Computing resources are centrally managed by an algorithm framework. After jobs are submitted, the algorithm framework schedules compute nodes in the CPU cluster and dispatches jobs to the compute nodes.
  - The GPU cluster runs deep learning framework jobs and provides computing resources such as GPU and graphics memory. Computing resources are centrally managed by an algorithm framework. After jobs are submitted, the algorithm framework schedules compute nodes in the GPU cluster. For a task that requires multiple workers and GPUs, a virtual network is automatically created to dispatch the jobs to the compute nodes in the virtual network.

- Computing framework layer: manages the CPU resources, GPU resources, and a basic runtime environment for algorithms, such as the MapReduce runtime library, MPI runtime library, PS runtime library, and TensorFlow framework.

  The deep learning framework TensorFlow supports the open-source version 1.4. The computing framework layer also optimizes the performance and I/O interfaces. You can use TensorFlow to read files from and write models to OSS buckets. When TensorFlow is running, you can start TensorBoard to display the status of parameter convergence during convolution.

- Model and algorithm layer: provides basic components such as data preprocessing, feature engineering, and machine learning algorithm components. All algorithm components come from the Alibaba Group algorithm system and have been tested on petabytes of service data.
- Business application layer: The Alibaba search system, recommendation system, Ant Financial, and other projects use Machine Learning Platform for AI for data mining. Machine Learning Platform for AI can be applied to industries such as finance, medical care, education, transportation, and security.

Based on this architecture, Machine Learning Platform for AI provides a Web-based visual algorithm experiment console. The Web GUI allows you to perform offline training, prediction, and evaluation, visualize models, deploy online prediction services, or release experiments to the scheduling system of DataWorks.

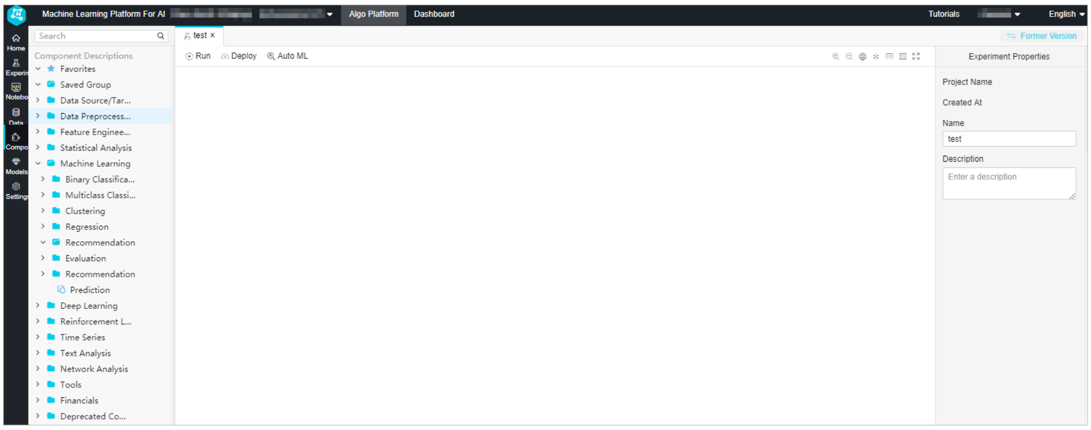
34.4. Features

34.4.1. Visualized modeling

Machine Learning Platform for AI provides the easy-to-use visual modeling feature, which allows you to view the logic of the procedure.

The visual modeling pages include the algorithm platform page and online model service page. Algorithm platform page shows the function section on the algorithm platform page. Online model service page shows the online model service page.

Algorithm platform page



Project Name

Created At

Name: test

Description: Enter a description

Sections on the algorithm platform page

The algorithm platform page includes the following sections:

- Features section: displays machine learning features and information, such as experiments, components, data sources, and models, in a tree structure.
- Canvas section: You can drag and drop components to the canvas to build a directional workflow in order to complete data mining tasks, such as the metadata collection, data processing, modeling,

and model deployment.

- Properties section: You can configure component parameters in this section.

Features section

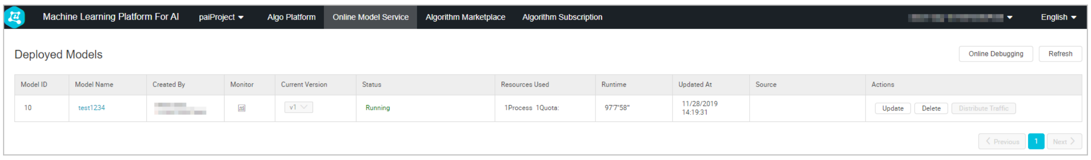
The features section on the algorithm platform page provides the following menus:

- Search: You can search data, tables, and experiments.
- Experiments: After you double-click the name of an experiment, the canvas displays the directional flowchart of the experiment. You can continue modifying the experiment.
- Data sources: allows you to view and manage all data tables.
- Components: provides multiple key features of machine learning, such as machine learning components.
- Models: allows you to manage all models.
- Developer tool: allows you to view the experiment runtime log and troubleshoot experiment issues based on returned error messages and alerts.

Online model service

In the upper section of the page, select Online Model Service to go to the online model service page. This page displays user-created online services. You can monitor or select an action for these algorithm services.

Online model service page



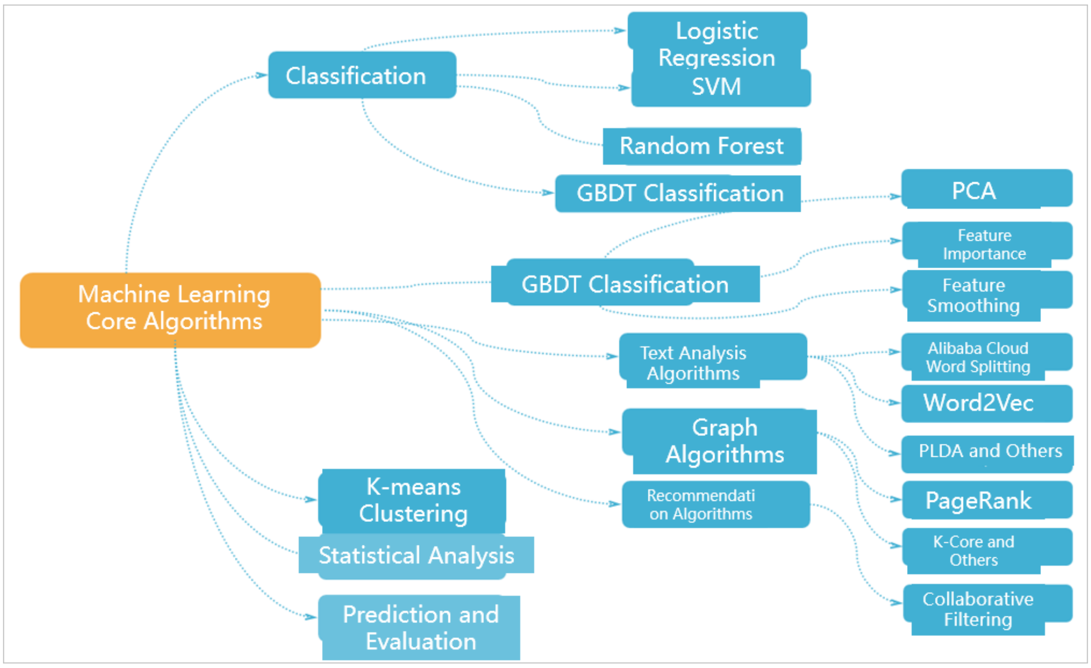
Deployed Models

Model ID	Model Name	Created By	Monitor	Current Version	Status	Resources Used	Runtime	Updated At	Source	Actions
10	test1234	[redacted]		v1	Running	1Process 1Quota:	97'7"58"	11/28/2019 14:19:31		Update Delete Distribute Traffic

34.4.2. All-in-one experience

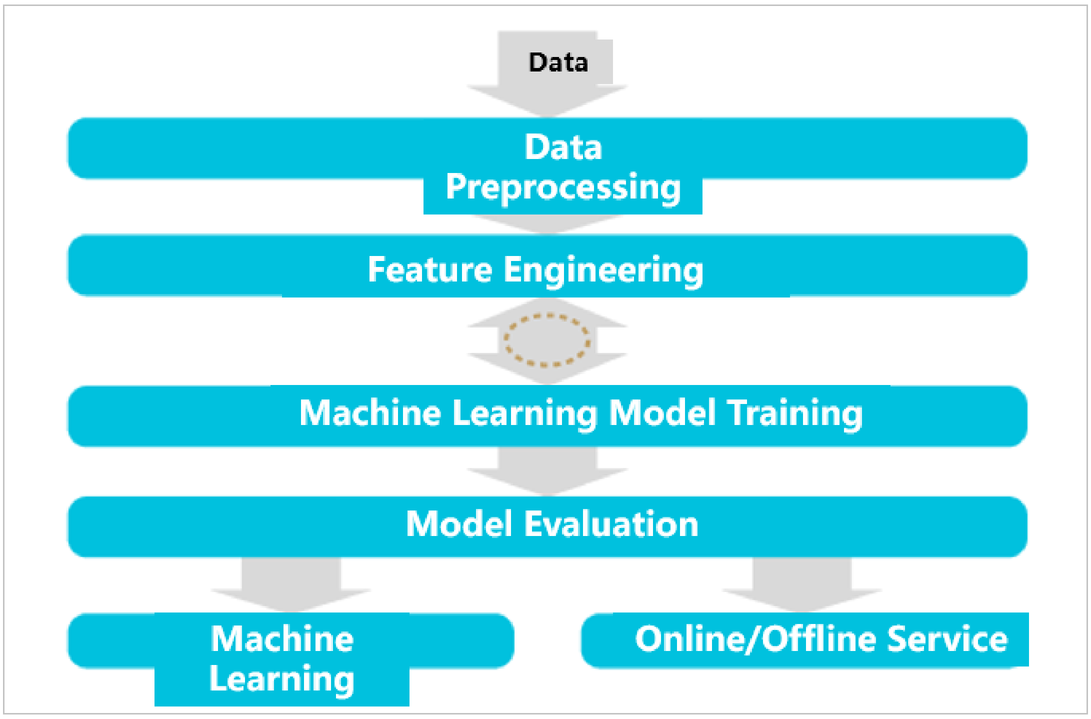
Machine Learning Platform for AI has a complete algorithm library, as shown in Algorithm library of Machine Learning Platform for AI.

Algorithm library of Machine Learning Platform for AI



Typically, you need to perform many operations to complete data mining or model training such as data extract, transform, and load (ETL), data preprocessing, feature engineering, model training, evaluation, and deployment, as shown in Algorithm development process of atypical model. Machine Learning Platform for AI provides an all-in-one development environment with a complete set of components and tools for you to complete the entire data mining or model training task, such as metadata processing and model deployment. With these basic components, you can import data to the platform, create an experiment, and create solutions to resolve issues in different scenarios, and save costs on environment switching.

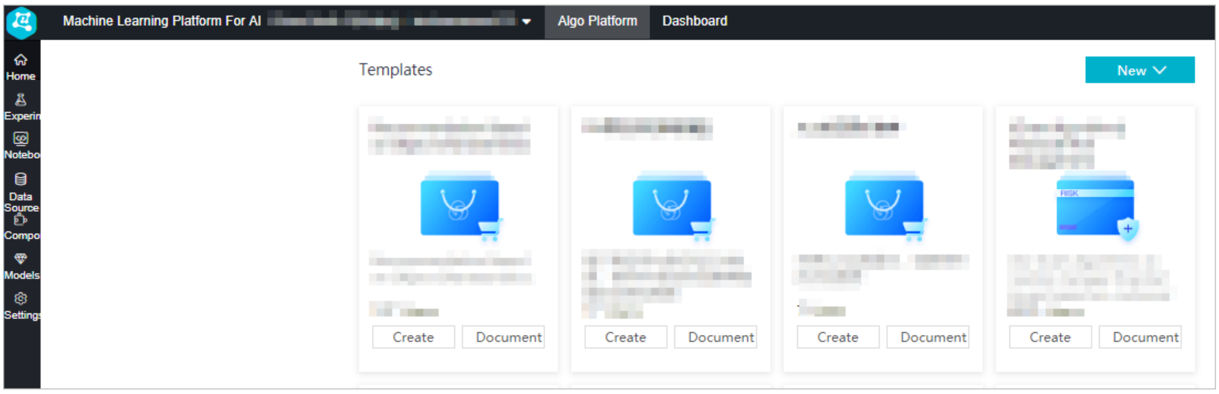
Procedure of developing algorithms for typical machine learning models



34.4.3. Multiple templates on the homepage

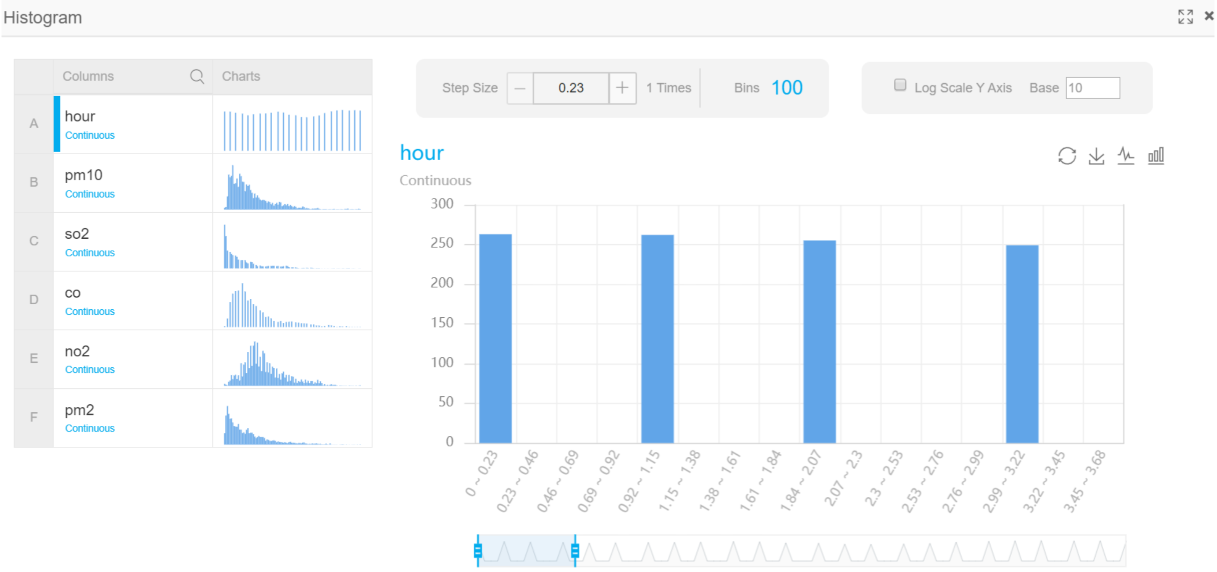
To help data analysts get started with the service, Machine Learning Platform for AI provides a set of experiment templates for scenarios such as product recommendations, text analysis, financial risk management, and weather prediction. These templates contain configurations and data that you can run the experiment with.

You can create experiments from the templates provided on the homepage. You can learn information about how an experiment is configured, how machine learning works, and how data is processed.



34.4.4. Data visualization

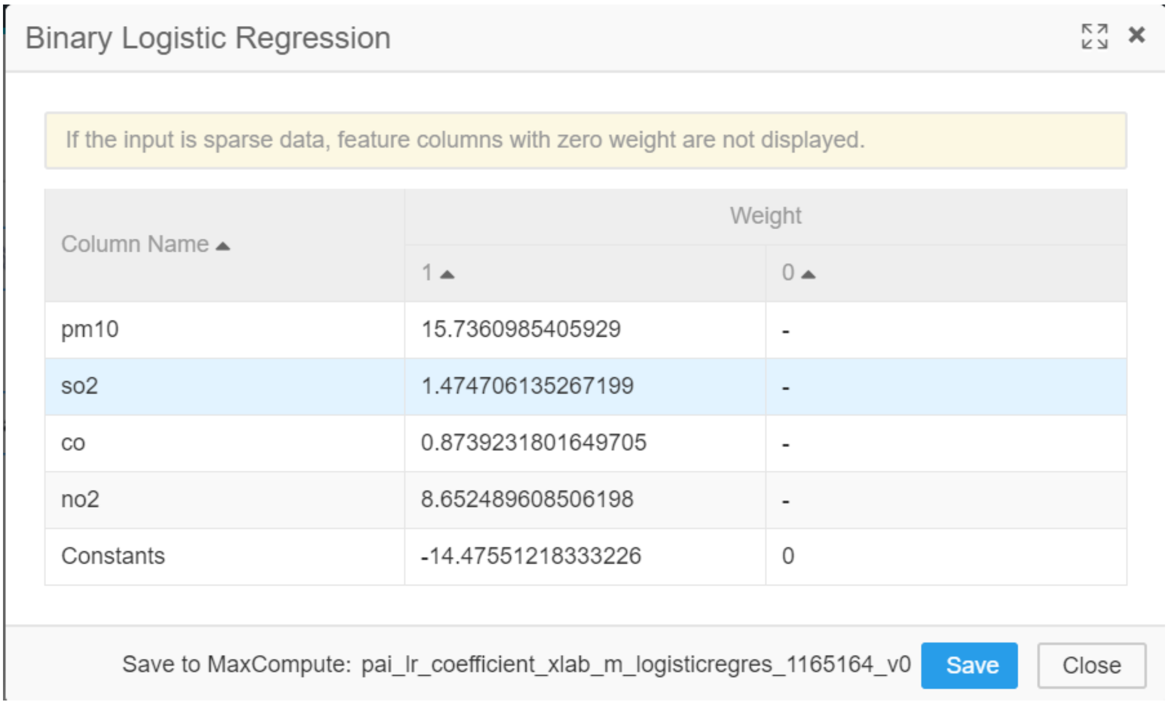
You can right-click an output component to view the visual output model. For example, you can view the model evaluation report and data analysis results. Visualized output can be displayed in multiple forms such as line charts, dot charts, and bar charts.



34.4.5. Model management

Visualized model management:

1. In the left-side navigation pane, click **Models**.
2. Expand the **Models** folder to locate the model built from a specified experiment.
3. Right-click the model and then select **Show Model**.



Binary Logistic Regression

If the input is sparse data, feature columns with zero weight are not displayed.

Column Name ▲	Weight	
	1 ▲	0 ▲
pm10	15.7360985405929	-
so2	1.474706135267199	-
co	0.8739231801649705	-
no2	8.652489608506198	-
Constants	-14.47551218333226	0

Save to MaxCompute: pai_lr_coefficient_xlab_m_logisticregres_1165164_v0

Note You can also right-click the model to perform other actions. For example, you can export the PMML file or deploy the model.

34.4.6. Algorithm components

The current version of Machine Learning Platform for AI on Apsara Stack provides up to 89 algorithm components. These components are classified into 11 categories. The following table lists the algorithm components.

Level 1	Level 2	Level 3	Description
1. Data source/target	1.1 Read MaxCompute table	N/A	Reads a MaxCompute table.
	1.2 Write MaxCompute table	N/A	Writes a MaxCompute table.
	2.1 Sampling and filtering	2.1.1 Weighted sampling	Collects samples according to the specified sampling fraction.
		2.1.2 Random sampling	Collects samples randomly.
		2.1.3 Filtering and mapping	Uses the SQL WHERE clause to filter data.
		2.1.4 Stratified sampling	Collects samples by stratum.
		2.2.1 JOIN	Uses the SQL JOIN clause to merge data.

Level 1	Level 2	Level 3	Description
2. Data preprocessing	2.2 Data merge	2.2.2 Merge columns	Merges two columns from two tables.
		2.2.3 Merge rows (UNION)	Uses the SQL UNION operator to merge rows.
	2.3 Others	2.3.1 Standardization	Standardizes a column in a table.
		2.3.2 Splitting	Splits data according to the specified ratio.
		2.3.3 Normalization	Normalization is a method to simplify computation. Normalization converts a dimensional expression into a dimensionless expression (scalar).
		2.3.4 Missing data imputation	Replaces a null or specified value with the maximum, minimum, average, or custom value.
		2.3.5 KV to Table	Converts key:value (KV) pairs to a regular table.
		2.3.6 Table to KV	Converts a regular table to KV pairs.
		2.3.7 Append ID column	Adds an auto-increment ID column to a table.
3. Feature engineering	3.1 Feature transformation	3.1.1 PCA	Dimension reduction algorithm.
	3.2 Feature importance evaluation	3.2.1 Linear model feature importance evaluation	Evaluates the importance of the features in a linear model.
		3.2.2 Random forest feature importance evaluation	
	4.1 Percentile	N/A	Calculates the percentile of a column.
	4.2 Data pivoting	N/A	Supports the data pivoting function.
	4.3 Covariance	N/A	Measures the covariance of two given values.
	4.4 Empirical probability density chart	N/A	Returns the nonparametric density based on the estimated probability density.

Level 1	Level 2	Level 3	Description
4. Statistical analysis	4.5 Chi-square goodness of fit test	N/A	Determines the differences between the observed and expected frequencies of each class for a multiclass nominal variable.
	4.6 Chi-square test of independence	N/A	Checks whether two factors (each has two or more classes) are mutually independent. The null hypothesis is that the two factors are independent of each other.
	4.7 Two sample T test	N/A	The two sample T test includes the independent sample T test and the paired sample T test. The two samples independent of each other are called independent samples. An independent sample T test checks whether two samples are significantly different from each other. A paired sample T test checks whether the mean values from two paired populations are significantly different from each other.
	4.8 One sample T test	N/A	One sample T test checks whether the mean of a normally distributed population differs significantly from the target value.
	4.9 Normality test	N/A	Determines whether observed values are normally distributed.
	4.10 Lorenz curve	N/A	Illustrates the distribution of wealth across a population.
	4.11 Whole table statistics	N/A	Calculates the statistical information of each column in a table, including the default value, maximum value, minimum value, variance, and deviation.
	4.12 Pearson coefficient	N/A	Calculates the Pearson coefficient of two numerical columns.
	4.13 Histogram	N/A	Shows metrics in a histogram.
	4.14 Scatter plot	N/A	A chart where data points are distributed on the Cartesian coordinate plane.
	4.15 Correlation coefficient matrix	N/A	Calculates a matrix of correlated coefficients for multiple columns.

Level 1	Level 2	Level 3	Description
5. Machine learning	5.1 Binary classification	5.1.1 Linear SVM	A supervised machine learning algorithm used to identify and classify models, and then perform regression analysis.
		5.1.2 Logistic regression for binary classification	A supervised machine learning algorithm that uses logistic regression to train a binary classification model.
		5.1.3 GBDT binary classification	GBDT is an iterative decision tree algorithm that calculates results based on the final conclusions of multiple decision trees.
	5.2 Multiclass classification	5.2.1 Logistic regression for multiclass classification	A linear regression algorithm used for multiclass classification.
		5.2.2 Random forest	A type of classifier that uses multiple trees and sample data to generate models and make predictions.
		5.2.3 KNN	For a row in a prediction table, this component selects K nearest records from the training table. It then adds the row to the class that is most common among the K records.
		5.2.4 Naive Bayes	A family of classification algorithms based on the theorem of Bayes and independent hypothesis of feature conditions.
	5.3 Regression	5.3.1 GBDT regression	An algorithm that uses the GBDT structure for regression.
		5.3.2 PS linear regression	An algorithm that supports a large amount of training data by using parameter servers.
		5.3.3 Linear regression	Analyzes the linear relationship between a dependent variable and multiple independent variables.
	5.4 Clustering	5.4.1 K-means clustering	Clustering similarity is calculated based on a central object obtained by using mean values of objects in different clusters.

Level 1	Level 2	Level 3	Description
	5.5 Evaluation	5.5.1 Binary classification evaluation	Evaluates a binary classification model and uses it to make predictions.
		5.5.2 Multiclass classification evaluation	
		5.5.3 Regression model evaluation	
		5.5.4 Clustering model evaluation	
		5.5.5 Confusion matrix	
	5.6 Prediction	5.6.1 Prediction	
	5.7 Collaborative recommendation	5.7.1 Collaborative filtering (etrec)	Etrec is an item-based collaborative filtering algorithm that uses two input columns and outputs the top K items with the highest similarity.
6. Time series	6.1 x13_arima	N/A	ARIMA is short for Autoregressive Integrated Moving Average Model. x13-arima is an ARIMA algorithm based on the open-source X-13ARIMA-SEATS seasonal adjustment.
	6.2 x13_auto_arima	N/A	Automatically selects an ARIMA model based on the Gomez and Maravall processes.
	7.1 Word splitting	N/A	An algorithm that is used to split words in the specified text. Currently, only Chinese is supported for the Taobao and Internet word splitting models.
	7.2 Word frequency statistics	N/A	After word splitting, words are listed in the same order of document IDs. The frequency of a word appears in each document is calculated.
	7.3 TF-IDF	N/A	A statistical method for evaluating the importance of a word to a document in a collection or corpus.
	7.4 PLDA	N/A	Outputs the probability density of the topics in each document.

Level 1	Level 2	Level 3	Description
7. Text analysis	7.5 Word2Vec	N/A	Converts words to vectors.
	7.6 Convert rows, columns, and values to KV pairs	N/A	Converts a set of row, column, and value to a KV pair (row, [col_id,value]).
	7.7 Text summarization	N/A	Uses machine learning algorithms to create a summary on a document.
	7.8 Keyword extraction	N/A	Extracts words from a document. The extracted words are most correlated to the meaning of the document.
	7.9 Sentence splitting	N/A	Splits sentences by punctuation.
	7.10 Deprecated word filtering	N/A	A preprocessing method in text analysis. This method is used to filter out the noise in word splitting results, such as of, yes, and ah. Custom dictionaries are supported.
	7.11 String similarity		String similarity calculation is a basic operation in machine learning. It is typically used in industries such as information retrieval, natural language processing, and bioinformatics. This algorithm supports these similarity calculation methods: Levenshtein distance, longest common substring, string subsequence kernel, cosine, and simhash_hamming. It also supports these input methods: string-to-string and top N.
	7.12 String similarity-top N	N/A	Checks whether a given string is one of the top N predictions.
	7.13 Semantic vector distance	N/A	Outputs words (sentences) that are nearest to a given word (sentence) based on the semantic vectors calculated by an algorithm component, such as Word2vec. For example, you can generate a list of words that are most similar to a given word based on the semantic vectors returned by the Word2vec component.

Level 1	Level 2	Level 3	Description
	7.14 N-gram counting	N/A	N-grams are generated based on words.The number of the corresponding N-grams in all corpora is counted.
	7.15 PMI	N/A	Counts the co-occurrence of all words in several documents and calculates the Pointwise Mutual Information (PMI) between every two words.
	7.16 Document similarity	N/A	Calculates the similarity between documents or sentences that are separated with spaces based on the similarity of strings. The document similarity is calculated in same way as string similarity calculation.
8. Deep learning	8.1 TensorFlow 1.4	N/A	Uses the PAI-Tensorflow framework to implement deep learning.
	8.2 Read OSS buckets	N/A	Specifies the authorization information and endpoint that are required when Machine Learning Platform for AI reads or writes the corresponding Object Storage Service (OSS) bucket.
	9.1 K-Core	N/A	The k-core of a graph is the subgraph that remains after all vertices with a degree less than or equal to K are removed.
	9.2 Single-source shortest path	N/A	Calculates the shortest path between two points.
	9.3 Page ranking	N/A	Calculates the rank of a Web page.
	9.4 Label propagation clustering	N/A	A graph-based semi-supervised machine learning algorithm. A node is labeled based on the labels of its neighboring nodes. The scale at which labels are propagated is determined by the similarity between the node and its neighbors. Labels are propagated iteratively and updated over time.
	9.5 Label propagation classification	N/A	A semi-supervised classification algorithm that uses the label information of labeled nodes to predict that of unlabeled nodes.

Level 1	Level 2	Level 3	Description
9. Network analysis	9.6 Modularity	N/A	A measure of the structure of networks. It measures the closeness of communities divided from a network structure. A value larger than 0.3 represents an obvious community structure.
	9.7 Maximal connected subgraph	N/A	A maximal connected subgraph of an undirected graph G is a connected subgraph of G, where all vertices are connected with the least edges.
	9.8 Vertex clustering coefficient	N/A	Calculates the peripheral density of nodes near a node in an undirected graph. The density of a star network is 0. The density of a fully meshed network is 1.
	9.9 Edge clustering coefficient	N/A	Calculates the peripheral density of each edge in an undirected graph.
	9.10 Counting triangle	N/A	Generates all triangles in an undirected graph.
	9.11 Tree depth	N/A	Generates the depth and tree ID of each node in a network composed of many trees.
10. Tools	10.1 MaxCompute SQL	N/A	Runs MaxCompute SQL statements.
	11.1 Binning	N/A	Performs data binning by using the equal frequency, equal width, or auto binning mode. The input is continuous or discrete features. The output is binning rules for all features.
	11.2 Data conversion module	N/A	Uses binning results to convert features that you input. The data conversion module supports normalization, Weight of Evidence (WoE), and discretization. Normalization outputs values between 0 to 1. WoE replaces feature values with WOE of bins. Discretization converts variables to dummy variables based on binning results. All output data is in KV format.

Level 1	Level 2	Level 3	Description
11. Financials	11.3 Scorecard training	N/A	The scorecard is a modeling tool commonly used in credit risk evaluation. Scorecard modeling performs original variable discretization through binning and uses linear models (such as logistic regression and linear regression) to conduct model training. The scorecard supports various features, including feature selection and score conversion. In addition, it allows you to add constraints to variables during model training.
	11.4 Scorecard prediction	N/A	The scorecard prediction component uses the model generated by the scorecard training component to predict scores.
	11.5 Population stability index	N/A	Population stability index (PSI) is an important metric to identify a shift in the population for credit scorecards, for example, the changes in the population within two months. A PSI value smaller than 0.1 indicates insignificant changes. A PSI value between 0.1 and 0.25 indicates minor changes. A PSI value larger than 0.25 indicates major changes in the population.

34.5. Scenarios

Machine Learning Platform for AI can be applied to the following scenarios:

Marketing

- Use cases: commodity recommendations, user profiling, and precise advertising.
- Example: Machine Learning Platform for AI associates user shopping behavior data with commodities to offer commodity recommendations and evaluate the recommendation results.

Finance

- Use cases: loan delivery prediction, financial risk management, stock trend prediction, and gold price prediction.
- Example 1: Agricultural loan delivery is used in a typical example of data mining. A lender uses machine learning to build an empirical model based on historical data such as the lendee's annual income, cultivated crop type, and debit and credit history. This model is then used to predict the lendee's capacity.
- Example 2: Users' credit card expense records are processed by a machine learning algorithm. After raw data binning and feature engineering transformation, data is used to build a linear model. The

final credit score of each user is determined by the model predictions, and can be used in a variety of loan and finance related credit checks.

Text

- Use cases: news classification, keyword extraction, text summarization, and text analysis.
- Example: A simple system for automatic commodity label classification is built using the text analysis function of Machine Learning Platform for AI.

  Take online shopping as an example. A commodity typically has labels for multiple dimensions. For example, the commodity description of a pair of shoes may be "Korean Girl Dr. Martens Women's Preppy/British-style Lace-up Dull-polish Ankle High Platform Leather Boots." A bag may be described as "Discount Every Day 2016 Autumn and Winter New Arrival Women's Korean-style Seashell-shaped Tassel Three-way Bag as a Messenger Bag, Hand Carry Bag, and Shoulder Bag."

  Each product description contains multiple dimensions such as the time, place of origin, and style. E-commerce platforms face the daunting challenge of how to classify hundreds of thousands of products based on these specified dimensions. The biggest challenge is determining which labels constitute the dimensions of each product. A label classification system can be built using a machine learning algorithm to automatically learn label terms. For example, the system can learn location-related labels such as Japan, Fujian, and Korea.

Unstructured data processing

- Use cases: image classification and image text extraction by using optical character recognition (OCR).
- Example: A prediction model can be quickly built for image recognition by using the TensorFlow deep learning framework. The TensorFlow deep learning framework can recognize images and return image classification results within half an hour. Image recognition by using deep learning can also be used in illicit image filtering, facial recognition, and object detection.

Other prediction cases

- Use cases: rainfall forecast, football match result prediction, microblog leader analysis, and social relationship chain analysis.
- Example: Air quality can be predicted by Machine Learning Platform for AI based on historical air quality index data such as PM 2.5, carbon monoxide concentration, and nitrogen dioxide concentration. The prediction results can then be used to determine which air quality index has the greatest impact on PM 2.5 levels.

34.6. Limits

This topic describes the limits of Machine Learning Platform for AI.

Item	Description
MaxCompute service deployment	The computing service of Machine Learning Platform for AI relies on MaxCompute to store tables and perform some SQL-related computations. Therefore, MaxCompute must be deployed before you can use the machine learning service.

Item	Description
OSS service deployment (deep learning)	The deep learning service of Machine Learning Platform for AI relies on OSS to store, read, and write data. Therefore, OSS must be deployed.
Limit to component use	For more information, see parameter configurations for each component.

34.7. Terms

This topic describes terms used in Machine Learning Platform for AI.

data mining

A broad definition that describes the use of algorithms to explore useful information from large amounts of data. Typically, data mining uses machine learning algorithms.

Alibaba Cloud DTplus

The big data platform of Alibaba Cloud. DTplus provides enterprises with a complete set of end-to-end big data solutions for fields such as enterprise data warehouses, BI, machine learning, and data visualization. These solutions help enterprises become more agile, smarter, and more perceptive in the data technology (DT) era.

table

Data storage units of MaxCompute. Tables used by machine learning are stored in MaxCompute. Logically, a table is a two-dimensional structure that consists of rows and columns. Each row represents a record. Each column represents a field of the same data type. One record can contain one or more columns. The schema of a table consists of column names and column types.

On Machine Learning Platform for AI, you can create a table, add the table to favorites, and import data to the table. The table is automatically stored in MaxCompute. To delete a table, you must log on to MaxCompute.

partition

Certain columns specified in a table when the table is created. In most cases, you can consider a partition as a directory in a file system.

Tables are stored in MaxCompute. MaxCompute uses each value in a partition column as a partition (directory). You can specify multiple hierarchies of partitions by using multiple table columns as table partitions. The relationships between partitions are similar to those between multiple hierarchies of directories.

When using data, if you specify the name of a partition, only the data in the specified partition is read. This removes the need to scan the entire table for data, improves processing efficiency, and minimizes costs.

lifecycle

The period of time that determines how long a table partition is retained since it was last updated. If a table (partition) is not updated within the specified time period, MaxCompute automatically deletes it.

sparse data format

Datasets in which most data entries are null or have a value of 0. Sparse data can be utilized effectively if efficient methods are used to explore the useful information that exists in the relatively incomplete sparse data set.

On Machine Learning Platform for AI, if the data of a feature in a sample is in the sparse format, you must convert the format to the LibSVM format, select **key:value, key:valuesparse data format** on the parameter setting page, and then upload the data.

feature

An attribute that is used to describe an object. For example, a person can be described by age, gender, occupation, and other attributes. Each of these attributes is a feature of the person.

On Machine Learning Platform for AI, a dataset is stored as a table. A column in the table is a feature of this dataset. The features of data are important to machine learning. Data and its features determine the upper limit of machine learning capabilities. Models and algorithms are used to help machine learning reach the upper limit. Therefore, features must be processed before a machine learning experiment can be executed. Typical feature processing methods include data preprocessing, feature selection, and dimension reduction.

dimension reduction

A method that is used to remove the dimensions that have minor impacts and extract key features from a large number of features. A dimension is the way something is observed. In machine learning, dimensions describe the features of a dataset. If a dataset has millions of features, the training model for machine learning will be complex and the training will take a long period of time In this case, dimension reduction is required. The dimension reduction algorithms on Machine Learning Platform for AI include PCA and LDA.

35.E-MapReduce (EMR)

35.1. What is E-MapReduce?

E-MapReduce (EMR) is short for Elastic MapReduce. EMR is an end-to-end big data processing and analytics system. It leverages open source big data ecosystems such as Hadoop, Spark, Kafka, and Storm to manage clusters, jobs, and data.

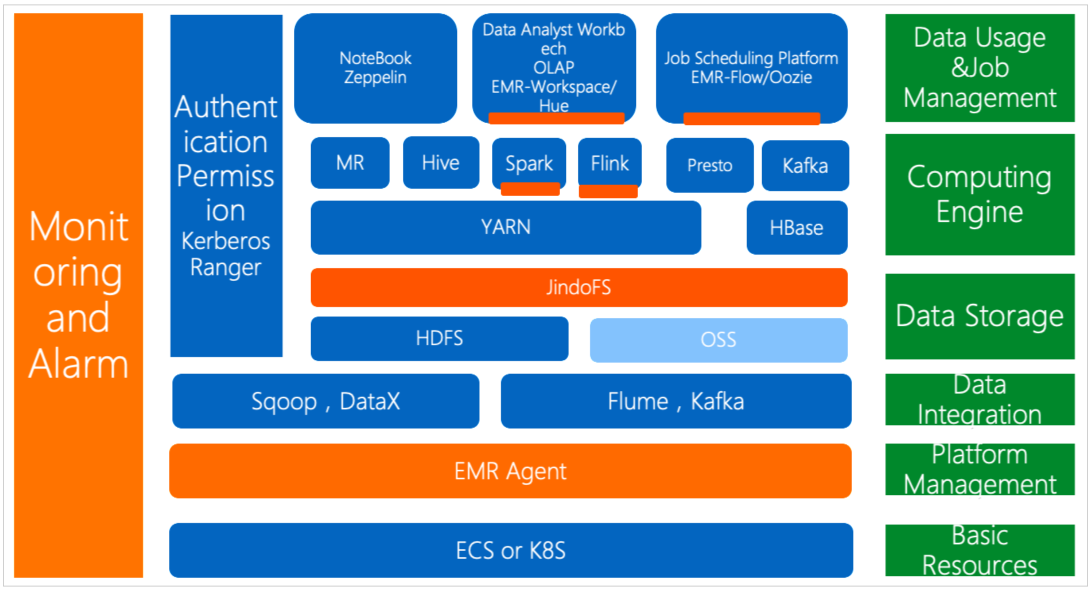
EMR is built in a virtual machine environment such as ECS instances or on physical machines. It allows you to analyze and process data by using peripheral systems in the Hadoop and Spark ecosystems, such as Apache Hive, Apache Pig, and HBase. You can also use EMR to export and import data from and to the data storage systems and database systems of Alibaba Cloud, such as Object Storage Service (OSS) and ApsaraDB for RDS.

35.2. Architecture

This topic describes the architecture of E-MapReduce (EMR).

The following figure shows the architecture.

EMR architecture



35.3. Benefits

This topic describes the benefits of E-MapReduce (EMR).

EMR provides an integrated solution to manage clusters, which frees you up from the complex management of user-created clusters. EMR also has the following benefits:

- Deep integration

  EMR is integrated with other Alibaba Cloud services such as Object Storage Service (OSS), Message Service (MNS), ApsaraDB for RDS, and MaxCompute. This enables these services to act as the input source or output destination of the Hadoop or Spark compute engine in EMR.

- Security

EMR is integrated with Resource Access Management (RAM), which allows you to use Alibaba Cloud accounts and RAM users to isolate permissions on services.

35.4. Features

This topic describes the features of E-MapReduce (EMR).

EMR provides the following features:

- Allows you to create various types of jobs such as Spark, Hadoop, Hive, Pig, Sqoop, Spark SQL, and Shell to meet your business needs, such as log analysis, data warehousing, business intelligence, machine learning, and scientific simulation.

  After you select a job type, you can define the commands to execute and the actions to follow after a job failure. You can copy, modify, or delete a job.

- Allows you to create execution plans.

  An execution plan is a set of jobs. You can run an execution plan on an existing cluster or a temporary cluster that is dynamically created. You can configure scheduling policies to determine whether to run an execution plan only once or on a schedule. An execution plan consumes as many resources as each job requires. This maximizes resource utilization and reduces costs. The flexibility of execution plans lies in the following aspects:

  - You can combine different types of jobs such as Hadoop, Spark, Hive, and Pig in the same execution plan.
  - You can run an execution plan once or periodically.

- Provides an interactive workbench.

  The interactive workbench allows you to write and execute Spark, Spark SQL, and Hive SQL tasks in the EMR console. After a task is complete, you can view the results in the workbench. You can use the workbench to process short-term, real-time results-oriented, and debugging tasks. We recommend that you use jobs and execution plans to process long-term scheduled tasks.

- Supports alerts.

  You can associate execution plans with alert groups. If you enable Alert Notification on the Execution Plan page, contacts in the associated alert group receive an SMS message after each execution plan is complete. The SMS message contains the name of the execution plan, job execution results (the numbers of successes and failures), the cluster name, and the duration of execution.

35.5. Scenarios

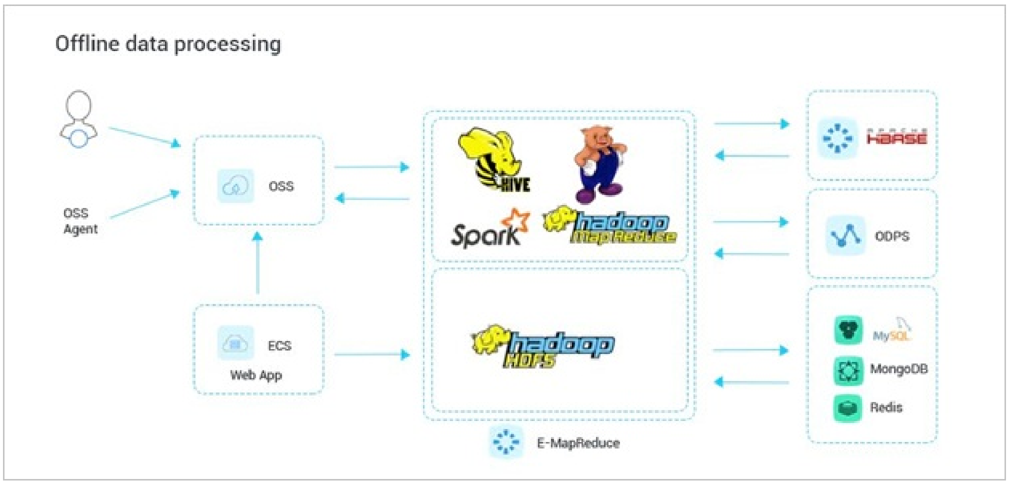
This topic describes the scenarios where EMR is used.

EMR is used in the following scenarios:

- Offline data analysis

  You can synchronize a large number of logs from business services such as games, web apps, and mobile apps to the data nodes of EMR. You can then use tools such as Hue and a mainstream computing framework such as Hive, Spark, and Presto to get a quick insight into the data. You can also use tools such as Sqoop to load data in ApsaraDB for RDS or other storage engines to EMR. Then, you can analyze the data and synchronize the results to ApsaraDB for RDS or other storage engines. This feature helps implement data visualization.

  Offline data analysis

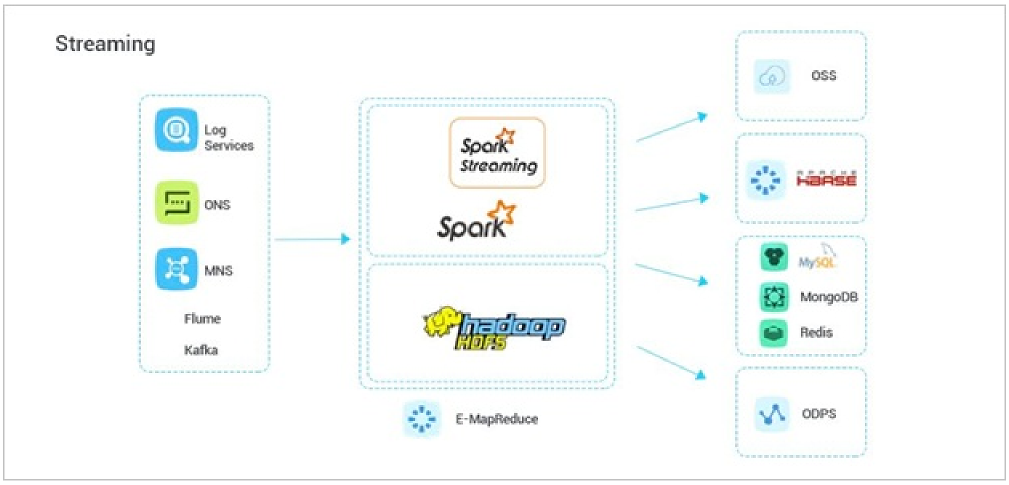


- Streaming data analysis

  EMR allows you to use and process real-time streaming data from services such as Log Service (Log), Message Queue (MQ), Message Service (MNS), and Apache Kafka based on Spark Streaming and Storm.

  EMR analyzes streaming data in fault-tolerant mode and writes analysis results into Object Storage Service (OSS) or HDFS.

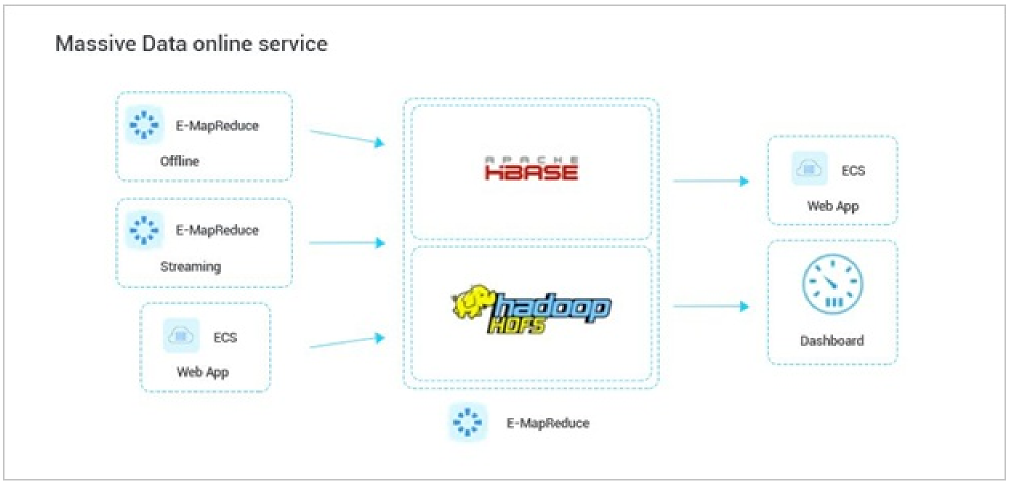
  Streaming data analysis



- Online analysis of a large volume of data

  EMR analyzes petabytes of structured, semi-structured, or unstructured data generated by web apps and mobile apps. This allows web apps or data visualization services to visualize data in real time based on the analysis results obtained from EMR.

  Online analysis of a large volume of data



35.6. Terms

This topic introduces terms used in EMR.

job

Similar to MaxCompute or Hadoop jobs, an EMR job is the basic unit used to process and analyze big data.

Hadoop

The following two core components of Hadoop are used in EMR:

- YARN

  YARN schedules tasks and manages cluster resources.

- HDFS

  HDFS is a distributed file system.

Hive

Hive is a Hadoop-based offline data processing system that provides an SQL-like interface to query data. It uses tables to store and manage data.

Spark

Spark is an in-memory distributed computing framework that supports offline and real-time computing, SQL statements, and machine learning.

Hue

Hue is a visualized platform used to manage open source components, such as Hadoop, Hive, Oozie, and HBase.

Oozie

Oozie is a job scheduler that supports workflow orchestration by building a directed acyclic graph (DAG). It supports multiple types of jobs.

Presto

Presto is a distributed SQL query engine that retrieves large datasets from one or more data sources.

Zeppelin

Zeppelin is a web-based notebook that enables interactive data analysis and collaborative documents with SQL and Scala.

ZooKeeper

ZooKeeper is an open source distributed application coordination service. It is an open source implementation of Google's Chubby and an important component of Hadoop and HBase. It mainly solves the consistency problem of distributed applications. Its services include configuration maintenance, domain name services, distributed synchronization, and group services.

Sqoop

Sqoop is a tool designed to transfer data between HDFS and relational databases.

Kafka

Kafka is a distributed messaging system that features high throughput, scalability, reliability, and performance. It is used in real-time computing, log processing, and data aggregation.

HBase

HBase is an open source, distributed, and column-oriented database. It is a component of the Apache Hadoop project. Different from typical relational databases, HBase is designed to store unstructured data. HBase is column-oriented rather than row-oriented.

Phoenix

Phoenix provides SQL-like statements for you to analyze HBase data.

MetaService

MetaService helps you access Alibaba Cloud resources in EMR clusters without an AccessKey pair.

Metadatabase

A metadatabase is a data repository that organizes, stores, and manages data based on data structures.

Kerberos

Kerberos is a third-party authentication protocol that is designed for TCP/IP networks. It uses symmetric cryptography based on the Data Encryption Standard (DES).

36.DataHub

36.1. What is DataHub?

DataHub is a platform designed to process streaming data. You can publish and subscribe to streaming data in DataHub and distribute the data to other platforms. DataHub allows you to analyze streaming data and build applications based on the streaming data.

DataHub collects, stores, and processes streaming data from mobile devices, applications, website services, and sensors. You can use your own applications or Apsara Stack Realtime Compute to process streaming data in DataHub, such as real-time website access logs, application logs, and events. The processing results such as alerts and statistics presented in graphs and tables are updated in real time.

Based on the Apsara system of Alibaba Cloud, DataHub features high availability, low latency, high scalability, and high throughput. DataHub is seamlessly integrated with Realtime Compute, allowing you to use SQL to analyze streaming data.

DataHub can also distribute streaming data to Apsara Stack services such as MaxCompute and Object Storage Service (OSS).

36.2. Benefits

High throughput

You can write terabytes (TB) of data into a topic and up to 80 million records into a shard every day.

Real-time processing

DataHub makes it easy to collect and process various types of streaming data in real time so you can react quickly to new information.

Ease of use

- DataHub provides a variety of SDKs for C++, Java, Python, Ruby, and Go.
- In addition to SDKs, DataHub provides RESTful APIs so that you can manage DataHub by using existing protocols.
- You can use collection tools such as Fluentd, Logstash, and Oracle GoldenGate to write streaming data into DataHub.
- DataHub supports structured and unstructured data. You can write unstructured data to DataHub, or create a schema for the data before it is written into the system.

High availability

- The processing capacity of DataHub is automatically scaled out without affecting your services.
- DataHub automatically stores multiple copies of data.

Scalability

You can dynamically adjust the throughput of each topic. The maximum throughput of a topic is 256,000 records per second.

Data security

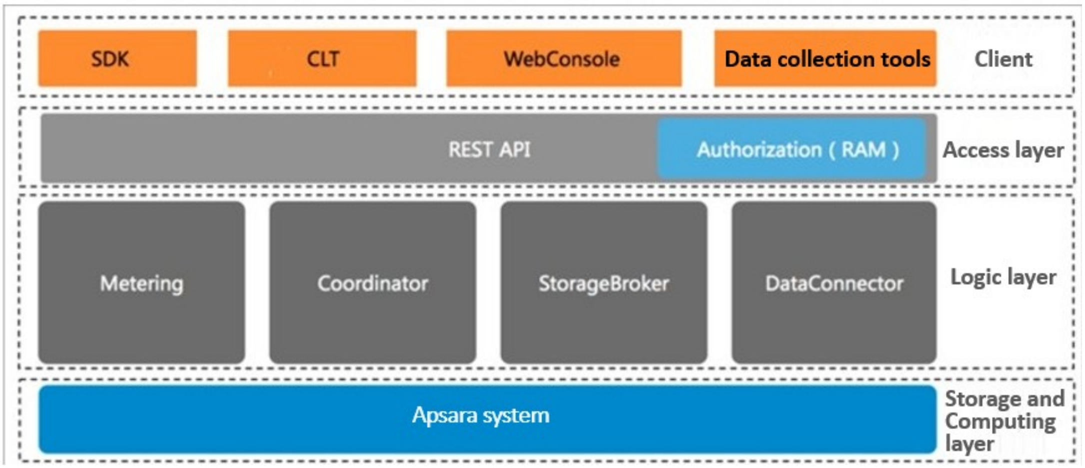
- DataHub provides enterprise-level security measures and isolates resources between users.

- It also provides several authentication and authorization methods, including whitelist configuration and RAM user management.

36.3. Architecture

Architecture shows the architecture of DataHub.

Architecture



The architecture of DataHub consists of four layers: **clients**, **access layer**, **logic layer**, and **storage and scheduling layer**.

Clients

DataHub supports the following types of clients:

- SDKs: DataHub provides SDKs in a variety of languages such as C++, Java, Python, Ruby, and Go.
- Command-line tools (CLTs): You can run commands in Windows, Linux, or Mac operating systems to manage projects and topics.
- Console: In the console, you can manage projects and topics, create subscriptions, view the shard status, monitor topic performance, and manage DataConnectors.
- Data collection tools: You can use Logstash, Fluentd, and Oracle GoldenGate (OGG) to collect data to DataHub.

Access layer

You can access DataHub by using HTTP and HTTPS. DataHub supports Resource Access Management (RAM) authorization and horizontal scaling of topic performance.

Logic layer

The logic layer handles the key features of DataHub, including project and topic management, data read and write, offset-based data consumption, traffic statistics, and data synchronization. Based on these key features, the logic layer is composed of the following modules: StorageBroker, Metering, Coordinator, and DataConnector.

- StorageBroker: provides data reads and writes in DataHub. This module adopts the log file storage model of Apsara Distributed File System, halving the read/write volume compared with the conventional write-ahead logging (WAL) model. This module stores three copies of data to ensure that no data is lost if a server fault occurs, and supports disaster recovery between data centers. It supports real-time data caching to ensure efficient consumption of real-time data and supports an independent read cache of historical data to enable concurrent consumption of historical data.

- Metering: supports shard-level billing based on the consumption period.
- Coordinator: supports offset-based data consumption and horizontal scaling of the processing capacity. It supports up to 150,000 QPS on a single node.
- DataConnector: supports automatic data synchronization from DataHub to other Apsara Stack services, including MaxCompute, OSS, AnalyticDB, ApsaraDB RDS for MySQL, Tablestore, and Elasticsearch.

Storage and scheduling layer

- Storage: Based on the log file storage model of Apsara Distributed File System, DataHub supports append operations and solid state drive (SSD) storage. Data in each shard is stored in a separate file based on the timestamp of the data.
- Scheduling: Based on Job Scheduler, DataHub assigns shards to nodes based on the traffic on each shard. This ensures that the shards do not occupy the CPU or memory of Job Scheduler. The number of partitions on a single node has no upper limit. DataHub supports failovers within milliseconds and hot upgrades.

36.4. Features

36.4.1. Data queue

DataHub automatically generates a cursor for each record in a shard. The cursor is a unique sequence of numbers. You can improve the performance of a topic by increasing the number shards in the topic.

36.4.2. Checkpoint-based data restoration

DataHub supports saving checkpoints for subscribed applications in the system. You can restore data from any checkpoint you saved if your subscribed application fails.

36.4.3. Data synchronization

Data in DataHub is automatically synchronized to other Alibaba Cloud services.

DataConnector

You can create a DataConnector to synchronize DataHub data in real time or near real time to other Alibaba Cloud services, including MaxCompute, OSS, Elasticsearch, ApsaraDB RDS for MySQL, AnalyticDB, and Table Store.

You can configure the DataConnector so that the data you write to DataHub can be used in other cloud platforms. At-least-once semantics is applied in data synchronization. This ensures that no data is lost, but may result in duplicated records in the destination platform if an error occurs during the synchronization process.

Destination platforms

The following table describes the platforms to which DataHub records can be synchronized.

Destination platforms

Destination platform	Timeliness	Description
MaxCompute	Near real-time. Latency: 5 minutes.	The column names and data types in the source topic must be the same as those in MaxCompute. The MaxCompute table must have one or more corresponding partition columns.
OSS	Real-time	Records are synchronized to the specified bucket in OSS and are saved as CSV files.
Elasticsearch	Real-time	Records are synchronized to the specified index in Elasticsearch. Records may not be synchronized in the order of the recording time. If you want to synchronize data in the order of the recording time, you must write the records with the same partition key into the same shard.
ApsaraDB RDS for MySQL	Real-time	Records are synchronized to the specified table in ApsaraDB RDS for MySQL.
AnalyticDB	Real-time	Records are synchronized to the specified table in AnalyticDB.
Table Store	Real-time	Records are synchronized to the specified table in Table Store.

36.4.4. Scalability

The throughput of each topic can be scaled by splitting or merging shards.

You can adjust the number of shards in a topic according to the service load.

For example, if the topic throughput cannot handle a surge in the service load during Double 11, you can split existing shards to up to 256 to increase the throughput to 256 MB/s.

As the service load decreases after Double 11, you can reduce the number of shards as needed by performing the merge operation.

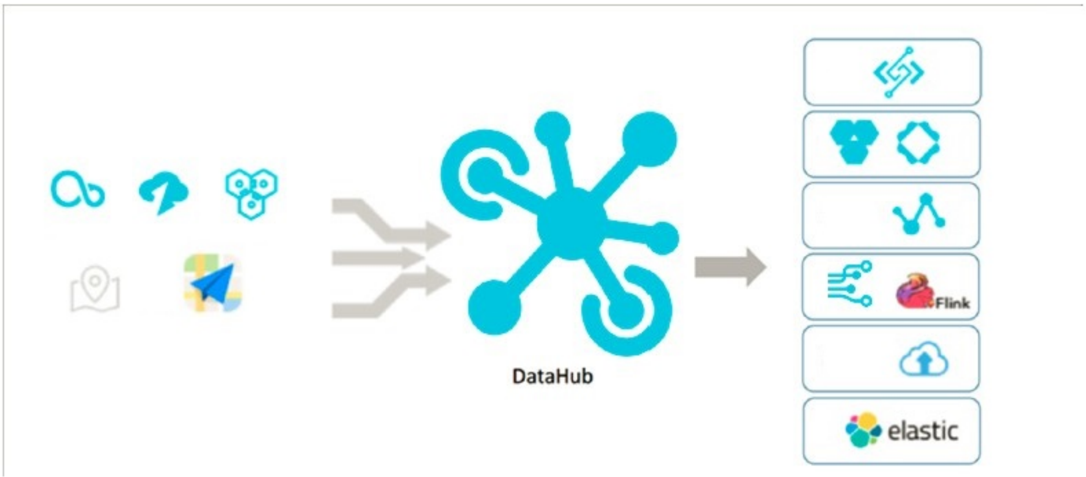
36.5. Scenarios

36.5.1. Overview

As a streaming data processing platform, DataHub can be used with various Alibaba Cloud products to provide one-stop data processing services.

36.5.2. Data uploading

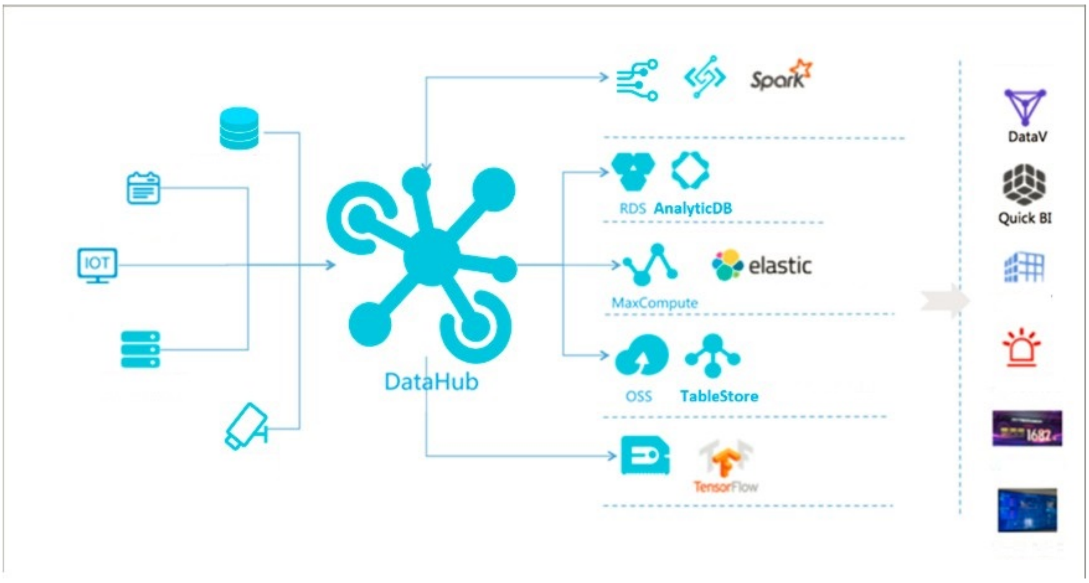
Data uploading



DataHub is connected to other Alibaba Cloud services, saving you the trouble of uploading the same data to different platforms.

36.5.3. Data collection

Data collection

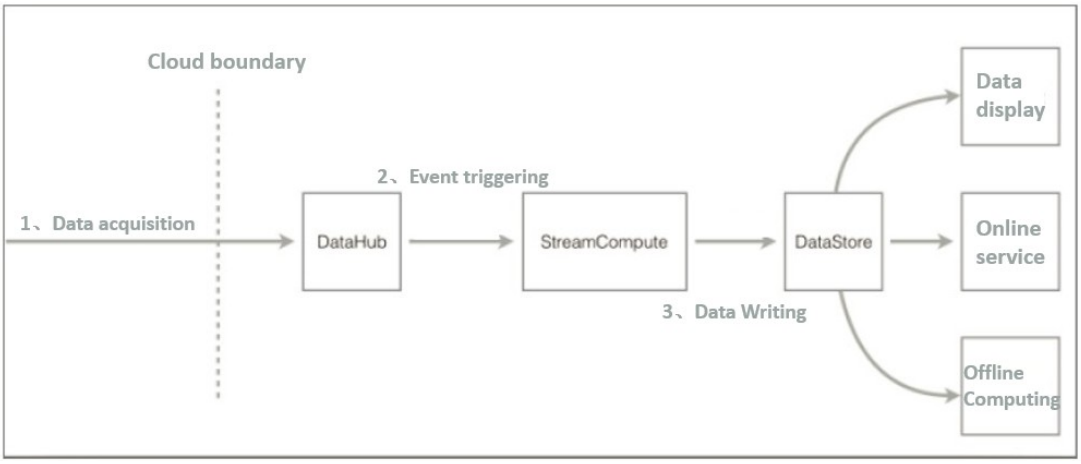


DataHub provides several types of data collection tools for you to write your data into DataHub. DataHub supports log collection from Logstash and Fluentd, and binary log collection from Data Transmission Service (DTS) and Oracle GoldenGate (OGG). DataHub also supports the collection of surveillance videos through GB28181.

36.5.4. Realtime Compute

Realtime Compute is a real-time computing engine of Alibaba Cloud, which allows you to use a language similar to SQL to analyze streaming data. Data can be transferred from DataHub to Realtime Compute or from Realtime Compute to DataHub.

DataHub and Realtime Compute



36.5.5. Data utilization

You can build an application to consume the data in DataHub, process the data in real time, and output the process results.

You can also use another application to process the streaming data output from the previous application to form a directed acyclic graph (DAG)-based data processing procedure.

36.5.6. Data archiving

You can create a DataConnector to periodically archive data in DataHub to MaxCompute.

36.6. Limits

Limits

Item	Limit	Description
Active shards per topic	(0,256]	Each topic can contain up to 256 active shards.
Total shards per topic	(0,512]	You can create up to 512 shards in each topic.
Http BodySize	Up to 4 MB	The size of the HTTP request body cannot exceed 4 MB.
String size	Up to 1 MB	The size of a string cannot exceed 1 MB.
Merge and split operations on new shards	5s	You cannot merge a shard with another shard or split the shard within the 5s after the shard is created.
Queries per second (QPS)	Up to 5,000	The write QPS limit for each shard is 5,000. Multiple queries in one batch are considered as one query.

Item	Limit	Description
Throughput	Up to 5 MB/s	Each shard provides a throughput of up to 5 MB/s.
Projects	Up to 100	You can create up to 100 projects with each account.
Topics per project	Up to 1,000	You can create up to 1,000 topics in each project. Contact the administrator if you need to create more topics.
Time-to-live (TTL) of records	[1,7]	The TTL of each record in a topic ranges from one to seven days.

36.7. Terms

project

A project is an organizational unit in DataHub and contains one or more topics. DataHub projects and MaxCompute projects are independent of each other. Projects that you create in MaxCompute cannot be used in DataHub.

topic

The smallest unit for data subscription and publishing. You can use topics to distinguish different types of streaming data. For more information about projects and topics, see Limits in Product Introduction.

time-to-live of records

The period that each record can be retained in the topic. Unit: day. Minimum value: 1. Maximum value: 7.

shard

A shard in a topic. Shards ensure the concurrent data transmission of a topic. Each shard has a unique ID. A shard can be in a different status. For more information about shard status, see the following table. Each active shard consumes server resources. We recommended that you create shards as needed.

Note Shard status

Status	Description
Activating	All shards in a topic are in the Activating state when the topic is created. You cannot perform read or write operations on shards because they are being activated.
Active	Read and write operations are enabled when a shard is in the Active state.
Deactivating	A shard is in the Deactivating state when it is being split or merged with another shard. You cannot perform read or write operations on the shard because it is being deactivated.
Deactivated	A shard is in the Deactivated state when the split or merge operation is completed. The shard is read-only when it is in the Deactivated state.

hash key range

The range of hash key values for a shard, which is in [Starting hash key,Ending hash key) format. The hashing mechanism ensures that all records with the same partition key are written to the same shard.

merge

The operation that merges two adjacent shards. Two shards are considered adjacent if the hash key ranges for the two shards form a contiguous set with no gaps.

split

The operation that splits one shard into two adjacent shards.

record

A unit of data that is written into DataHub.

record type

The data type of records in a topic. Tuple and blob are supported. A tuple is a sequence of immutable objects. A blob is a chunk of binary data stored as a single entity.

Note

- The following data types are supported in a tuple topic. Tuple data types

Type	Description	Value range
Bigint	An 8-byte signed integer. **Note** Do not use the minimum value (-9223372036854775808) because this is a system reserved value.	-9223372036854775807 to 9223372036854775807
String	A string. Only UTF-8 encoding is supported.	The size of a string cannot exceed 1 MB.
Boolean	One of two possible values.	Valid values: True and False, true and false, or 0 and 1.
Double	A double-precision floating-point number. It is 8 bytes in length.	-1.0 10^{308} to 1.0 10^{308}
Timestamp	A timestamp.	It is accurate to microseconds.

- In a blob topic, a chunk of binary data is stored as a record. Records written into DataHub are Base64 encoded.

37.Quick BI

37.1. What is Quick BI?

Quick BI is a flexible, lightweight self-service BI platform based on cloud computing.

Quick BI supports various data sources:

- MaxCompute (formerly ODPS) and AnalyticDB for PostgreSQL
- User-created MySQL databases that are hosted on ECS
- VPC data sources

Quick BI can analyze large amounts of data online in real time. Quick BI helps significantly reduce data retrieval costs and is easy to use with the support of intelligent data modeling tools. Drag-and-drop operations and various visual charts allow you to perform data pivoting, downloads, and data exploration, and prepare reports and BI portals.

Quick BI enables everyone to be both a data viewer and a data analyst to achieve data-driven operations for enterprises.

37.2. Benefits

Benefits of Quick BI can be summarized as follows.

High compatibility

Connects to various Alibaba Cloud data sources, such as MaxCompute and AnalyticDB for PostgreSQL.

Quick response

Responds in seconds to hundreds of millions of data queries.

Powerful capabilities

Allows users to easily create complex reports by using workbooks.

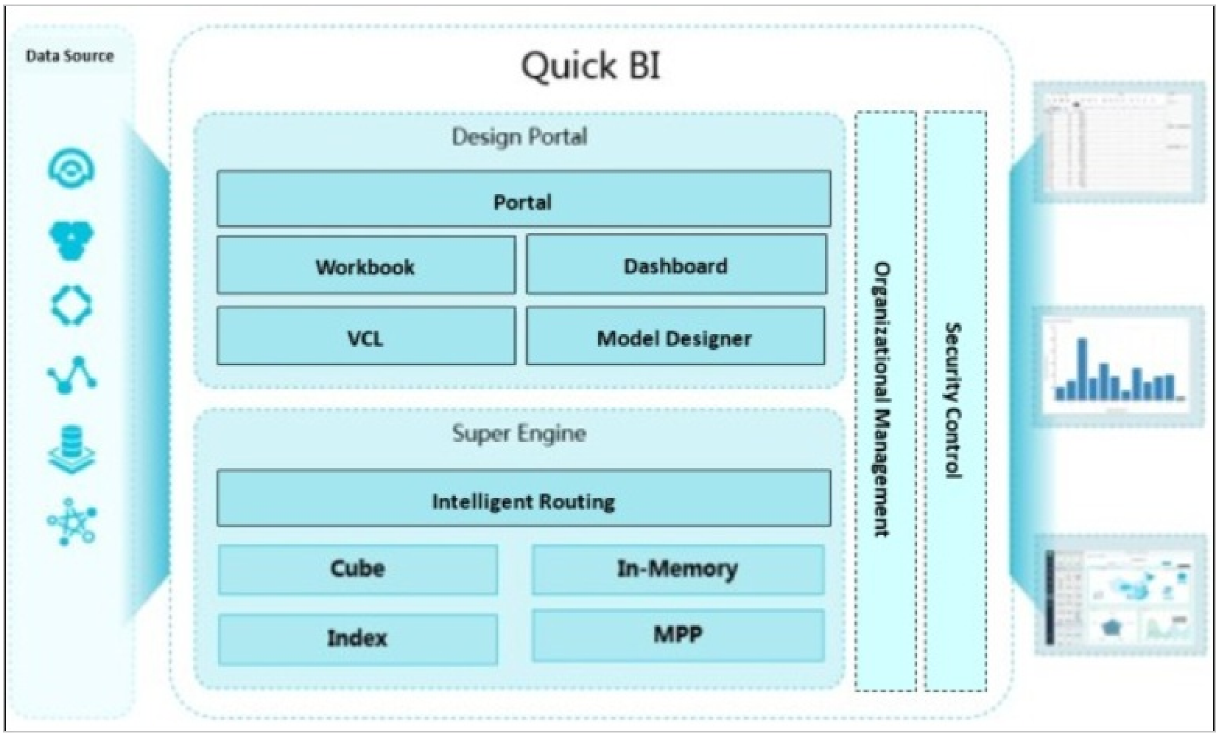
User-friendliness

Provides various data visualization functions and automatically identifies data properties to generate the most appropriate charts for users.

37.3. Architecture

This topic describes the architecture of Quick BI, and its major modules and their functions.

The following figure shows the architecture of Quick BI.



The following sections describe major Quick BI modules and their functions:

- Data connection module

  Connects to various Alibaba Cloud data sources, such as MaxCompute and AnalyticDB for PostgreSQL. This module provides APIs to query metadata or data from data sources.

- Data preprocessing module

  Provides lightweight extract, transform, load (ETL) processing for data sources. Quick BI supports custom SQL statements of MaxCompute. Quick BI will support data preprocessing for more data sources in the future.

- Data modeling module

  Takes charge of OLAP modeling of data sources and transforms data sources into multi-dimensional analysis models. It supports standard semantics such as dimensions (including dimensions of Date and Geo types), measures, and galaxy schemas. It also supports calculated fields, and allows you to process dimensions and measures by using SQL syntax for existing data sources.

- Workbook

  Provides workbook functions, including row and column filtering, standard and advanced filtering, subtotal and total calculation, and conditional formatting. This module also supports operations such as data export, text processing, and sheet processing.

- Dashboard

  Assembles visual charts into a dashboard. Various charts are supported: line chart, pie chart, vertical bar chart, funnel chart, hierarchy chart, bubble map, colored map, and kanban. This module provides five basic widgets: query control, tab, iFrame, image, and text area, and supports data interaction among charts.

- **BI portal**

  Assembles dashboards into a BI portal and supports built-in links (dashboards), external links (third-party links), and basic settings of templates and the menu bar.

- **Query engine**

  Queries data that is stored in data sources.

- **Organization permission management module**

  Configures permissions based on organization or workspace and manages workspace-specific user roles. This module allows you to grant your members different permissions on the same report.

- **Row-level permission control module**

  Controls row-level permissions and allows different members to view different parts of a report based on the permissions they are granted.

- **Share and publish module**

  Shares workbooks, dashboards, and BI portals with other members and publishes dashboards to the Internet.

37.4. Features

This topic describes the features of Quick BI.

Quick BI provides the following features:

Seamless integration with cloud databases

Supports various Alibaba Cloud data sources, such as MaxCompute and AnalyticDB for PostgreSQL.

Various charts

Provides diverse options for data visualization. Quick BI provides various built-in charts, such as vertical bar chart, line chart, pie chart, radar chart, and scatter chart to meet requirements in different business scenarios. Quick BI automatically identifies data properties and intelligently recommends visualization solutions.

Multidimensional data analysis

Supports Multidimensional data analysis. Quick BI is a web-based data analysis system. It supports drag-and-drop operations, data presentation similar to EXCEL files, one-click data import, and real-time data analysis. You can use Quick BI to analyze data from different perspectives without repetitive modeling.

Quick building of BI portals

Supports drag-and-drop operations and provides powerful data modeling capabilities and multiple visual charts to help you build BI portals in a short period of time.

Real-time analysis

Supports online analysis for large amounts of data without the need of preprocessing. This significantly improves the efficiency of data analysis.

Data permission management

Supports member permission management and row-level permission control. Different users can view different reports or view different parts of the same report.

37.5. Scenarios

37.5.1. Instant data analysis and effective decision-making

Quick BI can instantly analyze a large volume of data and make decisions.

Business goals:

- Convenient data retrieval

  Quick BI eliminates the reliance on IT professionals to write SQL statements for multidimensional data analysis.

- Convenient report generation and maintenance

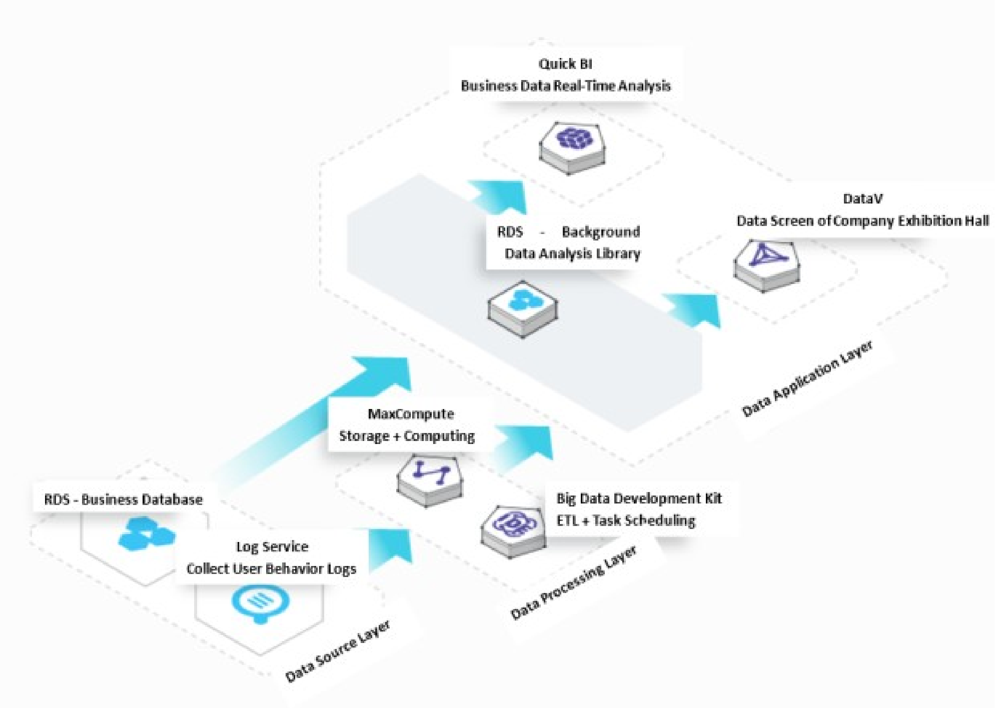
  Quick BI simplifies and shortens the process of delivering updates and new code to an analytics system.

- Low human resource costs

  Quick BI provides easy-to-use user interfaces, reducing your maintenance costs.

Recommended combination: relational database and Quick BI

Instant data analysis and effective decision-making



37.5.2. Integration with existing systems

The Quick BI report system can be integrated with your own systems, such as the OA system and internal management system, to efficiently present data.

Business goals:

- Easy adoption

  Quick BI is a user-friendly and easy-to-use service for users from different backgrounds, which satisfies data analysis needs of personnel in various departments.

- High efficiency for data visualization

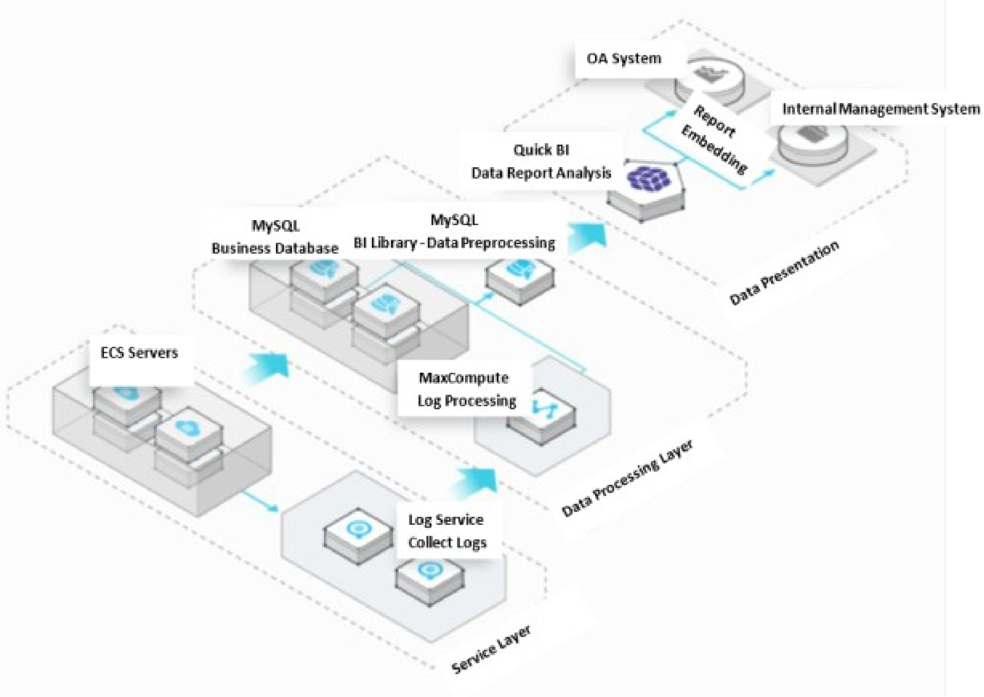
  Integration with existing systems allows for quick data analysis and improves the efficiency of viewing data.

- Unified management platform

  You can access and manage data by using a unified platform.

Recommended combination: relational database and Quick BI

Integration with existing systems



37.5.3. Permission control of transaction data

Quick BI controls transaction data permissions and row-level permissions.

Business goals:

- Row-level permission control

  You can easily create a report for all members and allow members to view only data related to their marketplaces.

- Dynamic business requirements

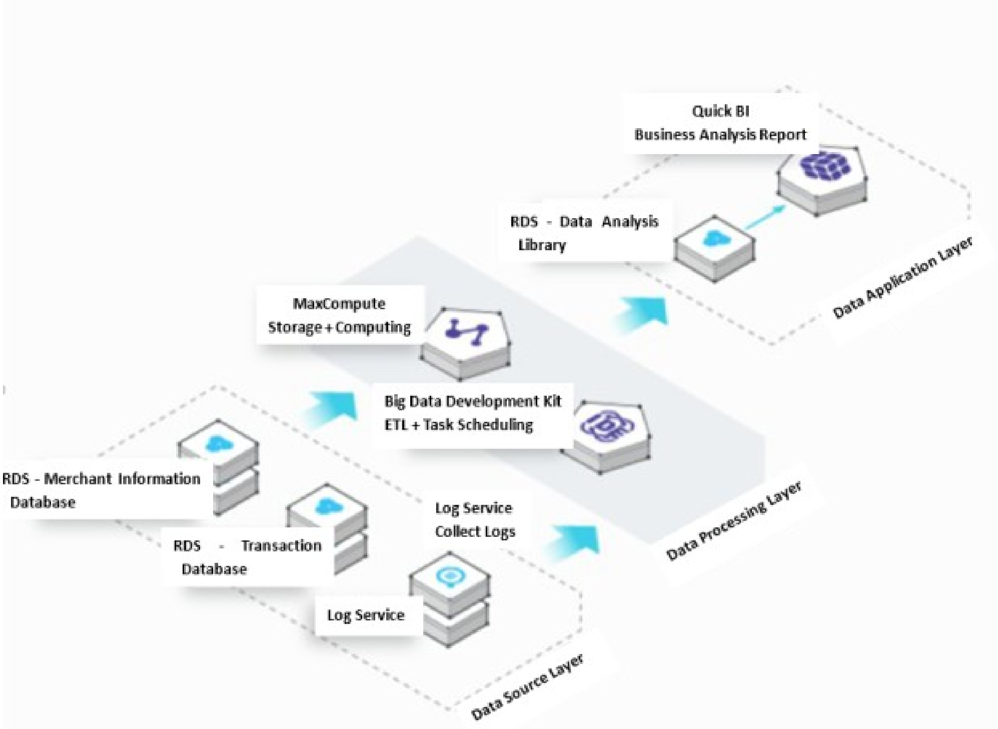
  Quick BI responds quickly to frequent changes in statistical indicators as business grows.

- Consistent computing performance across multiple data sources

  Quick BI leverages the BI capabilities of Alibaba Cloud to resolve the issues arising from cross-source data analysis and the computing performance bottleneck.

Recommended combination: Log Service, relational database, Quick BI, and MaxCompute

Permission control of transaction data


Quick BI
Business Analysis Report
RDS - Data Analysis Library
Data Application Layer
MaxCompute Storage+Computing
Big Data Development Kit ETL+Task Scheduling
Data Processing Layer
RDS - Merchant Information Database
RDS - Transaction Database
Log Service Collect Logs
Log Service
Data Source Layer

37.6. Limits

None.

37.7. Terms

This topic describes commonly used terms in Quick BI.

Data source

When you use Quick BI for data analysis, you must first specify the source of raw data. A data source is where data is stored. You can add data sources by using either of the following methods:

- Add data sources from cloud databases.
- Add data sources from user-created databases.

Dataset

You can use tables from data sources to create datasets. You can edit, move, or delete a dataset from the dataset list.

Dashboard

A dashboard adopts a flexible tile layout to allow you to create interactive reports. Moreover, dashboards provide data filtering and data query functions, and display data in multiple charts.

You can drag and drop fields or double-click the fields to add them to the fields of charts so that you can clearly view the data. Dashboards provide user-friendly interfaces, improving user experience.

Workbook

Workbooks display analyzed and processed data in a dataset. You can use workbooks in both personal and group workspaces. To analyze data in a workbook, you can select the dataset where the data is located and perform required operations.

BI portal

A BI portal is a set of dashboards organized in the form of menus. You can build a business analysis system by using BI portals. A BI portal can reference analysis results in Quick BI and external links.

38.Graph Analytics

38.1. What is Graph Analytics?

Graph Analytics is a visual analysis platform for relationship networks. Graph Analytics is widely used in Alibaba Group and Ant Financial for risk control including anti-fraud, anti-theft, and anti-money laundering solutions. Graph Analytics provides solutions for multiple industries, including public security protection, taxation, customs, banking, insurance, and the Internet.

Graph Analytics is designed to facilitate multi-source data integration, computing applications, visual analytics, and intelligent businesses. Based on relationship networks, Graph Analytics can visualize the properties of objects and reveal the relationship among objects.

Graph Analytics provides features including relationship networks, search networks, intelligent networks, information cubes, intelligent judgement, collaboration and sharing, and dynamic modeling. It visualizes data and integrates machine computing capabilities with human cognition. This allows you to gain insight into massive data and obtain information and knowledge directly and efficiently.

38.2. Benefits

This topic describes the features and technological advantages of Graph Analytics.

Performs massive data mining in real time

Graph Analytics can handle petabytes of data, tens of billions of nodes, hundreds of billions of edges, and trillions of records. Graph Analytics performs relationship mining and computing based on time and space metrics, and supports interactive responses in real time.

Understands the connectivity of things using the OLEP model

Graph Analytics uses the OLEP model to analyze objects, links, and real-world events, and integrates heterogeneous data based on their properties. As the foundation of Graph Analytics, the OLEP model is the key to connecting correlated links and objects.

Flexible business scenarios

Based on the OLEP model, Graph Analytics provides suitable business configurations and detection features to enable human-machine interaction. Applicable scenarios include public security protection, anti-fraud solutions, financing, and taxing.

Efficient visual analyses

Graph Analytics works on key issues to be improved in user experience and pain points in the data analysis business. Based on the analysis results, Graph Analytics provides iterative, visual analysis, and collaborative analysis services for users to build traceable links and paths among objects.

User-friendly intelligence

Graph Analytics helps business users analyze, scrutinize, and handle challenges in an accurate and intelligent manner. Graph Analytics provides deep training models, including the intimacy degree model, terror degree model, and the drug involvement model.

Robust analysis systems

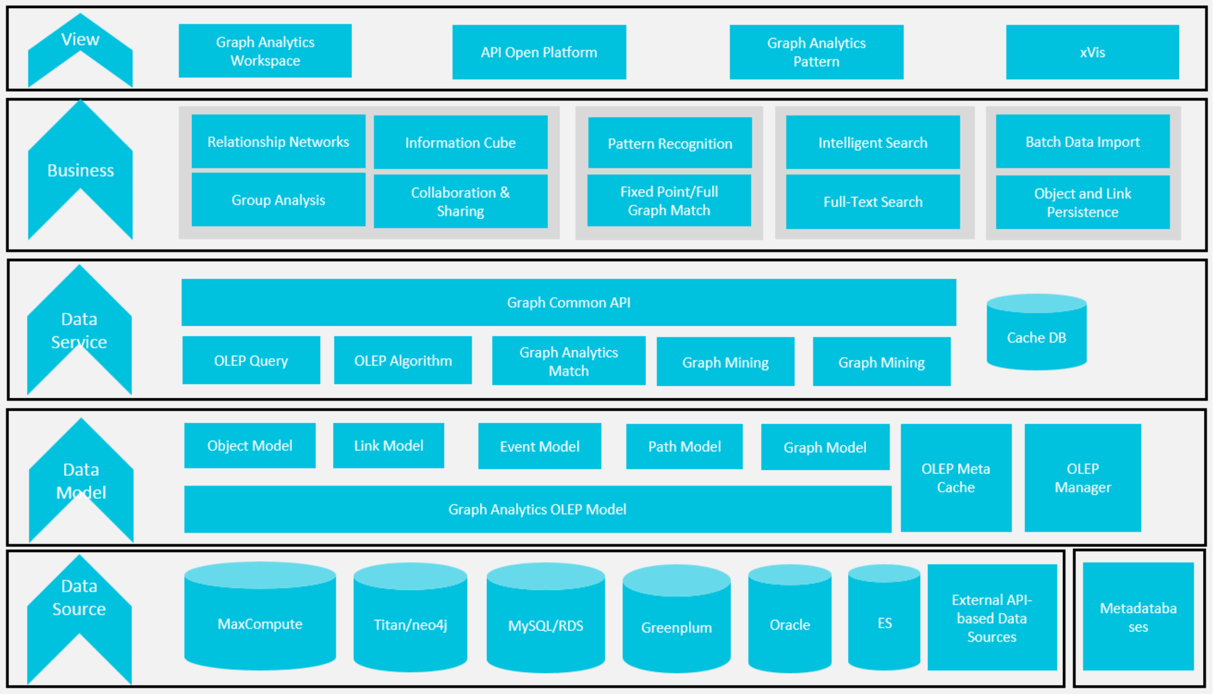
Tested by multiple key national projects, Graph Analytics is considered an important product and has impressed customers with its application in public security protection, anti-terror, and tariff services.

38.3. Architecture

38.3.1. System architecture

This topic describes the system architecture of Graph Analytics.

Graph Analytics provides multiple components and a multi-layer architecture, including the data source layer, data model layer, data service layer, business layer, and the view layer.



Data source layer

Based on the Alibaba Cloud Big Data platform, the data source layer can store and handle petabytes or exabytes of data. It provides powerful data integration, processing, analysis, and computing capabilities. The data source layer provides the following features:

- Supports open source graph databases, such as Titan and Neo4j.
- Supports open source relational databases, such as MySQL, RDS, Oracle, and Greenplum.
- Supports NoSQL databases, including Elasticsearch and KV HBase, a database where each row is a key/value pair.
- Supports external API-based data sources.
- Supports the integration, processing, and online calculation of data from multiple sources.

Data model layer

The data model layer supports the following features:

- Established based on ontological theories, the OLEP model studies the objects, relationships between natural objects, relationships between social objects, and event information.
- Various types of data are converted into nodes and links in the graph. Based on these nodes and

links, Graph Analytics builds paths and graph models to lay the foundation for a subgraph model, providing a standardized data model for data mining and graph algorithm calculation.

Data service layer

The data service layer provides link queries, relationship mining, and graph algorithms for you to analyze relationship networks. This layer supports pattern recognition and extracts graph structure data that is matched with the user-defined graph pattern.

Business layer

The business layer supports the following features:

- Graph Analytics provides an API to call application components at the analysis layer. These application components include relationship networks, search networks, information cubes, intelligent judgement, collaboration and sharing, and dynamic modeling.
- Supports intelligent networks, including pattern definition and pattern matching features.

View layer

The view layer refers to the Web layer of Graph Analytics. This layer displays the entire graph, and its features are as follows:

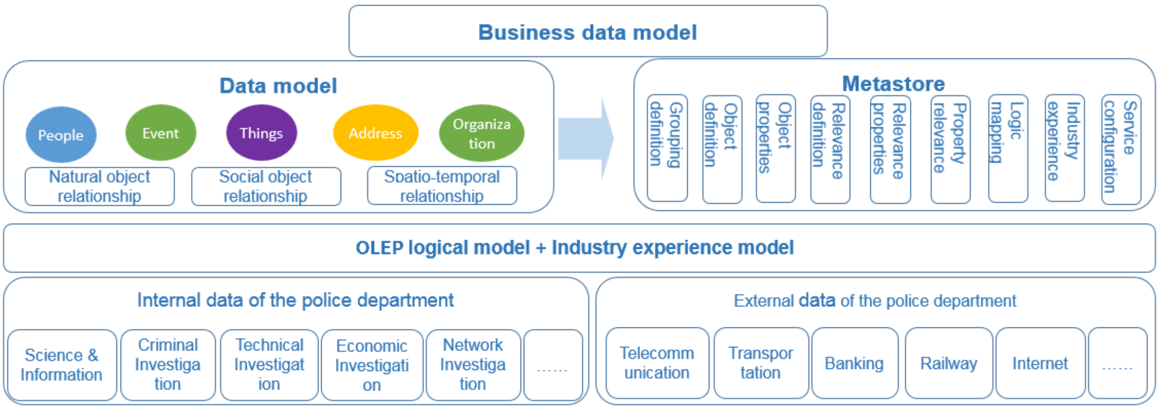
- Supports multiple layouts of relationship networks to fit with different business scenarios.
- Graph Analytics provides a diversified, visual, and interactive analysis interface and supports various terminals.
- Graph Analytics provides visual components and external APIs and supports third-party system integration.

38.3.2. OLEP model

OLEP model analyzes objects, links, and real-world events, and integrates heterogeneous data based on its properties. As the foundation of Graph Analytics, the OLEP model is the key to connecting correlated links and objects and building elaborate relationship networks.

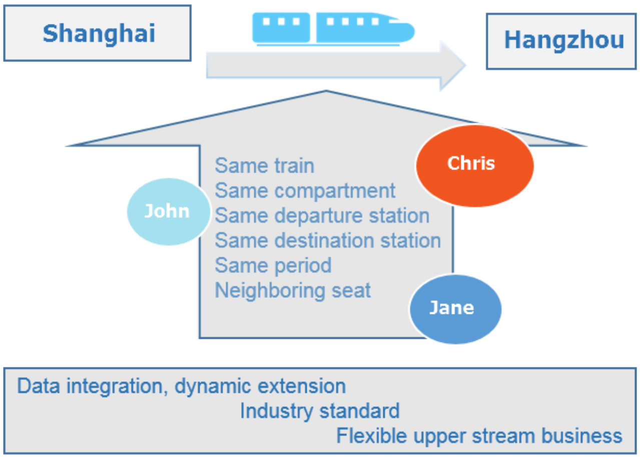
OLEP model structure

The public security industry uses both data within the security industry and external security data. In this scenario, Graph Analytics can leverage the physical data to build OLEP models and industry models, and map elements in these models to metadata definitions, including object definitions, object properties, link definitions, and link properties.



Example of high-speed rail OLEP model

Three people, John, Jane, and Chris are taking a high-speed train from Shanghai to Hangzhou. Using the OLEP model, you can analyze the travel data and determine whether they are on the same train or in the same carriage. You can also tell whether they are from the same source station or heading to the same destination.



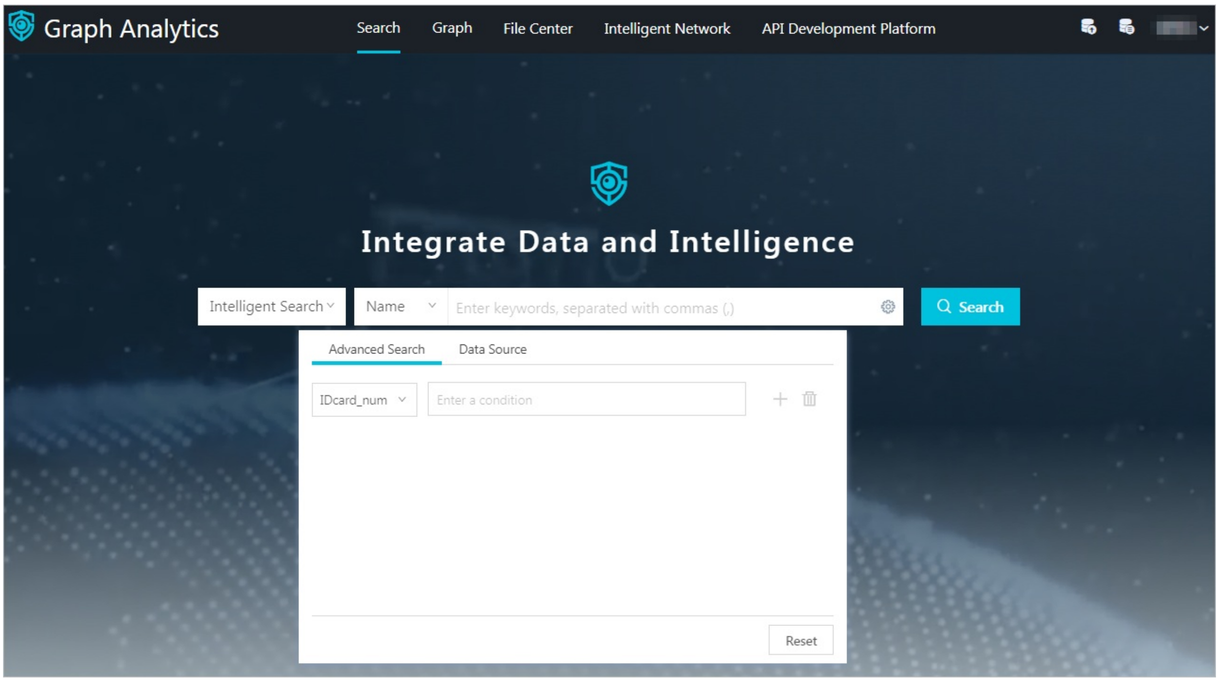
38.4. Features

38.4.1. Search module

This topic introduces the concepts, types, configurations of the Search module, and the relationship between this module and Graph.

Overview

As one of the two independent modules of Graph Analytics, the search module can help analysts quickly locate and view specific objects. The Search module is also the entrance of the relationship network, as it can introduce the retrieved object information into Graph for extended analyses.



Search types

In Graph Analytics, you can perform simple searches and advanced searches:

- Simple Search: You can use this feature to quickly search for objects that contain a certain type of keywords. Fuzzy search is supported. When you perform a simple search, you only need to select a keyword type and enter one or more keywords.
- Advanced Search: Supports fuzzy search and multiple search conditions. You can specify the search terms in Advanced Search in the same way you perform a simple search. You can specify the advanced correlated items for the selected search terms. This is similar to a combined search based on multiple keyword types. You can also specify the data source items to be searched, which is similar to specifying the search range.

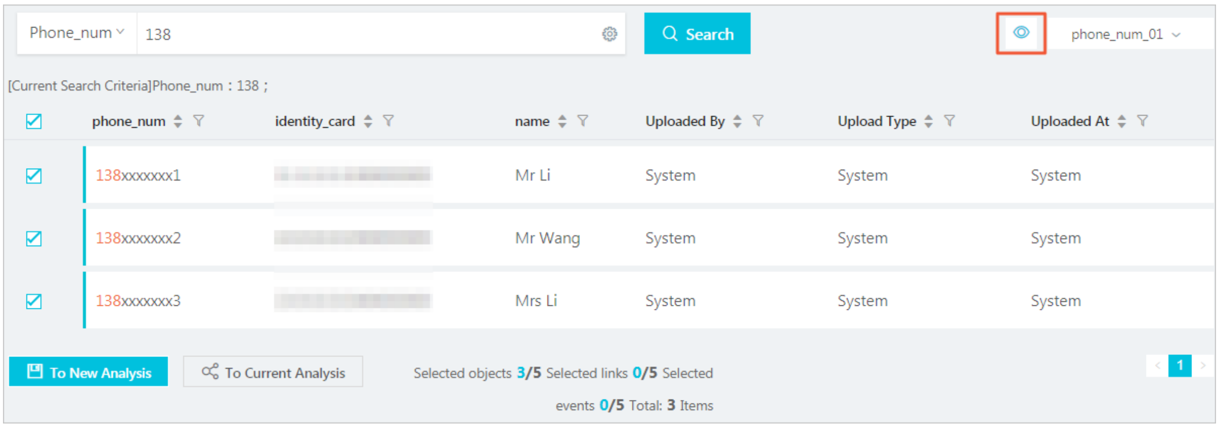
Search Configurations

Before you use the Search module in Graph Analytics, you must configure the search items and related items in advance:

- Simple Search: You must configure the search items in advance.
- Advanced Search: You must configure the search items and related items in advance.

View the search results and send the specified objects to the graph for analysis.

You can view the search results after the search is completed.



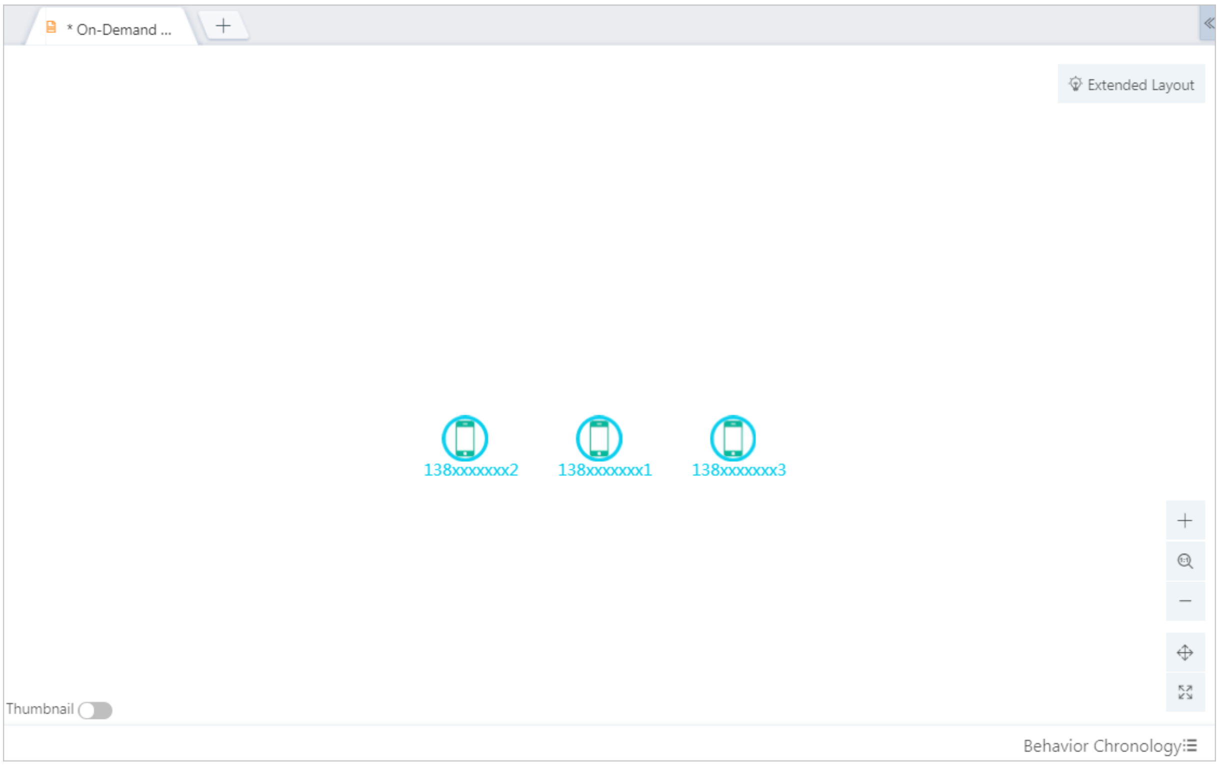
Phone_num 138 Search phone_num_01

[Current Search Criteria]Phone_num : 138 ;

☑	phone_num	identity_card	name	Uploaded By	Upload Type	Uploaded At
☑	138xxxxxxx1	[redacted]	Mr Li	System	System	System
☑	138xxxxxxx2	[redacted]	Mr Wang	System	System	System
☑	138xxxxxxx3	[redacted]	Mrs Li	System	System	System

To New Analysis To Current Analysis Selected objects 3/5 Selected links 0/5 Selected events 0/5 Total: 3 Items 1

Select an object in the search results and click **To New Analysis** or **To Current Analysis** to send the selected search results to the graph to perform an analysis.



38.4.2. Graph

This topic introduces the concepts and supported features of Graph.

Overview

Graph Analytics provides multiple analysis methods for you to easily obtain useful intelligence from complex networks. The features of Graph Analytics include link lookup, group analysis, common neighbor analysis, backbone analysis, lineage analysis, information cube, group statistics, label statistics, collaboration, and sharing.

Link extension

Link extension allows you to perform unlimited link extensions starting from any single object or a group of objects. This helps to achieve unlimited information association. The key to intelligence analysis is to discover related clues and intelligence from a large amount of unrelated information and convert the information into useful and actionable intelligence. Link extensions can be simple or advanced.

Group analysis

Analyzes the direct and indirect relationships between a group of objects of the same type or of different types.

Common neighbor analysis

Analyzes the objects that are commonly associated with two groups of objects, including groups of objects of the same type or of different types.

Path analysis

Analyzes the link path between two objects.

Backbone analysis

Locates the core backbone nodes in a group network using smart algorithms.

Lineage analysis

Displays the lineage relationship among people based on families (family IDs).

Information cube

- Behavior analysis

  Displays the frequency of an event in a chronological order.
- Chronology analysis

  Displays the details of each event in a chronological order.
- Behavior details

  Displays the details of events. The original data records are filtered according to specific rules.
- Object information

  Aggregates objects in a relationship network and classifies the objects by type.
- Statistics information

  Analyzes the relationships and objects in a relationship network, including object properties, link properties, and the distribution of objects.

Group statistics

Group statistics is used to measure the group distribution in Graph Analytics. A group is a group of object nodes. A group consists of multiple object nodes, with any two object nodes connected topologically. Nodes within a merged node are connected topologically.

Label statistics

Collates the label information of object nodes in a relationship network. Graph Analytics supports system labels and user labels. System labels, such as whitelists and blacklists, are defined by the service system for nodes. User labels are defined by each Graph Analytics user for nodes.

Graph layouts

Graph Analytics supports multiple layouts, including matrix layouts, ring layouts, horizontal layouts, vertical layouts, force-directed layouts, and hierarchical layouts.

Right-click operations

The information on the Graph page includes objects, links, events, and graphs in the mapped network. Objects (nodes) and links (edges) are the core elements. On the Graph page, all the analyses are based on the nodes and edges in the graph. The right-click operations focus on the main features of Graph Analytics.

Collaboration and sharing

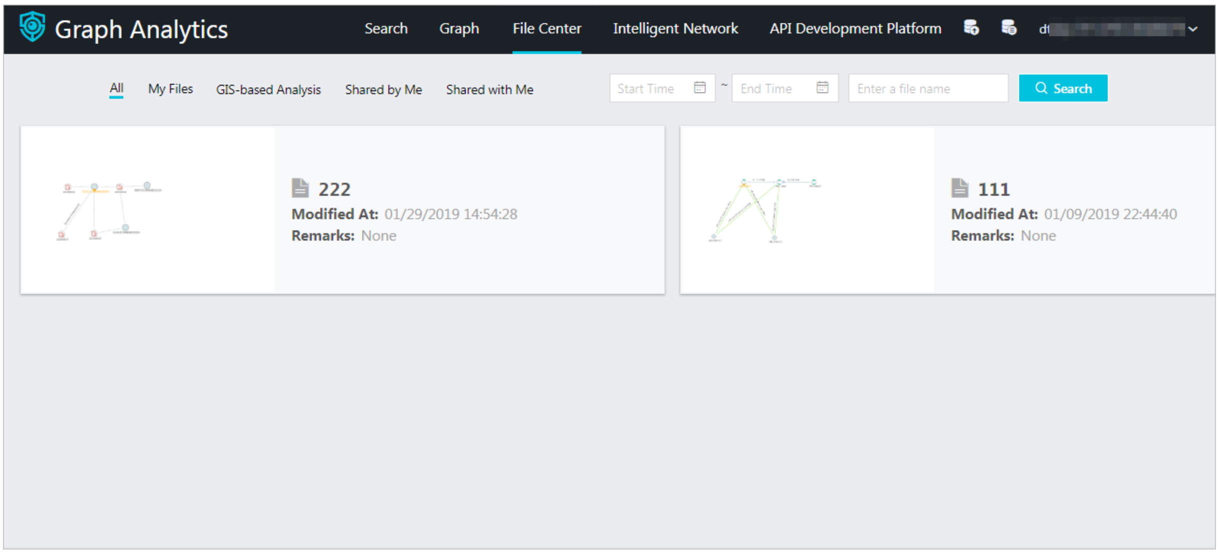
Collaboration and sharing is a new analysis mode provided by Graph Analytics. It allows you to share your analysis files with other users and perform a collaborative analysis. You can use this mode to pass on your ideas and experience to other users, and integrate others' experience and discoveries to achieve team collaboration.

38.4.3. File Center

This topic introduces the concepts and features of the Graph Analytics File Center.

Overview

File Center is the research and judgment workspace of Graph Analytics. It manages all analysis files related to the current user. You can view analysis files in **All**, **My Files**, **Shared by Me**, and **Shared with Me** pages.



222
Modified At: 01/29/2019 14:54:28
Remarks: None

111
Modified At: 01/09/2019 22:44:40
Remarks: None

All

This page displays analysis files related to the current user in the order of creation time, including personal files and shared items received by the current user. Personal files can be divided into unshared and shared files. Unshared files are created by the current user but not shared with any other users. Shared files are created by the current user and shared with other users. You can view the shared files in **Shared by Me**.

On the **All** page, you can perform the following operations on each analysis:

- My files (unshared): You can delete or rename the analysis files, and open, edit, and save the analysis files on the Graph page.
- My files (shared): You can delete the analysis files and open, edit, save, and publish analysis files on the Graph page. If an analysis file is saved on the Graph page and has not been published, a draft version is generated.
- Shared with Me: You can open, edit, save, and publish analysis files on the Graph page. If an analysis file is saved on the Graph page and has not been published, a draft version is generated.

My files

You can view all your personal directories and personal analysis files in the order of creation time. On the **My Files** page, you can add, delete, edit, and perform other operations on the catalogs and analysis files.

- Personal directories: You can create, rename, and delete personal directories.
- Personal files: You can rename, move, share, and delete personal files, and open, edit, and save analysis files on the Graph page.

Shared by Me

The Shared by Me page displays all the files shared by the current user in the order of time when the files were created. After you share an analysis file, the system automatically creates a directory with the same name as the source analysis on the **Shared by Me** page. By default, the directory has two files: the **initial file** and **automatically merged file**.

On the **Shared by Me** page, you can perform the following operations on the analysis files.

- Delete an analysis file.
- Modify sharing permissions.
- Merge multiple versions of the analysis file.
- On the Graph page, you can open, edit, save, and publish a version of the specified analysis file. If an analysis file is saved on the Graph page and has not been published, a draft version is generated.
- You can delete a version of the specified analysis file.

Shared with Me

The analysis files shared by other users are displayed in the order of creation time. After a member receives a shared analysis, the system automatically creates a directory with the same name as the source analysis on the **Shared with Me** page. By default, the directory has two files: the **initial file** and the **automatically merged file**.

On the **Shared with Me** page, you can perform the following operations on the analysis files.

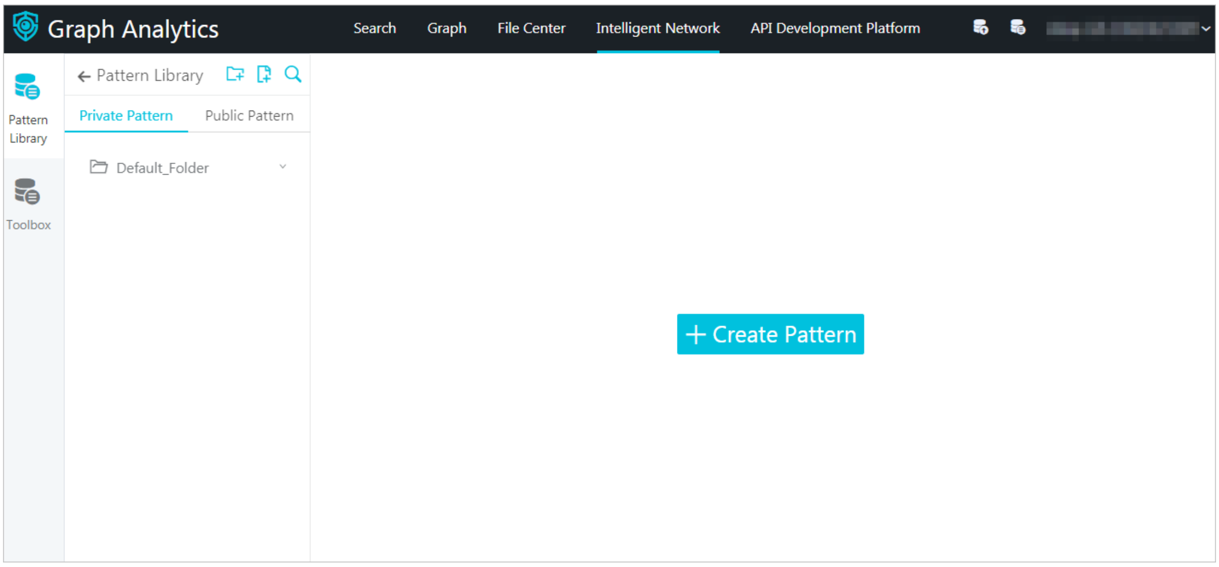
- Open, edit, save, and publish analysis files on the Graph page. If an analysis file is saved on the Graph page and has not been published, a draft version is generated.
- You can delete the draft version of the specified analysis file.

38.4.4. Intelligent network

This topic introduces the concept and functions of intelligent network in Graph Analytics.

Intelligent network overview

In Graph Analytics, you can use an intelligent network in a predefined mode to query subgraph data that has the same graph structure as a specific task. A pattern is the relationship graph structure model that is predefined in Intelligent Network. A task is created based on a pattern. It can be used to query the data with the same graph structure as the task in the data source.



Pattern

A pattern is a relationship graph structure model that is predefined in Intelligent Network. Patterns are divided into private patterns and public patterns.

- Private pattern: Only administrators and creators can use private patterns to create private tasks. Private patterns can be set to public patterns, but this is an irreversible operation.
- Public pattern: All users can use public patterns to create public or private tasks. Public patterns cannot be set to private patterns.

On the **Intelligent Network** page, you can create, view, modify, and delete patterns, and set private patterns to public patterns.

Task

A task is created based on a pattern. It can be used to query the data with the same graph structure as the task in the data source. Tasks are created based on the pattern and used to query data with the same graph structure as the task in the data source. You can modify the graph structure, filter conditions, and other information of the task.

Tasks are divided into private tasks and public tasks.

- Private task: Only administrators and creators can use private tasks. Private tasks created based on public patterns can be set to public tasks, but this is an irreversible operation.
- Public tasks: All users can use public tasks. Public tasks cannot be set to private tasks.

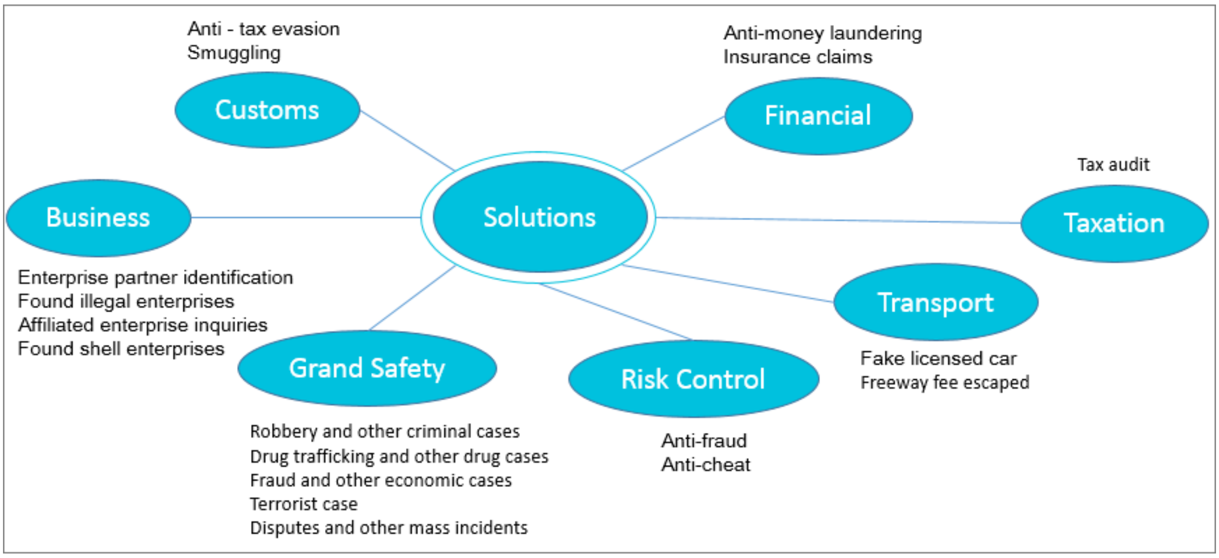
On the **Intelligent Network** page, you can create, view, modify, and delete tasks, and set private tasks to public tasks. After you execute the task, you can also view the execution results in **Graph**.

38.5. Scenarios

38.5.1. Scenario overview

This topic describes the main scenarios for Graph Analytics.

Graph Analytics provides solutions for customs, industry and commerce, transportation, taxation, finance, risk control, and security industries.

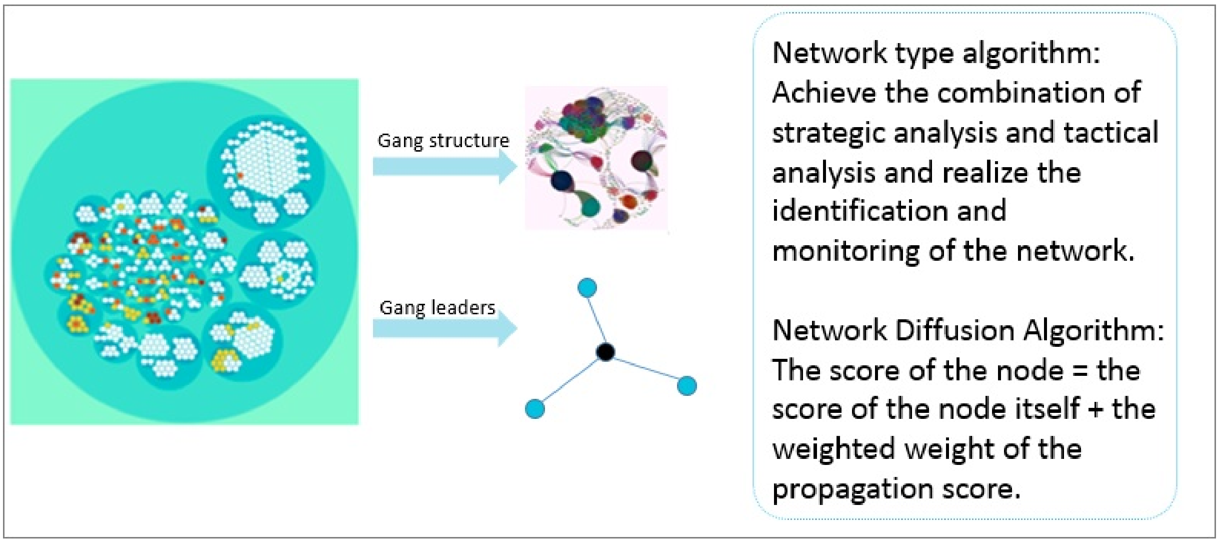


38.5.2. Intelligent relationship networks

Graph Analytics provides intelligent relationship networks to help you quickly analyze the relationships among multiple objects. This topic uses group relationship analyses and transfer transactions as examples.

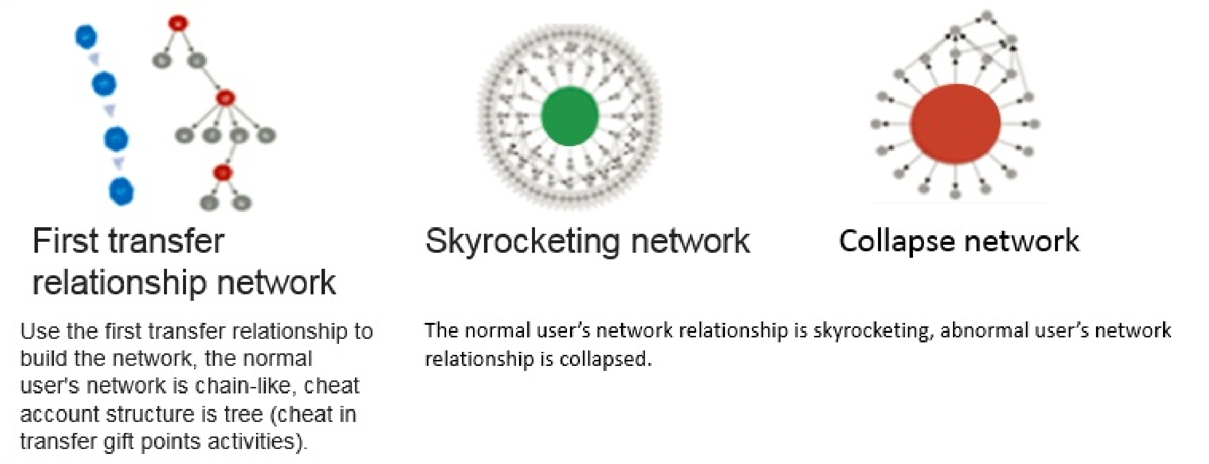
Gang relationship analysis

Graph Analytics can analyze the relationships among gang members, and illustrate the structure of the gang. Graph Analytics can use network topologies to locate key gang members in the relationship network, as shown in the following figure.


Gang structure
Gang leaders
Network type algorithm: Achieve the combination of strategic analysis and tactical analysis and realize the identification and monitoring of the network.
Network Diffusion Algorithm: The score of the node = the score of the node itself + the weighted weight of the propagation score.

Transaction analysis

Graph Analytics can detect potential abnormal transactions by analyzing the transactions between accounts. For example, Graph Analytics can detect market manipulation, as shown in the following figure.



Use the first transfer relationship to build the network, the normal user's network is chain-like, cheat account structure is tree (cheat in transfer gift points activities).

The normal user's network relationship is skyrocketing, abnormal user's network relationship is collapsed.

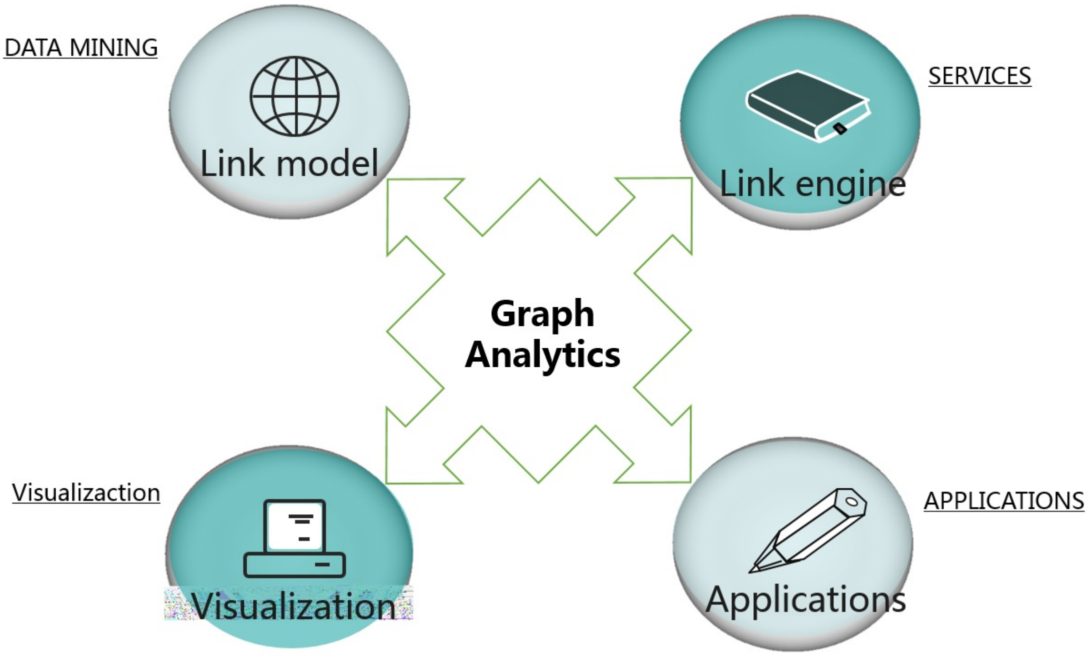
The initial transaction network is generated the first time a transaction relationship is established. In normal cases, the network is linear, but during any market manipulation, the initial transaction network is very intricate. For example, in the case of special offers where users can obtain extra points after completing a transaction, manipulation activities will generate an intricate transaction network. For transactions during market manipulation, you can select the buyer and the seller to build the initial transaction network. In this network, you can analyze the size and growth rate of the network and the proportions of modes.

Upward-trend networks and downward-trend networks: You can start from some of the most heavily funded nodes in the network and move down along the funding path to check the growth trend of the network. The growth trend of a normal network is upward, while the growth of a marketing cheating network is downward and all paths will eventually go to one account.

38.5.3. Industrial risk control

Graph Analytics has been widely used in Alibaba Group and Ant Financial for risk control, such as anti-fraud, anti-theft, and anti-money laundering solutions.

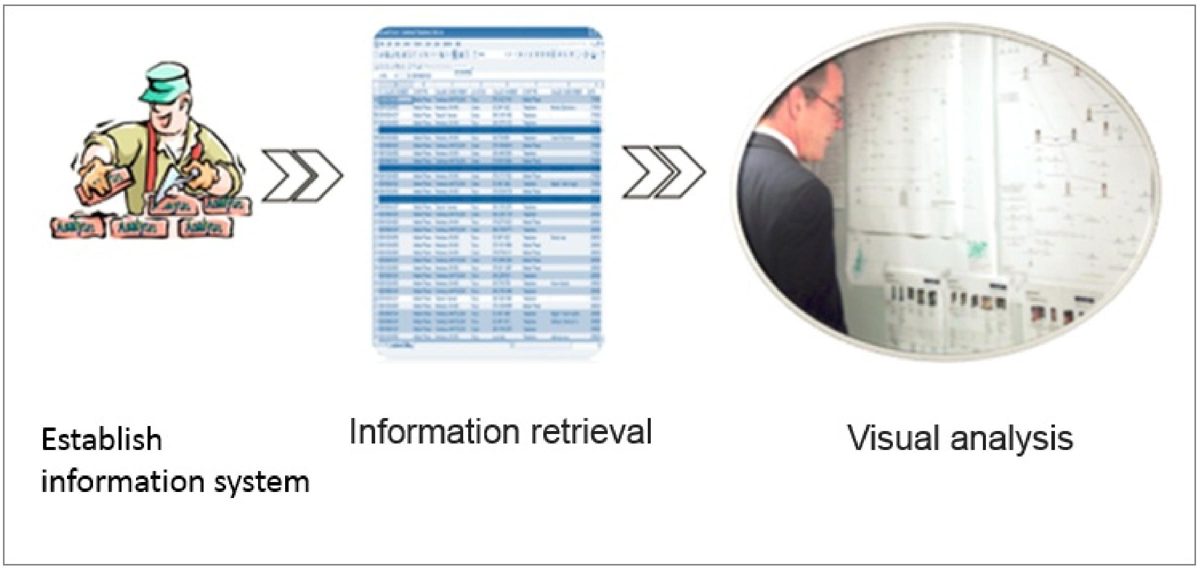
The application of Graph Analytics in industrial risk control is as follows.



- Link model: Graph Analytics creates a link model among humans, accounts, equipment, and the environment. Graph Analytics uses the data mining algorithm to identify the properties of each link, such as the strength, influence, and type of the link. Graph Analytics also identifies the key characters and studies their sub-groups.
- Link engine: Graph Analytics converts relationship data to standardized engine and interface services to benefit more businesses.
- Visualization: Graph Analytics displays the relationships among objects in an intuitive, user-friendly manner.
- Applications: Graph Analytics has gained insights from its application in multiple scenarios, including risk control and relationship network recommendation.

38.5.4. Public security protection

Customers in the public security industry can use Graph Analytics to build their own information systems to query, analyze, and visually display the security information.



38.6. Limits

None.

38.7. Terms

This topic introduces the basic concepts in Graph Analytics.

Object

Object refers to entities and things that exist in the real world. For example, people, mobile phone numbers, and cars. In Graph Analytics, each object needs a primary key as a unique identifier. For example, the primary keys of people, mobile phones, and cars are ID cards, mobile phone numbers, and license plate numbers, respectively.

Link

Link describes the interaction among multiple objects. In Graph Analytics, a link refers to the relationship built among objects. For example, the link between two mobile phone numbers can be phone calls and text messages. The direct link between a person and a mobile phone number can be that the person is the owner of this mobile phone number.

Event

Events are things that have an impact on specific entities. In Graph Analytics, an event refers to the behavior of an object. For example, people choosing to travel by car is an event.

Property

Properties of objects or links. In Graph Analytics, properties cannot be separated from objects or links. For example, properties of a person include height, weight, birthplace, and name. Properties of a mobile phone number include the registration location and the telecommunications operator of this phone number. Primary keys are also properties. For example, an ID card number is one of the properties of a person, and a mobile phone number is one of the properties of a mobile phone.

OLEP data

This module parses data into objects, properties, events, and links between objects to build a highly abstract OLEP model for relationship analysis.

Link lookup

An infinitely extended analysis that begins with any single object or a group of objects. Link lookup helps to build infinite information associations. The key to intelligence analysis is to discover related clues and intelligence from a large amount of unrelated information and convert the information into useful and actionable intelligence. Graph Analytics provides simple link lookup services and advanced link lookup services.

Group analysis

Analyzes the direct and indirect relationships between a group of objects of the same type or of different types.

Common neighbor analysis

Analyzes the objects that are commonly associated with two groups of objects, including groups of objects of the same type or of different types.

Path analysis

Analyzes the link path between two objects.

Backbone analysis

Locates the core backbone nodes in a group network using smart algorithms.

Lineage analysis

Displays the lineage relationship among people based on families (family IDs).

Information cube

- Behavior analysis

  Displays the frequency of an event in a chronological order.
- Chronology analysis

  Displays the details of each event in a chronological order.
- Behavior details

  Displays the details of events. The original data records are filtered according to specific rules.
- Object information

  Aggregates objects in a relationship network and classifies the objects by type.
- Statistics information

  Analyzes the relationships and objects in a relationship network, including object properties, link properties, and the distribution of objects.

Group statistics

Analyzes the distribution of groups in a network. A group consists of multiple object nodes, with any two object nodes connected topologically. Nodes within a merged node are connected topologically.

Label statistics

Collates the label information of object nodes in a relationship network. Graph Analytics supports two types of labels: system labels and user labels. System labels, such as whitelists and blacklists, are defined by the service system for specific nodes. User labels are added to specific nodes by users on the Graph Analytics platform.

Pattern

A pattern is the relationship graph structure model that is predefined in Intelligent Network. Patterns are divided into private patterns and public patterns.

- Private pattern: Only administrators and creators can use private patterns to create private tasks. Private patterns can be set to public patterns, but this is an irreversible operation.
- Public pattern: All users can use public patterns to create public or private tasks. Public patterns cannot be set to private patterns.

Task

Intelligent Network allows you to query subgraphs with the same graph structure as a task specified in a predefined pattern. Tasks are created based on the pattern and used to query data with the same graph structure as the task in the data source. You can modify the graph structure, filter conditions, and other information of the task. Tasks are divided into private tasks and public tasks.

- Private task: Only administrators and creators can use private tasks. Private tasks created based on public patterns can be set to public tasks, but this is an irreversible operation.
- Public tasks: All users can use public tasks. No public tasks can be converted to private tasks.

39.Apsara Big Data Manager (ABM)

39.1. What is Apsara Big Data Manager?

Apsara Big Data Manager (ABM) is an operations and maintenance platform tailored for big data products.

Currently, ABM supports the following products:

- MaxCompute
- DataWorks
- StreamCompute
- Quick BI
- DataHub
- Machine Learning Platform for AI

ABM supports operations and maintenance of big data products from perspectives such as business, services, clusters, and hosts. You can also install patches for big data products, customize alert configurations, and view O&M history through the ABM console.

ABM allows on-site Apsara Stack engineers to manage big data products with ease. For example, they can view performance metrics in real time, modify runtime configurations, and check and handle alerts in a timely manner.

39.2. Benefits

This topic describes the benefits of Apsara Big Data Manager in the following aspects: cluster health monitoring, resource usage analysis, and graphical O&M management.

Cluster health monitoring

Allows you to monitor and configure the devices, resources, and services that are used in the clusters of big data products, and collects performance metrics in real time for dynamic display.

Resource usage analysis

Collects the runtime statuses of cluster devices, resources, and services in real time, and supports data aggregation and analysis to help you evaluate the health status of the cluster. If the evaluation result indicates potential risks in a cluster, responsible engineers would be notified immediately.

Graphical management interface

Provides a graphical user interface for performance metrics visualization and common O&M operations.

39.3. Architecture

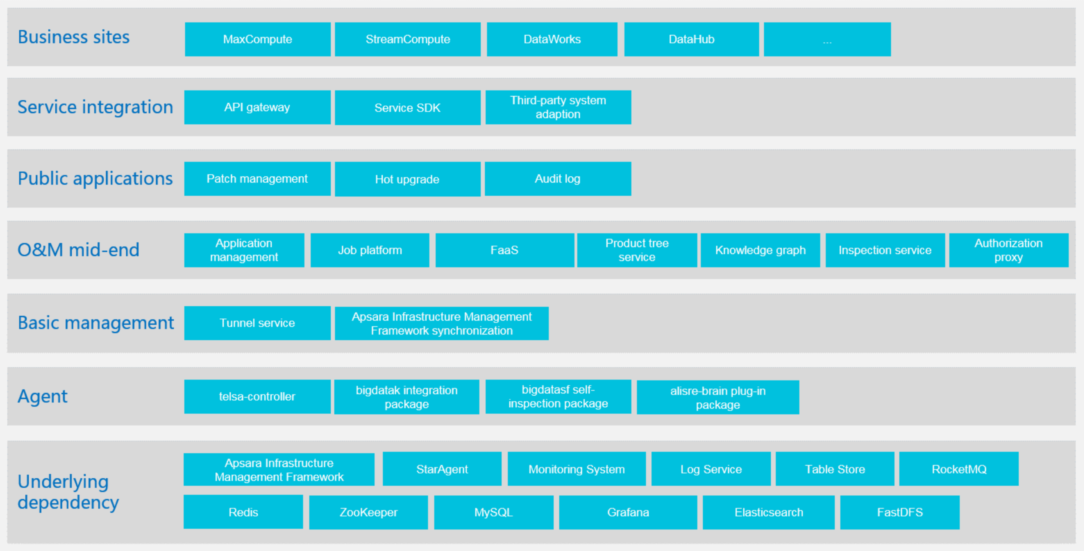
39.3.1. System architecture

This topic describes the system architecture of Apsara Big Data Manager (ABM) and the functions of each component.

ABM uses a microservice architecture that supports data integration, interface integration, and feature integration through a unified platform, and provides standard service interfaces. This architecture enables a consistent user interface, which means that O&M operations are the same for all products. This reduces training costs and lowers O&M risks.

The ABM system consists of the following components: underlying dependency, agent, basic management, O&M mid-end, public applications, service integration, and business sites.

Architecture



Underlying dependency

ABM depends on open source systems from Alibaba and third parties.

- Uses StarAgent and Monitoring System of Alibaba to run remote commands and remote data collection instructions.
- Uses ZooKeeper to coordinate primary and secondary services. This guarantees high availability of services.
- Uses RDS to store metadata, Redis to store cache data, and Table Store to store large amounts of self-test data. This improves service throughput.

Agent

The agent provides client SDKs, scripts, and monitoring packages to be deployed on managed servers.

O&M mid-end and basic management

The O&M mid-end and basic management components form the base of the ABM system. Each service in these two components provides different capabilities for business sites. This enables quick construction of business sites and makes the capabilities of each business site complete.

Public applications

Public applications are developed based on the O&M mid-end and designed with special purposes. These applications are adaptive to all big data products supported by ABM.

Service integration

Service integration links business sites with underlying components. It integrates interfaces of all internal services, adapts to various third-party systems, and provides a unified SDK for users.

Business sites

Business sites are built based on the O&M mid-end and cover all big data products, including MaxCompute, StreamCompute, DataWorks, and DataHub. Each business site provides comprehensive O&M capabilities for one product.

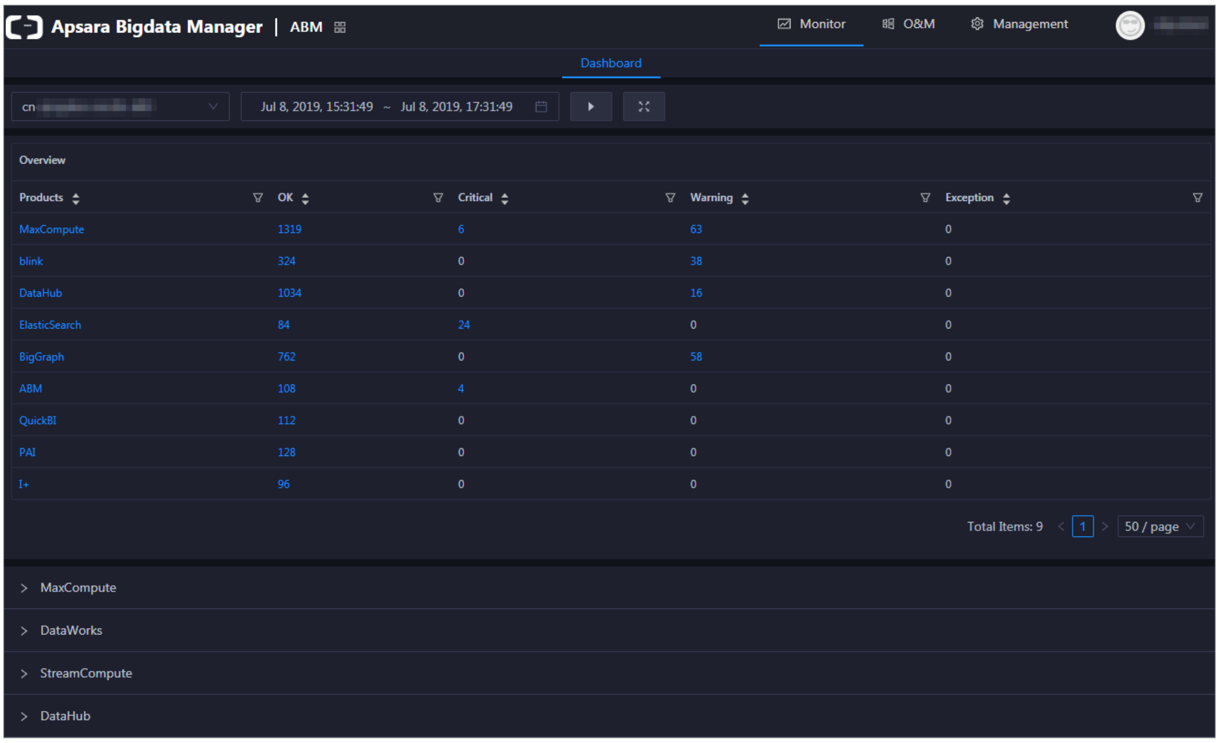
39.4. Features

39.4.1. Dashboard

The Dashboard module is the homepage of the ABM console. It displays key performance metrics for MaxCompute, DataWorks, StreamCompute, and DataHub, and provides alerts for all big data products. This allows you to understand the overall runtime performance of all big data products.

Dashboard page

After you log on to the ABM console, the **Dashboard** page appears by default. To return to the **Dashboard** page from any other page, click ⊞ in the upper-left corner and select **ABM**.


Apsara Bigdata Manager | ABM
Monitor
O&M
Management
[redacted]
Dashboard
cn [redacted]
Jul 8, 2019, 15:31:49 ~ Jul 8, 2019, 17:31:49

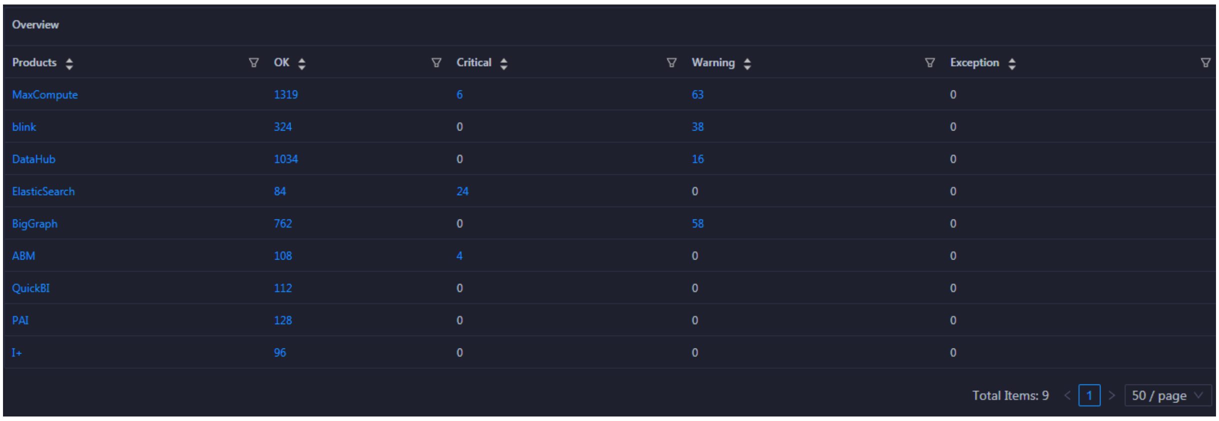
Overview

Products	OK	Critical	Warning	Exception
MaxCompute	1319	6	63	0
blink	324	0	38	0
DataHub	1034	0	16	0
ElasticSearch	84	24	0	0
BigGraph	762	0	58	0
ABM	108	4	0	0
QuickBI	112	0	0	0
PAI	128	0	0	0
I+	96	0	0	0

On the **Dashboard** page, you can select a region from the **Dashboard** drop-down list in the upper-left corner. You can then view the runtime performance of big data products in the selected region.

Alerts overview

In the **Overview** section, you can view the number of alerts about each big data product. You need to pay close attention to **Critical** and **Warning** alerts, which must be cleared in a timely manner.



Overview

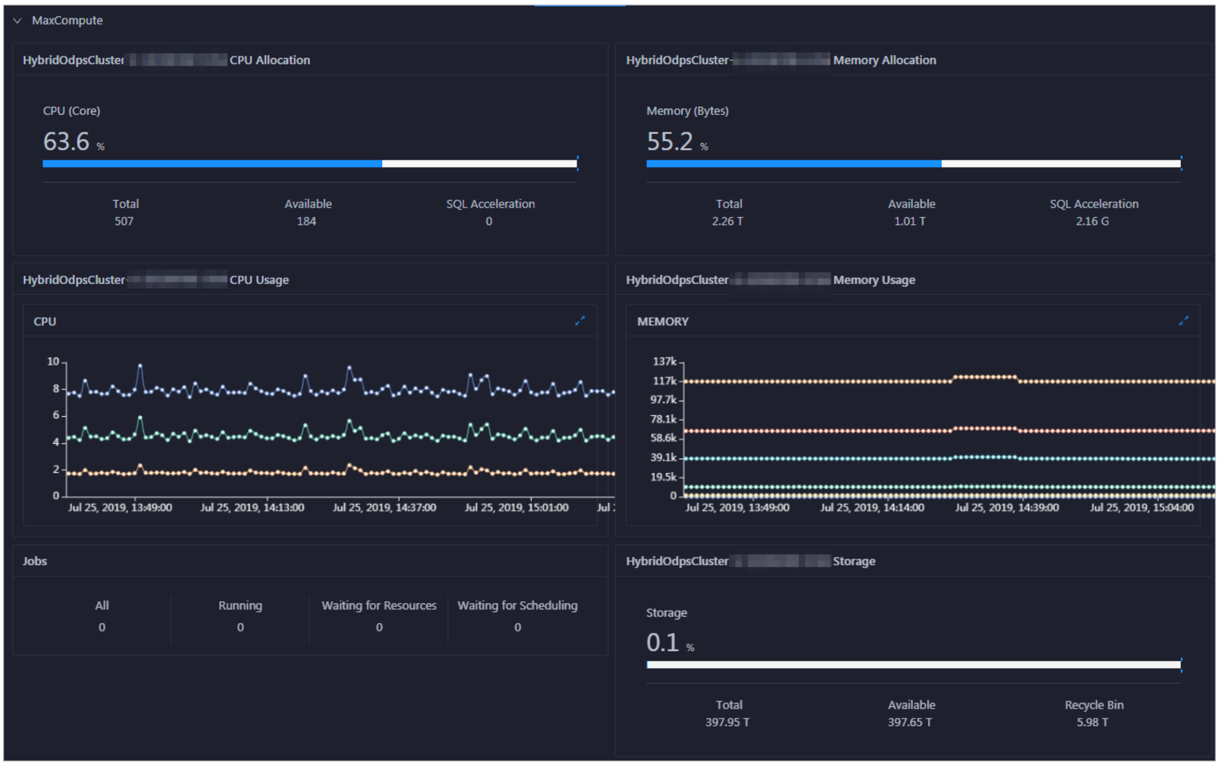
Products	OK	Critical	Warning	Exception
MaxCompute	1319	6	63	0
blink	324	0	38	0
DataHub	1034	0	16	0
ElasticSearch	84	24	0	0
BigGraph	762	0	58	0
ABM	108	4	0	0
QuickBI	112	0	0	0
PAI	128	0	0	0
I+	96	0	0	0

Total Items: 9 < 1 > 50 / page

In the **Overview** section, you can click a product name or alert count to go to the O&M page of the product.

MaxCompute metrics

The Dashboard page displays key performance metrics for MaxCompute. To view these metrics, click **MaxCompute** in the Overview section of the **Dashboard** page.



MaxCompute

HybridOdpsCluster[redacted]CPU Allocation

CPU (Core)
63.6 %

Total 507 | Available 184 | SQL Acceleration 0

HybridOdpsCluster-[redacted]Memory Allocation

Memory (Bytes)
55.2 %

Total 2.26 T | Available 1.01 T | SQL Acceleration 2.16 G

HybridOdpsCluster[redacted]CPU Usage

HybridOdpsCluster[redacted]Memory Usage

Jobs

All 0 | Running 0 | Waiting for Resources 0 | Waiting for Scheduling 0

HybridOdpsCluster[redacted]Storage

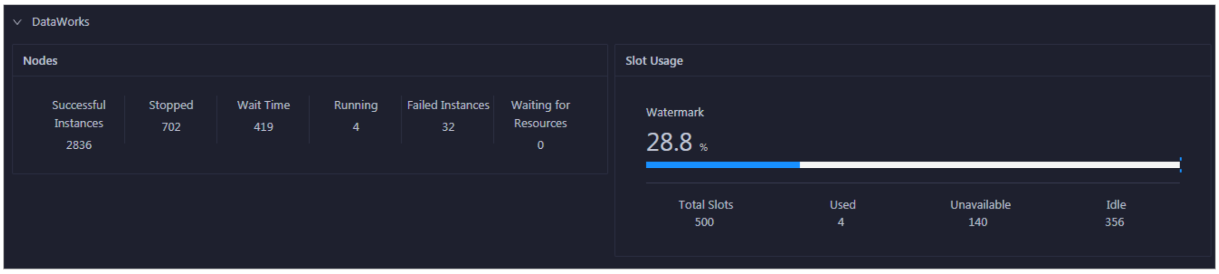
Storage
0.1 %

Total 397.95 T | Available 397.65 T | Recycle Bin 5.98 T

The **MaxCompute** section displays jobs overview, real-time capacity for control system, data traffic, compute resource usage, storage resource usage, and the trends of logical and physical CPU usage.

DataWorks metrics

The Dashboard page displays key performance metrics for DataWorks. To view these metrics, click **DataWorks** in the **Overview** section of the **Dashboard** page.



DataWorks

Nodes

Successful Instances	Stopped	Wait Time	Running	Failed Instances	Waiting for Resources
2836	702	419	4	32	0

Slot Usage

Watermark

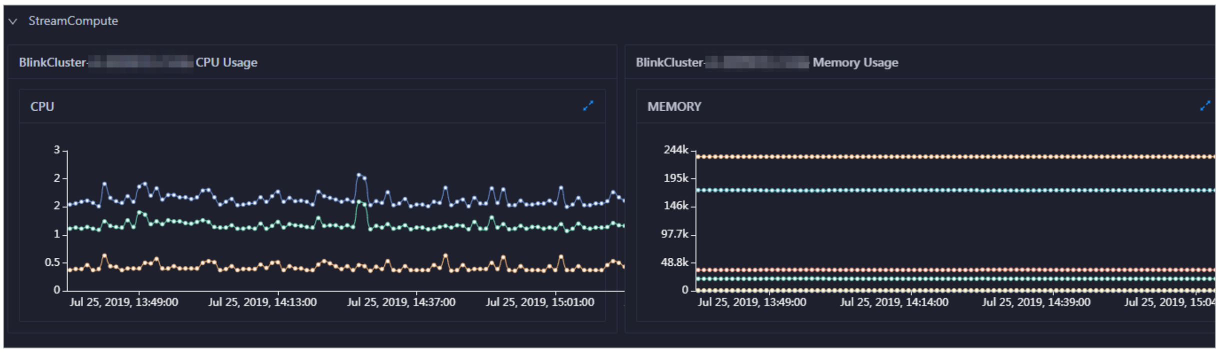
28.8 %

Total Slots	Used	Unavailable	Idle
500	4	140	356

The **DataWorks** section displays nodes overview, slot usage overview, and the trend of finished tasks.

StreamCompute metrics

The Dashboard page displays key performance metrics for StreamCompute. To view these metrics, click **StreamCompute** in the **Overview** section of the **Dashboard** page.



StreamCompute

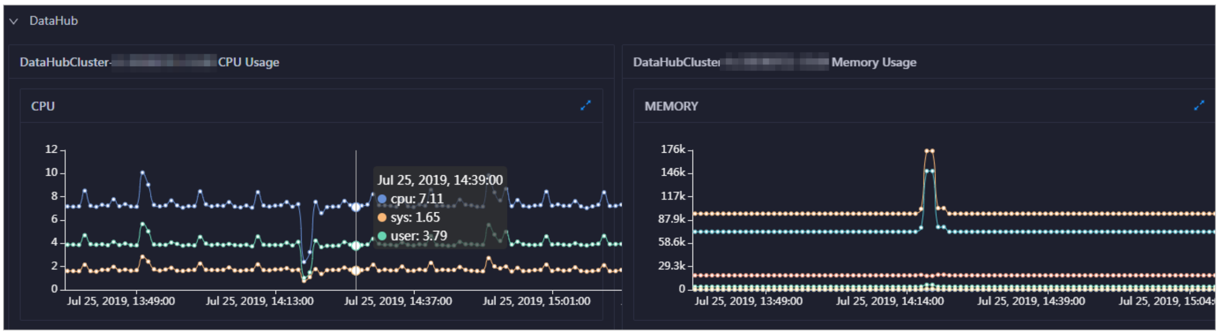
BlinkCluster-[redacted] CPU Usage

BlinkCluster-[redacted] Memory Usage

The **StreamCompute** section displays the following trends of cluster jobs: TPS, failover rate, CPU usage, and memory usage.

DataHub metrics

The Dashboard page displays key performance metrics for DataHub. To view these metrics, click **DataHub** in the **Overview** section of the **Dashboard** page.



DataHub

DataHubCluster-[redacted] CPU Usage

DataHubCluster[redacted] Memory Usage

The **DataHub** section displays read/write latency, the numbers of read/write records, read/write request rates, read/write throughputs, and the trends of CPU and memory usage.

39.4.2. Repository

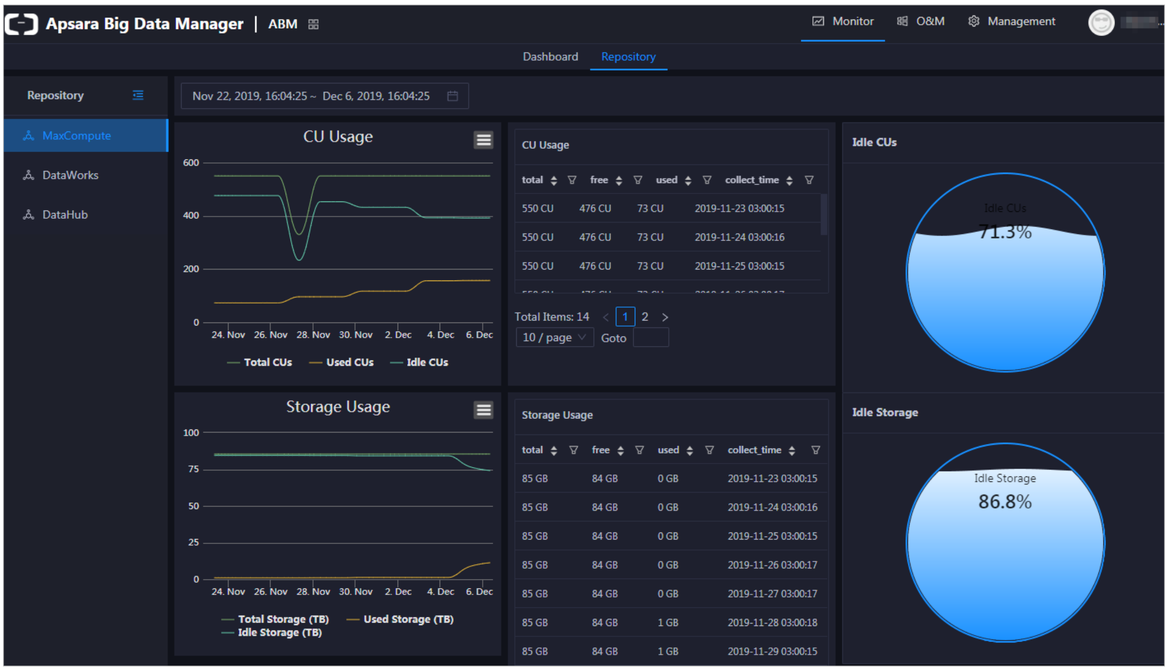
The Repository module displays the resource usage in MaxCompute, DataWorks, and DataHub clusters.

Repository page

1. Log on to the ABM console. The **Dashboard** page appears by default.

> **Note** After you log on to the ABM console, the **Dashboard** page appears by default. To return to the **Dashboard** page from any other page, click ⊞ in the upper-left corner and select **ABM**.

2. On the **Dashboard** page, click the **Repository** tab to go to the **Repository** page.



CU Usage

total	free	used	collect_time
550 CU	476 CU	73 CU	2019-11-23 03:00:15
550 CU	476 CU	73 CU	2019-11-24 03:00:16
550 CU	476 CU	73 CU	2019-11-25 03:00:15

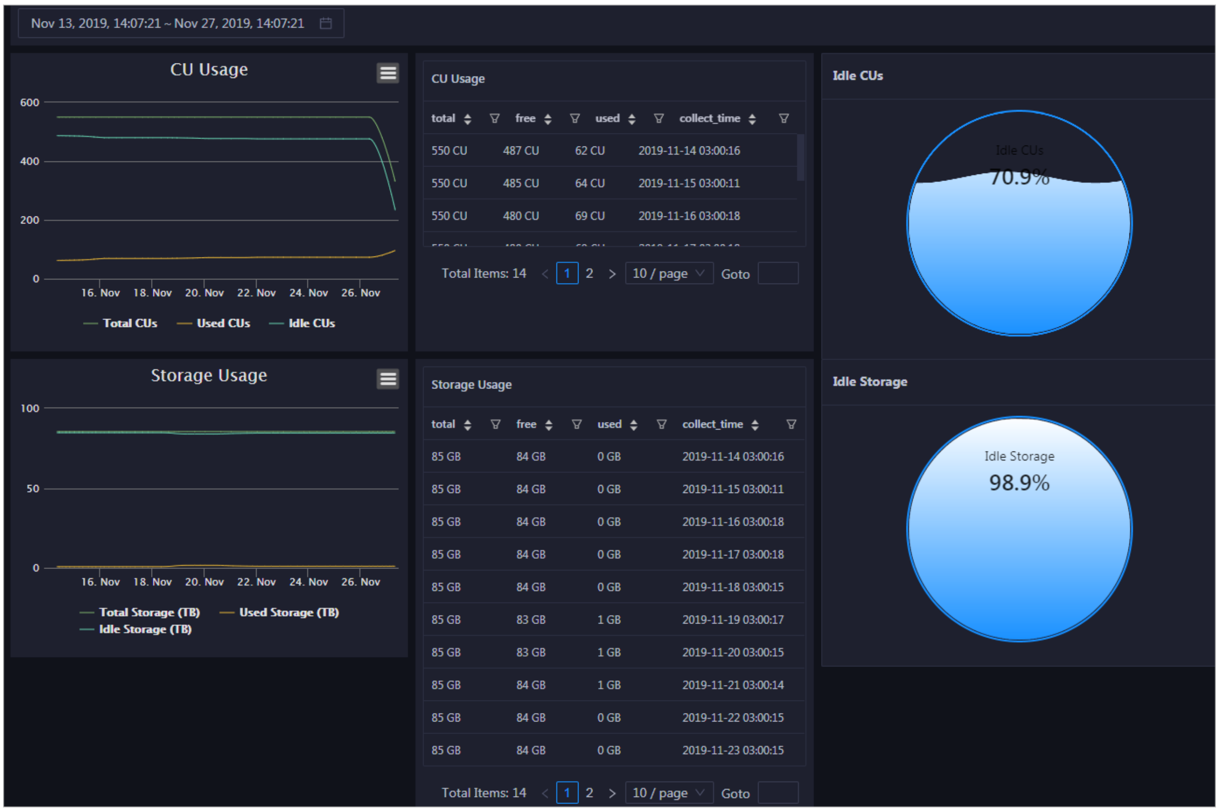
Total Items: 14 < 1 2 >
10 / page Goto

Storage Usage

total	free	used	collect_time
85 GB	84 GB	0 GB	2019-11-23 03:00:15
85 GB	84 GB	0 GB	2019-11-24 03:00:16
85 GB	84 GB	0 GB	2019-11-25 03:00:15
85 GB	84 GB	0 GB	2019-11-26 03:00:17
85 GB	84 GB	0 GB	2019-11-27 03:00:17
85 GB	84 GB	1 GB	2019-11-28 03:00:18
85 GB	84 GB	1 GB	2019-11-29 03:00:15

MaxCompute repository

In the left-side navigation pane of the **Repository** page, click **MaxCompute** to view the resource usage in MaxCompute.



CU Usage

total	free	used	collect_time
550 CU	487 CU	62 CU	2019-11-14 03:00:16
550 CU	485 CU	64 CU	2019-11-15 03:00:11
550 CU	480 CU	69 CU	2019-11-16 03:00:18

Total Items: 14 < 1 2 > 10 / page Goto

Storage Usage

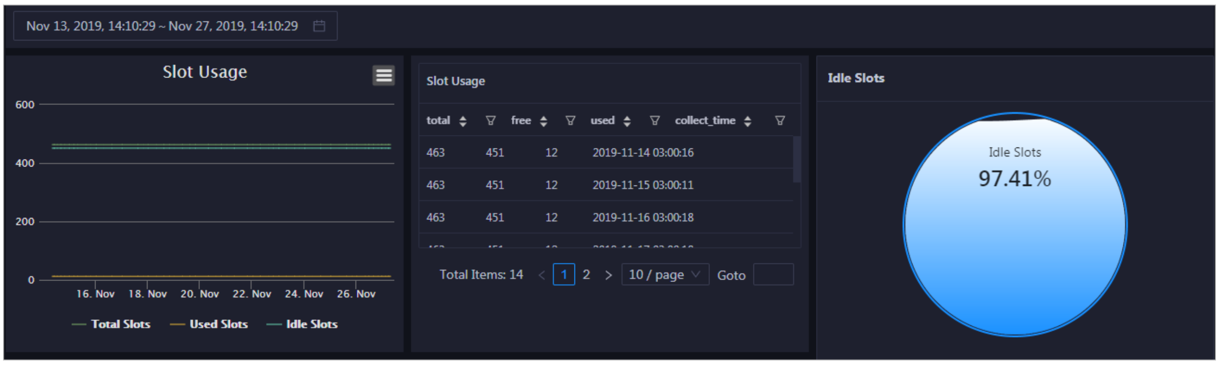
total	free	used	collect_time
85 GB	84 GB	0 GB	2019-11-14 03:00:16
85 GB	84 GB	0 GB	2019-11-15 03:00:11
85 GB	84 GB	0 GB	2019-11-16 03:00:18
85 GB	84 GB	0 GB	2019-11-17 03:00:18
85 GB	84 GB	0 GB	2019-11-18 03:00:15
85 GB	83 GB	1 GB	2019-11-19 03:00:17
85 GB	83 GB	1 GB	2019-11-20 03:00:15
85 GB	84 GB	1 GB	2019-11-21 03:00:14
85 GB	84 GB	0 GB	2019-11-22 03:00:15
85 GB	84 GB	0 GB	2019-11-23 03:00:15

Total Items: 14 < 1 2 > 10 / page Goto

This page displays the trends and details of CU and storage usage, and the percentages of idle CUs and storage.

DataWorks repository

In the left-side navigation pane of the **Repository** page, click **DataWorks** to view the resource usage in DataWorks.



Slot Usage

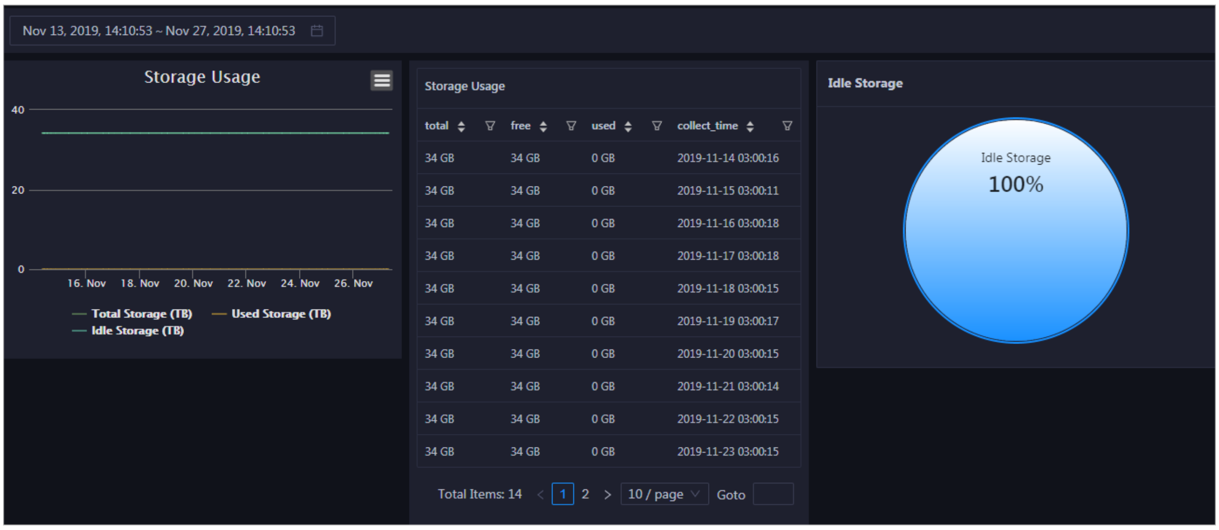
total	free	used	collect_time
463	451	12	2019-11-14 03:00:16
463	451	12	2019-11-15 03:00:11
463	451	12	2019-11-16 03:00:18

Total Items: 14 < 1 2 > 10 / page Goto

This page displays the trend and details of slot usage, and the percent of idle slots.

DataHub repository

In the left-side navigation pane of the **Repository** page, click **DataHub** to view the resource usage in DataHub.



Storage Usage

total	free	used	collect_time
34 GB	34 GB	0 GB	2019-11-14 03:00:16
34 GB	34 GB	0 GB	2019-11-15 03:00:11
34 GB	34 GB	0 GB	2019-11-16 03:00:18
34 GB	34 GB	0 GB	2019-11-17 03:00:18
34 GB	34 GB	0 GB	2019-11-18 03:00:15
34 GB	34 GB	0 GB	2019-11-19 03:00:17
34 GB	34 GB	0 GB	2019-11-20 03:00:15
34 GB	34 GB	0 GB	2019-11-21 03:00:14
34 GB	34 GB	0 GB	2019-11-22 03:00:15
34 GB	34 GB	0 GB	2019-11-23 03:00:15

Total Items: 14 < 1 2 > 10 / page Goto

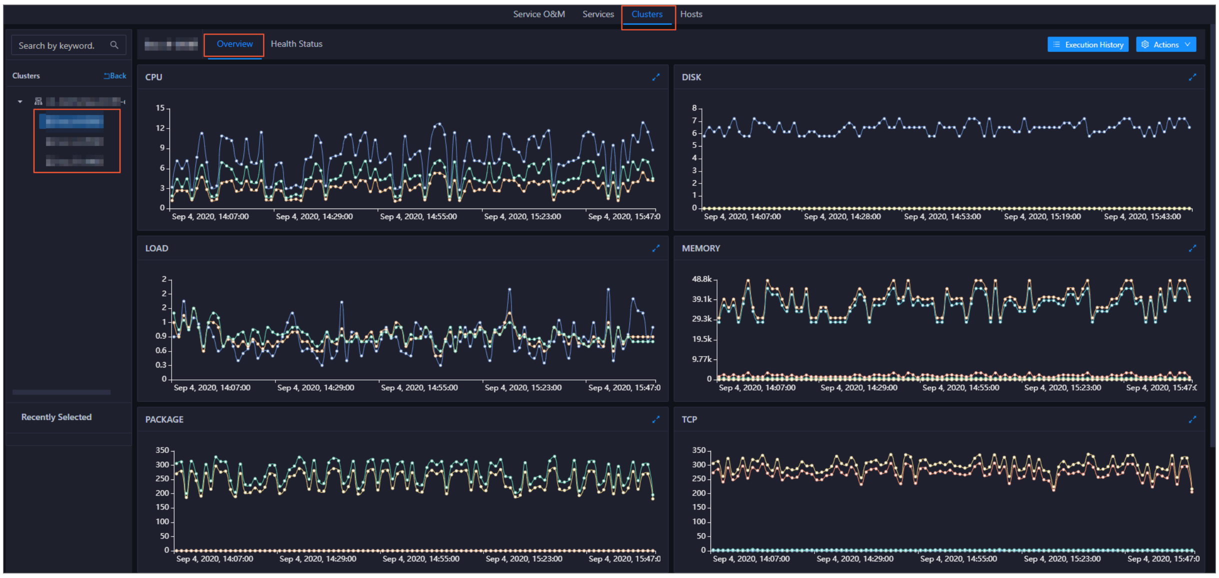
This page displays the trend and details of storage usage, and the percent of idle storage.

39.4.3. O&M

Apsara Big Data Manager (ABM) allows you to perform operations and maintenance (O&M) on big data services from the perspectives of business, clusters, services, and hosts. The big data services include MaxCompute, DataWorks, StreamCompute, and DataHub. It also provides tailored features for some services.

Clusters

The Clusters module is provided for all big data services. ABM provides two major features for cluster O&M: Overview and Health Status.



- Overview: shows the overall running information about a cluster. You can view the host status, service status, health check results, and health check history. You can also view the trend charts of CPU utilization, disk usage, memory usage, load, and packet transmission for the cluster.
- Health Status: shows all checkers for a cluster. You can query checker details and check results for

hosts in the cluster. The status of a checker can be CRITICAL, WARNING, or EXCEPTION.

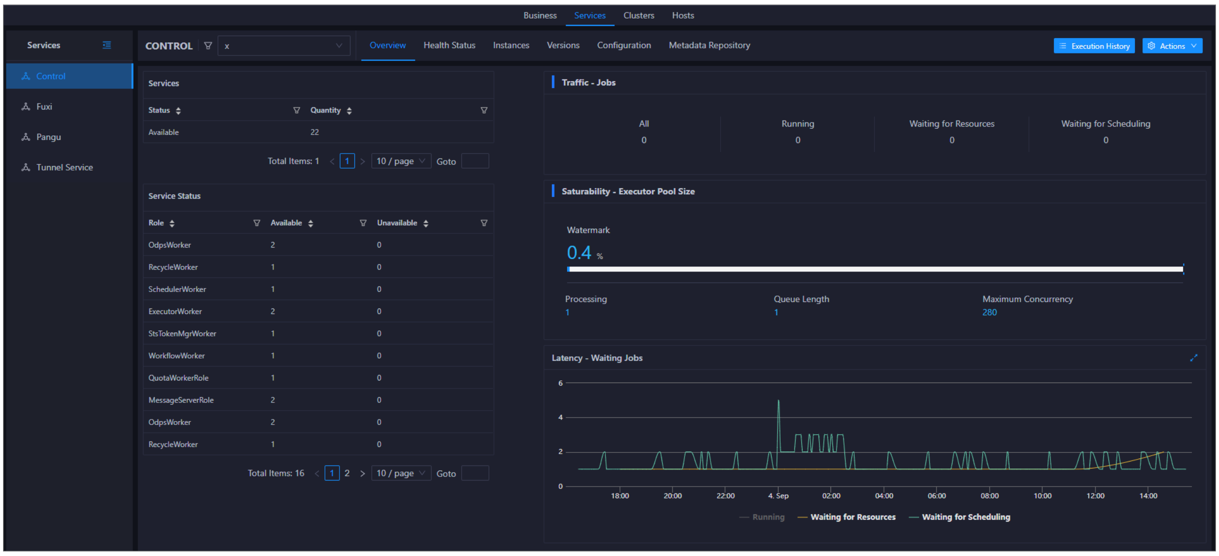
ABM also provides MaxCompute, StreamCompute, and DataHub with the following tailored features for cluster O&M:

- Servers: shows information about hosts in a cluster. You can view the CPU utilization, memory usage, root disk usage, packet loss rate, and packet error rate.
- Scale in Cluster or Scale out Cluster: allows you to scale in or scale out a MaxCompute cluster.
- Reverse Parse Request ID (exclusive for DataHub): allows you to reversely parse a request ID in DataHub to obtain the time that a job is run and the IP address of the host. This helps you query logs for troubleshooting.
- Delete Topic from Smoke Testing (exclusive for DataHub): allows you to delete topics from a DataHub test project and view the execution history.

Services

The Services module is provided for all big data services. ABM also provides MaxCompute, DataWorks, DataHub, and StreamCompute with tailored features for service O&M.

For MaxCompute, ABM supports O&M on the Control service, Job Scheduler, Apsara Distributed File System, and Tunnel service.



Services

Status	Quantity
Available	22

Total Items: 1 < 1 > 10 / page Goto

Service Status

Role	Available	Unavailable
OdpsWorker	2	0
RecycleWorker	1	0
SchedulerWorker	1	0
ExecutorWorker	2	0
StsTokenMgrWorker	1	0
WorkflowWorker	1	0
QuotaWorkerRole	1	0
MessageServerRole	2	0
OdpsWorker	2	0
RecycleWorker	1	0

Total Items: 16 < 1 2 > 10 / page Goto

Traffic - Jobs

All	Running	Waiting for Resources	Waiting for Scheduling
0	0	0	0

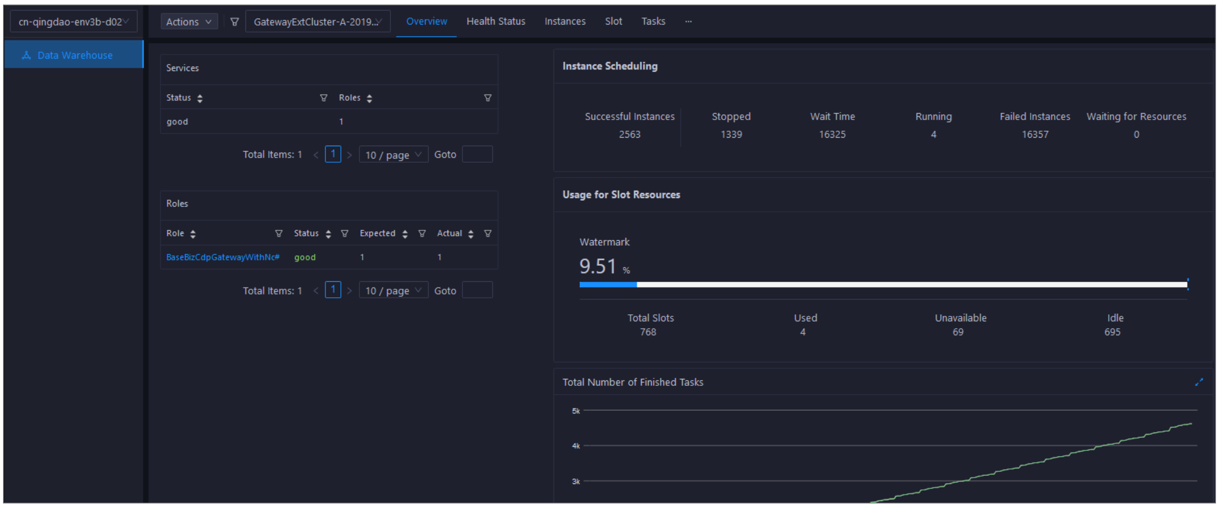
Saturability - Executor Pool Size

Watermark 0.4 %

Processing	Queue Length	Maximum Concurrency
1	1	280

- Control: shows general information about the service, health checkers, and instances. You can configure services for clusters and start or stop server roles.
- Fuxi: shows general information about Job Scheduler, health checkers, and instances. You can manage quota groups, set compute nodes to read-only or read-write, and add or remove compute nodes to or from a blacklist. You can also enable or disable SQL acceleration, and restart master nodes.
- Pangu: shows general information about Apsara Distributed File System, health checkers, and instances. You can set the storage node status to disabled or normal, set the disk status to error or normal, and change the primary master node. You can also clear the recycle bin, enable or disable data rebalancing, and run checkpoints on master nodes.
- Tunnel Service: shows general information about Tunnel and instances. You can also restart the Tunnel server.

For DataWorks, ABM supports O&M on data warehouses. Data Warehouse shows general information about the service, health status, instances, slots, and service configurations. You can also add or remove hosts to scale out or scale in a cluster.



Services

Status	Roles
good	1

Total Items: 1 < 1 > 10 / page Goto

Roles

Role	Status	Expected	Actual
BaseBizCdpGatewayWithNc#	good	1	1

Total Items: 1 < 1 > 10 / page Goto

Instance Scheduling

Successful Instances	Stopped	Wait Time	Running	Failed Instances	Waiting for Resources
2563	1339	16325	4	16357	0

Usage for Slot Resources

Watermark

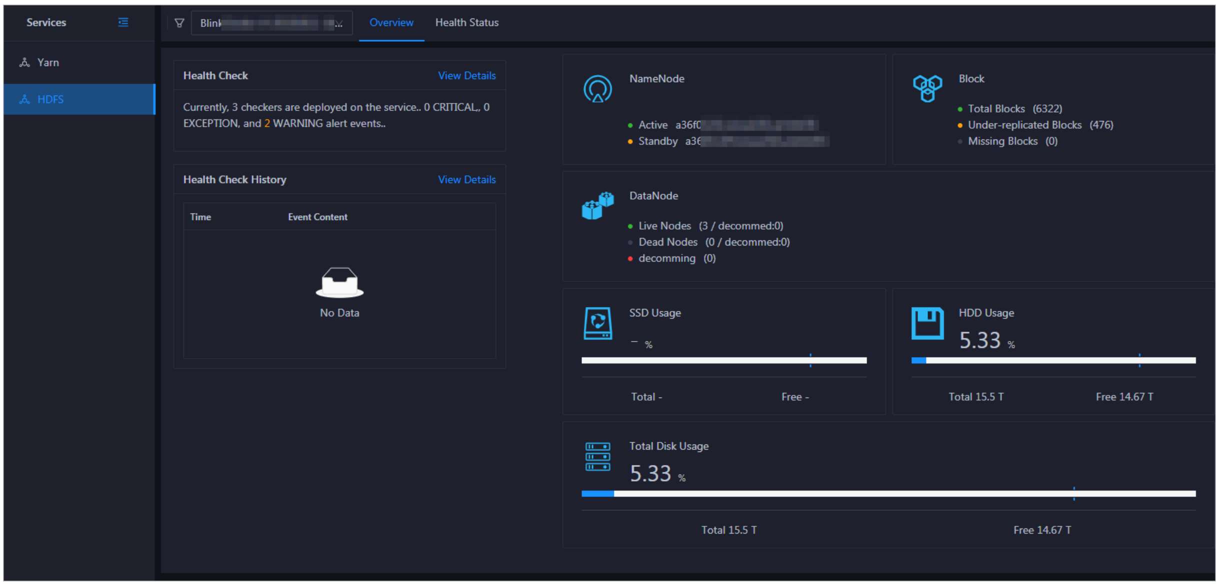
9.51 %

Total Slots	Used	Unavailable	Idle
768	4	69	695

Total Number of Finished Tasks

For DataHub, ABM supports O&M on the Control service, Job Scheduler, and Apsara Distributed File System, which are similar to those for MaxCompute.

For StreamCompute, ABM supports O&M on Blink, YARN and Hadoop Distributed File System (HDFS).



Health Check View Details

Currently, 3 checkers are deployed on the service.. 0 CRITICAL,, 0 EXCEPTION, and 2 WARNING alert events..

Health Check History View Details

Time	Event Content
No Data	

NameNode

- Active a36f0[redacted]
- Standby a36[redacted]

Block

- Total Blocks (6322)
- Under-replicated Blocks (476)
- Missing Blocks (0)

DataNode

- Live Nodes (3 / decommed:0)
- Dead Nodes (0 / decommed:0)
- decomming (0)

SSD Usage

– %

Total - Free -

HDD Usage

5.33 %

Total 15.5 T Free 14.67 T

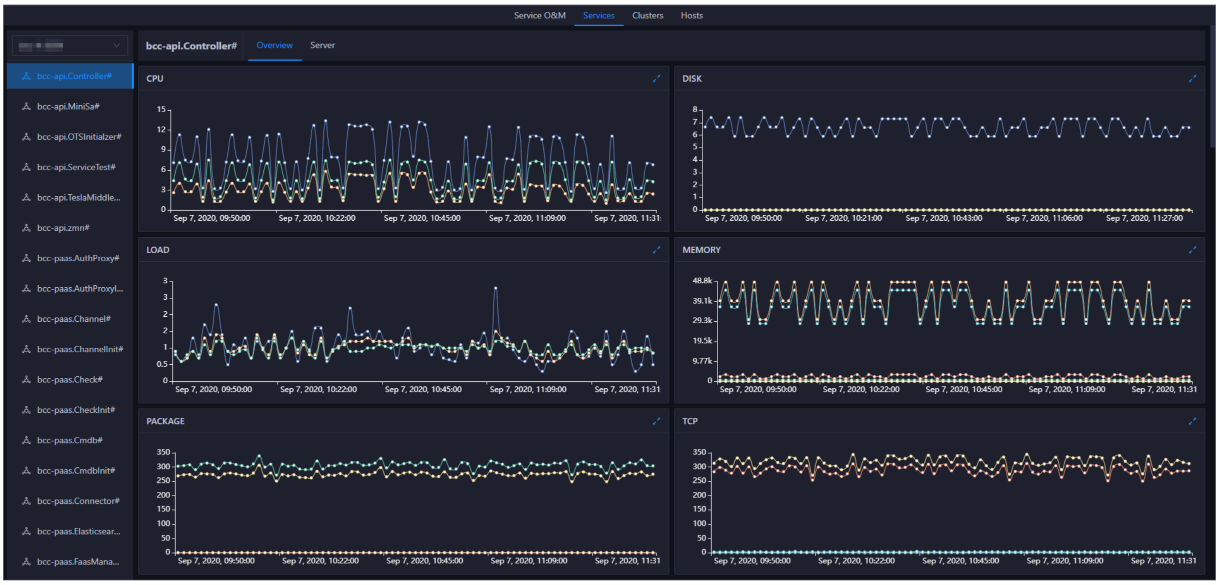
Total Disk Usage

5.33 %

Total 15.5 T Free 14.67 T

- Blink: shows information about the Blink service. You can view the service status, health check results, health check history, and core metrics such as TPS and Failover Rate.
- Yarn: shows information about the YARN service. You can view information about checkers, health check results, health check history, applications, containers, and nodes. You can also view logical CPU utilization and logical memory usage.
- HDFS: shows information about the HDFS service. You can view information about checkers, health check results, health check history, NameNode, blocks, and DataNode. You can also view SSD usage, HDD usage, and total disk usage.

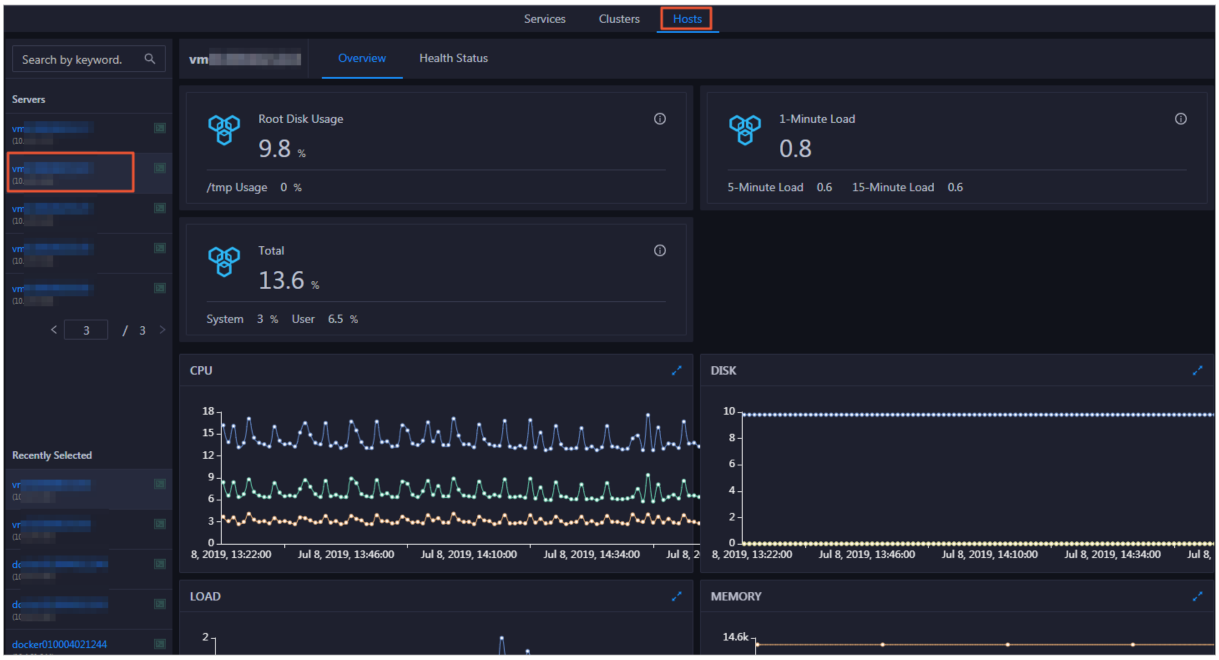
For other big data services, ABM provides information about all server roles in a cluster and the resource usage trend of each server role.



You can select a service from the left-side navigation pane to view the trend charts of CPU utilization, disk usage, memory usage, load, packet transmission, TCP connections, and root disk usage.

Hosts

The Hosts module is provided for all big data services. ABM provides two major features for host O&M: Overview and Health Status.



Root Disk Usage 9.8 %

/tmp Usage 0 %

1-Minute Load 0.8

5-Minute Load 0.6 15-Minute Load 0.6

Total 13.6 %

System 3 % User 6.5 %

- Overview: shows brief information about hosts in a MaxCompute cluster. You can view the server information, server role status, health check results, and health check history. You can also view the trend charts of CPU utilization, disk usage, memory usage, load, and packet transmission for the host.
- Health Status: shows all checkers for a cluster. You can query checker details and check results for hosts in the cluster. The status of a checker can be CRITICAL, WARNING, or EXCEPTION.

ABM also provides MaxCompute, DataHub, and StreamCompute with the following tailored features for host O&M:

- Charts: shows enlarged trend charts of CPU utilization, memory usage, disk usage, load, and packet transmission for a host. These trend charts are the same as those displayed on the Overview tab for the host.
- Services: shows the cluster to which a host belongs, the services running on the host, and the server roles.

Business

ABM provides MaxCompute, StreamCompute, Elasticsearch, and DataHub with tailored features for business O&M. For MaxCompute, ABM supports the following features: Projects, Jobs, and Business Optimization.

- Projects:
    - Project List: shows all projects and project details in a MaxCompute cluster. You can filter, query, and sort projects. You can also change the quota group of a project. If zone-disaster recovery is enabled, you can specify resource replication parameters and determine whether to enable resource replication for a project.
    - Authorize Package for Metadata Repository: allows you to authorize members of a project to access the metadata warehouse.
    - Encryption at Rest: allows you to encrypt the data stored in MaxCompute projects.
    - Disaster Recovery: allows you to view the cluster status when zone-disaster recovery is enabled for MaxCompute. You can enable the switchover between the primary and secondary clusters. You can also determine whether to run scheduled tasks to synchronize resources between the primary and secondary clusters.
- Jobs: shows information about jobs in a MaxCompute cluster. You can filter and search for jobs. You can also view operational logs, terminate a running job, and collect job logs.
- Business Optimization: provides the following features: File Merging, File Archiving, and Resource Analysis.

For DataHub, ABM provides information about projects and topics in DataHub clusters.

- Projects: shows all projects and project details, which include the project overview and related topics.
- Topics: shows all topics and topic details. You can view the topic overview and information about monitoring metrics, shards, subscriptions, DataConnectors, and schemas.

For StreamCompute, ABM provides information about projects, jobs, and queues in StreamCompute clusters.

- Projects: shows all projects.
- Jobs: shows all jobs and allows you to diagnose them to troubleshoot issues.
- Queues: shows all queues, queue resources, and tasks in a queue to help you analyze queues.

39.4.4. Management

The Management module allows you to manage configurations in a comprehensive manner. This module supports features such as job management, package management, hot update, health management, and operation auditing.

Job management

ABM executes jobs to implement O&M on big data services. Jobs are divided into two types: cron jobs and ordinary jobs. The system executes cron jobs based on a schedule. You can also manually execute cron jobs. Ordinary jobs are all manually executed.

ABM offers multiple O&M schemes to meet the needs of most scenarios. A scheme is a job template. You can easily generate and execute jobs by using a scheme.

ABM also provides an atom library that contains most common O&M operations. An atom is a template of an atomic step. When you generate a job from a scheme, you can directly use atoms as job steps.

Package management

This feature allows you to apply patches to the Docker containers of big data services. Docker is an application container engine. It allows you to quickly update product software by replacing only files that need to be updated.

Hot update

This feature allows you to update monitoring configurations and checkers without the need to interrupt services.

Health management

ABM provides a wide variety of built-in checkers for each big data service. These checkers are used to check service faults and send alerts. This helps you detect and fix faults in a timely manner.

Scheduling: Checkers run scheduling scripts on the servers of specific TianJi roles and generate raw alert data. Raw alert data provides information about the checker, server, alert severity, and alert content. Raw alert data is stored in the ABM database.

Monitoring: You can mount checkers to service pages in ABM. You can configure filter policies to display alerts about high priority checkers.

ABM allows you to customize the execution interval, runtime parameters, and mount point of a checker. You can also enable or disable a checker.

Operation auditing

This feature allows you to view the O&M history and details of all O&M operations. It also allows you to track and identify faults.

39.5. Scenarios

If you are using Apsara Stack and have deployed one or more big data products, you need to use Apsara Big Data Manager (ABM) to perform O&M operations on these big data products.

Apsara Stack Enterprise and big data products

If you are using Apsara Stack Enterprise and have deployed one or more big data products, such as MaxCompute and DataWorks, you need to use Apsara Big Data Manager (ABM) to perform O&M operations on these big data products.

39.6. Limits

None.

39.7. Concepts

This topic describes basic concepts of ABM.

Product

A group of clusters. A product provides services for users.

Cluster

A group of physical hosts. A cluster provides services logically and is used to deploy software of a product. A cluster belongs to only one product. You can deploy multiple services on a cluster.

Service

A group of software used to provide an independent feature. A service contains one or more service roles. You can deploy a service on multiple clusters.

Service role

One or multiple indivisible function units of a service. A service role contains one or more applications. If you deploy a service on a cluster, you must deploy all service roles of the service on hosts in the cluster.

Service role instance

A service role on a specific host. A service role can be deployed on multiple hosts. The service role on a specific host is called a service role instance.

Application

A software entity, which is the minimum unit for starting software. Generally, an application is an executable file or a Docker container. If you deploy a service role on a host, you must deploy all applications of the service role on the host.

Service tree

The overall organizational structure of a product. Each product is an independent entity consisting of a certain number of services. The hierarchy of a product's services forms a service tree.

Workflow

A packaged framework that consists of a sequence of processes predetermined based on specific rules. A workflow supports automatic execution. You can use workflows to perform repetitive tasks.

Job

A product O&M task created by users.

Atom

A template of an atomic step. Atoms can be used to create jobs.

Atomic step

An atom that is directly included as a step when you use schemes to create jobs.

Scheme

A job template. You can use schemes to create jobs.

40.Dataphin

40.1. What is Dataphin?

Dataphin is an intelligent engine for building big data platforms. It is designed to meet the requirements of big data development, management, and application across multiple industries. Dataphin combines technologies and methodologies. It provides all-in-one intelligent data development and management services, including data ingestion, data standardization, data modeling, data asset management, and data services.

Dataphin applies to different computing and storage environments. This enables you to use a single console to process data from various data sources. Dataphin allows you to import data, standardize data production, develop data by data modeling, and create a tag system by extracting tags from entities. You can also generate and manage data assets by using your business data and knowledge. Dataphin also provides multiple types of data services such as table query and intelligent voice search.

40.2. Benefits

This topic describes the benefits of Dataphin.

Dataphin provides the following benefits.

- Standard data: The definitions of dimensions, dimension attributes, business processes, and metrics are standardized based on dimensional modeling. This ensures the quality of data and accuracy of metrics.
- Efficient and automatic coding: Dataphin defines logical components for common data computing based on functional programming. It allows you to customize statistical metrics. You can create data models as required. Then, Dataphin automatically generates code to produce data.
- Optimal intelligent computing: You can create logical models from business perspectives. After you publish your logical models, Dataphin automatically generates the physical representations and code of the logical models. This simplifies data modeling and coding.
- All-in-one development: Dataphin integrates data ingestion, data modeling, scheduling and management, data search, and data exploration to help you develop data in a centralized and efficient manner.
- Systematic data catalog: Based on standardized modeling and efficient and automatic metadata extraction, Dataphin provides a standardized and readable business data catalog. The data catalog forms a data asset map and allows you to spend less time finding and using the data you require.
- Efficient data search: An overview of data assets is provided based on your metadata and the data from the Dataphin system database. You can search for tables and query data in a fast and intelligent manner.
- Visualized data assets: A business data asset map is built to represent your business system from different data perspectives, extract business data knowledge, and obtain more information about key business stages and data.
- Easy and reliable data utilization: Data elements can be used for data production after they are created. You can search and access logical tables that are created based on business themes with ease. This simplifies about 80% of query code.
- High efficiency: Dataphin provides end-to-end and intelligent data development and management tools to improve the data development efficiency. Developers can independently run the extract, transform, and load (ETL) procedure to meet data requirements. The patent protected OneData,

OneEntity, and OneService methodology allows you to abstract and customize models and metrics. Dataphin can also automatically generate code, aggregate data by theme, and provide data aggregation results.

- Low costs: Dataphin is based on metadata and driven by intelligent algorithms. Data can be automatically produced on the physical platform and logical plane in an intelligent manner. In addition to comprehensive analysis for data assets, Dataphin ensures the optimal allocation of computing and storage resources. This reduces the cost of data production and consumption.

40.3. Features

This topic describes the features of Dataphin.

- Support for computing engines: Dataphin supports multiple types of computing engines, including MaxCompute and Hadoop.
- Data ingestion: You can import and structure data from various data sources, including on-premises and Alibaba Cloud databases, unstructured data storage, and big data storage.
- Global design: You can define business units, theme domains, and projects based on a global view.
- Data standardization: Dataphin allows you to define data elements for data standardization by setting parameters in the console. You can define multiple statistical metrics at a time. Dataphin then processes the metrics to generate aggregate data.
- Data modeling: You can build logical data models in a visualized manner. Dataphin automatically generates the code representation of your data models. It also generates tasks to convert your logical data models to physical models. You will not be aware of the code and task generation process. Dataphin also allows you to customize coding for data development.
- Scheduling and management: You can schedule tasks and manage task running.
- Metadata management: Dataphin can automatically extract metadata in a standardized manner and process the metadata to create a metadata center.
- Asset analysis: Dataphin visualizes data assets and provides a data asset map. You can gain an overview of your data assets and locate and use the data that you require.
- Data security: Dataphin supports access control for projects, tables, and fields.
- Data service: Dataphin allows you to query logical and physical tables by theme.

40.4. Functions

40.4.1. Overview

This topic describes the modules of Dataphin.

Platform

This module helps you obtain up-to-date information about the entire system and global settings, and understand the system features to get started with ease. It also implements system management and control to ensure that all the other modules are running as expected.

Global design

You can design a data architecture based on a global view of your business and data. During the design, you can define namespaces, theme domains, and terms. You can also create projects as management units and add data sources.

Data ingestion

Based on the projects and physical data sources defined during global design, the data ingestion module can extract data of various business systems and types and load it to the target database. During this process, data is synchronized and integrated, and the source data layer is built based on various data cleansing policies.

Data standardization

Based on the data architecture defined in global design and the source data layer built by data ingestion, you can create data elements such as statistical metrics. You can use these data elements to ensure that clear and standardized data will be produced.

Data modeling

You can use the data elements created for data standardization to design data models. After the data models are submitted and published, Dataphin automatically generates code and scheduling tasks to complete data production at the common dimensional model layer in a fully managed manner.

Coding

Dataphin provides a code editor for you to configure and submit code tasks.

Resource and function management

Dataphin allows you to manage resource packages such as JAR packages and other type of files to meet data processing requirements. You can search for and use built-in functions and create user-defined functions to meet specific requirements for functional processing.

Data distilling

Based on the source data layer and common dimensional model layer, Dataphin can focus on objects. You can set parameters in a visualized manner to identify and map the IDs of objects, extract behaviors of the objects, and define tags, thereby achieving data integration and data mining. Dataphin can then generate code and scheduling tasks to complete data production in the data distilling center in a fully managed manner.

Scheduling and management

Dataphin supports policy-based scheduling and management of tasks generated by modeling, coding, and data distilling. You can deploy, run, and manage data production tasks, and view and manage task dependencies. This module ensures that all tasks can run as expected without interruption.

Metadata center

Dataphin allows you to collect, parse, and manage metadata of the source data layer, common dimensional model layer, and data distilling center.

Data asset management

Based on the metadata center, this module supports deep metadata analysis and data asset management. It shows asset distribution and metadata details. This makes it easy for you to search for data assets and obtain information about data assets in more detail.

Security management

This module supports quality and security management, including data standardization, analysis result display, process management, monitoring and alerts, and end-to-end tracing from data sources to applications. These features allow you to locate asset optimization problems and provide solutions.

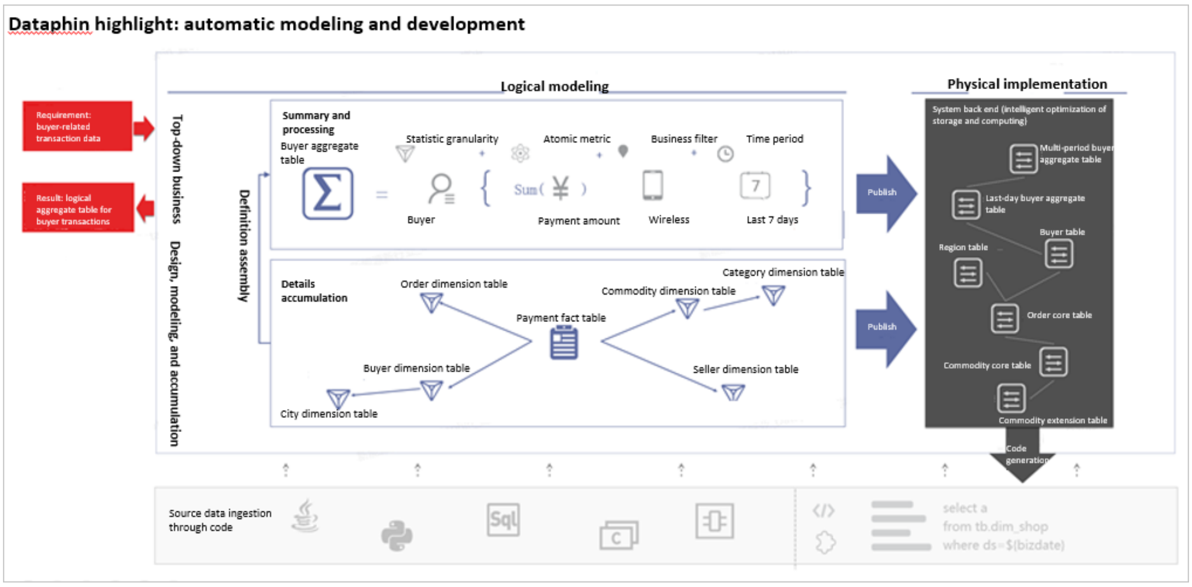
Ad hoc query

This module allows you to execute custom SQL statements to query data. It uses the search and analysis engine to find data in physical tables and theme-based logical tables. Theme-based logical tables are also known as data models or logical models.

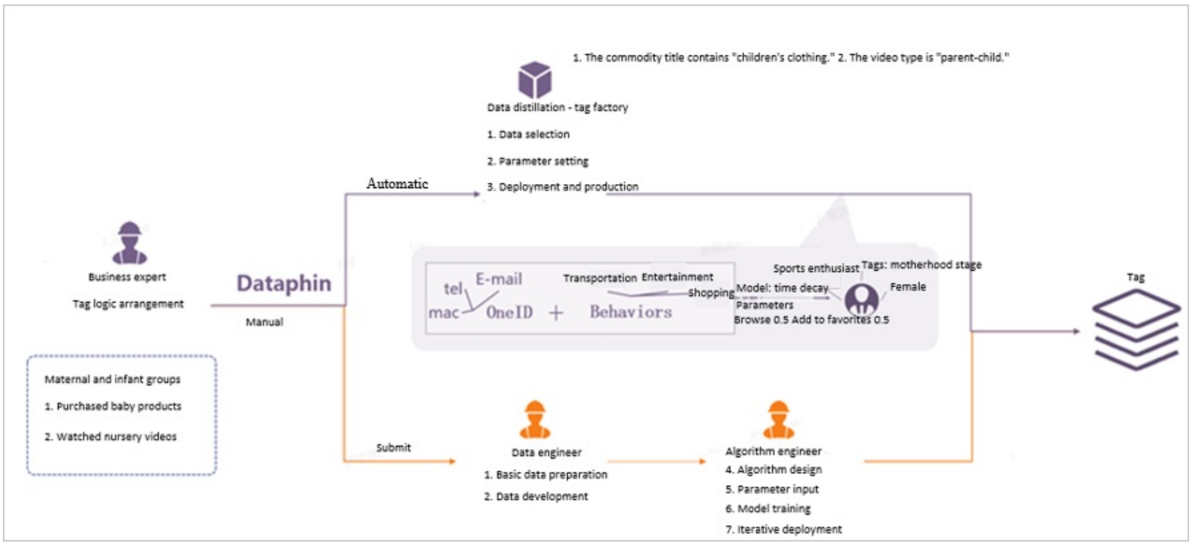
40.4.2. Resolved issues

With Dataphin, you can resolve the following issues:

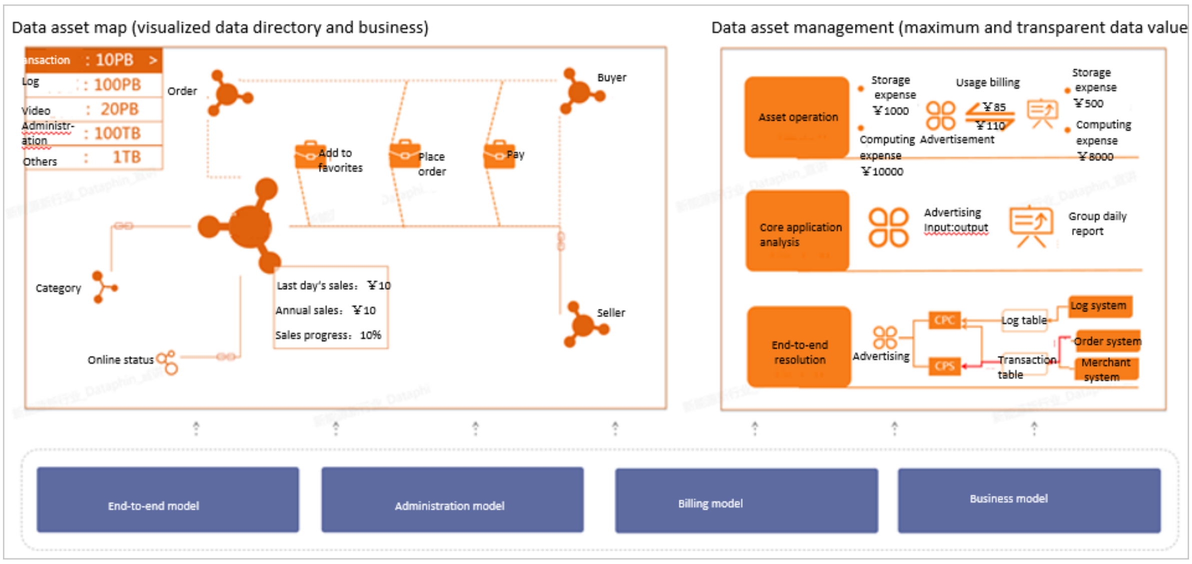
- Modeling: You can build data models by using a graphical user interface rather than writing SQL code. The system then automatically publishes the models and generates tasks to produce data. All metrics and standards are clearly defined.



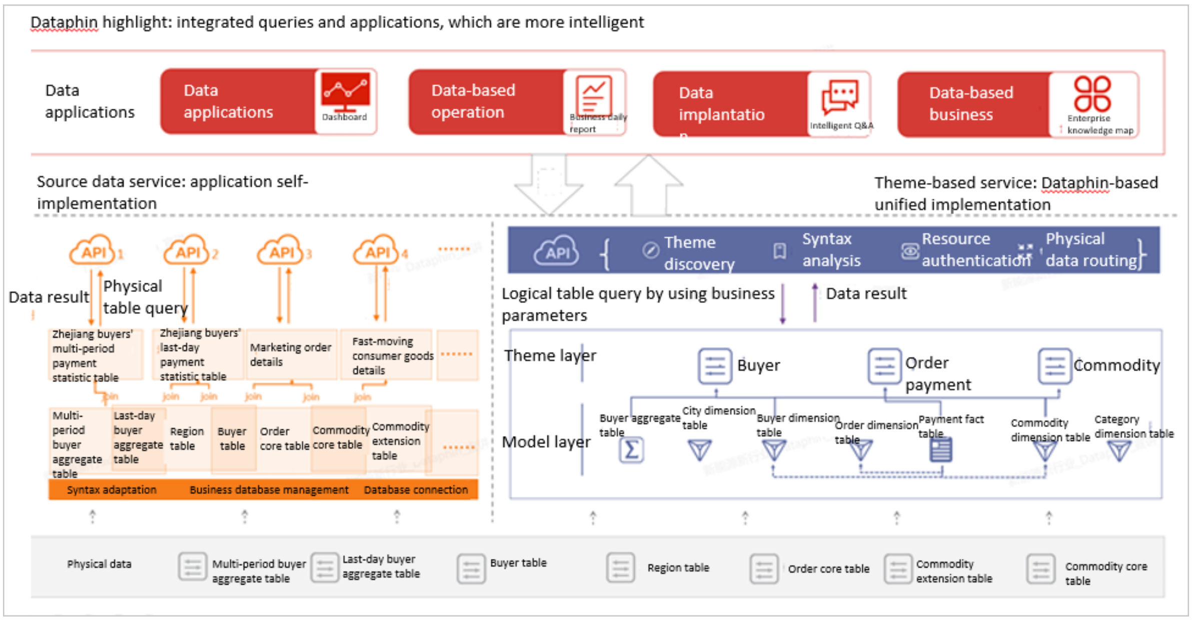
- Data distilling (coming soon): You can extract master business data and build a data management platform (DMP) based on entities. This will include three steps and involve customizing parameters, ID recognition, and automatic tag creation following a standard process.



- Data asset management: You can create and manage data assets, gain a deep understanding of data assets from a unique perspective, and get more value from your business data.



- Ad hoc queries: Dataphin supports theme-based queries for logical tables. This ensures quick data query and locating, and greatly simplifies SQL query statements. This also ensures that data is produced in a standard, regular, and clear manner. The standardized output data can be used by several business applications.



40.4.3. Console

As the basis of Dataphin, the Dataphin console guarantees that all Dataphin members can develop data in a controllable, orderly, and smooth manner. In this console, you can configure global settings, such as account management and computing management. The Dataphin console supports both Chinese and English, and provides introductions and entrances to various modules on its homepage. This helps the super administrator get the whole picture of Dataphin and members of other roles quickly access modules.

Account management

The Dataphin console allows you to manage member accounts to guarantee secure use of Dataphin. You can connect your enterprise account system to Dataphin. Then, the users who need to use Dataphin can be added to Dataphin as members. Users with the highest privileges can manage the accounts and permissions of other users.

Computing management

- As a Platform as a Service (PaaS), Dataphin enables you to select a computing engine type and configure connection settings for your data sources. This makes Dataphin compatible with various environments at the Infrastructure as a Service (IaaS) layer. In this way, Dataphin can develop and compute data in a uniform and stable manner.
- Dataphin supports two major types of computing engines: MaxCompute and Hadoop. Dataphin can automatically collect and parse the metadata of these two types of engines. For more information about how to collect and deploy metadata, see Metadata warehouse.

Homepage

- The Dataphin console provides shortcuts to functional modules, projects, and the scheduling center on the homepage. You can also find an overview of the scheduling center and projects on the homepage.
- The homepage classifies modules based on the workflow in Dataphin, which consists of data warehouse planning, data R&D, data asset management, and theme-based data services. The workflow helps you learn about Dataphin features before you get started, and enables you to quickly access specific modules.

Language

To help users from different countries and regions use Dataphin, the Dataphin console selects Chinese or English based on the language of your operating system.

40.4.4. Global design

Based on a global view of your business and data, you can design an architecture for your data warehouse. This a fundamental step in data development. The architectural design ensures that data is manageable and controllable. The data systems defined and designed during data development, distilling, and management meet mid- and long-term business requirements. The produced business data is service-oriented, theme-based, and easy to use.

The global design involves the following:

- Data warehouse architecture definition based on business characteristics includes business unit management and access control, data domain management and access control, and management of the defined global objects.
- Project definition based on requirements for independent data management and collaborative development includes member management and the management of basic project information and computing resources.
- Data source configuration based on computing resources for projects and requirements for business data includes data source management.

Data warehouse architecture

The data warehouse architecture defines logical namespaces (business units), theme domains (data domains), and terms (global objects) based on business characteristics. This standardizes data definitions during architectural design management and data development control.

Projects

A project is a physical namespace used to isolate users from resources. Projects are created to meet the requirements for independent management of data development projects and efficient management of data resource quality. Data development constraints can be configured for each project.

Physical data sources

Dataphin supports data source creation, modification, and other features that allow you to register and cancel the registration of databases. The data source types supported by Dataphin include MaxCompute, MySQL, SQL Server, and PostgreSQL. Data sources can be used as the source storage or target storage for data synchronization. Some special types of data sources (such as MaxCompute) can serve as the computing engine for projects to function as the computation and storage base.

40.4.5. Data ingestion

The source data layer is built through data ingestion. Before ingesting data, you need to select a business data storage system as the data source. Then, you need to formulate data synchronization, cleansing, and structuring polices to satisfy your data requirements in terms of storage, accuracy (up-to-date), and quality.

Data ingestion is an important initial stage in data development. The data synchronization suite of Dataphin is developed based on several years of industry practice. In the past, Alibaba has overseen the synchronization and exchange of many types of data including business and log data. This helps achieve efficient ingestion of raw business data. The data transmission channel can collect and analyze metadata to check the amount and content of data that has been transmitted. The flexible management of custom error tolerance mechanisms is also supported. This helps achieve high-quality data synchronization.

Data source configuration

You can import and manage multiple data sources. The data source list allows you to manage imported data sources and add various different types of additional data sources. Currently, data sources that can be used for data synchronization include MaxCompute, MySQL, SQL Server, PostgreSQL, and Hive.

Data synchronization

You can select source data and target data, configure parameters for incremental or full synchronization, and identify mappings between source data fields and target data fields. You can also configure the data transfer rate and the number of concurrent sync tasks. With these configurations, synchronization tasks can be generated and scheduled.

40.4.6. Data standardization

In most cases that involve traditional development, specific and important data creation and development (such as data modeling and metric definition), depend on the developer's professional capabilities. Without a uniform naming convention, standards for development and designs are transferred based on individual and changing documents. This may cause a series of problems such as metric name conflicts or repeated calculation.

Based on the OneData methodology, Dataphin standardizes the definition of important data elements such as dimensions, business processes, and metrics. This ensures unique computing logic and names, and eliminates metric ambiguities during the initial stages of architectural design. In addition, Dataphin provides form-based interfaces for you to create multiple metrics at a time. This lowers the requirements of data development and increases overall development efficiency. This also allows business users with limited data analysis expertise to carry out development work by using Dataphin.

Data standardization involves defining five types of data elements: dimensions, business processes, atomic metrics, business filters, and derived metrics. Dataphin helps you design a data architecture by creating business units and data domains. You can extract standard data elements and reuse data elements based on the data architecture. Standard data elements include data warehouse themes (such as granularity that is composed of dimensions) and metric creation elements (such as atomic metrics and business filters).

40.4.6.1. Dimensions

- A dimension is unique within a business unit and it exclusively belongs to a data domain. This standardizes naming and theme classification.
- You can create dimensions by adding additional attributes to an existing dimension, which is used as a parent dimension.
- Dataphin supports the creation of various types of dimensions, including common, common (hierarchy), enumeration, and virtual dimensions.
- Dataphin allows you to view and manage the list of dimensions created in a specific business unit or a specific project. You can also view and modify each dimension.

View and manage the dimension list

Dataphin allows you to view the list of dimensions created in a specific project. You can view the name, creator, and publishing status of each dimension. You can search for a specific dimension in the list, and then modify, unpublish, or delete the dimension.

View and manage dimensions

Dataphin provides form-based interfaces for you to view, create (using a standard template), and modify dimensions. A dimension is a key concept of business. You need to specify the following information when creating a dimension:

- Basic information: the data domain (to which the dimension belongs), name, display name, and description. The name is prefixed with `dim_` by default to distinguish the name from other names.
- Logic information: The logic information is used to describe and define the scope of the dimension. This is to ensure that the dimension is accurate and unique when you later need to add dimension attributes. The required configurations vary by dimension type.

Quick view of dimensions: You can click a dimension in the left-side navigation pane to view basic information of the dimension and then perform supported operations on the dimension. This does not affect your previous operations.

40.4.6.2. Business processes

A business process is a collection of the smallest unit of behaviors or events that occur in a business activity. For example, the smallest unit of behavior can be to create an order or browse a web page. The behaviors occurring in a business process, such as paying for an order and browsing a web page, are recorded in a fact table. The fact table models a particular business process.

Similar to dimension, business process is a key concept in the OneData methodology used for designing the data architecture. It works with dimensions to define the data architecture. Dataphin supports standard definition for business processes. This allows you to check the overall business data of your organization and easily categorize fact tables by business process.

To ensure that a fact-based model is built in a unified and standard manner, a business process is unique within a business unit and it exclusively belongs to a data domain. This standardizes naming and theme classification.

Dataphin allows you to view and manage the list of business processes created in a specific business unit or a specific project. You can also view and modify each business process.

View and manage the business process list

Dataphin allows you to view the list of business processes created in a specific project. You can view the name, creator, and publishing status of each business process. You can search for a specific business process in the list, and then modify or delete the business process.

View and manage business processes

Dataphin provides form-based interfaces for you to view, create (using a standard template), and edit business processes. A business process is a key concept of business. You need to specify the following information when creating a business process: data domain (to which the business process belongs), business process name, display name, and description.

40.4.6.3. Atomic metrics

An atomic metric is an abstraction of computing logic. To eliminate definition and development inconsistency, Dataphin introduces the concept of "Design to Code". When a metric is defined, the statistical criteria (computing logic) is also defined. Re-engineering of the ETL process is not required, which increases development efficiency and ensures the consistency of statistical results. Based on the complexity of computing logic, Dataphin categorizes atomic metrics into native atomic metrics and composite metrics. An example of a native atomic metric is payment amount. A composite metric is created based on the combination of atomic metrics. For example, the average sales per customer is calculated by dividing the total sales by the number of customers.

An atomic metric is unique within a business unit and has only one source logical table. The computing logic of an atomic metric is defined based on the fields of the source logical table model. This ensures that all statistical metrics are created in a unified and standard manner. The data domain of each logical table linked to the source logical table is retrieved to trace the data domains to which the atomic metric belongs. For example, an atomic metric may belong to multiple data domains. This ensures that names and logic are normalized and themes are classified in a standard manner.

View and manage the atomic metric list

Dataphin allows you to view the list of atomic metrics created in a specific project. You can view the name, creator, and publishing status of each atomic metric. You can search for a specific atomic metric in the list, and then modify, unpublish, or delete the atomic metric.

View and manage atomic metrics

- Native atomic metrics

To ensure standard creation of atomic metrics, Dataphin allows you to define an atomic metric that is only based on a logical table and its model. You can select a source table. Select a field from the snowflake or star schema that contains the source table, and define the computing logic for the atomic metric based on the field.

- Composite metrics

  Composite metrics are calculated based on multiple atomic metrics. For example, you can obtain the payment conversion rate metric based on several atomic metrics. You can first define two atomic metrics: the number of customers who pay for orders and the number of customers who place orders. The payment conversion rate metric is expressed as the number of customers who place orders divided by the number of customers who pay for orders.

40.4.6.4. Business filters

An atomic metric is the standardized definition of computing logic, and a business filter is the standardized definition of a query condition. Similar to an atomic metric, a business filter is unique within a business unit and has only one source logical table. The computing logic of a business filter is defined based on the fields of the source logical table model. This ensures that all statistical metrics are created in a unified and standard manner. The data domain of each logical table linked to the source logical table is retrieved to trace the data domains to which the business filter belongs. For example, a business filter may belong to multiple data domains. This ensures that names and logic are normalized and themes are classified in a standard manner.

View and manage the business filter list

Dataphin allows you to view the list of business filters created in a specific project. You can view the name, creator, and publishing status of each business filter. You can search for a specific business filter in the list, and then modify, unpublish, or delete the business filter.

View and manage business filters

To ensure the standard creation of business filters, you can only define a business filter based on the source logical table and the models associated with the table. You can select a source table. Select fields from the snowflake or star schema that contains the source table, and define the computing logic for the business filter based on the fields.

40.4.6.5. Derived metrics

Derived metrics are commonly used statistical metrics. To create derived metrics in a standard, regular, and clear manner, each derived metric is a calculation based on the following criteria:

- Atomic metric: statistical criteria, that is, the computing logic.
- Business filter: the scope of business to be measured. It is used to filter the records that comply to specific business rules.
- Statistical period: a period during which statistics are collected, for example, the last 1 or 30 days.
- Granularity: a statistical object or perspective that defines the level of data aggregation. It can be considered as a grouping condition for aggregation, that is, GROUP BY clauses in SQL statements. Granularity is a combination of dimensions. For example, if a derived metric is a seller's turnover in a province, the granularity is the combination of the seller and the region dimensions.

By combining the preceding parts, multiple derived metrics can be quickly created at a time while ensuring that the definitions and computing logic are clear without any duplication. This metric creation method is simple, available to all users, and does not require a high level of technical expertise. For example, business users can also complete metric creation. A derived metric is a concept that is based on the same level as a field. Each derived metric is unique and defined at the specified granularity level.

View and manage the derived metric list

Dataphin allows you to view the list of derived metrics created in a specific project. You can view the name, creator, and publishing status of each derived metric. You can search for a specific derived metric in the list, and then modify, unpublish, or delete the derived metric.

View and manage derived metrics

To standardize the creation of derived metrics, the scope and objects to be measured must be determined based on the statistical computing logic. Therefore, you must select an atomic metric and the granularity, statistical period, and business filter related to the atomic metric. Then, you can follow a standard process to create multiple derived metrics at the same time.

- Select statistic granularity

  Granularity is a combination of dimensions. The dimensions in the selection box are all the dimensions linked to the logical table model where the atomic metric resides. This provides a strong basis for useful and practical calculation at the specified granularity.

- Select a statistical period

  Dataphin provides default statistical periods and also allows you to add custom statistical periods on the Planning page.

- Select a business filter

  A business filter is a constraint or filter condition defined for a logical table. You may need to obtain a group or a type of business data. For example, you want to define metrics for the same statistical scope and computing logic for different statistical periods, such as the last one day, seven days, and 30 days. Dataphin allows you to define multiple levels of granularity, statistical periods, and business filters. These elements can be combined to create multiple derived metrics. This ensures standard metric creation and improves development efficiency.

40.4.7. Modeling

40.4.7.1. Overview

Dataphin provides systematic modeling and development functions to deeply implement the data warehouse theory. You can create business dimensions and business processes by using a top-down approach, and then enrich dimension tables, fact tables, aggregate tables, and the application data store layer. This process allows you to produce standardized data assets, which provides you with layered business data. The data standardization process can also optimize computation and storage.

40.4.7.2. Logical dimension tables

A logical dimension table contains details about a dimension. Dataphin allows you to view and manage the list of created logical dimension tables, and to view and modify a specific logical dimension table.

View and manage the logical dimension table list

Dataphin allows you to view the list of logical dimension tables created in a specific project. You can view the name, creator, creation time, and publishing status of each table. You can search for a specific logical dimension table in the list, and then modify or delete the table.

You can view details about a specific logical dimension table. You can view the primary key, dimension-associated fields, and attributes in the logical table. You can also view the star schema and snowflake schema containing this logical dimension table. If an inheritance relationship is defined, you can view settings of the parent and child dimension tables. You can also publish a logical dimension table after unlocking and modifying the table, zoom in or zoom out from the canvas, and view the published version. Dataphin provides a graphical user interface for you to configure a specific logical dimension table. You can define dimension attributes, associate dimensions with the table, and add child dimensions. Other supported operations include configuring the logical table conversion settings, viewing table details, and customizing the scheduling policy for the logical table conversion task.

40.4.7.3. Logical fact tables

Dataphin supports using logical fact tables to model a specific business process (such as placing an order and paying for a commodity) or a state measure (such as account balance and inventory). A logical fact table is created in an optimized schema that is similar to a snowflake schema. Apart from measures and dimension-associated fields, this type of schema allows a fact table to also contain fact attributes. This reduces the complexity of the model design and makes it more user-friendly.

View and manage the logical fact table list

Dataphin allows you to view the list of logical fact tables created in a specific project. You can view the name, creator, and publishing status of each table. You can search for a specific logical fact table in the list, and then modify, unpublish, or delete the table.

View and modify logical fact tables

Dataphin allows you to view details about a specific logical fact table model on a form-based interface. You can view the dimension-associated fields, measures, fact attributes in the logical fact table, and the logical dimension tables associated with the table. You can also publish a logical fact table after unlocking and modifying the table, zoom in or zoom out from the canvas, and view the published version. Dataphin provides a graphical user interface for you to configure a specific logical fact table model. The configurations include defining basic information, primary key, and fields, configuring the logical table conversion settings, and customizing the scheduling policy for the logical table conversion task.

40.4.7.4. Logical aggregate tables

The logical aggregate table model is an important data warehouse model. It contains two types of elements. The first type of element refers to various statistical values used to describe statistic granularity. The statistical values form a derived metric, for example, the sales in the last seven days. Granularity is a combination of several dimensions, such as the province and the product line dimensions. The second type of element refers to the attributes of the dimensions that constitute granularity. Examples of attributes are province name, product line name, product line level.

View and manage the logical aggregate table list

Dataphin allows you to view the list of logical aggregate tables created in a specific project. You can view the name and creation time of each table. You can search for a specific logical aggregate table in the list, and then modify, unpublish, or delete the table.

View and modify logical aggregate tables

A logical aggregate table can be created by aggregating the derived metrics defined following a standard process. You can also associate the logical aggregate table with fields of physical tables generated by code tasks.

40.4.7.5. Coding automation

After a logical dimension table, logical fact table, or logical aggregate table is published, Dataphin automatically designs the corresponding physical model, generates code and tasks to produce required data. Multiple tasks are usually generated to convert a logical table to a physical model. If you want to view the task running logic, go to the Scheduling page.

40.4.8. Coding

40.4.8.1. Overview

Coding is an important data development method. This method can be used to achieve the same goal as building data models on graphical user interfaces. Dataphin allows you to edit scripts by using the coding method supported by your computing engine. You can submit the scripts to the scheduling system, which schedules the code tasks to produce data. You can also view historical versions of each code task. Multiple types of scripts are supported, such as SQL, Shell, and MapReduce scripts. The requirements for coding and configuration vary by script type. The requirements include syntax requirements and requirements for scheduling configuration. After a script is submitted and published, Dataphin creates a code task to run and produce data. In a directed acyclic graph (DAG), a task is also called a node. Dataphin supports the following operations for code task management: create, view, modify, and delete code tasks, edit scripts, configure task scheduling policies, publish tasks, and manage task versions.

40.4.8.2. Code editor

The code editor provides an online code editing interface to complete data development tasks. It supports SQL, MapReduce, Spark, and Shell programming.

40.4.8.3. Task scheduling configuration and publishing

Scheduling configuration

You can configure the scheduling policy for one-time and recurring tasks. Tasks with a scheduling policy configured can be published. The system can check the integrity of task scheduling configurations. Only tasks with a complete scheduling configuration can be published. All published tasks are recurring tasks. You can choose **Scheduling > Recurring Tasks** and view the published recurring tasks in the left-side navigation pane.

Publish

Members of a project can publish tasks if they have required permissions. Only a scheduling configuration with complete parameter settings, valid dependencies, and no circular dependencies can be published to create tasks. This guarantees that stable and orderly data production can be completed on schedule.

40.4.8.4. Code management

Dataphin supports various code operations to facilitate code file management and use. You can create, delete, update, rename, and view code files, and place code files in specific folders to categorize the code files.

Manage files

Dataphin allows you to edit, delete, unpublish, and rename each code file. You can also view the publishing status, creator, and creation time of each code file. This facilitates easy creation, clear display, and systematic management of code files.

Manage folders

When there are many code files, sort them in different folders to save and display these files in an orderly manner. You can create, rename, and delete folders, and move historical and new code files to specified folders for better management. Dataphin also supports hierarchical folder structures.

40.4.8.5. Collaborative programming

Manage node versions

Dataphin allows you to view historical task node versions. You can view the version number, submitter, submission time, and description. You can also view the code of each version to identify differences in code. Dataphin supports multiple node types, including MaxCompute_SQL, MaxCompute MR, and Shell.

Collaborative development

To achieve more efficient development by allowing collaboration between multiple developers, Dataphin provides a script locking mechanism, which prevents conflicts during collaborative development. This mechanism ensures that a line of code can only be edited by one user at a time. A user can steal the lock of another user to obtain the script editing permission. The user whose lock is stolen can obtain editing permission again by stealing the lock.

40.4.9. Resource and function management

40.4.9.1. Overview

Resource and function management assists code development. Data developers can upload local resources and configure task nodes for calling these resources to meet specific data processing requirements. These developers can also complete common data processing by using the built-in functions in the programming language supported by the computing engine. If a data logic (such as data conversion in compliance with a business logic) requires frequent processing and this cannot be achieved with the built-in functions, developers can define custom functions based on self-uploaded resources.

40.4.9.2. Resource management

Dataphin allows the data developers of a project to add, edit, and perform other operations on resources in the project. You can name and upload resource files, and then copy the resource file name to reference the resource file in the code. You can also delete unnecessary resource files.

Create and upload resource files

By default, the following types of local resource files can be uploaded: XLS, DOC, TXT, CSV, JAR, Python, and other types (such as ZIP packages). New file types that are different from these types can be quickly added in three days by using the standard interface. Each resource file name is unique within a project. The file name and resource package cannot be changed after a resource file is 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 copy and paste a resource file name to a specific position in the code editor, and write a statement to call this resource.

Update resources

You can update the description of managed resources and delete existing resources to save storage space.

40.4.9.3. Function management

You can search, use, and manage functions. Functions are classified into two types: built-in functions of the system and user defined functions based on uploaded resources such as JAR and Python packages. You can extend user defined functions by referencing standard functions.

Create user defined functions

Each user defined function must have a unique name within its project and cannot be renamed after being registered.

Reference functions

You can click Copy to copy the name of a built-in function or a user defined function, and then paste the name to a specific position in the code editor. Then, write a statement in the format of the sample command to process data.

Update functions

You can update user defined functions by editing related information (except name) and delete unnecessary user defined functions.

40.4.10. Scheduling and management

The scheduling center allows you to perform management work during the later stages of data development. The scheduling center provides the list of all data processing tasks and task instances. Data processing tasks include recurring and one-time tasks. Task instances include instances of the data processing tasks and retroactive data generation tasks. The scheduling center also provides the directed acyclic graphs (DAGs) showing task dependencies, task instance dependencies, and instance status. You can set the task running sequence, schedule specific nodes in a DAG, achieve optimal allocation of resources, and discover abnormal tasks. This ensures that all the tasks are run on schedule. The scheduling center also reports alerts during task running to ensure that errors can be handled in time. The scheduling center allows you to view and manage tasks.

Task list

You can view the lists of recurring and one-time tasks created in a specific project and the DAGs showing task dependencies.

Recurring tasks

You can view the recurring task list, search for specific tasks, and view the dependencies of each task. You can switch between different projects to view and search for tasks in a specific project. You can search for tasks by task node name or task node ID. You can also filter the task nodes that you own and nodes published the current day. This helps narrow down the scope of tasks or find specific tasks that you want to manage.

One-time tasks

You can view the one-time task list, search for specific tasks, and view details of each task. You can switch between different projects to view and search for tasks in a specific project. You can search for tasks by task node name or task node ID. You can also filter the task nodes that you own and nodes published the current day. This helps narrow down the scope of tasks or find specific tasks that you want to manage.

Task instance management

You can view the lists of recurring, one-time, and retroactive data generation task instances created in a specific project while viewing details of each task instance.

Recurring task instances

You can view the instance list, search for specific instances, and view details of each instance. You can view the running status and details of each recurring task instance. The details include task node ID, node name, task owner, task start time, end time, and run duration. You can switch between different projects to view and search for task instances in a specific project. You can search for task instances by task node name or task node ID. You can also filter the task instances that you own, instances with errors, and incomplete instances. This helps narrow down the scope of instances or find specific instances that you want to manage.

One-time task instances

You can view the instance list, search for specific instances, and view details of each instance. You can view the running status and details of each one-time task instance. The details include task node ID, node name, task owner, task start time, end time, and run duration. You can switch between different projects to view and search for task instances in a specific project. You can search for task instances by task node name or task node ID. You can also filter the task instances that you own and instances that run the current day. This helps narrow down the scope of instances or find specific instances that you want to manage.

Retroactive data generation instances

You can view the list of created retroactive data generation task instances and details of each instance. The details include the data timestamp, status, and run duration. You can also view the node ID, node name, and owner of the task for which you generate retroactive data. Dataphin also supports search and filter for retroactive data generation task instances.

Logical tables

You can search for and view logical tables and their conversion tasks. You can also view the fields of each logical table. You can switch between a logical table task and a logical table task instance to view details. By default, the DAG on the right of the logical table task list shows all conversion task nodes of the current logical table and the dependencies between the nodes, including indirect dependencies. By default, the DAG on the right of the logical table task instance list shows all conversion task instances of the current logical table and their status. The status may be running, success or failed.

40.4.11. Metadata warehouse

Dataphin provides powerful metadata management capabilities. It can collect and extract metadata from MaxCompute, Hadoop, Hive, MySQL, PostgreSQL, and Oracle data sources. It supports real-time tracing of metadata in the preceding computing and storage engines, and builds a unified metadata model by extracting metadata from different types of storage engines. Dataphin supports the rapid enrichment of multiple types of metadata and provides diverse metadata that complies with unified standards. This provides a rich source of stable metadata to catalog and handle data.

The metadata warehouse is the core foundation of data asset management. We recommend that you ensure that the following items are available or guaranteed when building the metadata warehouse:

- Metadata collection standard: A unified data development standard is required to ensure the consistency of metadata for modeling, data table creation, and data lineage. This improves the availability of metadata for data retrieval and data services.
- Metadata accuracy (up-to-date) and quality: The metadata output time and quality must be guaranteed to improve the accuracy of the data in the data asset module and the efficiency of data retrieval performed by developers.
- Metadata model system: A unified public metadata model is used to ensure compatibility with various types of data and deliver a comprehensive data map service.

40.4.12. Data asset management

After data acquisition, integration, processing are complete, you can systematically manage data assets. Based on OneData and data assets methodologies, Dataphin designs the data use principle and provides core technologies, including metadata acquisition, extraction, and processing technologies. You can classify and manage data in the form of assets, monitor data quality, and optimize resources. This allows you to minimize the cost of data, obtain the maximum value from data, and use this value to benefit your business.

Data asset management is implemented by using a series of core technologies. The real-time event subscription service provides real-time metadata update for tables and tasks. The rules engine ensures efficient and accurate judgment of data governance rules and the creation of health scoring models. Dynamic log analysis supports analyzing numerous daily operational logs for production tasks and daily machine management logs. Graph computing supports the analysis and creation of data lineage. The Onelog data tracking technology ensures the consistency of metadata between the data production, service, and consumption phases. You can access metadata during each of the three phases. The metadata import and processing architecture (in the form of a plug-in) supports management for data from different computing and storage engines. This architecture provides a set of services including data collection, analysis, governance, application, and operation. It is developed by Alibaba and based on the extensive experience with mass data management. It covers the entire data lifecycle, including data creation, management, application, and destruction.

Based on the data catalog established through an analysis of enterprise data assets, the data map module provides a search engine and data profiling (both derived from user behavior data). This allows you to efficiently retrieve an enterprise's data assets.

Asset overview

Dataphin can display the structure of the enterprise data assets that are created based on OneData. Components in different shapes represent business entities, whereas lines of different styles represent business links between these entities. This helps to visualize the structure of the data for a business unit.

Asset map

An asset map summarizes the relationships between dimensions and business processes in a data domain of a business unit to show the composition of your enterprise data. In addition, the asset map provides efficient, fast, and accurate data search and exploration based on your self-initiated behaviors, such as searches, access history, and favorites.

40.4.13. Security management

40.4.13.1. Overview

The wide use of big data services makes data security an important issue. In China, the Cyber Security Law of the People’s Republic of China was implemented on June 1, 2017. The Cyber Security Law encourages the development of network data security precautions and utilization technologies. EU General Data Protection Regulation (GDPR) was enacted on May 25, 2018. It aims to enhance the protection of data such as personal information. Dataphin focuses on intelligent development and management of data and places great importance on data security management. It provides comprehensive data security protection throughout the entire lifecycle (from data production to destruction). The protection is implemented by data access control, data isolation, and data security level classification. Other data protection methods include privacy compliance, data masking, and auditing of data security.

Data access control and data isolation require the highest priority in data security management. Dataphin provides management of data access permission requests, approvals, and lifecycle. It supports data isolation for multi-tenancy and field level access control, and offers a data access authorization model based on access control lists (ACLs).

Dataphin establishes a comprehensive data security guarantee system covering the entire lifecycle of data. This system provides technologies and management measures to protect data from the perspectives of data access behaviors, data content, and data environment. During big data development and management, Dataphin works with the Alibaba Cloud data security management system to provide an "available but invisible" environment for secure big data exchange. Dataphin also supports field level access control, control of permission request approval processes, and tracing and auditing of data use behaviors. All these combined methods help to guarantee data security during the storage, transfer, and use of big data.

Dataphin offers a hierarchical permission control system and a full range of management, covering the request, approval, assignment, handover, and authentication of data access permissions.

40.4.13.2. Permission types

Dataphin provides data access control based on user roles and resources. This allows you to use Dataphin and access data in a secure and controllable manner.

Role privileges

Dataphin provides account management mechanisms to obtain the super administrator and system members for centralized management of user operations. This controls the access methods of users at the platform level. Dataphin also allows you to control resource access at the organizational level by using project management. This access control method is role-based access control. It assigns specific roles a set of data resource permissions. Users acquire permissions through the roles to which the users are assigned.

Resource permissions

Dataphin provides a data access control mechanism to centrally manage user operations on project data resources. When each project is independently managed, and system members are isolated from resources, cross-project resource access can be controlled. This helps achieve data sharing by allowing users to use data of a specific project in another project without data migration.

40.4.13.3. Permission management

Permission requests

Data developers can find the required data table on the Data Map page and view the metadata details of this table. However, if they want to query data in the table, they must apply for permissions.

In a permission request process, Dataphin displays information about the requested data table by default, including the table type and the business unit to which the table belongs. Field metadata of the table is also displayed. Dataphin supports permission requests that follow the principle of least privilege. Specifically, requests for field-level permissions are supported. Multiple options of permission validity period are provided. You can customize a date range or select 30 days, 90 days, 180 days, or 1 year as the validity period. You can describe the purposes for which you intend to use the requested permissions. The approver can determine whether to grant you the permissions based on the description.

Request management

Dataphin allows you to view your requests and the status of the requests. You can click **Details** to view details of a request and click **Cancel** to cancel a request. After your request is approved, you can view your permission details, including the accessible fields.

Permission approval

After a permission request is submitted, the system randomly assigns the ticket to an administrator of the project to which the requested data table belongs. The administrator needs to approve the request. Approvers can view details about the submitted requests on the **My Approvals** tab and decide whether to approve or reject the request.

Permission handover

Users must hand over their permissions before shifting to another position or leaving the company. This ensures that related data and data production tasks can be handed over to appropriate staff. On the **My Permissions** page, you can click **Revoke** to hand over your permissions to the project administrator. Then, Dataphin reclaims the permission.

40.4.14. Ad hoc query

Dataphin supports high-performance ad hoc queries based on the OneService engine. Dataphin supports both traditional simple query and theme-based query methods, and enables code simplicity and fast query.

Syntax

- Dataphin supports offline queries on all logical tables. The intelligent engine selects the optimal physical table based on factors such as the output time and query performance.
- Dataphin supports join queries based on snowflake schemas. This makes it simpler to write SQL queries.
- Dataphin supports queries on physical tables, logical tables, and combinations of physical tables and logical tables.
- Dataphin supports multiple computing engines (each with unique syntax), such as MaxCompute SQL and Hive SQL.
- Dataphin provides intelligent code completion, precompilation, and beautification for SQL statements.
- Dataphin can manage permissions and authenticate users for access to fields in a logical or physical table.

Query implementation

You can enter any query statements in a query script. The script editor provides intelligent prompts based on the input content, quickly locates the required data table or field, and verifies the validity of the script syntax.

40.5. Scenarios

A retail group plans to launch a marketing program for members on New Year's Eve and wants to invite a celebrity for a promotional event. For this purpose, its business team needs to analyze the members' reaction to promotional offers for each quarter to determine the total amount of coupons to issue. In addition, the team also studies the members' celebrity preferences to determine whom to invite and the key commodities to promote.

The group has imported all transaction data and commercial-related music and video data into a MaxCompute database. Dataphin needs to calculate the promotion-based sales amount for each member and the celebrities each member follows. The group will then determine the activity plan.

40.6. Limits

None.

40.7. Concepts

Business unit

A business unit is used to define the name and business space of a data warehouse. If your business only involves retail, and the systems in the business are less isolated, you only need to build one business unit: retail.

Global object

A global object is a global concept. By defining global objects, you can universally reference the definitions of global concepts and ensure consistency throughout the entire system.

Project management

A project is a physical space division that allows users to isolate developers from resources. After setting a name for a project, you can start data modeling and development in the project.

Physical data source

You can register your physical databases to Dataphin. Physical databases serve as the underlying data sources for projects and data synchronization.

Dimension

A dimension is a statistical object. It is an entity that actually exists. By creating a dimension, you can standardize your business entities (or master data) during architectural design to ensure that they are unique.

Business process

A business process is a collection of all events in a business activity. By creating a business process, you can standardize a type of transaction event in business to ensure that it is unique.

Logical dimension table

One logical dimension table corresponds to one dimension. A logical dimension table stores dimension attributes that describe facts. Logical dimension tables are used to extract details of common objects from business data.

Logical fact table

A logical fact table models a specific business process and provides detailed information of transactions in the business process. Logical fact tables are used to extract details of common transactions from business data.

Atomic metric and business filter

An atomic metric and business filter are the computing logic and attributive limitation commonly used in business. An atomic metric and business filter are expressions formulated based on fields in a logical table. These are reusable common data elements extracted to calculate aggregate data.

Derived metric

A derived metric is a commonly used statistical metric. It is used to aggregate the data of an object group in a specific range during a time period. Therefore, a derived metric is defined by the time period (statistical period), statistical object (statistic granularity), range (business filter), and calculation method (atomic metric). After specifying the preceding elements, you need to set a name and a display name for the derived metric to complete metric creation. For example, you can define the promotion-based sales amount for each member in a quarter (Q1, Q2, Q3, and Q4) as a derived metric. You can also add other conditions as required.

41.Elasticsearch (on ECS)

41.1. What is Elasticsearch?

Elasticsearch is a distributed search and data analytics service based on Lucene. It provides a distributed multi-tenant search engine that supports full text queries. This engine is based on a RESTful Web interface. Elasticsearch is developed based on Java. It is released as an open source product that complies with the Apache license terms and conditions. Elasticsearch is a mainstream search engine for enterprises. Elasticsearch is designed to serve cloud computing for real-time search. It is stable, reliable, fast, and easy to install and use.

Apsara Stack Elasticsearch provides two open source versions: Elasticsearch V5.5.3 and Elasticsearch V6.3.2. Apsara Stack Elasticsearch is designed to serve users in data search, data analytics, and other scenarios. Based on open source Elasticsearch, Apsara Stack Elasticsearch also supports enterprise-class permission management.

The default plug-ins provided by Apsara Stack Elasticsearch include but are not limited to the following:

- **IK analyzer**: an open source and lightweight Chinese analysis kit based on Java. The IK analyzer plug-in is very popular in open source communities for Chinese tokenization.
- **Smart Chinese analysis plug-in**: the default Lucene Chinese tokenizer.
- **ICU analysis plug-in**: a Lucene ICU tokenizer. ICU is a set of stable, tested, powerful, and easy to use libraries, providing Unicode and globalization support for applications.
- **Japanese (Kuromoji) analysis plug-in**: a Japanese tokenizer.
- **Stempel (Polish) analysis plug-in**: a French tokenizer.
- **Mapper attachments type plug-in**: an attachment-type plug-in which can parse files of different types into strings based on the Tika library.

41.2. Benefits

This topic describes the benefits provided by Apsara Stack Elasticsearch.

Apsara Stack Elasticsearch provides the following benefits:

- Near-real-time data search and analytics

  Supports distributed storage and search of large volumes of data, enables real-time analytics and search of petabytes of data, and responds within a few milliseconds.

- Auto scaling

  Allows you to scale out Elasticsearch clusters. Physical machines can be easily expanded without interrupting services.

- Visualized data search and analytics

  Provides a user-friendly operations and management (O&M) platform for indexes and clusters that are used for full-text search. You can use this platform to monitor the status of indexes and servers in real time. This platform supports web-based display of basic server metrics.

- SQL queries

  Allows you to execute SQL statements and use various combinations of conditions to search for data. After data is stored in Elasticsearch, you do not need to create additional indexes.

- Access control
  - Allows you to manage clusters in a centralized manner and perform dynamic configuration and management, resource isolation, and resource usage statistics.
  - Allows you to manage multiple levels of permissions and tenants in the Apsara Stack Cloud Management (ASCM) console.
  - Supports the management of data access permissions, including logon permissions, table creation permissions, read/write permissions, and whitelist-related permissions.
  - Allows you to use the ASCM console to manage administrative permissions, including administrator classification.
  - Allows you to use the ASCM console to manage user permissions in a centralized manner. You can manage the access control features of all components in the system. You can also block common users from querying access control details and simplify access control for administrators. This improves the usability and user experience of access control.
- Data backup
  - Allows you to back up full or incremental data and restore data from snapshots.
  - Allows you to back up data for clusters in different data centers. This meets the requirements for mutual data backup among multiple data centers.
  - Allows you to manage data backup processes in a visualized manner.
- Easy deployment and maintenance

  Supports automated deployment with no O&M costs and provides a system monitoring module.
- Visualized data analytics

  Integrates the Kibana module for visualized data analytics and background management.
- Chinese tokenization

  Integrates mainstream plug-ins that are used for Chinese tokenization, such as the third-party IK analyzer plug-in and the plug-in developed by DAMO Academy.
- Scalability

  Allows you to scale out an Elasticsearch cluster to hundreds of nodes and upgrade or downgrade these nodes as needed.
- Data snapshots

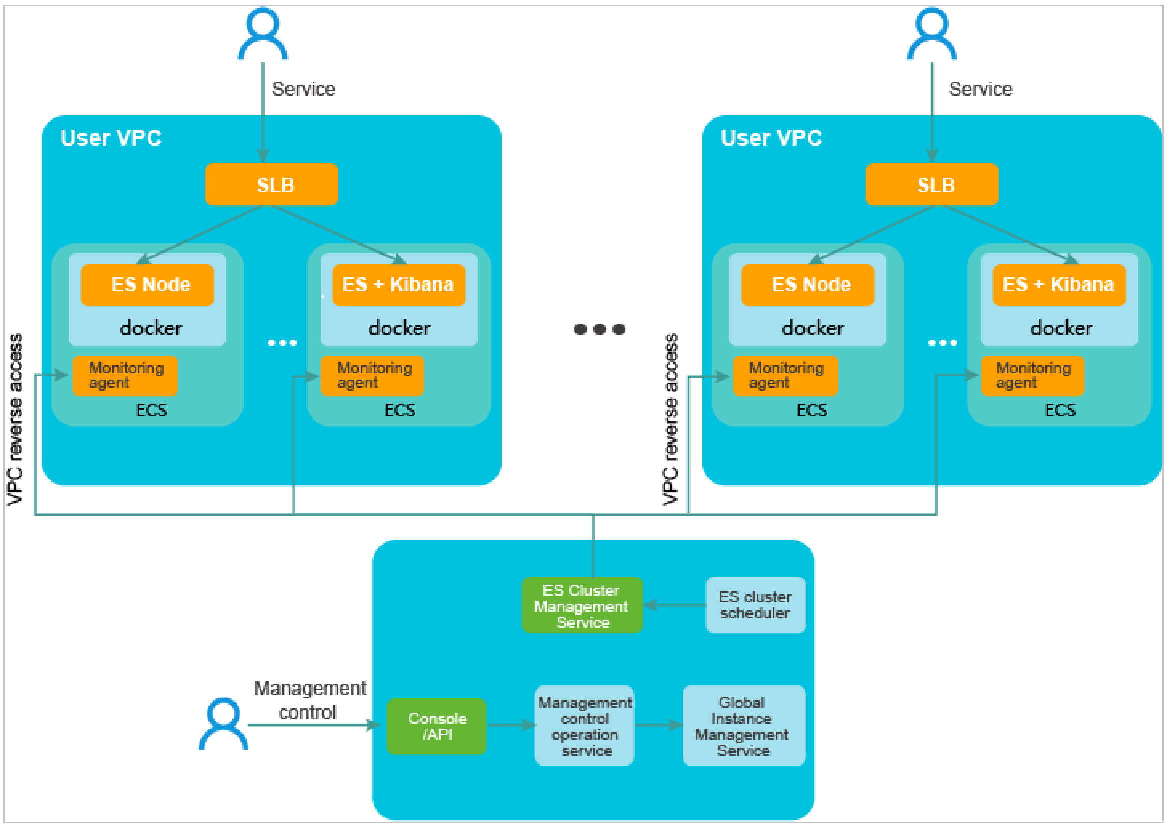
  Supports snapshot backup and restoration by using Object Storage Service (OSS). The data snapshot technology of the distributed file system allows data to be backed up and quickly restored from snapshots. This ensures data reliability.
- Technical support

  Offers 24/7 technical support and provides product documentation and training services.

41.3. Architecture

This topic describes the architecture of Apsara Stack Elasticsearch.

The following figure shows the architecture. In this figure, the procedure of creating an Apsara Stack Elasticsearch cluster is used as an example.



You can submit the configuration of the Elasticsearch cluster that you want to create from the Apsara Stack Cloud Management (ASCM) console or by calling the Elasticsearch API.

1. Select an Elastic Compute Service (ECS) instance. This ECS instance is used as an Elasticsearch node and provides storage space.
2. The governance service retrieves the instance and storage space information from ECS, saves your request to the database, and then submits the request to the global instance management service.
3. The global instance management service creates a configuration file for the Elasticsearch cluster based on the request type and submits the file to the Elasticsearch cluster management service.
4. The Elasticsearch cluster management service is an offline processing system that runs a task state machine based on the request type. The task state machine runs until the task reaches its desired state.

   When you create an ECS instance, the Elasticsearch cluster management service labels the ECS instance, connects it to a Virtual Private Cloud (VPC), and configures load balancing. Then, the service designates the cluster scheduler to manage the ECS instance. The cluster scheduler creates Elasticsearch and Kibana processes on the ECS instance.

   The Elasticsearch and Kibana processes run in containers on the ECS instance. The monitor agent, an independent process, collects monitoring metrics and sends them to Cloud Monitor by using Log Service. Elasticsearch clusters are isolated by VPCs. The governance service uses port mapping to establish reverse connections to your clusters for cluster management.

41.4. Features

This topic describes the features of Apsara Stack Elasticsearch.

- Distributed data search and analytics engine

  Data search: For example, you can search for data on a website or in an IT system.

  Data analytics: For example, you can obtain the top 10 best-selling toothpaste brands over the last seven days on e-commerce websites or obtain the top 3 news sections that have the most visits over the last month on news websites.

- Full-text search, structured search, and data analytics

  Full-text search: For example, you can search for commodities whose names contain the toothpaste keyword.

  Structured search: For example, you can search for commodities that are categorized as household chemicals.

  Data analytics: For example, you can count the number of commodities under each commodity category.

- Near-real-time processing of large amounts of data

  Distributed architecture: Elasticsearch automatically distributes large amounts of data to multiple servers.

  Processing of large amounts of data: Elasticsearch uses these servers to store and search for data.

  Near-real-time processing: Elasticsearch responds to a data search or analytics request within a few seconds.

- Support for ECS instances with local disks

  Q5PN54S1 is supported. You are allowed to create ECS instances of the ecs.d2-gab.4xlarge instance type that use local disks. Such an ECS instance offers 16 vCPUs, 64 GiB of memory, and six 1.2-TiB SATA disks. This ECS instance provides a total of 6.4 TiB of storage space.

  Scenario: Users, such as the Ministry of Public Security, can use such ECS instances to create Elasticsearch clusters. These ECS instances can be reused.

- Meeting the full-text search requirements of the Ministry of Public Security

  Scenario: full-text search in the public security industry.

  Usage: Use standard SDK for Java of the Ministry of Public Security or a RESTful API to access Elasticsearch clusters. You can use an automated test script to test the interface that corresponds to this feature.

41.5. Scenarios

This topic describes the common scenarios of Apsara Stack Elasticsearch.

- Website search based on the powerful full-text search feature.

  For example, you can submit a question about a program exception on the program exception forum. Then, other readers will answer your question or discuss with you about the question. The full-text search feature also enables you to search for related questions or answers.

- Log or transaction data search. Elasticsearch can be used to analyze your business development trends, retrieve log data, and analyze performance bottlenecks, running, or the development of your business system.

For example, Elasticsearch can be used to analyze the logs of user behavior, such as clicks, views, favorites, and comments. It can also be used to analyze social media data, such as comments about news on news websites. Comments from the public are pushed to the authors of the news.

- Warning. Elasticsearch continuously monitors and analyzes a metric and sends an alert when a specific threshold is exceeded.

  You can specify a price threshold for a commodity on a commodity price tracking website. If the price of the commodity is less than the threshold, you can receive a notification. For example, you can track the price of the Colgate toothpaste family pack. If the price of the family pack is less than CNY 50, the system sends you a notification.

- Business information analytics. You can use Elasticsearch to extract key information from millions of gigabytes of big data.

  For example, in the open source code management platform, users can search more than one hundred million code lines.

41.6. Limits

This topic describes the limits of Apsara Stack Elasticsearch, including the limits of disks and specifications.

Disk size limits

- Number of replicas: Each index must have a minimum of one replica.
- Indexing overheads: In most cases, indexing overheads are 10% greater than those of source data. The overheads of the _*all* parameter are not included.
- Space reserved by the Linux operating system: By default, the Linux operating system reserves 5% of the disk space for critical processes, system recovery, and disk fragments.
- Elasticsearch overheads: Elasticsearch reserves 20% of the disk space for internal operations, such as segment merging and logging.
- Security threshold overheads: Elasticsearch reserves at least 15% of the disk space as the security threshold.

Note

- Minimum required disk space = Volume of source data × 3.4

  ```
  Minimum required disk space = Volume of source data × (1 + Number of replicas) × Indexing ove
  rheads/(1 - Linux reserved space)/(1 - Elasticsearch overheads)/(1 - Security threshold overhead
  s)
  = Volume of source data × (1 + Number of replicas) × 1.7
  = Volume of source data × 3.4
  ```

- We recommend that you do not enable the *_all* parameter unless it is required by your businesses.
- Indexes that have this parameter enabled incur large overheads. Based on test results and user feedback, we recommend that you calculate the disk space of an Elasticsearch cluster by using the following formula:

  ```
  Minimum required disk space = Volume of source data × (1 + Number of replicas) × 1.7 × (1 + 0.5
  )
   = Volume of source data × 5.1
  ```

Specification limits

The performance of an Apsara Stack Elasticsearch cluster is determined by the specifications of each node in the cluster. Based on test results and user feedback, we recommend that you determine node specifications based on the following rules:

```
Maximum number of nodes per cluster = Number of vCPUs per node × 5
```

The maximum volume of data that a node in an Elasticsearch cluster can store depends on the scenario. Examples:

- Acceleration or aggregation on data queries

  ```
  Maximum volume of data per node = Memory per node (GiB) × 10
  ```

- Log data importing or offline data analytics

  ```
  Maximum volume of data per node = Memory per node (GiB) × 50
  ```

- General scenarios

  ```
  Maximum volume of data per node = Memory per node (GiB) × 30
  ```

Reference specifications

Specification	Maximum number of nodes	Maximum disk space per node in query scenarios	Maximum disk space per node in logging scenarios	Maximum disk space per node in general scenarios
2 vCPUs and 4 GiB of memory	10	40 GiB	200 GiB	100 GiB

Specification	Maximum number of nodes	Maximum disk space per node in query scenarios	Maximum disk space per node in logging scenarios	Maximum disk space per node in general scenarios
2 vCPUs and 8 GiB of memory	10	80 GiB	400 GiB	200 GiB
4 vCPUs and 16 GiB of memory	20	160 GiB	800 GiB	512 GiB
8 vCPUs and 32 GiB of memory	40	320 GiB	1.5 TiB	1 TiB
16 vCPUs and 64 GiB of memory	50	640 GiB	2 TiB	2 TiB

Shard limits

By default, an index is split into five shards. You can plan shards for each index of your Elasticsearch cluster as needed.

- For nodes with low specifications, we recommend that the size of each shard does not exceed 30 GB. For nodes with high specifications, we recommend that the size of each shard does not exceed 50 GB.
- For log analytics or extremely large indexes, we recommend that the size of each shard does not exceed 100 GB.
- The total number of shards and replicas is the same as or a multiple of the number of nodes.
- We recommend that you configure a maximum of five shards for an index on a node.

Resource limits

- Number of nodes: 2 to 50
- Disk size: 160 GiB to 2,048 GiB
- Specifications:
  - elasticsearch.sn2ne.xlarge (4 vCPUs and 16 GiB of memory)
  - elasticsearch.sn2ne.2xlarge (8 vCPUs and 32 GiB of memory)
  - elasticsearch.sn2ne.4xlarge (16 vCPUs and 64 GiB of memory)

41.7. Terms

This topic introduces the terms related to Apsara Stack Elasticsearch.

cluster

An Elasticsearch cluster consists of multiple nodes. Among these nodes, one node is elected as the dedicated master node. Each cluster has only one dedicated master node. Dedicated master node is a concept inside a cluster. One of the concepts in Elasticsearch is decentralization. Decentralization means that no dedicated master nodes exist. It is a concept outside a cluster. Communication with any node inside a cluster is equivalent to communication with the cluster.

shard

An index can be divided into multiple shards. These shards can be distributed among different nodes to support distributed searches. When you create an index, you must specify the number of shards for the index. After the index is created, you cannot change the number.

replica

Replicas refer to replica shards for an index. Each index can have more than one replica. Replicas provide the following benefits:

- Improved fault tolerance: When a shard on a node is damaged or lost, you can restore the shard from its replicas.
- Improved search efficiency: Elasticsearch automatically balances the load of queries among replicas.

recovery

Data recovery (or data redistribution) is the process of redistributing shards for a node. This ensures the integrity of data when the node joins or leaves a cluster, or when the node recovers from a failure.

gateway

A gateway is used to store snapshots of indexes. By default, a node stores all the indexes in its memory. When the node memory is full, the node stores the indexes on local disks. When an Elasticsearch cluster is rebooted, its index data is restored from the snapshots that are stored on the gateway. Elasticsearch supports multiple types of gateways, including the local file system (default), distributed file system, Hadoop Distributed File System (HDFS), and Amazon Simple Storage Service (S3).

discovery.zen

discovery.zen is an automatic node discovery mechanism. Elasticsearch is a peer to peer (P2P) system that sends broadcasts to discover nodes. Nodes communicate with each other by using multicast and P2P technologies.

transport

Transport refers to the method that is used by an Elasticsearch cluster or the nodes in the cluster to communicate with clients. By default, TCP is used. You can integrate plug-ins into Elasticsearch to use other protocols, such as HTTP over JSON, Thrift, Servlet, Memcached, and ZeroMQ.

42.Elasticsearch (on k8s)

42.1. What is Apsara Stack Elasticsearch?

Open source Elasticsearch is a Lucene-based, distributed, real-time search and analytics engine. It is a product released under the Apache License. Elasticsearch is a popular search engine for enterprises. It provides distributed services, allowing you to store, query, and analyze large amounts of datasets in near real time. Elasticsearch is typically used as a basic engine or technology to support complex queries and high-performance applications.

Apsara Stack Elasticsearch provides fully-managed Elasticsearch services. It supports multiple versions of open source Elasticsearch and is compatible with all open source Elasticsearch features. Apsara Stack Elasticsearch offers an optimized kernel and provides the multi-tenancy, high availability, and auto scaling features. In addition to the features of open source Elasticsearch, Apsara Stack Elasticsearch allows you to create a cluster in a visualized manner, use Migration Assistant to migrate data, manage repositories, create snapshots, manage plug-ins, and perform O&M operations.

42.2. Benefits

This topic describes the benefits provided by Apsara Stack Elasticsearch.

Apsara Stack Elasticsearch provides the following benefits:

- Near-real-time data search and analytics

  Supports distributed storage and search of large volumes of data, enables real-time analytics and search of petabytes of data, and responds within milliseconds.

- Auto scaling

  Allows you to scale out Elasticsearch clusters. Physical machines can be easily expanded without interrupting services.

- Visualized data search and analytics

  Provides a user-friendly operations and management (O&M) platform for indexes and clusters that are used for full-text search. You can use this platform to monitor the status of indexes and servers in real time. This platform supports web-based display of basic server metrics.

- SQL queries

  Allows you to execute SQL statements and use various combinations of conditions to search for data. After data is stored in Elasticsearch, you do not need to create additional indexes.

- Access control

  - Allows you to manage clusters in a centralized manner and perform dynamic configuration and management, resource isolation, and resource usage statistics.
  - Allows you to manage multiple levels of permissions and tenants in the Apsara Stack Cloud Management (ASCM) console.
  - Supports the management of data access permissions, including logon permissions, table creation permissions, read/write permissions, and whitelist-related permissions.
  - Allows you to use the ASCM console to manage administrative permissions, including administrator classification.

  - Allows you to use the ASCM console to manage user permissions in a centralized manner. You can manage the access control features of all components in the system. You can also block common users from querying access control details and simplify access control for administrators. This improves the usability and user experience of access control.

- Data backup
  - Allows you to back up full or incremental data and restore data from snapshots.
  - Allows you to back up data for clusters in different data centers. This meets the requirements for mutual data backup among multiple data centers.
  - Allows you to manage data backup processes in a visualized manner.

- Data migration
  - Allows you to use Migration Assistant to migrate full or incremental data.
  - Allows you to manage data migration processes in a visualized manner.

- Easy deployment and maintenance

  Supports automated deployment with no O&M costs and provides a system monitoring module.

- Visualized data analytics

  Integrates the Kibana module for visualized data analytics and background management.

- Chinese tokenization

  Integrates mainstream plug-ins that are used for Chinese tokenization, such as the third-party IK analyzer plug-in and the plug-in developed by DAMO Academy.

- Scalability

  Allows you to scale out an Elasticsearch cluster to hundreds of nodes and upgrade or downgrade these nodes as needed.

- Data snapshots

  Supports snapshot backup and restoration by using Object Storage Service (OSS). The data snapshot technology of the distributed file system allows data to be backed up and quickly restored from snapshots. This ensures data reliability.

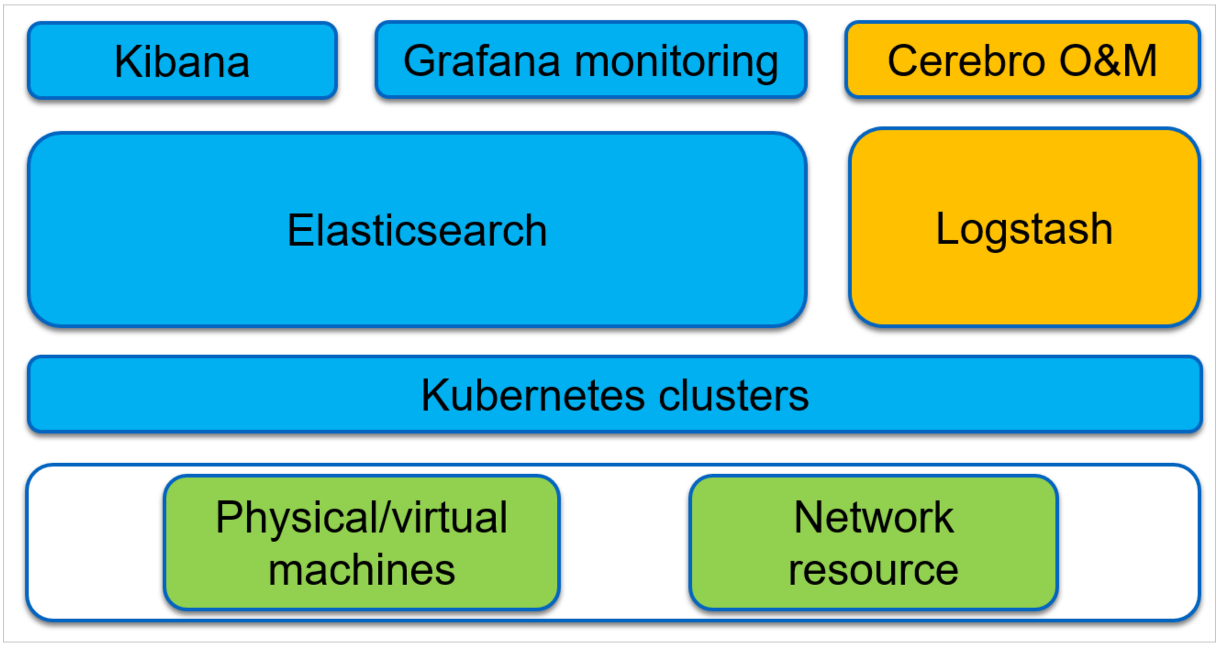
- Technical support

  Offers 24/7 technical support and provides product documentation and training services.

42.3. Architecture

This topic describes the architecture of Apsara Stack Elasticsearch.

The following figure shows the architecture.



Apsara Stack Elasticsearch is deployed on Kubernetes clusters. Kubernetes clusters can be deployed on physical or virtual machines. Then, you can create an Elasticsearch cluster, activate Kibana, and enable Grafana-based monitoring with one click on the operations and maintenance (O&M) platform. You can also activate Logstash on a Kubernetes cluster to import data and use Cerebro to perform O&M operations on Elasticsearch.

42.4. Features

The following table describes the features of Apsara Stack Elasticsearch.

Feature	Description
Multi-tenancy	You can create multiple Elasticsearch clusters in the Apsara Stack Operations (ASO) console. When you create a cluster, you can configure the logon password and number of nodes for it. You can also configure the number of CPUs, memory, and disk specifications for the nodes.
Search of petabytes of data	The distributed storage and index search technology allows you to analyze and search petabytes of data in near real time.
Auto scaling	You can scale out Elasticsearch clusters. Servers can be scaled without interrupting services.
Near-real-time search	After you import data into Elasticsearch clusters, you can analyze and search for data only a few milliseconds later.
Full-text search for large volumes of data	Apsara Stack Elasticsearch stores structured data and keyword information from full-text databases. It allows you to perform multidimensional information matching, filtering, inverted indexing, and full-text search on the data and information.

Feature	Description
Monitoring	Apsara Stack Elasticsearch provides a user-friendly operations and maintenance (O&M) platform for indexes and clusters that are used for full-text search. You can use this platform to monitor the status of indexes and servers in real time. This platform supports web-based display of basic server metrics.
API operations	Apsara Stack Elasticsearch supports API operations provided by open source Elasticsearch. These operations are used for data import, index creation, or data search.
Custom plug-ins and dictionary extension	You can extend tokenizers and dictionaries as needed. Third-party or custom tokenizers are supported. This meets the requirements for custom full-text search.
Aggregation	Apsara Stack Elasticsearch supports the aggregate operator pushdown feature, which facilitates the aggregation and analytics of search results.
Sorting	You can sort search results by the degree of correlation or by field, such as Time.
Comprehensive search	Apsara Stack Elasticsearch can respond to a comprehensive search request within milliseconds, a batch search request within seconds, and a comparison collusion request within minutes.
Chinese tokenization	You can use the Chinese tokenization plug-in developed by DAMO Academy to achieve exact search.
Data backup	The data snapshot technology is designed based on the distributed file system. This technology allows you to create snapshots and quickly restore data from the snapshots. This ensures data reliability. You can also use miniOSS to back up and restore snapshots.

42.4.1. Multi-tenancy

Apsara Stack Elasticsearch supports the multi-tenancy feature. This feature allows multiple businesses on the cloud platform to share an Elasticsearch service. It also ensures data isolation among the businesses.

Data isolation solution: Apsara Stack is deployed by using the Kubernetes container technology. Each tenant can create an Elasticsearch cluster. Each tenant can configure node and server specifications for their clusters based on actual business requirements. This ensures high data isolation and security.

42.4.2. Access control and security management

This topic describes the access control and security management features of Apsara Stack Elasticsearch.

The following methods are used to implement security management:

- Password protection, role-based access control, and IP address filtering: prevent unauthorized access.
- Identity authentication and SSL/TLS encryption: ensure data integrity.
- Maintenance auditing and tracking: allows you to obtain operations that are performed on your Elasticsearch clusters and information about personnel who perform the operations.

You can perform access control from the following perspectives:

- Add, delete, modify, and query permissions on indexes
- Access permissions on specific fields in a specified index
- Read permissions on specific documents in a specified index

42.5. Scenarios

This topic describes the common scenarios of Apsara Stack Elasticsearch.

- Website search based on the powerful full-text search feature.

  For example, you can submit a question about a program exception on the program exception forum. Then, other readers will answer your question or discuss with you about the question. The full-text search feature also enables you to search for related questions or answers.

- Log or transaction data search. Elasticsearch can be used to analyze your business development trends, retrieve log data, and analyze performance bottlenecks, running, or the development of your business system.

  For example, Elasticsearch can be used to analyze the logs of user behavior, such as clicks, views, favorites, and comments. It can also be used to analyze social media data, such as comments about news on news websites. Comments from the public are pushed to the authors of the news.

- Warning. Elasticsearch continuously monitors and analyzes a metric and sends an alert when a specific threshold is exceeded.

  You can specify a price threshold for a commodity on a commodity price tracking website. If the price of the commodity is less than the threshold, you can receive a notification. For example, you can track the price of the Colgate toothpaste family pack. If the price of the family pack is less than CNY 50, the system sends you a notification.

- Business information analytics. You can use Elasticsearch to extract key information from millions of gigabytes of big data.

  For example, in the open source code management platform, users can search more than one hundred million code lines.

42.6. Limits

This topic describes the limits of Apsara Stack Elasticsearch, including the limits of disks and specifications.

Disk size limits

- Number of replicas: Each index must have a minimum of one replica.
- Indexing overheads: In most cases, indexing overheads are 10% greater than those of source data. The overheads of the *_all* parameter are not included.
- Security threshold overheads: Elasticsearch reserves at least 15% of disk space as the security

threshold.

- Data compression ratio: You can use the default or optimal compression algorithm of Elasticsearch. The actual data compression ratio is determined by the algorithm that is used. In most cases, we recommend that you use 1.0 as the compression ratio. If you choose to use the optimal compression algorithm, we recommend that you use 0.5 as the compression ratio. You can determine whether to store the _source field in source data, set the tokenization type for fields, index documents, and enable DocValue based on your business scenarios. You can also optimize existing indexes on a regular basis.
- Space reserved by the operating system: By default, the operating system reserves 5% of disk space for critical processes, system recovery, and disk fragments.
- Elasticsearch overheads: Elasticsearch reserves 20% of disk space for internal operations, such as segment merging and logging.

In most cases, the index expansion rate is 3.4 times the size of source data.

```
Total disk space = Volume of source data × (1 + Number of replicas) × Indexing overheads/(1 - Security threshold overheads) × Compression ratio/(1 - Space reserved by the operating system)/(1 - Elasticsearch overheads)
= Volume of source data × (1 + Number of replicas) × 1.7
= Volume of source data × 3.4
```

In low-cost storage scenarios, the index expansion rate is 1.36 times the size of source data.

```
Total disk space = Volume of source data × (1 + Number of replicas) × (1 + Indexing overheads)/(1 - Security threshold overheads) × Compression ratio/(1 - Space reserved by the operating system)
 = Volume of source data × 1.36
```

Specification limits

The maximum volume of data that a node in an Elasticsearch cluster can store depends on the scenario. Examples:

- General scenarios

  ```
  Maximum volume of data per node = Memory per node (GiB) × 30
  ```

- Acceleration or aggregation on data queries

  ```
  Maximum volume of data per node = Memory per node (GiB) × 16
  ```

- Log data importing or offline data analytics

  ```
  Maximum volume of data per node = Memory per node (GiB) × 50
  ```

Reference specifications

Specification	Maximum number of nodes	Maximum disk space per node in general scenarios	Maximum disk space per node in query scenarios	Maximum disk space per node in logging scenarios
4 CPUs and 16 GiB of memory	100	512 GiB	256 GiB	2 TiB

Specification	Maximum number of nodes	Maximum disk space per node in general scenarios	Maximum disk space per node in query scenarios	Maximum disk space per node in logging scenarios
8 CPUs and 32 GiB of memory	200	1 TiB	512 GiB	4 TiB
16 CPUs and 64 GiB of memory	300	2 TiB	1 TiB	8 TiB

Shard limits

By default, an index is split into five shards. You can plan shards for each index of your Elasticsearch cluster as needed.

- For nodes with low specifications, we recommend that the size of each shard does not exceed 30 GB. For nodes with high specifications, we recommend that the size of each shard does not exceed 50 GB.
- For log analytics or extremely large indexes, we recommend that the size of each shard does not exceed 100 GB.
- The total number of shards and replicas is the same as or a multiple of the number of nodes.
- We recommend that you configure a maximum of five shards for an index on a node.

Resource limits

- Number of nodes: 3 to 300
- Disk size: 1 GiB to 8,000 GiB
- Specifications:
  - 1:4 specifications: 2 CPUs and 8 GiB of memory, 4 CPUs and 16 GiB of memory, 8 CPUs and 32 GiB of memory, and 16 CPUs and 64 GiB of memory
  - 1:2 specifications: 4 CPUs and 8 GiB of memory, 8 CPUs and 16 GiB of memory, 16 CPUs and 32 GiB of memory, and 32 CPUs and 64 GiB of memory

42.7. Terms

This topic introduces the terms related to Apsara Stack Elasticsearch.

cluster

An Elasticsearch cluster consists of one or more nodes. All nodes in a cluster work together to store data. This enables the cluster to provide joint indexing and search capabilities. Each cluster has a unique name. The default cluster name is elasticsearch. Before a node joins a cluster, the name of the cluster is required.

You must make sure that clusters in different environments use different names. Otherwise, you may add nodes to the wrong cluster.

A cluster that contains only one node is allowed.

node

A node runs on a server in an Elasticsearch cluster. Nodes are used to store data and support indexing and query activities in the cluster. Same as a cluster, each node has a unique name. By default, a random UUID is assigned to a node as its name when the node is started. UUID is short for universally unique identifier. You can also assign a custom name to the node. Node names are required to complete management work. You must determine which node runs on a specific server based on the name of the node.

You can add a node to a cluster with a specified name. By default, nodes are added to a cluster named elasticsearch. If these nodes can discover each other in a network, a cluster named elasticsearch is automatically created after you start the nodes.

The number of nodes that a cluster can contain is not limited. If no Elasticsearch nodes are running in your network, after you start a node, a single-node cluster named elasticsearch is created.

index

An index is a set of documents that have similar features. It is similar to a relational database. For example, you can create three indexes to store customer data, commodity catalog data, and order data, respectively. In most cases, a name is assigned to an index to identify the index. Index names must be in lowercase. When you index, query, update, or delete a document, you must specify the name of the index to which the document belongs.

type

A type is a logical class or partition of an index. In Elasticsearch V5.X, you are not allowed to create multiple types in an index. In Elasticsearch V6.X, you can create only one type in an index. In Elasticsearch V7.X, the names of index types can only be _doc. For example, the type of an index can be user or blog.

document

A document is a basic information unit that can be indexed. It is similar to a row in a table of a relational database. For example, you can create a document for a customer or commodity. A document is a JSON object. The number of documents that are stored in an index is not limited and these documents must be indexed.

shard

An index can be divided into multiple shards. These shards can be distributed among different nodes to support distributed searches. When you create an index, you must specify the number of shards for the index. After the index is created, you cannot change the number.

replica

Replicas refer to replica shards for an index. Each index can have more than one replica. Replicas provide the following benefits:

- Improved fault tolerance: When a shard on a node is damaged or lost, you can restore the shard from its replicas.
- Improved search efficiency: Elasticsearch automatically balances the load of queries among replicas.

recovery

Data recovery (or data redistribution) is the process of redistributing shards for a node. This ensures the integrity of data when the node joins or leaves a cluster, or when the node recovers from a failure.

river

A river is used to import data from external storage, such as databases, to Elasticsearch. You can use river plug-ins in Elasticsearch to synchronize data. A river plug-in reads data from the river and creates indexes for the data in Elasticsearch. Apache CouchDB, RabbitMQ, Twitter, and Wikipedia river plug-ins are supported.

gateway

A gateway is used to store the snapshots of indexes. By default, a node stores all the indexes in its memory. When the node memory is full, the node stores the indexes on local disks. When an Elasticsearch cluster is rebooted, its index data is restored from the snapshots that are stored on the gateway. Elasticsearch supports multiple types of gateways, including the local file system (default), distributed file system, Hadoop Distributed File System (HDFS), and Amazon Simple Storage Service (S3).

discovery.zen

discovery.zen is an automatic node discovery mechanism. Elasticsearch is a peer to peer (P2P) system that sends broadcasts to discover nodes. Nodes communicate with each other by using multicast and P2P technologies.

transport

Transport refers to the method that is used by an Elasticsearch cluster or the nodes in the cluster to communicate with clients. By default, TCP is used. You can integrate plug-ins into Elasticsearch to use other protocols, such as HTTP over JSON, Thrift, Servlet, Memcached, and ZeroMQ.