Alibaba Cloud
Apsara Stack Enterprise

Technical Whitepaper

Version: 1807

Issue: 20180731

Legal disclaimer

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

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.
5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, 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 the Alibaba Cloud website, product programs, or content shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion , or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos , marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).

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

Generic conventions

Table -1: Style conventions

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

Contents

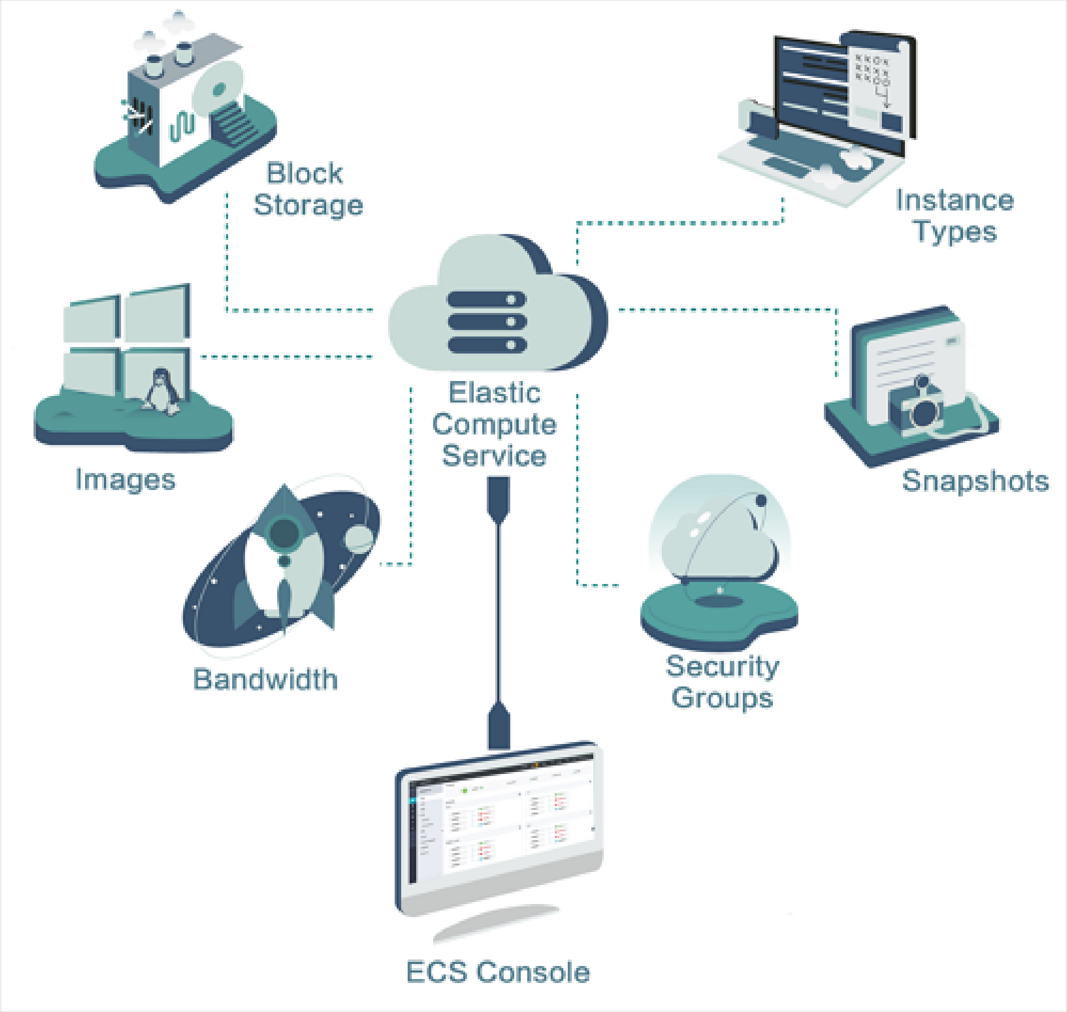
1 Elastic Compute Service (ECS)

1.1 What is ECS

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

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

Figure 1-1: Concept of an ECS instance



1.2 Architecture

ECS is made up of a virtualization platform, a distributed storage and control system, and an O&M and monitoring system.

1.2.1 Virtualization platform and distributed storage

Virtualization is the foundation of ECS. Alibaba Cloud adopts the KVM virtualization technology to virtualize the physical resources, thus providing elastic computing services via the virtualized resources.

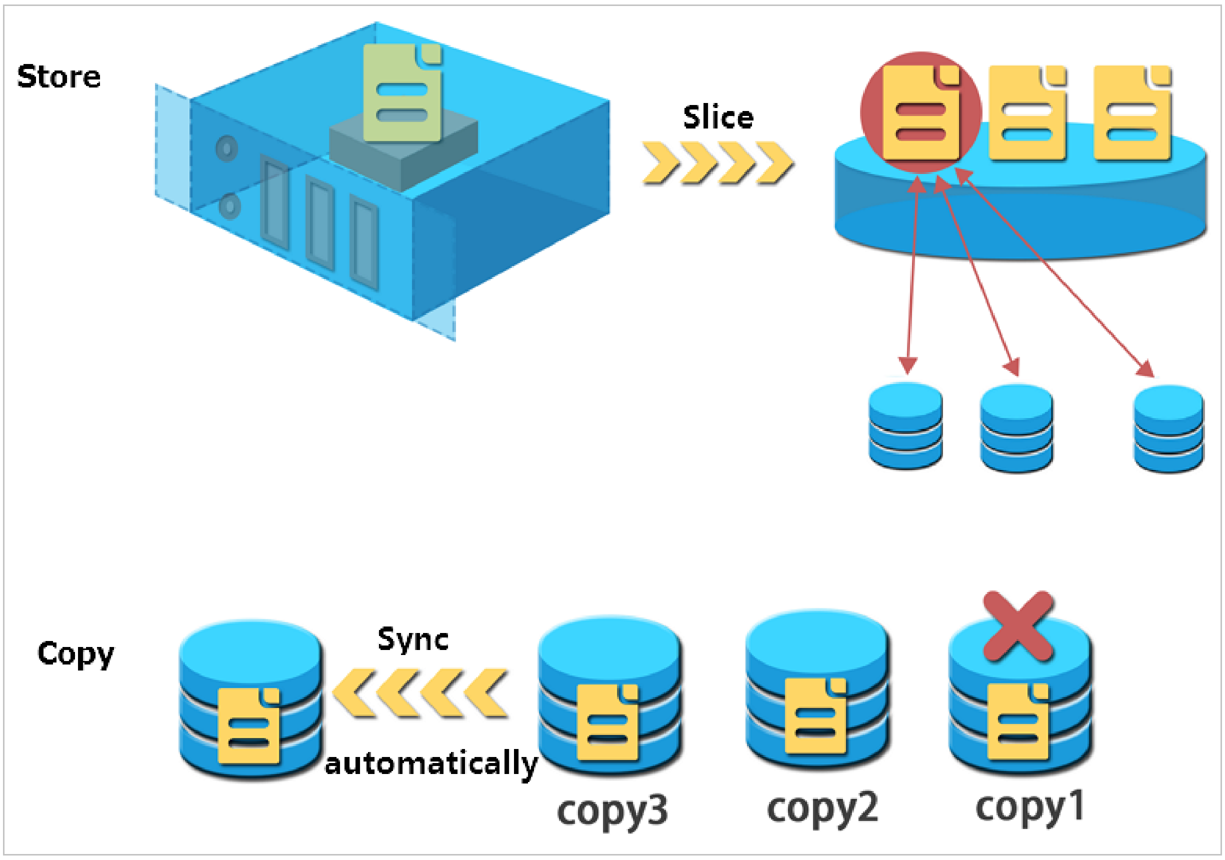
ECS contains two key virtualized modules: one for computing resources, and the other for storage resources.

- Computing resources refer to the CPU, memory, bandwidth, and other components of a physical server and are virtualized before being assigned to ECS. The computing resources of an ECS instance can only be located on the same physical server. If the resources of a physical server are used up, you have to create ECS instances on another physical server. With the resource QoS, ECS instances on the same physical server do not affect the running of each other.

- The storage module uses a large-scale distributed storage system. After the storage resources of an entire cluster are virtualized, the resources are bundled together and provided as an external service. Data for a single ECS instance is saved throughout the entire cluster. In the distributed storage system, all data is saved in triplicate. This way, if one copy is damaged, the data can be automatically recovered from another copy.

Triplicate technology is shown in *Figure 1-2: Distributed storage utilizing triplicate technology*

Figure 1-2: Distributed storage utilizing triplicate technology



1.2.2 Control system

As the core of the ECS platform, the control system determines the physical server where an ECS instance starts. In addition, all the information and functions of ECS are processed and maintained through the control center in a centralized way.

The control system is composed of the following four modules:

- **Data collection**

  This module collects data from the entire virtualization platform, including usage information for computing resources, storage resources, and network resources. The data collection module allows you to centrally monitor and manage the usage of cluster resources. Furthermore, it serves as the basis for resource scheduling.
- **Resource scheduling system**

  This module determines where an ECS instance starts. When you create an ECS instance, it rationally schedules the ECS instance based on physical server resource loads. This module can determine where to restart an instance if any fault occurs in an ECS instance.
- **ECS management module**

  This module manages and controls the ECS instances, for example, starting, stopping, and restarting ECS instances.
- **Security control module**

  This module monitors and manages the network security of the entire cluster.

1.3 Features

As the core of the elastic computing products, ECS is designed to provide the computing services for users. It taks only a few minutes to create and start an ECS instance. Moreover, once an ECS instance is created, it has specific system configuration. Compared to the traditional servers, ECS helps you improve the efficiency of delivering services considerably.

ECS instances are used the same way as the traditional hosted physical servers. You have full control over your ECS instances and can perform operations on them through the remote approach or the API approach (console).

The computing capabilities of ECS instances can be expressed in terms of virtual CPUs and virtual memory. ECS disk storage capabilities are measured by the capacity of available cloud disks. Unlike the traditional servers, ECS allows you to make more flexible machine configuration based on your needs. That is, if the current ECS instance configuration cannot meet the business needs, you can change the configuration at any time.

The ECS life cycle begins with ECS instance creation and ends after you release it. Once an ECS instance is released, all its data is irrevocably deleted.

The ECS console of Apsara Stack offers easy access to the following information areas:

- **Resources**

  In this area, you can view the number of instances created and the number of running instances. You can also view the quantity and distribution of ECS instances in respective zones .
- **Instances**

  In this area, you can do the following:

  - View and manage the created instances.
  - Start, stop, restart, release, and log on to VNC.
  - Replace system disks, reset your password, and change the configuration.
  - View the basic and configuration information of the instances.
- **Disks**

  In this area, you can do the following:

  - View and manage the created disks.
  - Reinitialize a disk, create snapshots, set an automatic snapshot policy for disks, release disks, and attach or detach disks.
  - View the basic information of disks.
- **Images**

  In this area, you can do the following:

  - View and manage information of the created or shared images.
  - Copy, share, and delete images.
- **Snapshots**

  In this area, you can do the following:

  - View and manage the created snapshots.
  - Roll back a disk, create custom images, and delete snapshots.
- **Automatic snapshot policy**

  In this area, you can do the following:

  - View and manage the configured automatic snapshot policy.
  - Configure the automatic snapshot policy in batch.
  - Change and delete the automatic snapshot policy.
- **Security group**

  In this area, you can do the following:

  - View and manage the created security groups.
  - Create, change and delete (individually or in batch) security groups.
  - View the instances and rules in a security group.
- **Elastic Network Interface**

  In this area, you can do the following:

  - View and manage the created Elastic Network Interfaces (ENIs).
  - Create, change and delete ENIs.
  - Bind and unbind instances.
- **Deployment set**

  In this area, you can do the following:

  - View and manage the created deployment sets.
  - Create, change and delete the deployment sets.
  - View the basic information of deployment sets.

2 Object Storage Service (OSS)

2.1 What is OSS

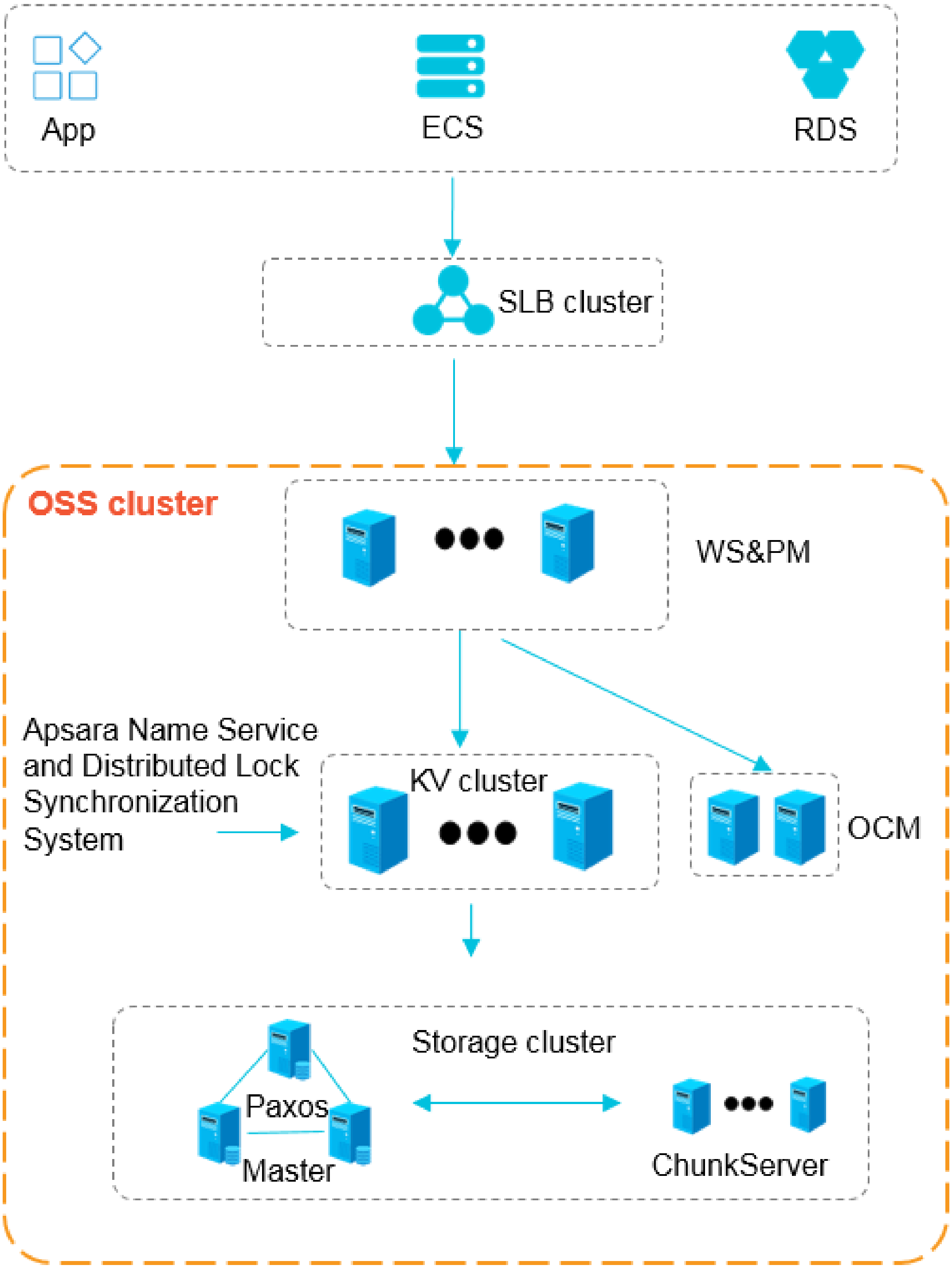
Alibaba Cloud Object Storage Service (OSS) is a storage service that enables you to store, back up, and archive any amount of data in the cloud. OSS is a cost-effective, highly secure, and highly reliable cloud storage solution. It uses RESTful APIs and is designed for 99.999999999% (11 nines) durability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

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

2.2 Architecture

Object Storage Service (OSS) is a storage solution built on the Alibaba Cloud Apsara platform. It is based on infrastructure such as the Apsara distributed file system and distributed job scheduling system, and provides distributed scheduling, high-speed networks, and distributed storage features. *Figure 2-1: Architechture of OSS* shows the architecture of OSS.

Figure 2-1: Architechture of OSS



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

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

2.3 Features

The OSS console supports the following functions and operations:

- Bucket overview: If you use HTTP to access OSS,all of your buckets are displayed by default .
- Set and query bucket access permissions, such as:
  - Private: Only the creator or an authorized user of the bucket can read and write objects in the bucket. No other users can access any objects in the bucket without authorization.
  - Public-read: Only the creator of the bucket can perform write operations on the objects in the bucket. Other users (including anonymous users) can perform read operations on the objects.
  - Public-read-write: Anyone (including anonymous users) can perform read and write operations on the objects in the bucket. The fees incurred by these operations are borne by the creator of the bucket. Use this permission with caution.
- Create/Delete buckets: Each user can create up to 10 buckets. If this limit is reached, further attempts to create buckets return an error message. The name of a new bucket must comply with the bucket naming conventions. A bucket is successfully created when the bucket name meets the naming conventions, and the system returns a successful message. If the name of bucket to create already exists, and the requester is its owner, the original bucket is retained (that is, it is not overwritten). If the name of a bucket to create already exists and the requester is not its owner, an error message is returned. To delete a bucket, you must have permission to delete the bucket, and the bucket must be empty.

List all objects in a bucket

This operation lists all the objects in the bucket of the specified name. To perform this operation, you must have the relevant permissions for the specified bucket. If the bucket does not exist, an error message is returned.

You can specify the prefix of the objects to be returned. You can also set the maximum number of objects to be returned. A maximum of 1,000 objects can be returned.

Upload/Delete object

This operation uploads an object to the specified bucket. The object upload succeeds if you have the corresponding permissions for the specified bucket. If an object of the same name already exists in the bucket, the new object overwrites the original object. You can delete a specified object, provided you have the corresponding permissions to delete the object.

Get object or object metadata

To retrieve specific information contained in an object or the object's metadata, you must have the corresponding permissions to obtain such information of this object.

Access object

OSS allows you to access objects using URLs.

Log and monitoring operations

When the server access logging feature is activated for a bucket, OSS pushes the resulting logs on an hourly basis. You can then query bucket, traffic, and request log information in the OSS console.

OSS VPC access control

You can create tunnels (Single Tunnel) between OSS and VPC to access resources stored in OSS from VPC.

3 Table Store

3.1 What is Table Store

3.1.1 Technical background

Data Technology era

As Internet usage continues to rise, data collected by applications across various industries and fields is offering significant insight into features and trends of data, including:

- The amount of data that needs to be stored and processed is increasing due to rising use of applications, such as microblogging, social events, image sharing, access logs and more.
- Structured data storage poses high requirements for concurrent writing as phones and other mobile devices usage become mainstream, and IoT device usage increase.
- Without schema, data tends to be half structured and data fields change dynamically.
- User access is characterized by hot spots and peak hours. For example, during promotional activities, user access rates surge to peak levels.
- Constant access to mobile Internet and availability requirements for Internet applications make users unable to comprehend unstable services caused by failures, or even planned service downtime.
- A large amount of data significantly raises the requirements for compute analysis.

Challenges of traditional IT software solutions

Traditional IT software solutions face the following challenges:

- Scalability

  Traditional software, such as relational databases, are incapable of handling such fast-growing data, leading to bottlenecks existing in both data writing throughput and access efficiency. In traditional database solutions, databases and tables are partitioned manually and statically, which is a time-consuming process. Furthermore, if nodes must be added to increase capacity , you need to repartition and migrate existing data, during which you cannot fully guarantee service performance, stability, or availability.
- Data model alteration

  In a traditional database, data is processed in accordance with a schema, and the number of columns in data is fixed. This means that any frequent changes to the settings of table schema and column number (that may result from Internet application requirements) will greatly impact service availability.

- Quick resizing

  In traditional solutions, business access load is considered as being stable, and the system is not required to quickly scale resources out or in. However, if data needs repartitioning and migration, the workload is intensive and inefficient. Additionally, once business loads return to normal, additional machines need to be removed, and data migrated again, in order to avoid costs incurred by low resource utilization. The entire process is time-consuming and complex.

- O&M guarantee

  With traditional software solutions, service recovery in case of hardware (network devices or disks) failure, hardware replacement, software upgrade, configuration tuning and update need to be performed manually. To make these processes transparent to applications and avoid service availability decrease, customers need a special engineering team to achieve system operation and maintenance. Therefore, workload caused from recruitment and fund investment is a huge challenge to enterprises.

- Computer bottleneck

  In existing business systems, we normally use OnLine Transaction Processing (OLTP) systems to process and analyze data, such as MySQL, Microsoft SQL Server, and other relational database systems. These relational database systems are adept at transaction processing , and maintain strong consistency and atomicity in data operations, thereby supporting frequent data insertion and modification. However, if the data volume exceeds the system processing capability, and reaches tens of millions or even billions of data records, or a complex computation process is needed, OLTP database systems will no longer be sufficient.

3.1.2 Table Store technologies

Table Store is a NoSQL data storage service built on Alibaba Cloud's distributed operating system , Apsara. Table Store partitions tables and dispatches data shards to different nodes to improve scalability. In event of a single hardware failure, Table Store quickly detects the faulty node using the heartbeat mechanism and migrates data shards from the defective node to a healthy node to continue service, thereby achieving rapid service backup.

Data partitioning and load balancing

The first column of primary keys in each row of a table is referred to as a Partition Key. Based on the Partition Key range, the system partitions a table into multiple shards that are evenly dispatched to different storage nodes. If data in a partition increases and exceeds a threshold, the partition is automatically divided into two small partitions to process data and access load. The two partitions are then dispatched to different nodes to achieve linear scaling.

Table Store can manage tables at the PB-level, and support millions of concurrent accesses.

Automatic backup

In the storage engine of Table Store, each node serves a number of data shards in different tables . A master node monitors shard distribution and dispatching, and the health of each service node . If a service node fails, the master node migrates data shards on this faulty node to other healthy nodes. The migration is logically performed, and does not involve physical entities, so services can rapidly recover in case of single node failure (achieving full restoration of services within several minutes).

Intra-city and remote disaster backup

To meet security and availability requirements of businesses, Table Store provides intra-city and remote disaster backup. Disaster backup is precise to the instance-level, which means any table operation on an active instance (including insertion, update, and deletion) is synchronized to the table of the same name in the standby instance. The synchronization duration between active and standby instance data depends on network environment of the active/standby cluster. In an ideal network environment, the synchronization duration is at the millisecond-level. Before manual switchover, you must stop resource access to the active cluster and wait for all data to be completely backed up. After the switchover, do not perform another switchover for one hour, and you must clear original cluster data and reset the standby cluster.

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

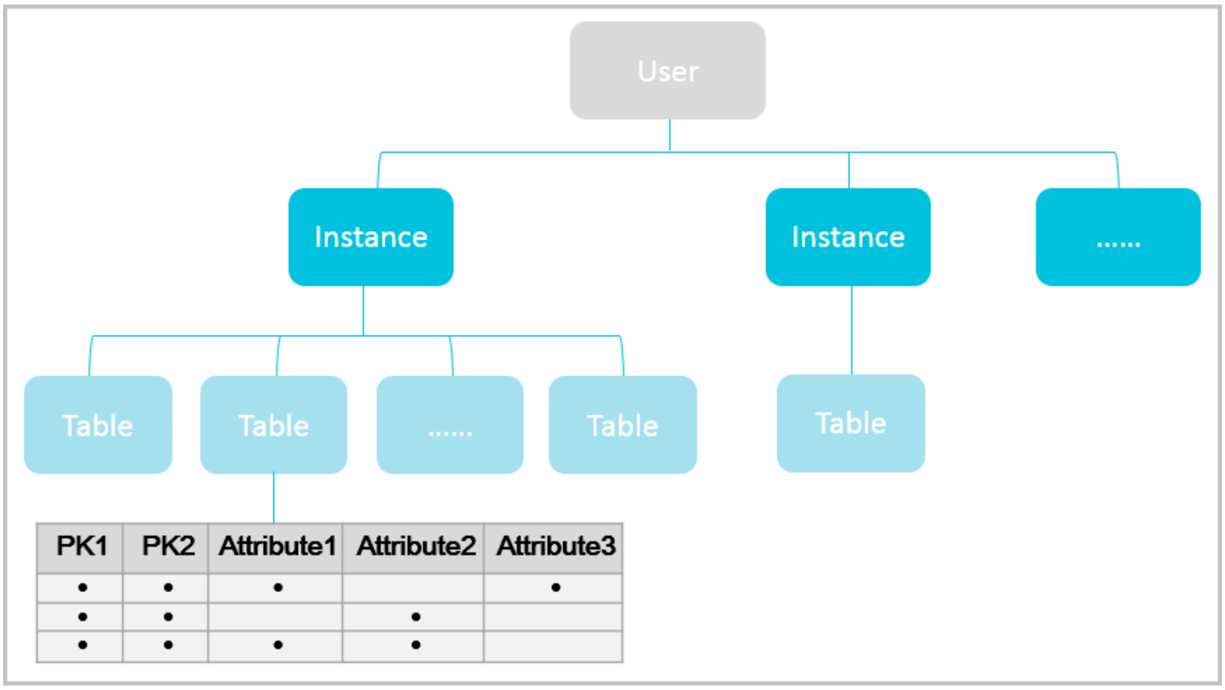
The RTO of Table Store is less than 2 minutes, the RPO less than 5 minutes, and the RCO is 1.

3.2 Functional performance

3.2.1 Users and instances

The following figure shows Table Store architecture in relation to the user and an instance.

Figure 3-1: User and instance architecture



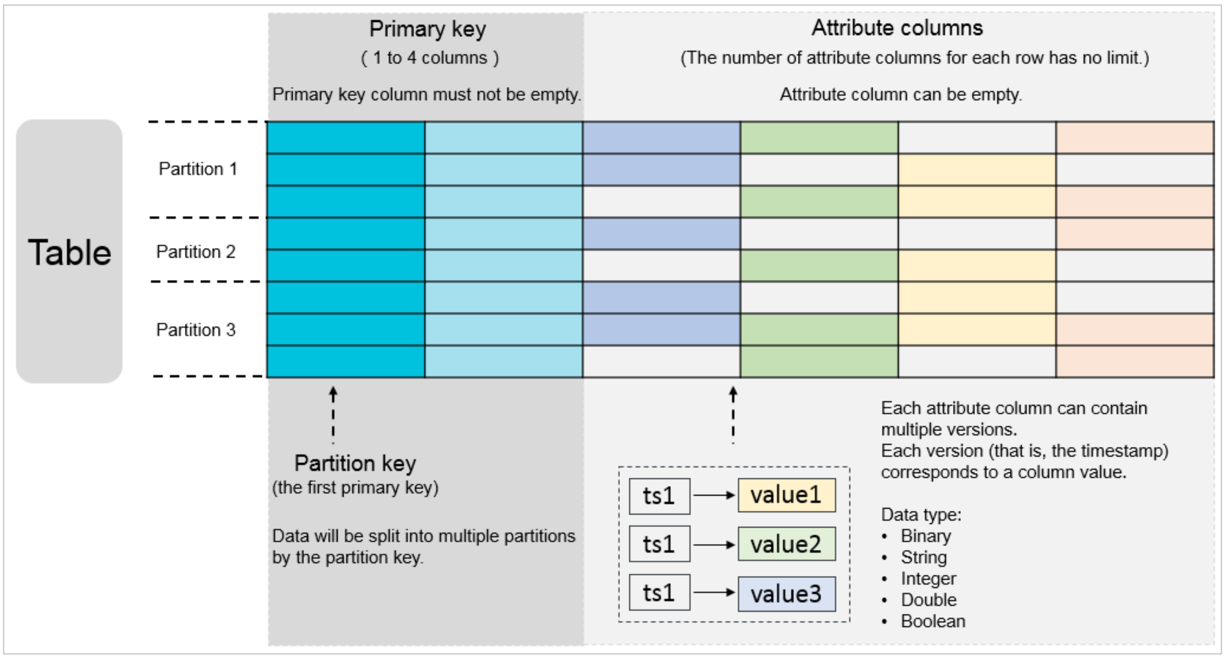
PK1	PK2	Attribute1	Attribute2	Attribute3
•	•	•		•
•	•		•	
•	•	•	•	

- User operations can be audited in fine-granularity.
- Users can organize resources using instances. A user can create multiple instances and use each instance to create and manage multiple data tables.
- An instance is a basic unit of multi-tenant isolation.
- Different users can be granted different permissions for better managed security.

3.2.2 Data table

The following figure shows the data table structure of Table Store.

Figure 3-2: Data table structure


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

- A data table is a basic unit of resource allocation.
- A table is a set of rows. A row consists of a primary key and attributes.
- A table partitions data according to the size of the first primary key column.
- All rows in a table must have the same quantity of primary key columns with the same names.
- The quantity, names, and data types of attribute columns in a row can be different.
- A table can contain a maximum of 1,024 columns.
- A table can contain rows of data numbering in the hundreds of billions.
- A table's capacity can reach PB level.

3.2.3 Data partitioning

- Data in a table is partitioned according to the size of the first primary key column.
- In the first primary key column, the rows of which the values are within the same partition area will be allocated to the same partition.
- To improve load balancing, Table Store splits and merges partitions according to specific rules.
- Data under the same partition key should not exceed 1 GB.

3.2.4 Common commands and functions

Commands

- ListTable: lists all tables under an instance.

- CreateTable: creates a table.
- DeleteTable: deletes a table.
- DescribeTable: gets attributes of a table.
- UpdateTable: updates the reserved read/write throughput of a table.

Functions

- GetRow: reads data from a row.
- PutRow: inserts a row.
- UpdateRow: updates data of a row.
- DeleteRow: deletes one row of data.
- BatchGetRow: reads multiple rows in one or more tables in batches.
- BatchWriteRow: inserts, updates, or deletes multiple rows in one or more tables in batches.
- GetRange: reads table data within a certain range.

3.2.5 Authorization and permission control

Table Store permission

In addition to access control and private network support, Table Store supports the following permission control:

- Authorizes access to tables.
- Controls authorization through APIs.
- Supports IP limit, HTTPS, multi-factor authentication (MFA), access time limit, and other conditions to implement authentication.
- Supports temporary access authorization (STS).
- Supports virtual private cloud (VPC) access control.

Cloud Console

- Supports account logon and authentication using the cloud platform.
- Provides graphic instance creation, management, and deletion functions.
- Provides graphic table creation, management, deletion, and reserved read/write throughput adjustment functions.
- Displays table monitoring information.

3.3 Benefits

Built on Alibaba Cloud's Apsara distributed operating system, Table Store is a NoSQL database service that enables you to store and access massive amounts of structured data in real time. Table Store organizes data into instances and tables, and achieve seamless scaling by using data partitioning and load balancing. It shields applications from faults and errors occurring on the underlying hardware platform, providing fast recovery capability and high service availability . Additionally, Table Store manages data with multiple data backups to solid state disks (SSDs), enabling quick data access and high data reliability. When using Table Store, you only pay for the resources you reserve and use, and do not need to handle complex issues such as cluster resizing, upgrade and maintenance of database software and hardware.

Table Store comes with the following features:

- Scalability

  There is no limit on the amount of data stored in Table Store. As data increases, Table Store increases shards so that more storage space is allocated, thereby improving concurrent access capabilities.
- Data reliability

  Table Store stores multiple data backup copies and enables fast recovery in case of a backup failure, delivering data reliability of at least 99.99999999%.
- High availability

  With automatic failure detection and data migration, Table Store shields machine-related and network-related hardware faults from applications, delivering high availability of at least 99.9%.
- Ease of management

  Applications do not require tedious and complex O&M tasks, such as shard management, software/hardware upgrade, configuration update, and cluster resizing.
- Access security

  Table Store provides multiple permission management mechanisms and performs identity authentication for each application request to prevent unauthorized data access, ensuring strict data access security.
- High consistency

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

- Flexible data models

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

4 Network Attached Storage (NAS)

4.1 What is NAS

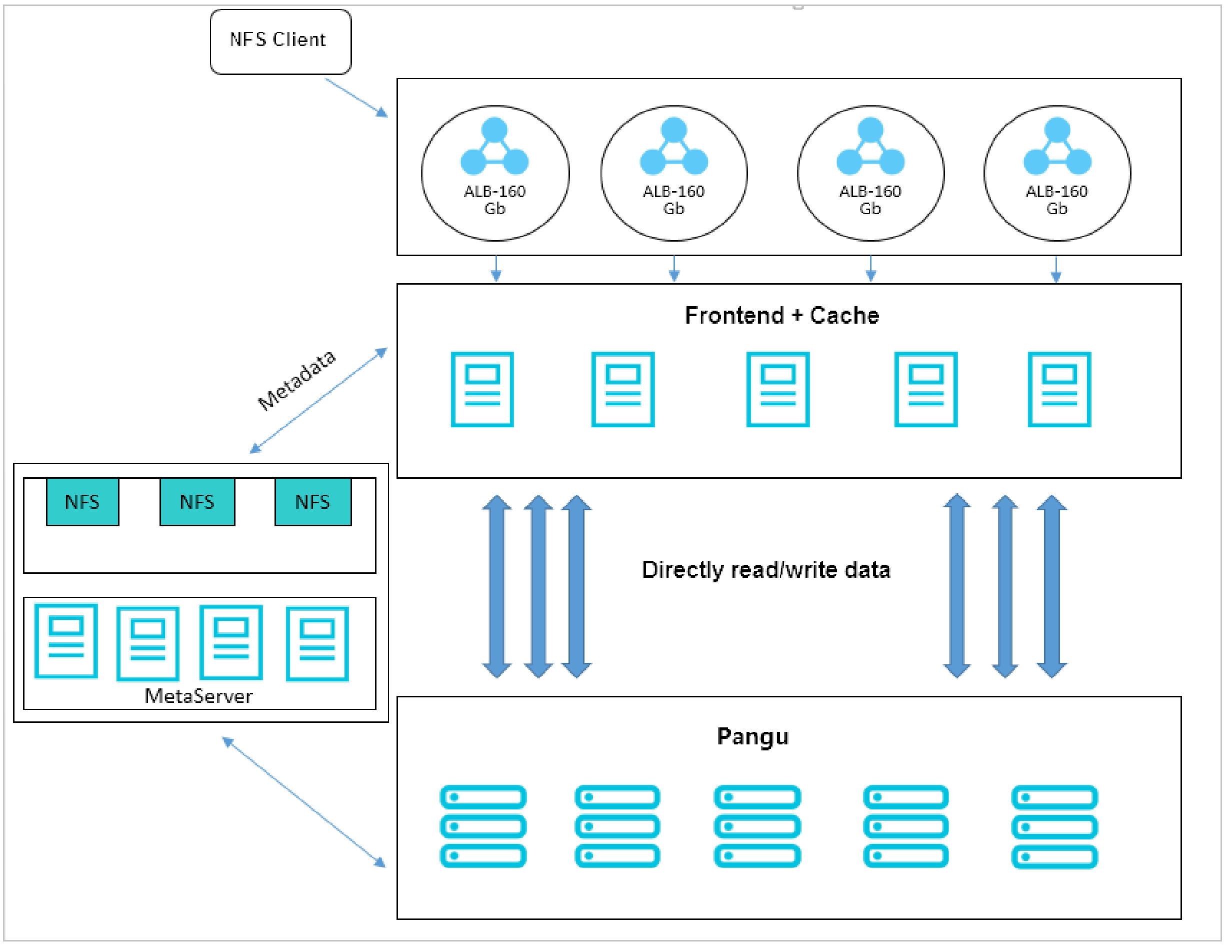
Alibaba Cloud Network Attached Storage (NAS) is a highly reliable, highly available file storage service for Alibaba Cloud ECS, E-HPC, and Container Service. The service features a distributed file system with unlimited capacity and performance scaling ability. It supports a single namespace and allows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

4.2 Architecture

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

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

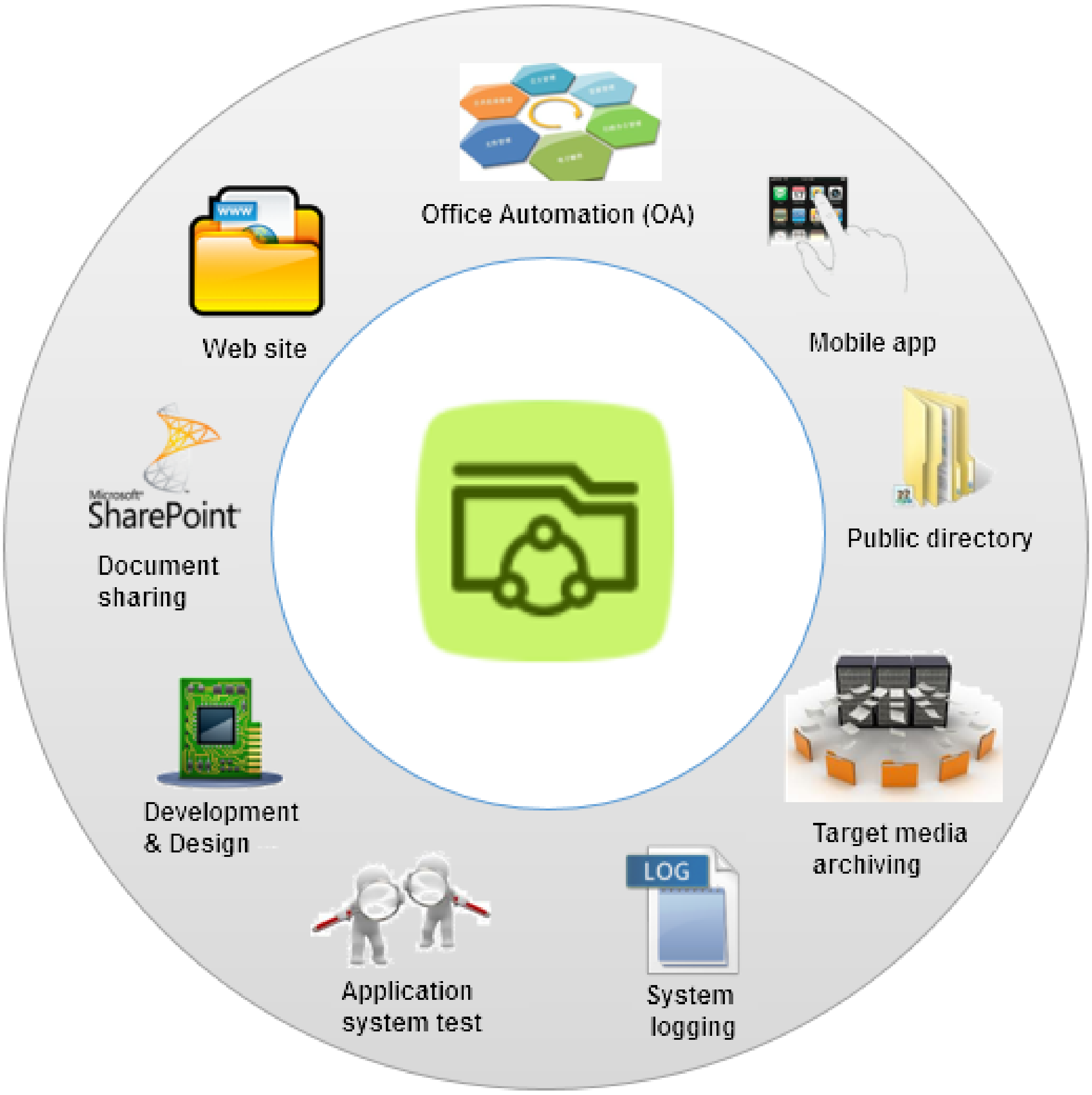
Figure 4-1: Architecture



4.3 Features

Alibaba Cloud NAS supports the NFSv3 and NFSv4 protocols. Your applications can use this service without any modifications. Alibaba Cloud NAS can meet various file storage needs, including business file sharing, backend file storage for office automation systems, enterprise database backup and storage, system log storage and analysis, website data storage and distribution, and data storage during system development and testing.

Figure 4-2: Features



4.4 Benefits

Alibaba Cloud NAS provides the following benefits:

- Shared file system

  You can mount the same file system on 10,000 clients using the NFSv3 or NFSv4 protocols to achieve data sharing.

- High performance

  The maximum throughput of the cluster can reach 20 Gbit/s and the IOPS can reach more than 20,000.

- Scalability

You can purchase storage capacity as demand increases. The maximum capacity of a file system can reach 10 PB. Each file system can store a maximum of 1 billion files, and the maximum file size is 32 TB.

- High availability

  Based on Apsara distributed file system, Alibaba Cloud NAS maintains three copies for each data file to achieve high availability and guarantee data reliability.

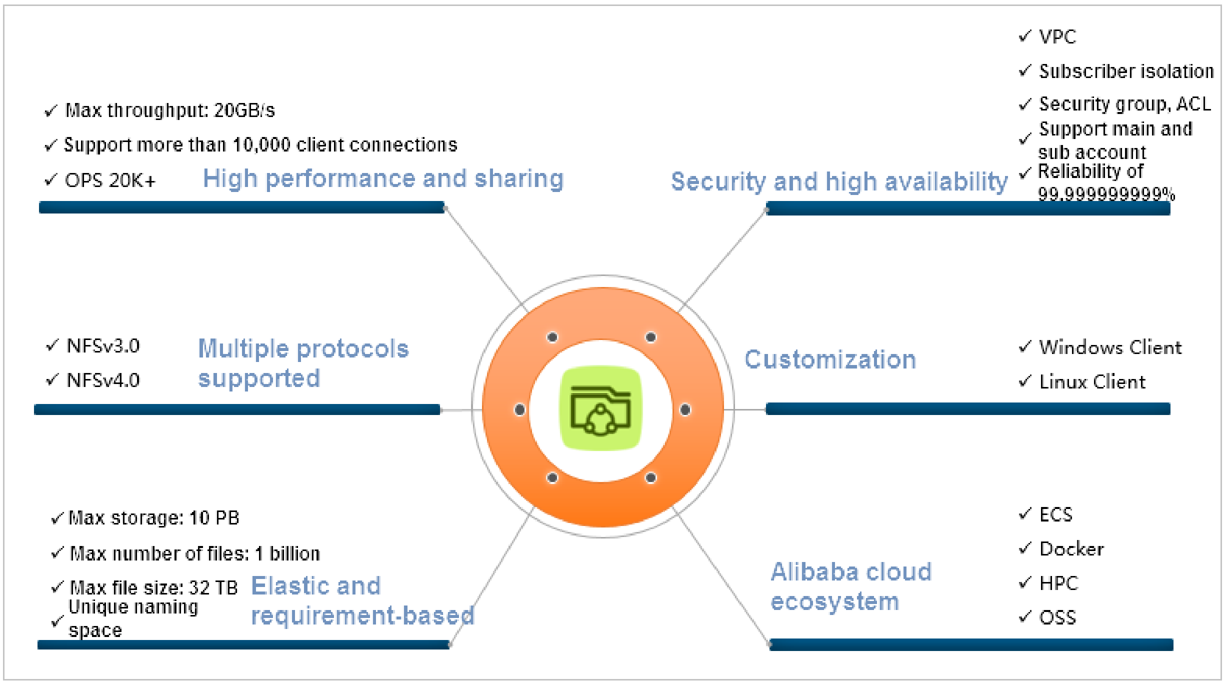
- Security

  Multiple security mechanisms such as VPC, security group, ACL, and account authorization are implemented to safeguard user data.

- Global namespace

  File data is distributed across the whole NAS cluster using a single namespace.

Figure 4-3: Benefits


✓ Max throughput: 20GB/s
✓ Support more than 10,000 client connections
✓ OPS 20K+
High performance and sharing
Security and high availability
✓ VPC
✓ Subscriber isolation
✓ Security group, ACL
✓ Support main and sub account
✓ Reliability of 99.9999999999%
✓ NFSv3.0
✓ NFSv4.0
Multiple protocols supported
Customization
✓ Windows Client
✓ Linux Client
✓ Max storage: 10 PB
✓ Max number of files: 1 billion
✓ Max file size: 32 TB
✓ Unique naming space
Elastic and requirement-based
Alibaba cloud ecosystem
✓ ECS
✓ Docker
✓ HPC
✓ OSS

5 Relational Database Service (RDS)

5.1 What is ApsaraDB for RDS?

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

ApsaraDB for MySQL

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

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

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

ApsaraDB for SQL Server

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

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

ApsaraDB for PostgreSQL

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

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

ApsaraDB for PPAS

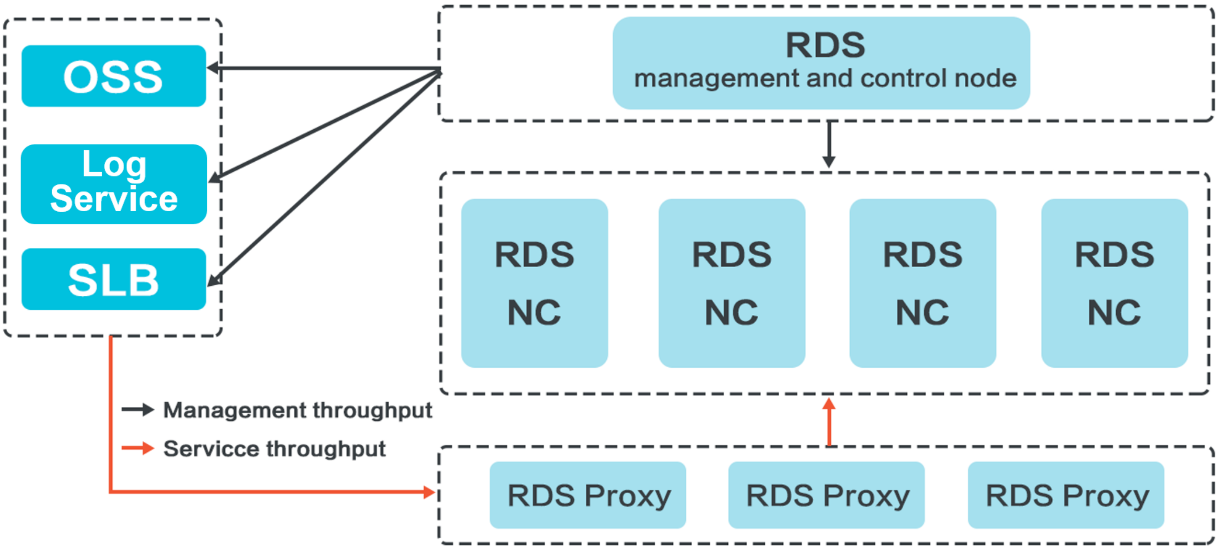
Postgres Plus Advanced Server (PPAS) is a stable, secure, and scalable enterprise-class relational database. Based on PostgreSQL, the world's most advanced open source database, PPAS brings enhancements in terms of performance, application solutions, and compatibility. It also provides the capability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

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

5.2 Architecture

The RDS system architecture is as follows.

Figure 5-1: RDS system architecture



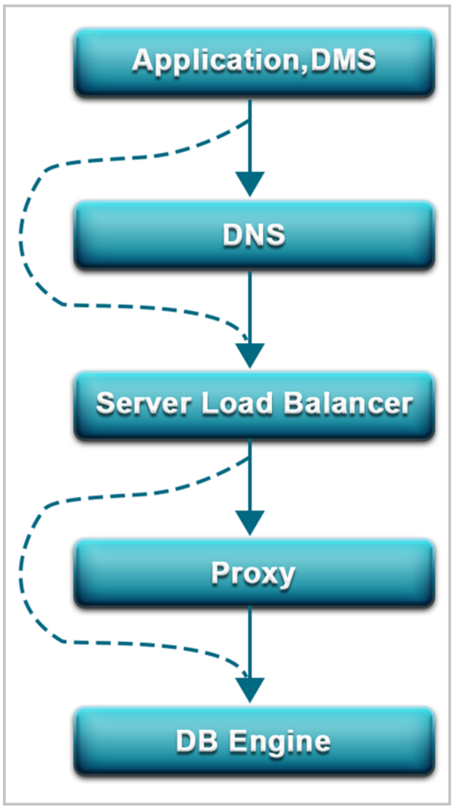
5.3 Features

High-availability RDS provides six core services: data link service, scheduling service, backup service, high-availability service, monitoring service, and migration service.

5.3.1 Data link service

The data link service mainly provides data operations, including adding, deleting, modifying, and querying table structures and data.

Figure 5-2: RDS data link service



5.3.1.1 DNS

The DNS module can dynamically resolve domain names to IP addresses, to prevent IP address changes from affecting the performance of RDS instances.

For example, assume that the domain name of an RDS instance is test.rds.aliyun.com, and the IP address corresponding to this domain name is 10.1.1.1. If a program's connection pool is configured as test.rds.aliyun.com or 10.1.1.1, the ApsaraDB for RDS instance can be accessed as normal.

After a zone migration or version upgrade is performed for this RDS instance, the IP address may change to 10.1.1.2. If the domain name configured in the program connection pool is test. rds.aliyun.com, the instance can still be accessed. However, if the IP address configured in the connection pool is 10.1.1.1, the instance will become inaccessible.

5.3.1.2 SLB

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

For example, assume that an RDS instance has an intranet IP address of 10.1.1.1 and the corresponding Proxy or DB Engine runs on 192.168.0.1. Normally, the SLB module redirects access traffic for 10.1.1.1 to 192.168.0.1.

If 192.168.0.1 goes wrong, 192.168.0.2, which works in hot standby mode, takes over for 192. 168.0.1. Now, the SLB module redirects access traffic from 10.1.1.1 to 192.168.0.2 and the RDS instance continues to provide service as usual.

5.3.1.3 Proxy

The Proxy module performs a number of functions, including data routing, traffic detection, and session persistence. This model is still evolving.

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

5.3.1.4 DB engine

RDS fully supports mainstream database protocols, as detailed in the following table:

Table 5-1: Database protocols supported by RDS

RDBMS	Version
MySQL	5.6 (including read-only instances)
MS SQLServer	2008R2
PostgreSQL	9.4

RDBMS	Version
PPAS	9.3
ORACLE	SQL syntax and stored procedures

5.3.1.5 DMS

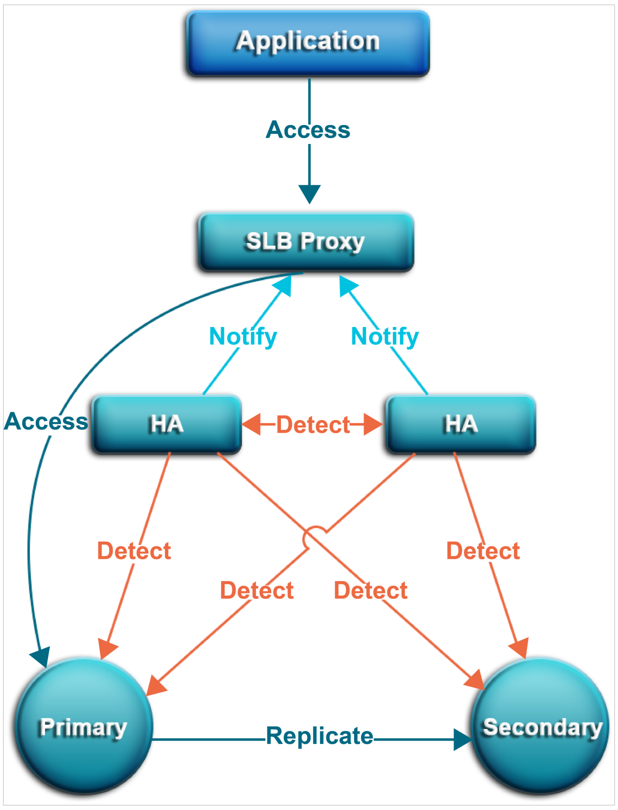
Data Management Service (DMS) is a web service designed to access and manage cloud data. DMS provides a range of functions including data management, object management, data transfer management, and instance management. DMS currently supports MySQL, MS SQL Server, PostgreSQL, ADS, and other data sources.

5.3.2 High-availability service

The high-availability service is mainly designed to ensure the availability of the data link service. It is also responsible for handling internal database exceptions.

In addition, the high-availability service is provided by multiple HA nodes, ensuring the high availability of the service.

Figure 5-3: RDS high-availability service



5.3.2.1 Detection

The Detection module detects whether the master node and slave node of the DB Engine are providing services normally.

The HA node uses heartbeat information at an interval of 8 to 10 seconds to determine the health status of the master node. This information, combined with the health status of the slave node and heartbeat information from other HA nodes, allows the Detection module to eliminate any risk of misjudgment caused by exceptions such as network jitter. As a result, switchover can be completed within 30 seconds.

5.3.2.2 Repair

The Repair module maintains the replication relationship between the master and slave nodes of the DB Engine. It can also correct any errors that may occur in either node during normal operations. The following are examples:

- It can automatically restore master/slave replication in case of abnormal disconnection.
- It can automatically repair table-level damage to the master or slave node.
- It provides on-site saving and automatic repair for master/slave node crashes.

5.3.2.3 Notice

The Notice module informs the Server Load Balancer (SLB) or Proxy module about status changes to the master and slave nodes to ensure that you can continue to access the correct node.

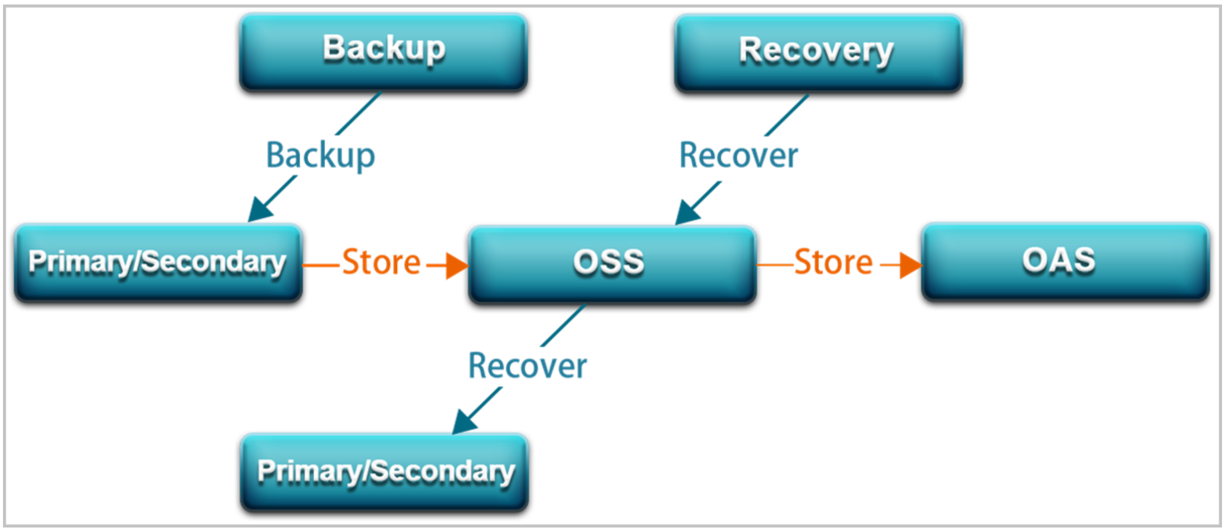
For example, the Detection module discovers problems with the master node and instructs the Repair module to fix these problems. If the Repair module fails to resolve a problem, it directs the Notification module to switch over traffic. The Notice module forwards the switching request to the SLB or Proxy module, which will redirect all traffic to the slave node.

At the same time, the Repair module creates a new slave node on another physical server and synchronizes this change back to the Detection module. The Detection module starts to recheck and determine the health status of the instance.

5.3.3 Backup service

The backup service supports offline data backup, dump, and recovery.

Figure 5-4: RDS backup service



5.3.3.1 Backup

The Backup module compresses and uploads data and logs from both the master and slave nodes to Object Storage Service (OSS). When the slave node is operating properly, backup is always initiated on the slave node so as not to affect the services on the master node. However, if the slave node is unavailable or damaged, the Backup module will create a backup on the master node.

5.3.3.2 Recovery

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

- Master node rollback: Rolls back the master node to its status at a specified time point if an operation error occurs.
- Slave node repair: Creates a new slave node to reduce the risk of an irreparable fault on the slave node.
- Read-only instance creation: Creates a read-only instance from a backup.

5.3.3.3 Storage

The Storage module is responsible for uploading, dumping, and downloading backup files.

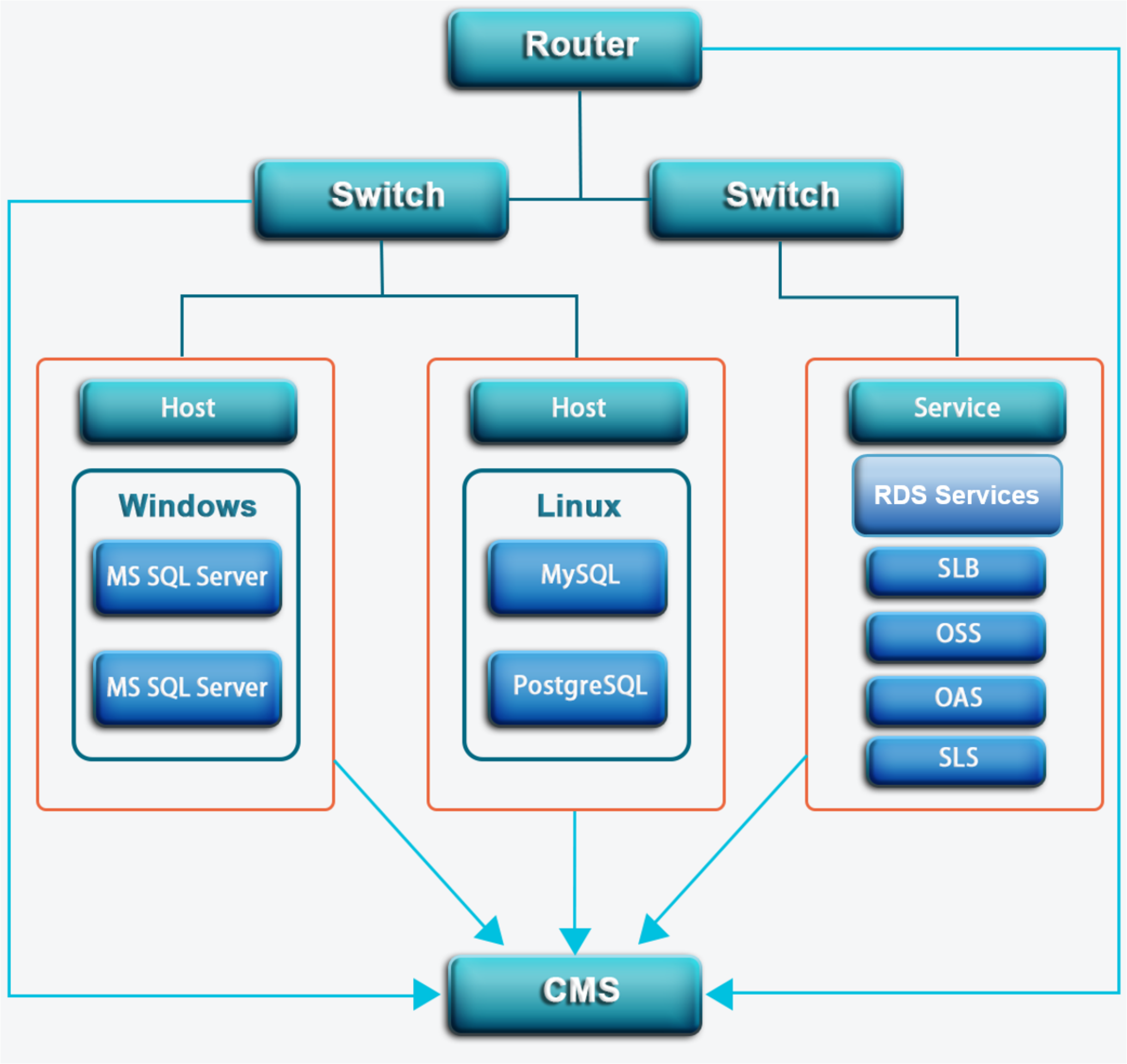
Currently, all backup data is uploaded to OSS for storage, and you can obtain temporary links to download this data as needed.

In certain scenarios, the Storage module also allows you to dump backup files from OSS to Archive Storage for cheaper and long-term offline storage.

5.3.4 Monitoring service

The monitoring service tracks the status of services, networks, operating systems, and instances.

Figure 5-5: RDS monitoring service



5.3.4.1 Service

The Service module is responsible for tracking the status of services.

For example, the Service module monitors whether SLB, OSS, and other cloud products on which RDS depends operate properly. The monitored indexes include functionality and response time. It also uses logs to determine whether the internal RDS services are operational.

5.3.4.2 Network

The Network module is responsible for tracking the status of the network layer. The following are examples:

- Monitor the connectivity between ECS and RDS.
- Monitor the connectivity between RDS physical machines.
- Monitor the packet loss rates of routers and VSwitches.

5.3.4.3 OS

The OS module tracks the status of hardware and operating system kernel. The following are examples:

- Hardware overhaul: The OS module constantly checks the operating status of the CPU, memory, motherboard, and storage devices. It predicts faults and automatically submits repair reports in advance.
- Operating system kernel monitoring: The OS module tracks all database calls and analyzes the causes of slow calling or call errors based on the kernel status.

5.3.4.4 Instance

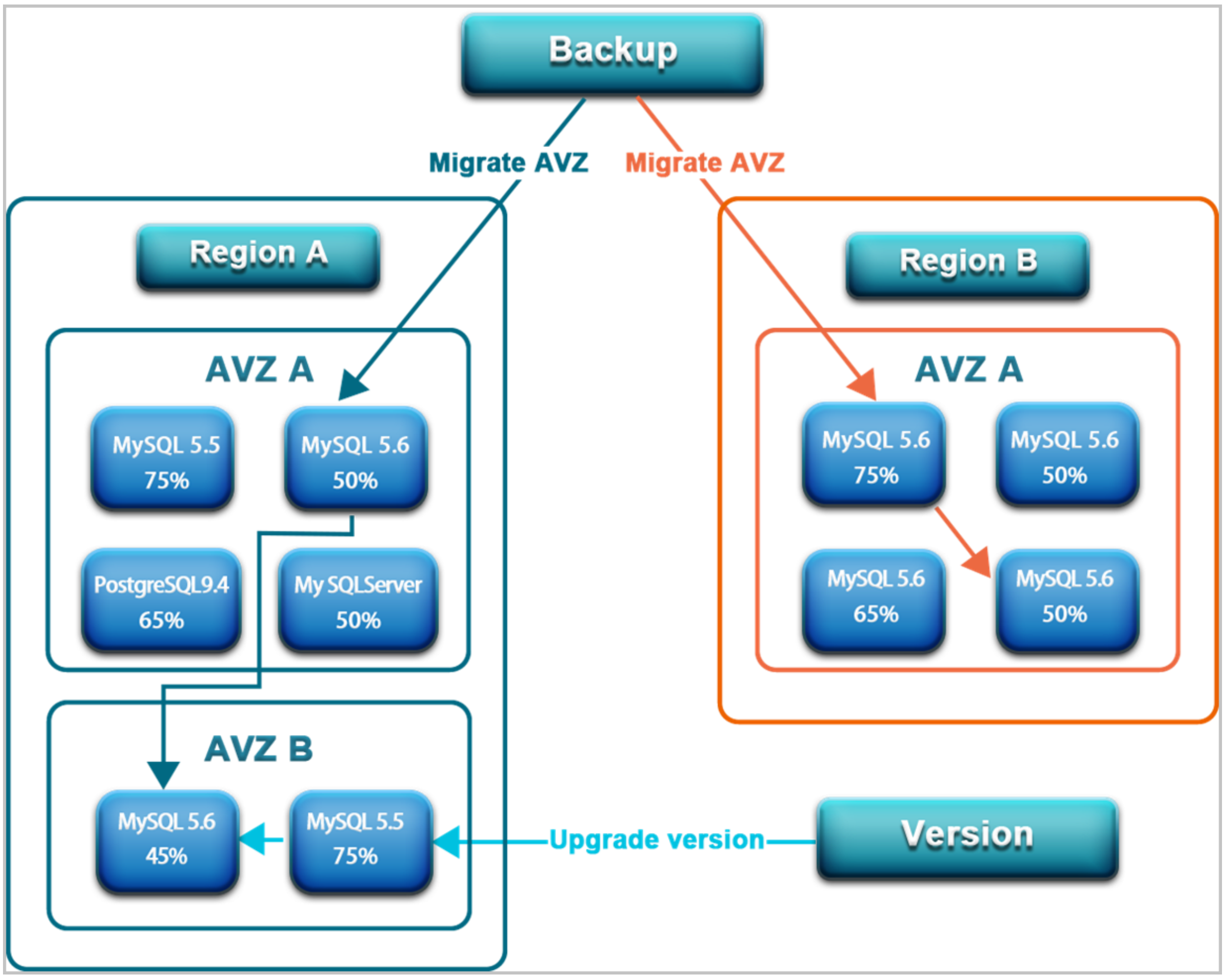
The Instance module collects information on RDS instances. The following are examples:

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

5.3.5 Scheduling service

The scheduling service allocates resources and manages the instance version.

Figure 5-6: RDS scheduling service



5.3.5.1 Resource

The Resource module allocates and integrates the underlying RDS resources. This is to activate and migrate instances from the perspective of users.

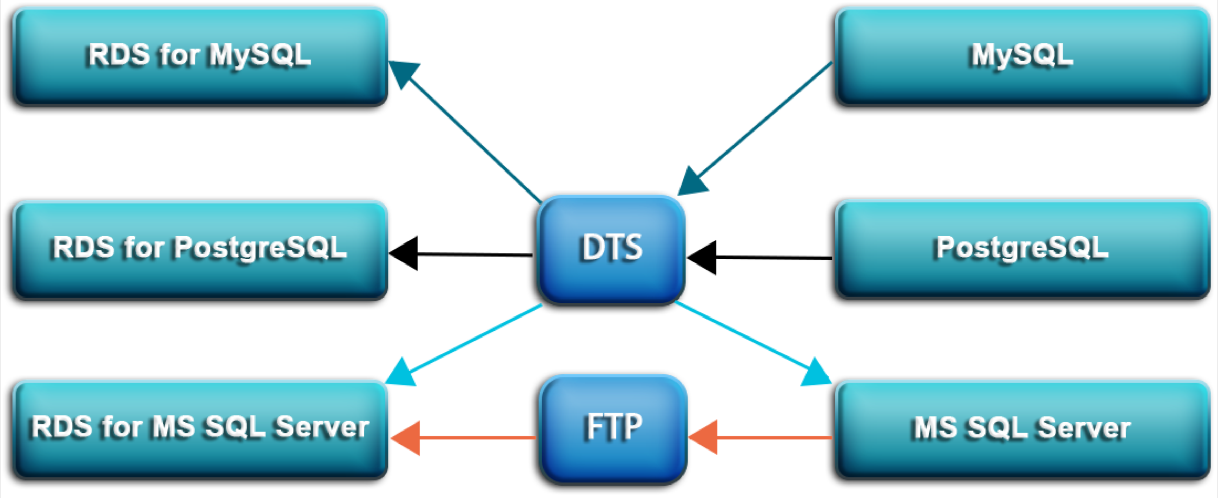
For example, when you create an instance through the RDS console or an Open API, the Resource module will calculate and find the most suitable physical server to carry the traffic. This is similar to cross-zone RDS instance migration.

After lengthy instance creation, deletion, and migration operations, the Resource module calculates the degree of resource fragmentation in a zone and regularly initiates resource integration to improve the service carrying capacity of the zone.

5.3.6 Migration service

The migration service helps you migrate data from self-built databases to RDS.

Figure 5-7: RDS migration service



5.3.6.1 FTP

The FTP module supports the full migration of RDS for MS SQL Server data to the cloud.

After you back up a self-built MS SQL Server database, you can use the FTP client to upload backup files to the dedicated FTP module provided by RDS. The FTP module will restore the backup files to the specified RDS instance.

6 ApsaraDB for Redis

6.1 What is ApsaraDB for Redis

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

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

6.2 Features

The high-availability ApsaraDB for Redis service provides four core services:

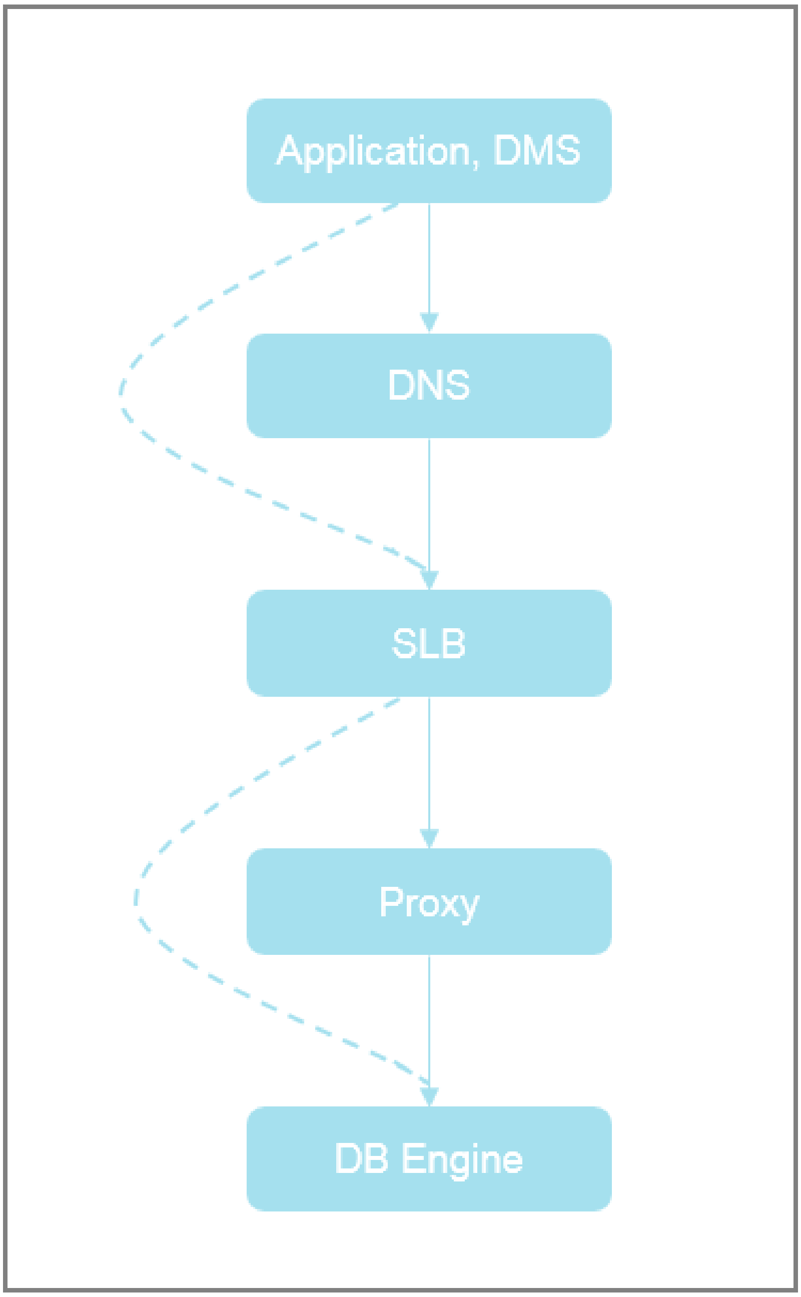
- Data link service
- High-availability service
- Monitoring service
- Scheduling service

6.2.1 Data link service

The data link service allows you to control data, such as adding, deleting, modifying, and querying data.

You can connect to the ApsaraDB for Redis service either through an application or through the GUI Data Management Tool (DMS) provided by ApsaraDB for Redis.

Figure 6-1: Data link service



6.2.1.1 DNS

The DNS module supports the dynamic resolution of domain names into IP addresses to avoid ApsaraDB for Redis instances unreachable due to the change of IP addresses.

For example, assume that the domain name of an ApsaraDB for Redis instance is `test.kvstore.aliyun.com`, and the IP address corresponding to this domain name is `10.1.1.1`.

You can access the ApsaraDB for Redis instance if `test.kvstore.aliyun.com` or `10.1.1.1` is added to the connection pool of the application you are using.

If the ApsaraDB for Redis instance is migrated to another host upon failover or its version is upgraded, the IP address may change to **`10.1.1.2`**.

If the domain name **`test.kvstore.aliyun.com`** has been added to the application's connection pool, you can still access the instance.

If the IP address **`10.1.1.1`** has been added, however, the instance is unreachable.

6.2.1.2 SLB

The Server Load Balancer module provides instance IP addresses to avoid ApsaraDB for Redis instances unreachable due to the change of physical servers.

For example, assume that an ApsaraDB for Redis instance has an intranet IP address of **`10.1.1.1`** and the corresponding Proxy or DB Engine runs on the host at **`192.168.0.1`**. Normally, the Server Load Balancer module redirects all traffic destined for **`10.1.1.1`** to **`192.168.0.1`**.

If **`192.168.0.1`** fails, its hot-standby host at **`192.168.0.2`** takes over for **`192.168.0.1`**. In this case, the Server Load Balancer module redirects the traffic for **`10.1.1.1`** to **`192.168.0.2`**, and the ApsaraDB for Redis instance provides services normally.

6.2.1.3 Proxy

The Proxy module provides data routing, traffic detection, and session persistence functions. More functions will be available soon.

- Data routing: ApsaraDB for Redis supports a cluster-based architecture and implements complex query and partition policies for distributed routes.
- Traffic detection: This reduces the risks from cyberattacks that make use of Redis vulnerabilities.
- Session persistence: This prevents database connection interruptions if any fault occurs.

6.2.1.4 DB engine

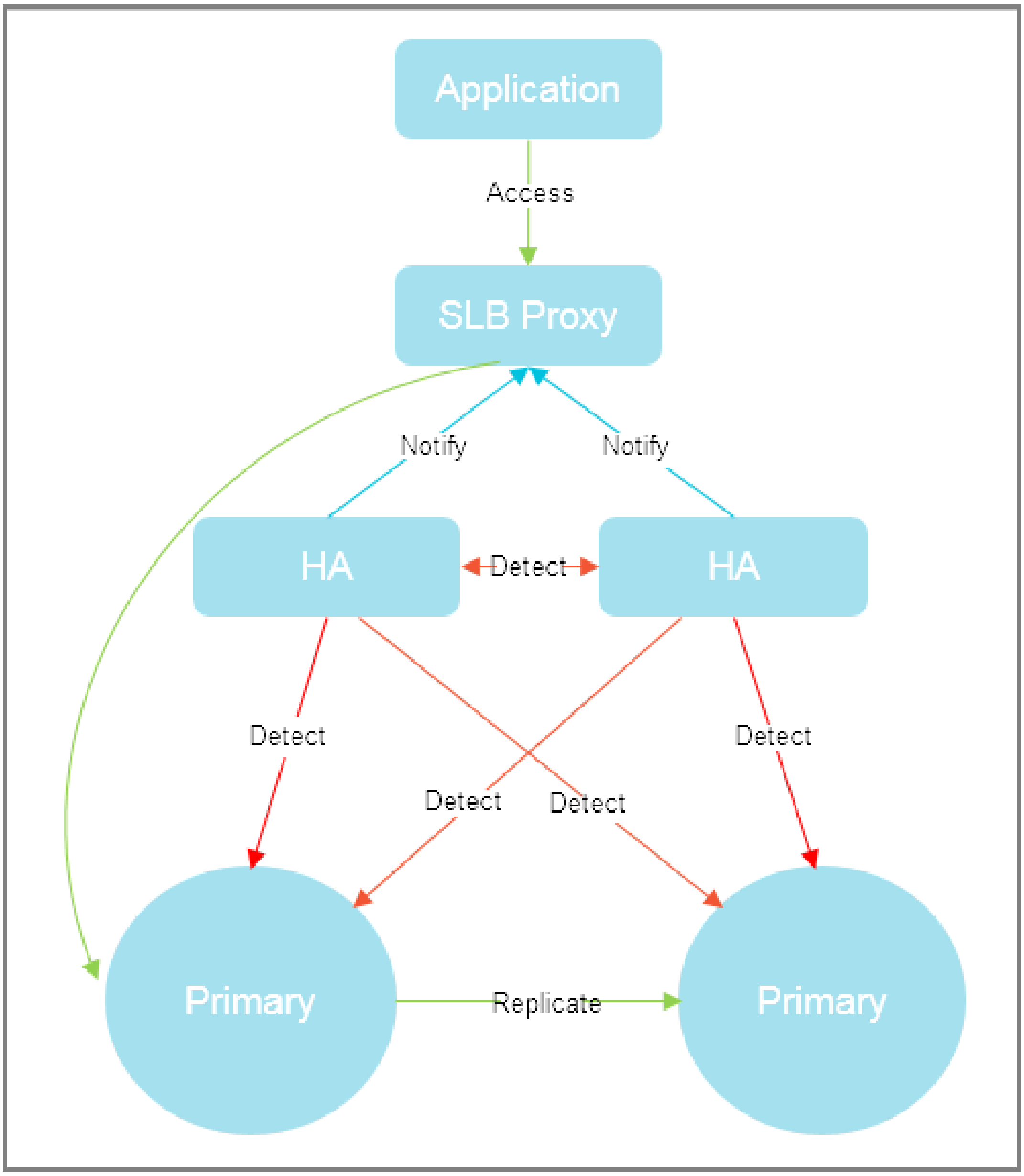
Standard protocols supported by ApsaraDB for Redis:

Engine	Version
Redis	Compatible with V2.8 and V3.0 Geo Edition

6.2.2 High-availability service

The high-availability service ensures the availability of the data link service and processes internal database exceptions.

The high-availability service is provided by multiple HA nodes, ensuring the high availability of the service itself.



6.2.2.1 Detection

The Detection module detects whether the master node and slave node of the DB Engine are providing services normally.

The HA node uses heartbeat information, acquired at an interval of 8 to 10 seconds, to determine the health status of the master node. This information, combined with the health status of the slave node and heartbeat information from other HA nodes, allows the Detection module to eliminate any risk of misjudgment caused by exceptions such as network jitter. As a result, switchover can be completed within 30 seconds.

6.2.2.2 Repair

The Repair module maintains the replication between the master and slave nodes of the DB Engine. It also repairs any errors that may occur in either node during daily operations, for example:

- It can automatically restore master/slave replication in case of disconnection.
- It can automatically repair table-level damage to the master or slave node.
- It can save and automatically repair crashes of the master or slave node.

6.2.2.3 Notice

The Notice module informs the Server Load Balancer or Proxy of status changes to the master and slave nodes to ensure that you can continue to access the correct node.

For example, the Detection module discovers that the master node encounters an exception and instructs the Repair module to fix it. If the Repair module fails to resolve the problem, it directs the Notice module to initiate traffic switching. The Notice module then forwards the switching request to the Server Load Balancer or Proxy, which begins to redirect all traffic to the slave node.

At the same time, the Repair module creates a new slave node on another physical server and synchronizes this change back to the Detection module. The Detection module starts to recheck the health status of the instance and discovers it is healthy.

6.2.3 Monitoring service

The monitoring service tracks the status of Apsara for Redis instances in terms of services, networks, operating systems, and instances.

6.2.3.1 Service-level monitoring

The independent Service module monitors Apsara for Redis instances in terms of services.

For example, the Service module monitors ApsaraDB for Redis-dependent Alibaba Cloud services such as Server Load Balancer, including function implementation and response time.

6.2.3.2 Network-level monitoring

The Network module monitors Apsara for Redis instances in terms of networks, for example:

- Connectivity between ECS and ApsaraDB for Redis
- Connectivity between ApsaraDB for Redis hosts
- Packet loss rates for routers and VSwitches

6.2.3.3 OS-layer monitoring

The OS (operating system) module monitors Apsara for Redis instances in terms of hardware and kernel of the operating system, for example:

- Hardware overhaul: The Operating System module constantly checks the operation statuses of the CPU, memory, motherboard, and storage. It predicts the possibilities of a fault and automatically submits a repair request in advance.
- Operating system kernel monitoring: The Operating System module tracks all database calls and uses the kernel status to analyze the causes of call slowdowns or errors.

6.2.3.4 Instance-level monitoring

The Instance module collects information of ApsaraDB for Redis instances, for example:

- Instance availability
- Instance capacity and performance metrics

6.2.4 Scheduling service

The scheduling service allocates resources. It integrates and allocates the underlying ApsaraDB for Redis resources. To you, this is the same as instance activation and migration.

For example, when you create an instance using the console, the scheduling service will calculate the most suitable physical server to carry the traffic.

After lengthy instance creation, deletion, and migration operations, the scheduling service calculates the degree of resource fragmentation in a zone and initiates resource integration regularly to improve the service carrying capacity of the zone.

7 ApsaraDB for MongoDB

7.1 What is Apsara for MongoDB?

ApsaraDB for MongoDB is fully compatible with the MongoDB protocol and provides stable, reliable, and auto-scaling database services. It offers a full range of database solutions, such as disaster recovery, backup, recovery, monitoring, and alarms.

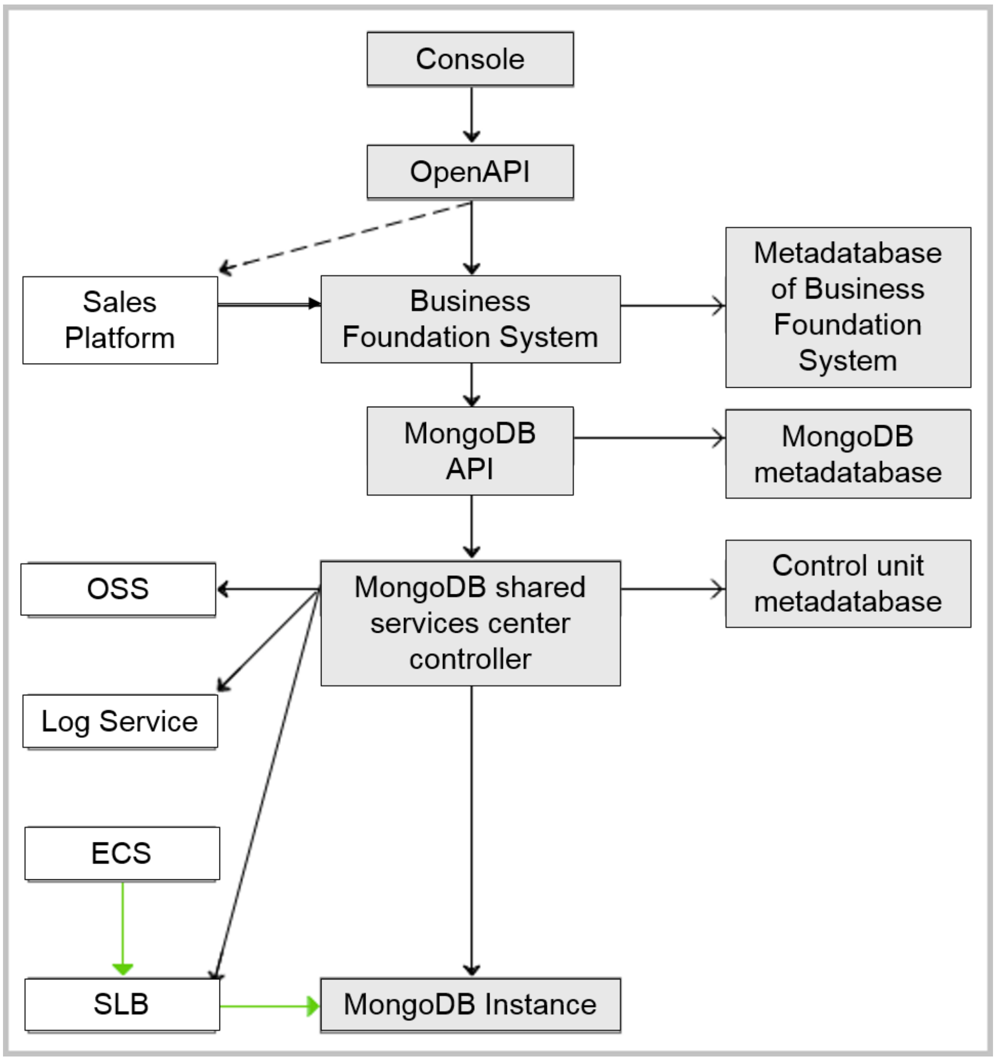
By default, ApsaraDB for MongoDB comes as a three-node MongoDB replica set. The primary node supports both read and write operations; the secondary node supports only read operations; the hidden node is used for high availability.

ApsaraDB for MongoDB provides various security features:

- Access control
- Network isolation
- Data backup
- Version maintenance
- Maintenance team permission control

7.2 Architecture

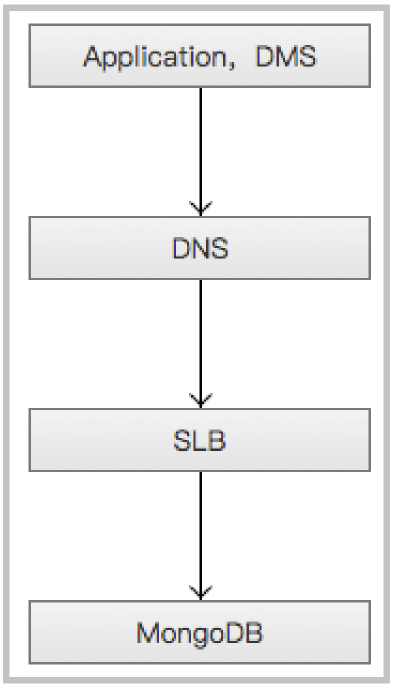
ApsaraDB for MongoDB provides six core services: data link service, scheduling service, backup service, high-availability service, monitoring service, and migration service.


Console
OpenAPI
Sales Platform
Business Foundation System
Metadatabase of Business Foundation System
MongoDB API
MongoDB metadatabase
OSS
MongoDB shared services center controller
Control unit metadatabase
Log Service
ECS
SLB
MongoDB Instance

7.3 Functions

7.3.1 Data link service

The data link service provides support for data operations.



DNS

For example, the domain name of a MongoDB instance is mongodb.aliyun.com, and the IP address corresponding to this domain name is 10.10.10.1. If either mongodb.aliyun.com or 10.10.10.1 is configured in the connection pool of a program, the instance can be accessed.

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

SLB

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

For example, the intranet IP address of a MongoDB instance is 10.1.1.1, and the corresponding MongoDB instance runs on 192.168.0.1. Normally, the SLB module redirects all traffic destined for 10.1.1.1 to 192.168.0.1. If 192.168.0.1 fails, another address in hot standby status, 192.168.0.2, takes over for 192.168.0.1. In this case, the SLB module redirects all traffic destined for 10.1.1.1 to 192.168.0.2, and the MongoDB instance continues to offer its services normally.

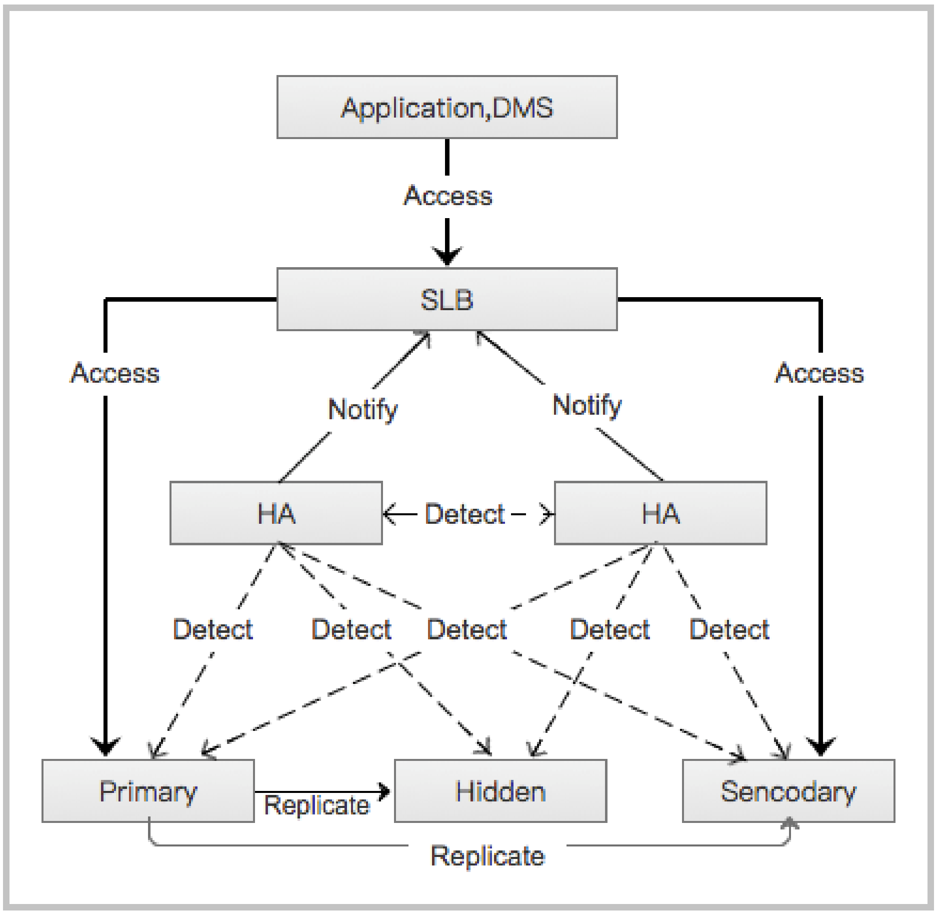
DMS

Data Management Service (DMS) is used for cloud-based data processing. It supports data management, and schema management.

7.3.2 High-availability service

The high-availability service guarantees the availability of the data link service and processes internal database exceptions.

Additionally, the high-availability service is composed of HA nodes to ensure high availability of itself.



Detection

The Detection module checks whether the primary, secondary, and hidden nodes of MongoDB offer their services normally. The HA (High Available) node uses heartbeat information, acquired at an interval of 8 to 10 seconds, to check the health status of the primary node. This information, combined with the heartbeat information of the secondary and hidden nodes, allows the Detection module to eliminate any risk of misjudgment caused by exceptions such as network jitter and allows that the exception switchover can be completed within 30 seconds.

Repair

The Repair module maintains the replication relationship among the primary, secondary, and hidden nodes of MongoDB. It can also repair or reconstruct any faulty node.

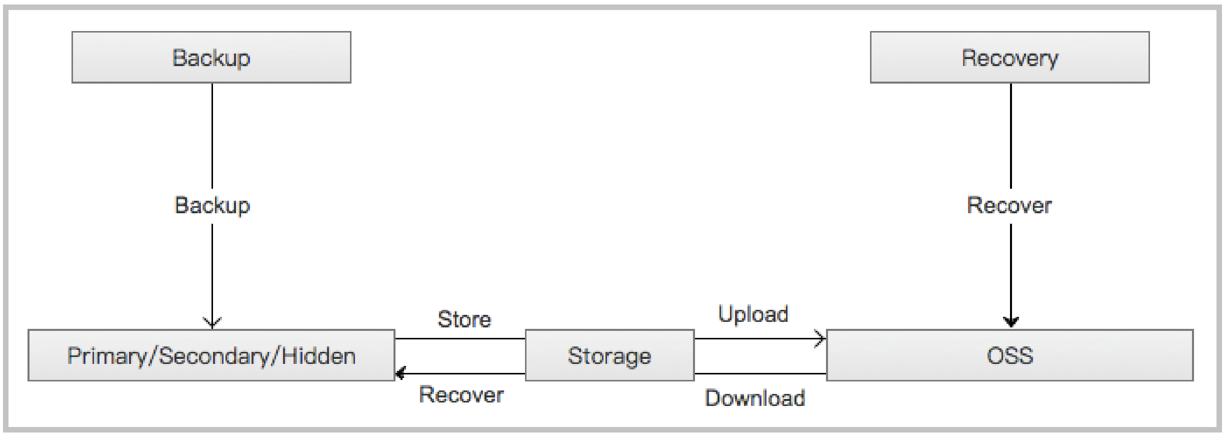
Notice

The Notice module informs the SLB of status changes to guarantee that you can continue to access the correct node.

For example, the Detection module discovers that the primary node has an exception and directs the Notification module to initiate traffic switching. The Notification module then forwards the switching request to the SLB, which begins to redirect traffic that used to flow to the primary to the secondary node and also redirect traffic that used to flow to the secondary to the hidden node. In this case, the secondary node becomes the primary node, and the hidden node becomes the secondary node. Simultaneously, the Repair module attempts to fix the original primary node and turn it into a new hidden node. If the fixing fails, the Repair module creates a new hidden node on another physical server and synchronizes this change back to the Detection module. The Detection module then incorporates this new information and starts to recheck the health status of the instance.

7.3.3 Backup service

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



Backup

The Backup module backs up, compresses, and then uploads instance data and logs to OSS. Backup is always performed on the hidden node and therefore does not affect the primary and secondary nodes.

Recovery

The Recovery module recovers a backup file stored on OSS to the target node.

Primary node roll-back: This can be used to restore a node to the state that it was in at a specific point in time.

Secondary/hidden node repair: This can be used to automatically create a new secondary node to reduce risks when an irreparable failure occurs to the secondary node.

Storage

The Storage module is responsible for uploading, dumping, and downloading backup files. Currently all backup data is uploaded to OSS for storage, and you can obtain temporary links to download their data as needed.

7.3.4 Monitoring service

The monitoring service monitors statuses of the services, networks, OS, and instances.

Service

The Service module tracks service-level statuses. It monitors whether the SLB, OSS, SLS, and other cloud products on which MongoDB depends are normal, including their functionality and response time. It also uses logs to determine whether the internal MongoDB services are operating normally.

Network

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

OS

The OS module tracks statuses at the hardware and OS kernel layer, for example:

- Hardware overhaul: Constantly checks the operational status of the CPU, memory, main board , and storage, pre-judges whether a fault will occur, and automatically submits a repair report in advance.
- Hardware overhaul: Constantly checks the operational status of the CPU, memory, main board , and storage, pre-judges whether a fault will occur, and automatically submits a repair report in advance.

Instance

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

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

7.3.5 Scheduling service

The scheduling service mainly implements resource allocation and instance version management.

Resource

The Resource module allocates and integrates the underlying MongoDB resources, namely, activation and configuration change for instances.

For example, when you create an instance through the *MongoDB console* or APIs, the Resource module determines which physical server is best suited to carry traffic. After lengthy instance creation, deletion, and migration operations, the Resource module calculates the degree of resource fragmentation in a zone and initiates resource integration regularly to improve the service carrying capacity of the zone.

Version

The Version module is applicable to version upgrades of MongoDB instances. For example, MongoDB major version upgrade from 3.2 to 3.4 and minor version upgrade including releases about source code bug fixing or customized kernel optimization.

7.3.6 Migration service

The migration service can migrate data from a local database to MongoDB.

Data Transmission indicates a data flow service provided by Alibaba Cloud for data interaction among data sources. Currently, both full and incremental MongoDB data migration are supported.

- Full migration: Data Transmission migrates all the object data of the source database to the target instance.
- Incremental migration: During incremental migration, the incrementally updated data of the local MongoDB instance is synchronized to ApsaraDB for MongoDB, and the local MongoDB and ApsaraDB for MongoDB enter the dynamic synchronization process. Incremental migration achieves smooth migration from local MongoDB to ApsaraDB for MongoDB without interrupting the normal service provision by local MongoDB.

8 ApsaraDB for Memcache

8.1 Overview

ApsaraDB for Memcache is a high-performance, highly reliable distributed in-memory database service that can scale smoothly. ApsaraDB for Memcache is based on the Apsara distributed file system and high-performance storage. It offers a full set of database solutions, including hot standby, fault recovery, business monitoring, data migration, etc.

In addition, as a cloud computing service, ApsaraDB for Memcache's hardware and data are deployed on cloud, and it provides comprehensive infrastructure planning, network security, and system maintenance services. These ensure that users can fully focus on their own business innovations.

8.2 Functions

The 6 core services of Memcache include:

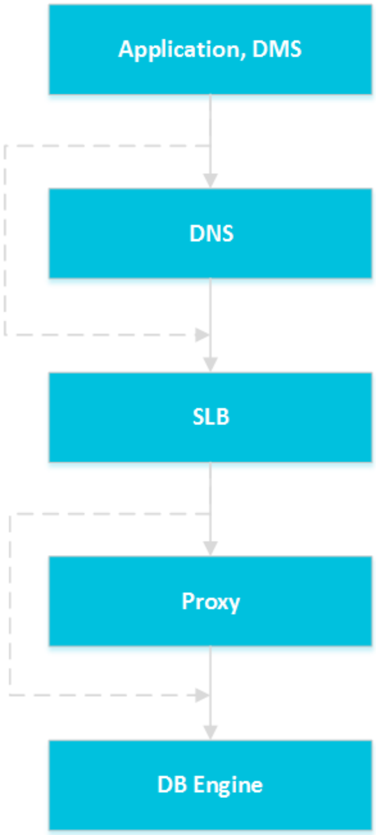
- Data link service
- Scheduling service
- Backup service
- High availability service
- Monitoring service
- Migration service

8.2.1 Data link service

Data link service offers data operations, such as adding, deleting, modifying and querying data.

You may connect to Memcache using applications, or you can use a data management tool (DMS) provided by Memcache for gui-based data management.

Figure 8-1: Data link service



8.2.1.1 DNS

The DNS module supports the dynamic resolution of domain names to IP addresses. It prevents IP address changes from affecting the performance of Memcache instances.

For example, consider a Memcache instance with an intranet IP of 10.1.1.1, and a correspond ing Proxy or DB Engine running on 192.168.0.1: Normally, the SLB module redirects all traffic destined for 10.1.1.1 to 192.168.0.1.

If 192.168.0.1 fails, another hot standby address, 192.168.0.2, takes over for 192.168.0.1. The SLB module redirects all traffic destined for 10.1.1.1 to 192.168.0.2, and the Memcache instance continues to offer its services normally.

8.2.1.2 SLB

The SLB module provides instance IP addresses to prevent physical server changes from affecting the performance of Memcache instances.

For example, consider an ApsaraDB for Memcache instance with an intranet IP of **`10.1.1.1`**, and a corresponding Proxy or DB Engine running on **`192.168.0.1`**: Normally, the SLB module redirects all traffic destined for **`10.1.1.1`** to **`192.168.0.1`**.

If **`192.168.0.1`** fails, another hot standby address, **`192.168.0.2`**, takes over for **`192.168.0.1`**. The SLB module redirects all traffic destined for **`10.1.1.1`** to **`192.168.0.2`**, and the Memcache instance continues to offer its services normally.

8.2.1.3 Proxy

The Proxy module provides data routing, traffic detection, and session persistence. However, its functions continue to expand.

- Data routing: Memcache data routing supports cluster architectures and allows complex query and partitioning strategies for distributed routing.
- Traffic detection: this reduces the risk of network attacks directed against Memcache.
- Session persistence: this prevents database connection interruptions if any failures occur.

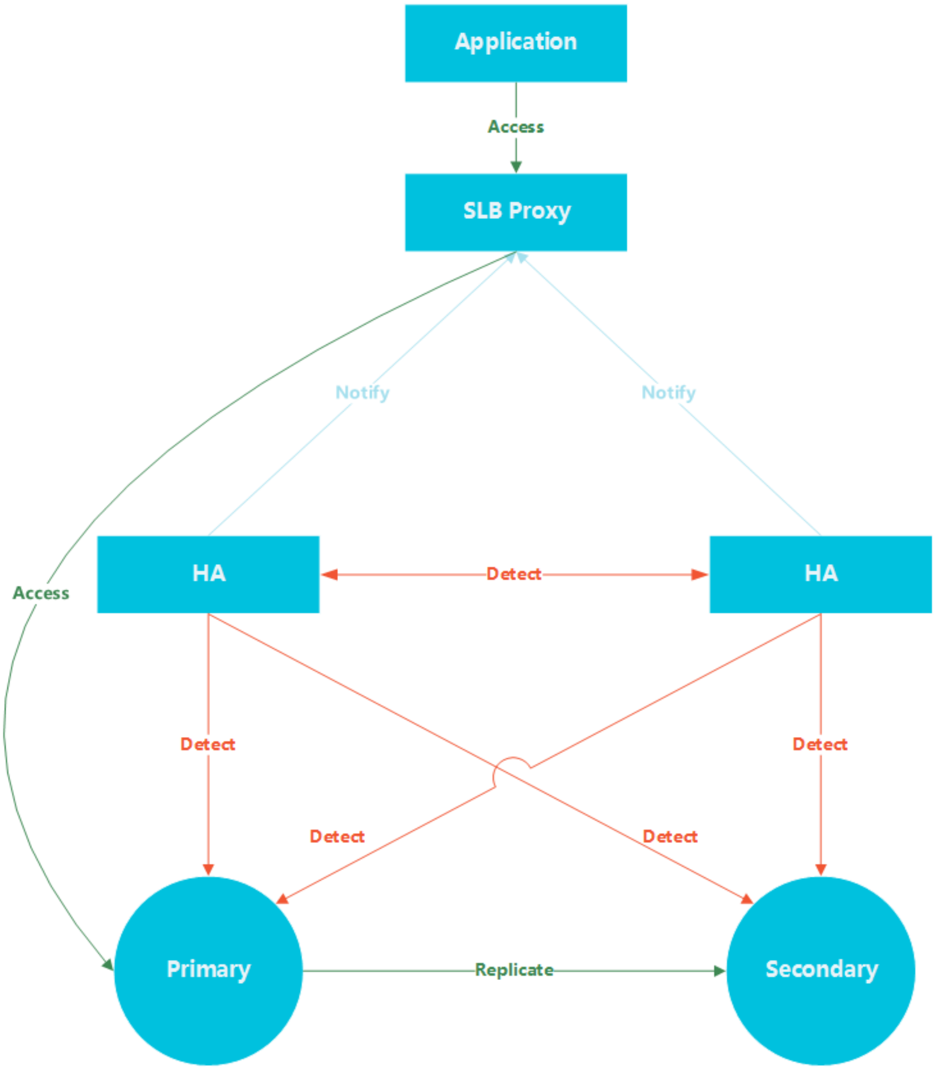
8.2.1.4 DB Engine

ApsaraDB for Memcache supports mainstream protocols and direct connections with a variety of clients.

8.2.2 High availability service

The high-availability service guarantees the availability of the data link services and processes any internal database exceptions. In addition, the high availability service is provided by multiple HA nodes which are highly available.

Figure 8-2: High availability service



8.2.2.1 Detection

The Detection module checks whether the master and slave nodes of the DB Engine offer their services normally.

The HA (High Available) node uses heartbeat information, acquired at an interval of 8 to 10 seconds, to check the health status of the master node. This information is combined with the health status of the standby node and heartbeat information from other HA nodes. It allows the Detection module to eliminate any risk of misjudgment caused by exceptions, such as network jitter, and allows the exception switchover to be completed within 30 seconds.

8.2.2.2 Repair

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

- Automatic restoration of master/slave replication in case of disconnection
- Automatic repair of table-level damage to a master or slave node
- On-site saving and automatic repair if a master or slave node crash

8.2.2.3 Notice

The Notice module informs the SLB or Proxy of status changes to the master and slave nodes to guarantee that users can continue to access the correct node.

For example, the Detection module discovers that the master node has an exception and instructs the Repair module to fix it. If the Repair module fails to resolve the problem, it directs the Notification module to initiate traffic switching. The Notification module then forwards the switching request to the SLB or Proxy, which begins to redirect all traffic to the slave node.

Simultaneously, the Repair module creates a new slave node on another physical server and synchronizes this change back to the Detection module. The Detection module then incorporates this new information and starts to recheck the health status of the instance.

8.2.3 Monitoring service

ApsaraDB for Memcache provides multilevel monitoring services across the service, network, operating systems, and instance layers to ensure status tracking.

8.2.3.1 Service-level monitoring

The Service module tracks the service-level status.

For example, it monitors whether other cloud products, such as SLB, on which Memcache depends, are normal. This includes their functionality and response time.

8.2.3.2 Network-level monitoring

The Network module tracks the network level status.

For example, the connectivity between ECS and ApsaraDB for Memcache; the connectivity between ApsaraDB for Memcache and physical machines; and the packet loss rate of VRouters and VSwitches.

8.2.3.3 OS-level monitoring

The OS (operating system) module tracks status at the hardware and OS kernel layer, for example:

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

8.2.3.4 Instance-level monitoring

The Instance module collects ApsaraDB for Memcache instance-level information, for example, available information for instances, instance capacity, and performance indicators.

8.2.4 Scheduling service

The scheduling service implements resource allocation, such as allocating and integrating the underlying Memcache resources. For users, this includes creating and migrating instances.

For example, when you create an instance through the console, the scheduling module determines which physical server is best suited to carry the traffic.

After lengthy instance creation, deletion, and migration operations, the scheduling service calculates the degree of resource fragmentation in a zone, and initiates resource integration regularly to improve the service carrying capacity of the zone.

9 Server Load Balancer (SLB)

9.1 What is Server Load Balancer

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

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

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

Components

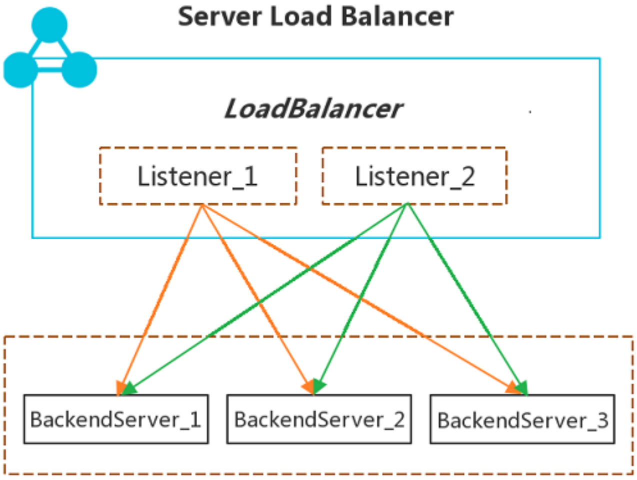
Server Load Balancer consists of the following components:

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

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

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

Figure 9-1: SLB components



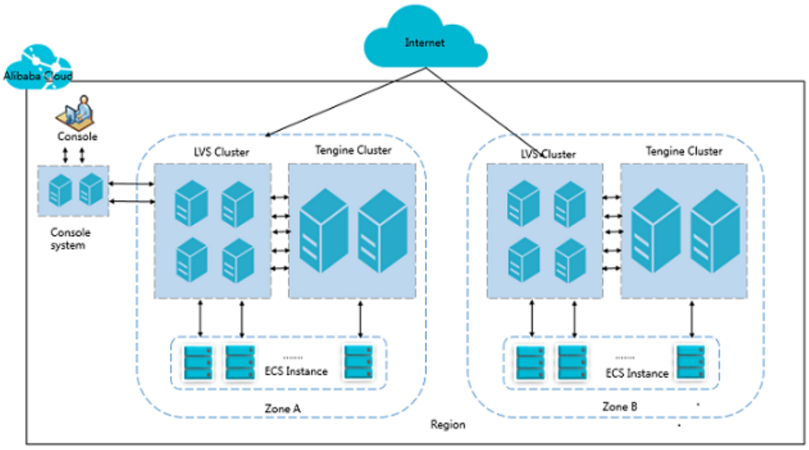
9.2 Architecture

Server Load Balancer is deployed in clusters.

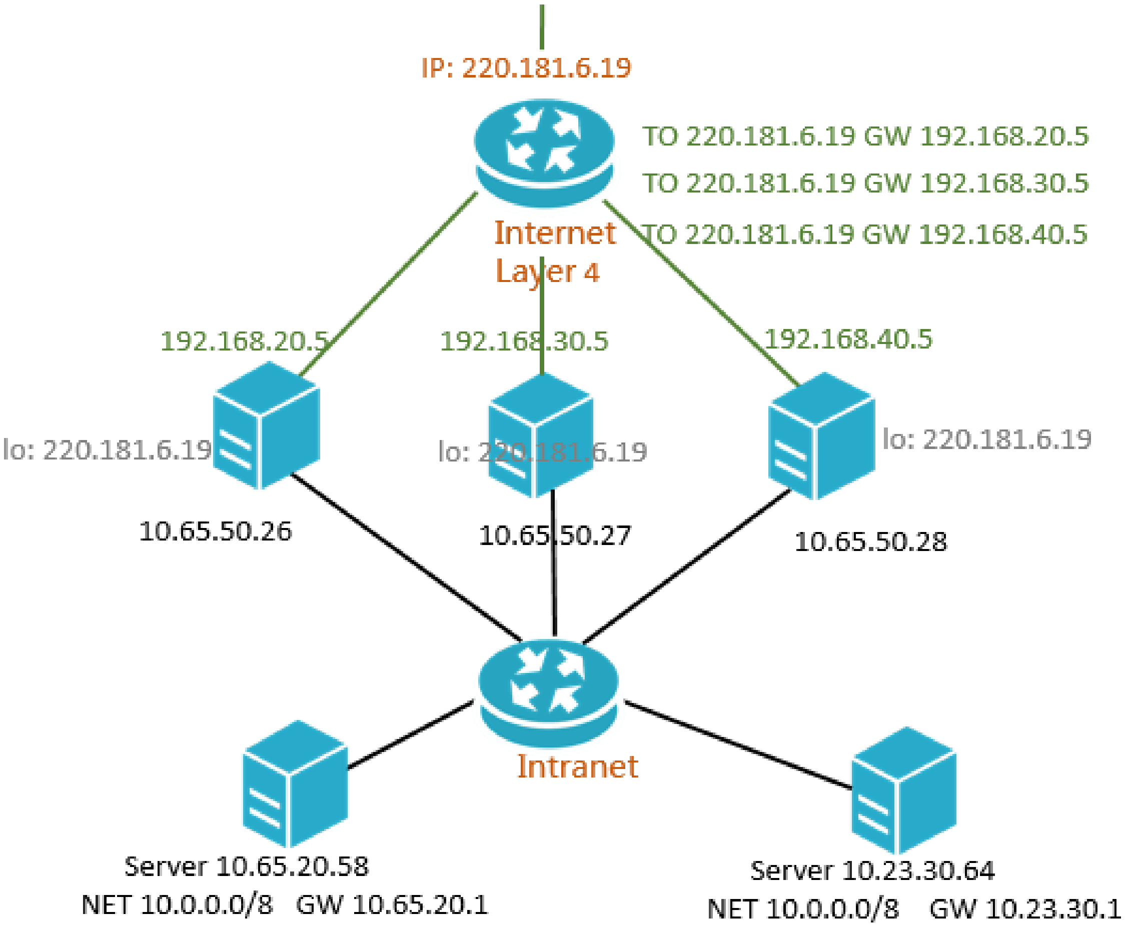
Apsara Stack provides the layer-4 (TCP protocol and UDP protocol) and layer-7 (HTTP protocol and HTTPS protocol) load balancing services. Deployed in clusters, Server Load Balancer can synchronize sessions to protect the ECS instances from single points of failure (SPOFs). This improves redundancy and guarantees the service stability.

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

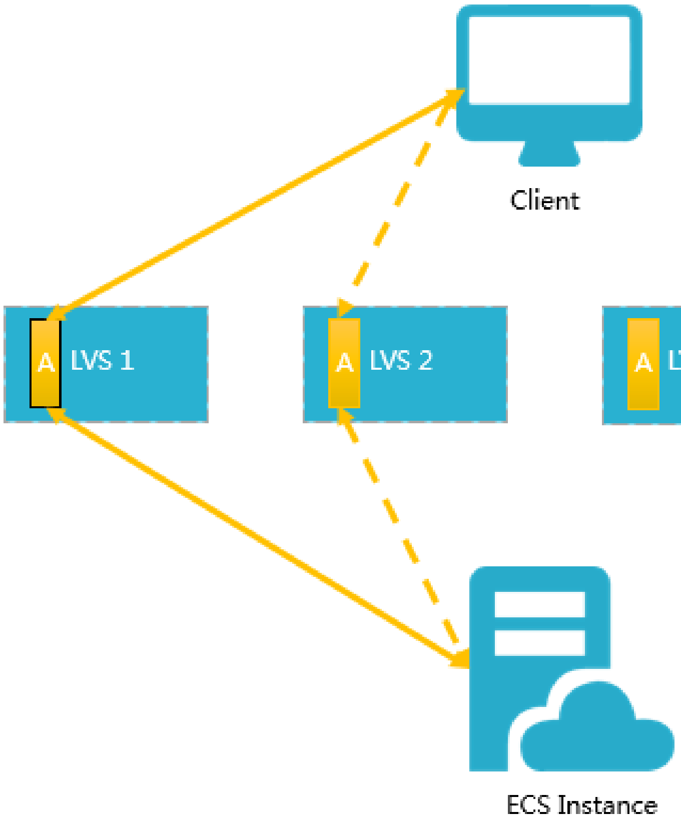
Figure 9-2: SLB architecture



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



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

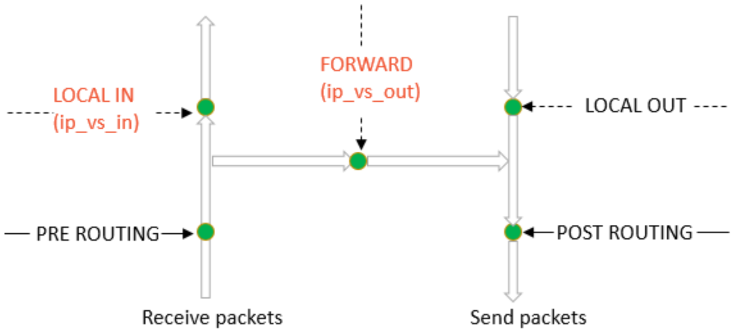


9.3 LVS in layer-4 Server Load Balancer

Problems in standard LVS

LVS is the most popular open-source layer-4 load balancing software in the world, founded by Dr. Zhang Wensong in May 1998. It achieves network load balancing for Linux platforms. LVS is a kernel module implemented on the basis of netfliter framework of Linux (same as iptables), which is known as *IPVS (IP Virtual Server)*. It hooks into netfilter at the NF_IP_LOCAL_IN and NF_IP_FORWARD points.

Figure 9-3: IPVS



In a large-scale cloud computing network, standard LVS has the following drawbacks:

- Drawback 1: LVS supports three packet forwarding methods: NAT, DR, and TUNNEL. When deploying these forwarding modes in a network with multiple VLANs, the network topology becomes complex and poses high O&M costs.
- Drawback 2: LVS lacks the DDoS defense compared with commercial load balancing equipment, for example, F5.
- Drawback 3: LVS uses PC servers and the Virtual Router Redundancy Protocol (VRRP) of Keepalived to do the master-slave deployment. Therefore, its performance cannot be extended .
- Drawback 4: The configurations and health check performance of the keepalived software (widely used in LVS) are insufficient.

LVS customized features

To to solve these problems, Alibaba Cloud added to following customized features to LVS. The URL for Alibaba Cloud LVS is *https://github.com/alibaba/LVS*.

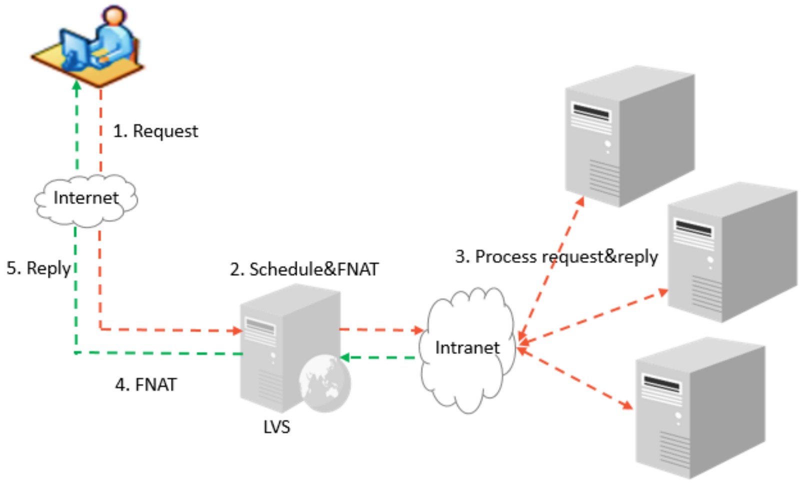
- Customization 1: A new packet forwarding method, FULLNAT, so that LVS load balancer and real servers can be in different vlans.
- Customization 2: Defense modules such as SYNPROXY against synflooding attack.
- Customization 3: Support for LVS cluster deployment.
- Customization 4: Optimization of keepalived performance.

FULLNAT technology

- The main principle is as follows: The module introduces local address (internal IP address), IPVS translates cip-vip to lip-rip, in which lip and rip both are internal IP addresses. This means that the load balancers and real servers can communicate across vlans.

- All inbound and outbound data flows are transferred through LVS. 10 GB NIC is used to ensure the network bandwidth.
- Only TCP protocol is supported by the FULLNAT method.

Figure 9-4: FULLNAT forwarding



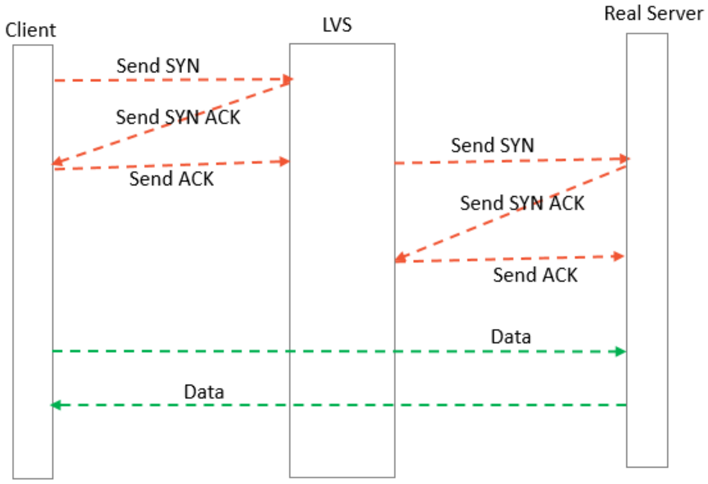
SYNPROXY technology

The main principle is as follows: Based on TPC syncookies, LVS uses a proxy to initiate a TCP three-way handshake.

The proxy process is as follows:

1. A client sends an SYN packet to LVS.
2. LVS constructs an SYN+ACK packet with a special sequence number and sends this packet to the client.
3. The client sends back an ACK response to LVS. LVS checks whether the ack_seq value in the ACK response is valid. If so, LVS establishes a three-way handshake with the real server

Figure 9-5: LVS proxy of three-way handshake



To defend against ACK, FIN, and RST flood attacks, LVS checks the connection table and discards any requests for connections which are undefined in the table.

Cluster deployment

The main principle is as follows: An LVS cluster communicates with the uplink switches over the OSPF protocol. The uplink switches use equal-cost multi-path (ECMP) routing to route traffic to the LVS cluster. Then, the LVS cluster forwards the traffic to the servers.

The cluster deployment ensures the stability of layer-4 Server Load Balancer by supporting the following characteristics:

- Robustness: The LVS and the uplink switches use OSPF as the heartbeat protocol. A VIP is added to all LVS nodes in the cluster. The switches can discover the failure of any LVS node and remove it from the ECMP routing list.
- Fexibility: If the traffic from a VIP exceeds the capacity that the current LVS cluster supports, you can scale up the cluster horizontally.

keepalived optimization

Improvements made by Alibaba Cloud to the Keepalived software include:

- Changing the asynchronous network model from select to epoll.
- Optimizing the reload process.

Benefits of layer-4 Server Load Balancer

As described in the preceeding sections, layer-4 Server Load Balancer has following characteristics:

- High availability: The LVS cluster ensures redundancy and prevents SPOF.
- Security: Together with Alibaba Cloud Security, LVS's intrinsic defenses provide near real-time defensive capabilities.
- Health check: LVS performs health checks on ECS instances and automatically blocks abnormal ones. Once the faulty ECS instance recovers, LVS unblocks it automatically.

9.4 Tengine in layer-7 Server Load Balancer

Tengine is a Web server project initiated by Alibaba. Based on Nginx, Tengine adds a wide range of advanced features dedicated for high-traffic websites. Nginx is one of the most popular open-source layer-7 load balancing software.

The URL for Alibaba Cloud Tengine is *http://tengine.taobao.org/*.

Customized features

For cloud computing scenarios, Tengine customizes the following features:

- Inherits all features of Nginx 1.4.6 and is fully compatible with Nginx configurations.
- Supports the dynamic shared object (DSO) module. This means you do not need to recompile the wholeTengine to add a module.
- Provides enhanced load balancing capabilities, including a consistent hash module and session persistence module. In addition, it can actively perform health checks on backend servers and automatically enable or disable the servers based on their status.
- Monitors system loads and resource usage to protect the system.
- Provides an enhanced attack protection (access speed limiting) module.
- Provides user-friendly error messages to help find the abnormal servers.

Using Tengine as its basic load balancing module, Layer-7 Server Load Balancer has the following features:

Benefits of layer-7 Server Load Balancer combined with Tengine

Using Tengine as the basic module, layer-7 Server Load Balancer has the following characteristics:

- High availability: The Tengine cluster ensures redundancy and prevents SPOF.
- Security: Tengine provides multidimensional protection against HTTP flooding attacks.
- Health check: Tengine performs health checks on ECS instances and automatically blocks abnormal ones. Once the faulty ECS instance recovers, Tengine automatically recovers the ECS instance.

- Supports session persistence.
- Supports consistent hash scheduling.

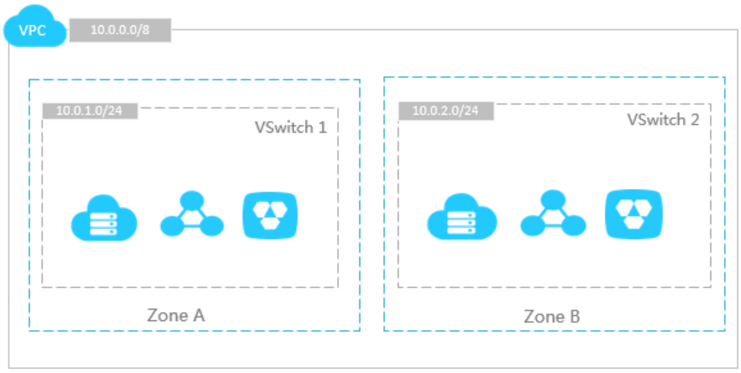
10 Virtual Private Cloud (VPC)

10.1 What is VPC

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

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

Figure 10-1: Virtual Private Cloud



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

- CIDR block

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

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

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

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

- VRouter

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

- VSwitch

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

10.2 Architecture

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

Background information

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

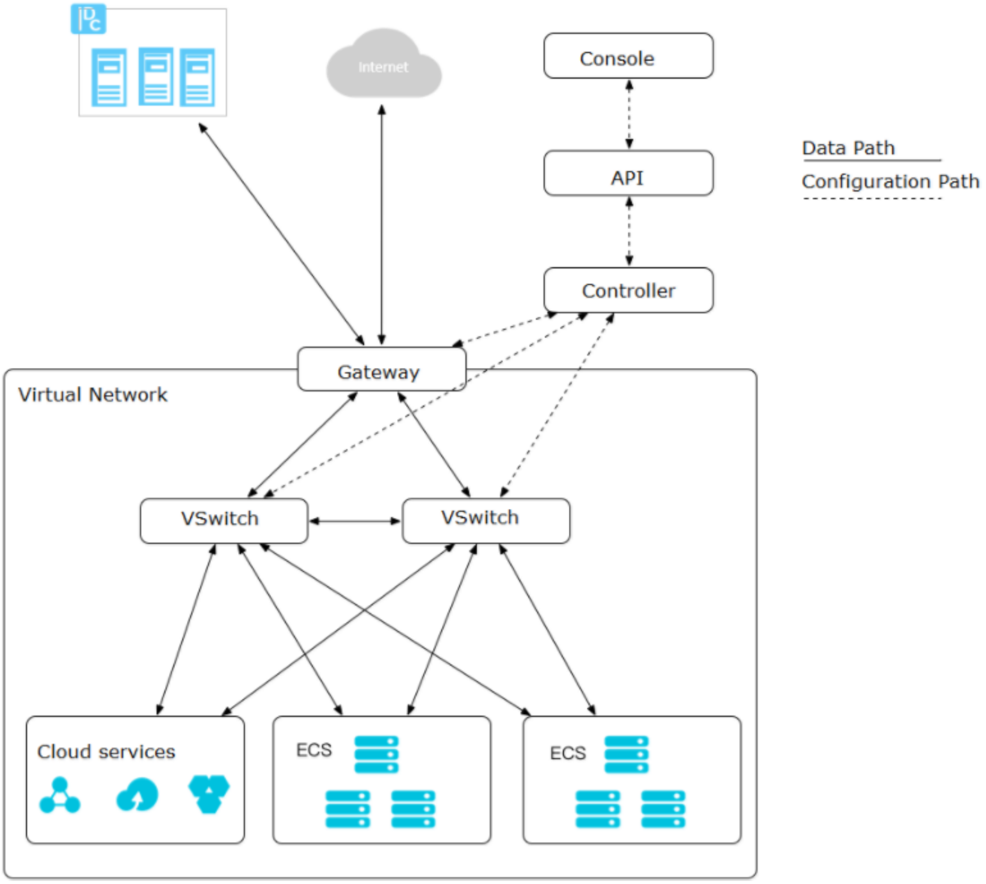
The earlier solutions combined the virtual machine's network with the physical network to form a flat network architecture, such as the large layer-2 network. With the increase of virtual network scalability, problems are getting more serious for the earlier solutions. These problems include ARP spoofing,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

VPC basis

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

Logical architecture

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


IDC
Internet
Console
API
Controller
Data Path
Configuration Path
Gateway
Virtual Network
VSwitch
VSwitch
Cloud services
ECS
ECS

11 Log Service (Log)

11.1 What is Log Service

Log Service is a one-stop solution designed for the log scenario to provide functions such as the collection/subscription, dump and query of massive log data.

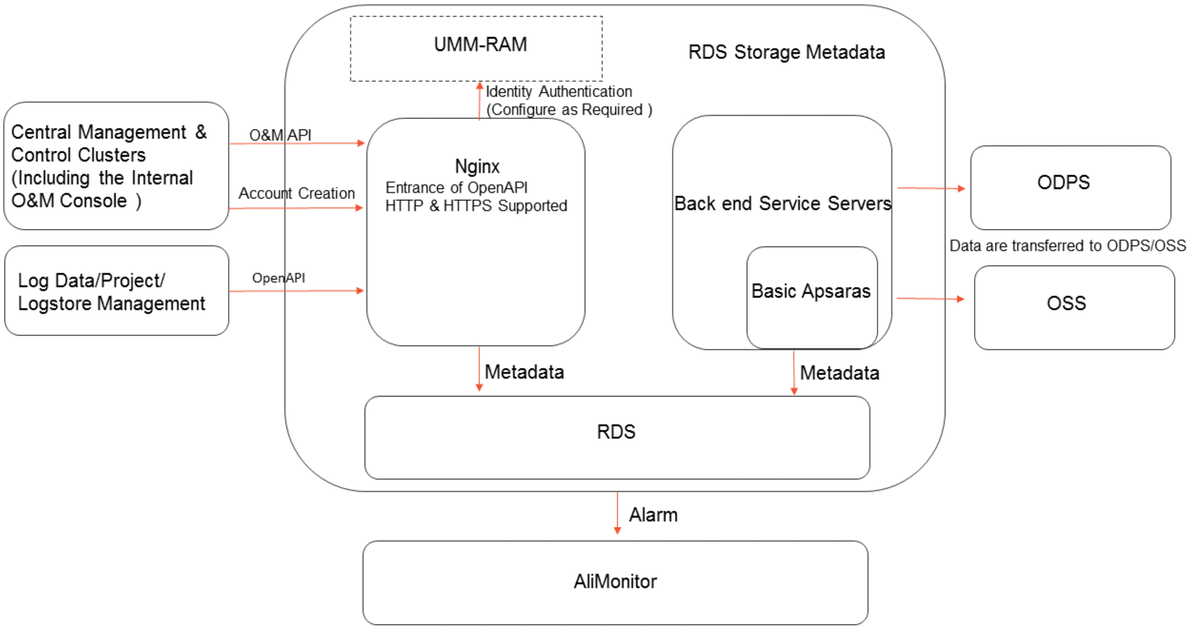
- Real-time collection and consumption: Real-time collection of massive data from multiple channel, using the client, API, Tracking JS, Library and other methods. After the data is written , it can be read in real time. Such as Spark Streaming, Storm, Consumer Library and other interfaces can be used for real-time processing of data.
- Log shipping: Data in the LogHub can be shipped through a certain rules, directory mapping and field development to large-scale storage system.
- Log data indexing and querying: With an index created for the LogHub, real-time and mass storage query engine,LogSearch retrieves logs by time, keywords, context, and other dimensions.

Log service provide fuctions with flexibility and flexibility,and support PB-level data.

11.2 Architecture

The Log Service system architecture is shown in the following diagram.

Figure 11-1: Architecture



- Consoles and OpenAPI are on the left of the architecture and interact with external modules.
- Data transfer service is on the right of the architecture. By using this service, Log Service transfers log data to MaxCompute or OSS.
- The following core modules are in the middle of the architecture:
  - UMM-RAM account module
  - RDS storage metadata
  - Nginx acting as the front end server
  - Log Service background consisting of back end service servers

11.3 Product components

Logtail

Logtail helps you quickly collect logs through the following features：

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

Frontend servers

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

- HTTP and REST protocols
- Horizontal scaling
  - Support horizontal scaling When traffic increases

  - Frontend machines can be quickly added to improve processing capabilities.
- High throughput, low latency
  - Pure asynchronous processing, a single request exception will not affect other requests.
  - Lz4 compression is adopted to increase the processing capabilities of individual machines and reduce network bandwidth consumption.

Backend servers

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

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

11.4 Features

Real-time Log Collection (LogHub)

Real-time collection and consumption. Uses 30+ methods to collect massive data for real-time downstream consumption.

- Using Logtail to collect logs: Stable and reliable, secure, available for all platforms (Linux, Windows, and Docker), high performance, and low resource utilization.
- Using API/SDK to collect logs: Flexible and convenient, scalable, available in 10 + languages and mobile terminals.

- Cloud product log collection: Support logs from Elastic Compute Service (ECS). One key implementation, convenient and efficient.
- Other methods: Syslog, Unity3D, Logstash, Log4j, Nginx, etc.

Real-time Log Consumption (LogHub)

Stream computing, collaborative consumption library, multiple-language support.

- Comprehensive functions: Compatible with 100% of Kafka functions while offering ordering, elastic scaling, time-frame-based seek, and other functions.
- Stable and reliable: Any written data can be consumed; 99.9% availability or better; multiple data copies; elastic scaling within seconds; low cost.
- Easy to use: Support Spark Streaming, Storm, Consumer Library (an automatic load balancing programming mode), SDK subscriptions, and more.

Logshipper

Stable and reliable log shipping. Ship LogHub data to storage services for storage and big data analysis.

- MaxCompute: Ship logs to MaxCompute for analysis.
- Table Store: Ship logs to Table Store.

Logsearch

Real-time data indexing and querying. Create indexes for LogHub data with a time and keyword-based search function.

- Large scale: Real-time indexing of PB-level data volumes (data can be queried within 1 second of writing); query over a billion log entries per second.
- Flexible queries: Support keyword, fuzzy, cross-topic, and context queries.

11.5 Product value

Help to quickly build the solutiona of massive log data.

Typical Log Service application scenarios include: Data collection, real-time computing, data warehousing and offline analysis, product operation and analysis, and O&M and management.

- Data Collection and Consumption
- ETL/Stream Processing
- Data Warehouse
- Event Sourcing/Tracing

- LogManagement

12 Apsara Stack Security

12.1 What is Apsara Stack Security

Apsara Stack Security is a comprehensive Apsara Stack security solution that provides cloud security with network security, host security, application security, data security and security management dimensions.

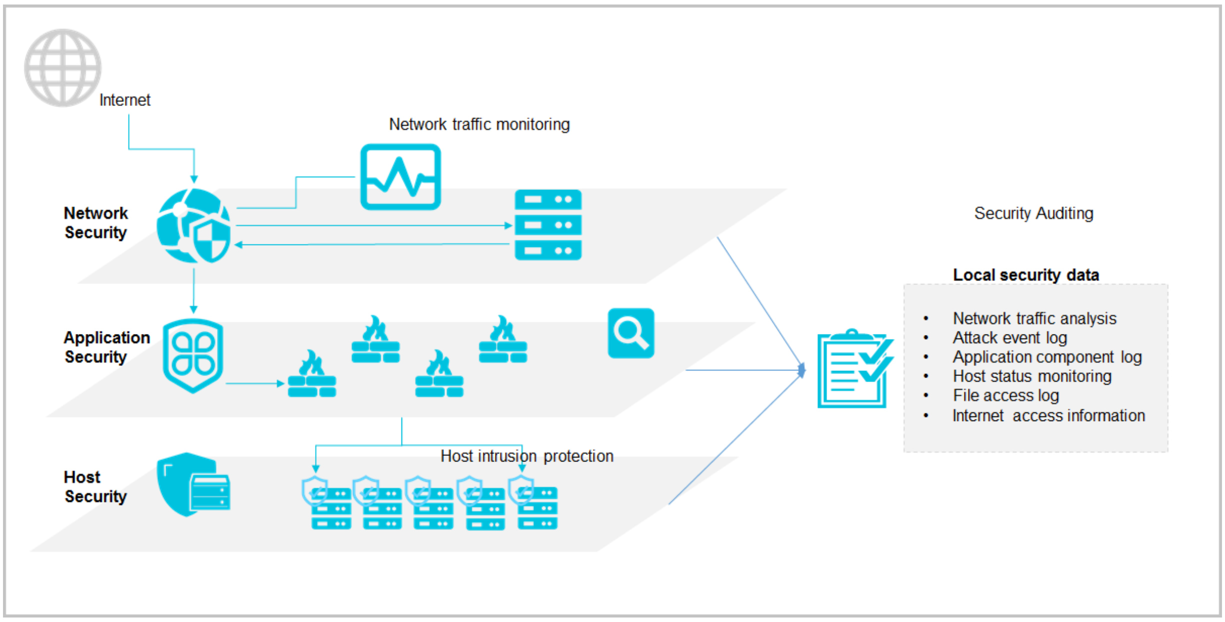
In the cloud computing environment, the traditional border security protection that relies on the detection technology cannot guarantee the security of businesses on the cloud. Apsara Stack Security combines the powerful data analysis capabilities of cloud computing platform with professional security operation team, to provide a multi-level and integrated security protection service.

12.2 Architecture

Apsara Stack Security Basic Edition

Figure 12-1: Structure of Apsara Stack Security Basic Edition shows the structure of Apsara Stack Security Basic Edition in Apsara Stack Enterprise.

Figure 12-1: Structure of Apsara Stack Security Basic Edition



- **Network Traffic Monitoring**: This module is deployed on Apsara Stack network boundaries. It allows you to inspect and analyze each packet passing through the Apsara Stack network by traffic mirroring. The analysis results then are referenced by other Apsara Stack Security modules.

- **Host Intrusion Detection System (HIDS)**: This module is used to detect the integrity of the key folders in the host and send alerts for abnormal processes, ports, and network connections on the host.
- **Server Guard Basic Edition**: This module is deployed on an ECS instance, which detects and removes web Trojans, blocks brute force password cracking attacks, and sends alerts for abnormal logons.
- **Security Audit**: This module is used to collect database logs, host logs, console operation logs on the user and O&M sides, and network device logs in the Apsara Stack platform.

12.3 Features

12.3.1 Apsara Stack Security Basic Edition

12.3.1.1 Network Traffic Monitoring

The network traffic monitoring module is able to monitor attacks within milliseconds after they occur. By performing in-depth analysis on the traffic packets mirrored from the Apsara Stack boundaries, this module can detect various attacks and abnormal behaviors in real time and coordinate with other protection modules to implement defenses. In addition, the network traffic monitoring module provides a wealth of information output and basic data support throughout the Apsara Stack Security defense system.

Features

The network traffic monitoring module provides the following functions:

Function	Function description
Traffic statistics	Collects inbound and outbound traffic of the interconnection switch (ISW) using a bypass in traffic mirroring mode and generates a traffic diagram.
Abnormal traffic detection	Detects abnormal traffic that exceeds the threshold using a bypass in traffic mirroring mode and leads the traffic for DDoS Cleaning. The traffic rate (Mbit/s), packet rate (PPS), HTTP request rate (QPS), or number of new connections can be set as the threshold.
Web application attack protection	Conducts network-layer interception and bypass blocking for common web application attacks, including SQL injection, code and command execution, script Trojan, file inclusion, and usage of upload and common

Function	Function description
	CMS vulnerabilities, based on embedded web application attack detection rules.

Key concepts

The network traffic monitoring module performs in-depth packet analysis on the traffic and detects various attacks and abnormal behaviors in real time. Additionally, it reports security events to the Apsara Stack cloud security central console and interacts with other protection systems. Furthermore, the network traffic monitoring module provides comprehensive support using a wealth of output information and basic data, and can integrate with the entire Apsara Stack Security system.

The network traffic monitoring module processes data in the sequence of collection, gathering, and output, and uses sockets for data exchange.

- Collection: The module collects traffic data using multiple high-performance PCs with dual-port 10GE network cards.
- Gathering: Traffic from an IP address may pass through multiple collectors. Therefore, traffic data must be consolidated to generate usable information.
- Output: The module stores and outputs the consolidated traffic data.

The network traffic monitoring module collects traffic mirrored from the following data center portals:

- Traffic mirrored from the 10 GE switch portal
- Traffic output from the splitter and shunt

The network traffic monitoring module outputs the following information:

- Alert information
- Traffic information, which is sent to the database
- HTTP logs, which are sent to the Situation Awareness analysis engine platform (Only applies to Apsara Stack Security Advanced Edition)
- Layer-4 and layer-7 attacks detection results, which are sent to the cloud security central console

Performance indicators

A single network traffic monitoring device provides a processing capability for 10 Gbit/s traffic.

12.3.1.2 Host Intrusion Detection System

The host intrusion detection system (HIDS) module collects information and performs detection through the client deployed on the physical servers. It detects file tampering, abnormal processes , abnormal network connections, suspicious port monitoring, and other behaviors on all servers in the Apsara Stack environment. This helps you immediately detect potential server security risks.

Features

The HIDS module provides the following functions:

Function	Function description
Key directory integrity check	Checks integrity of files in specified folders of the host system, detects tampering in time, and generates change alerts. The specified directories include '/etc/init.d'.
Abnormal process alert	Detects startup of abnormal processes in time, and generates alerts accordingly. Abnormal processes, such as XOR DDoS, Bill Gates malware bot family, and minerd, can be detected.
Abnormal port alert	Detects new port monitoring in time, and generates alerts accordingly.
Abnormal network connection alert	Detects active connections with external networks in time, and generates alerts accordingly.

Performance indicators

To ensure the physical server performance, the HIDS client is subject to the following performance constraints:

- **Normal working status**: The CPU usage of the HIDS client is 1%, and the memory usage is 50 MB.
- **Peak working status**: The CPU usage of the HIDS client is 10%, and the memory usage is 80 MB. If the peak value of CPU usage or memory usage is exceeded, the detection program automatically stops.

12.3.1.3 Server Guard Basic Edition

Server Guard Basic Edition provides security protection measures such as brute-force cracking protection, webshell detection and removal, and remote logon alert. It provides security protection measures for ECS instances by means of log monitoring, file analysis, and feature scanning. Server Guard is divided into clients and servers. Server Guard clients work with Server Guard servers to monitor attack behavior at the system layer and application layer and detect hacker intrusions in real time.

Features

Server Guard Basic Edition provides the following functions:

Function	Function description
Webshell detection and removal	Accurately detects and removes webshell scripts compiled using ASP, PHP, or JSP on the ECS instances by means of rule matching, and allows you to manually isolate webshell scripts.
Interception of brute-force password cracking	Detects and intercepts the brute-force password cracking behaviors initiated by hackers in real time, and monitors brute-force password cracking of the SSH and RDP services.
Remote logon alert	Analyzes and records users' frequently-used logon locations to identify frequently-used logon regions (accurate to the city), and generates alerts for suspicious logon behaviors in non-frequently-used logon regions.

Key concepts

The client of Server Guard Basic Edition consists of the webshell feature database and webshell isolation module.

- The webshell feature database is used to check whether a file is consistent with the features in the feature database. If it is consistent, the file is sent to the Server Guard server, which will further analyze whether the file is a Trojan based on additional feature databases.
- The webshell isolation module is used to isolate a file if the Server Guard server determines that the file is a Trojan.

The server of Server Guard Basic Edition consists of the Aegis-server and Defender modules.

- Aegis-server consists of the communication module and client check module. It interacts with Aegis-client to collect information about Trojan and patch files, and reports information about the remote logon, brute-force password cracking, and successful brute-force password cracking to Defender.
- Defender analyzes information about the remote logon and brute-force password cracking, and checks whether brute-force password cracking is successful.

The Server Guard server provides APIs for the Apsara Stack cloud security center console to obtain information, parses and analyzes security events, and displays analysis results. You can issue commands to isolate or ignore Trojan files.

Performance indicators

To ensure the ECS host performance, the client of Server Guard Basic Edition is subject to the following performance constraints:

- **Standard working status**: The CPU usage of the client of Server Guard Basic Edition is 1%, and the memory usage is 50 MB.
- **Peak working status**: The CPU usage of the client of Server Guard Basic Edition is 10%, and the memory usage is 80 MB. If the peak value of CPU usage or memory usage is exceeded, the client of Server Guard Basic Edition automatically stops.

12.3.1.4 Security Audit

The security audit module is an integrated solution that meets the basic requirements for information system classified security protection. It implements behavior log collection, storage , analysis, and alert functions at the physical server layer, network equipment layer, and cloud computing platform application layer.

Features

The security audit module provides the following functions:

Function	Function description
Raw log collection	Collects the RDS SQL logs, host syslogs, console operation logs on the user and O&M sides, and network equipment syslogs. **Note:** The network equipment logs must be manually collected.
Audit query	Allows you to query audit logs by audit type, audit object, operation type, operation risk level, alert, or creation time. In addition, full text retrieval of audit logs is supported.
Policy setup	Allows you to configure audit rules using the following parameters: Initiator, Target, Command, Result, and Cause. In addition, the module can identify high-risk operations in raw logs, and generate alerts accordingly.

Key concepts

Security audit is completed by the Auditlog component. Auditlog collects the logs of network equipment, ECS instances, physical servers, ApsaraDB for RDS, and APIs. Apsara Stack Security calls the Auditlog API to obtain the audit logs, audit policies, and audit events, analyzes the audit logs and events, and displays analysis results in graphs to help you learn the risks and threats faced by the system.

Benefits

The security audit module has the following features and advantages:

- **All-encompassing behavior logs**

  The module covers multiple cloud computing services and physical hosts. It can collect behavior information from various perspectives, ensuring a full audit coverage. The log collection center supports centralized and synchronized collection of behavior logs in quasi-real time.
- **Reliable log storage**

  Log storage is based on cloud computing storage services and clustered using triplicate technology (that is, in three copies). This ensures secure and stable storage. The storage space can be quickly expanded.
- **Real-time query of massive data**

  By creating a full-text index for massive volumes of log data, this module provides fast retrieval and query capability for large volumes of data. It supports simultaneous indexing of 50 billion rows of log data.

12.3.2 Apsara Stack Security Advanced Edition

In addition to all functional modules of the Basic Edition, the Advanced Edition also contains the following modules: Server Guard Advanced Edition, DDoS Cleaning, Web Application Firewall (WAF), Cloud Firewall, Bastion Host, and Situation Awareness.

For more information about the functional modules of the Basic Edition, see *Apsara Stack Security Basic Edition*.

12.3.2.1 Server Guard Advanced Edition

Server Guard Advanced Edition provides security protection measures such as vulnerability management, baseline check, intrusion detection, and asset management for ECS instances by means of log monitoring, file analysis, and feature scanning. The Server Guard Advanced Edition module is divided into clients and servers. Server Guard clients work with Server Guard servers to monitor attack behaviors and vulnerability information at the system layer and application layer, protecting the security of ECS instances in real time.

Features

Server Guard Advanced Edition provides the following extended functions:

Note:
Server Guard Advanced Edition includes all functions of Server Guard Basic Edition.

Function	Function description
Linux-CVE vulnerability management	Performs exact match on CVE vulnerabilities in the Linux system based on the official CVE vulnerability database, periodically scans the system to detect vulnerabilities, and generates vulnerability repair commands.
Windows vulnerability management	Subscribes to the Microsoft official updates to support vulnerability detection of the Windows system, synchronizes the Microsoft official patch files to the Server Guard server, and supports vulnerability repair and rollback.
Web-CMS vulnerability management	Identifies web directory files on the ECS instances at the source code level, accurately identifies CMS vulnerabilities, and uses proprietary vulnerability patches to support vulnerability repair and rollback.
Configuration-type and component-type vulnerability management	Accurately identifies software high-risk configuration vulnerabilities and component-type vulnerabilities such as ImageMagick.
Account security baseline check	• Detects SSH, RDP, FTP, MySQL, PostgreSQL, and SQLServer accounts with weak passwords. • Detects potentially risky accounts of ECS instances, such as suspicious hidden accounts and cloned accounts. • Checks the password policy compliance of the Linux server. • Detects the null-password accounts of ECS instances.
Database security baseline check	Checks whether the Redis service on the server is opened to the public network, whether vulnerabilities such as authorized access exist

Function	Function description
	, and whether abnormal data is written to key files of the system.
System security baseline check	• Checks whether the scheduled tasks of the Linux server contain suspicious self-startup items. • Checks the self-startup items on the Windows server. • Checks the system shared configurations. • Checks the SSH logon security policy settings of the Linux server. • Checks the account-related security policies on the Windows server.
Webshell detection and removal	Accurately detects and removes webshell scripts compiled using ASP, PHP, or JSP on the ECS instances by means of rule matching , and allows you to manually isolate webshell scripts.
Interception of brute-force password cracking	Detects and intercepts brute-force password cracking behaviors initiated by hackers in real time, and monitors brute-force password cracking of SSH and RDP services.
Remote logon alert	Analyzes and records users' frequently-used logon locations to identify frequently-used logon regions (accurate to the city), and generates alerts for suspicious logon behaviors in non-frequently-used logon regions.
Asset grouping	Divides ECS instances into a maximum of four groups, and supports filtering by region or online status.
Host Fingerprints	• Listener ports: Periodically collects information about listener ports on the server. • Accounts: Periodically collects system account information on the server. • Processes: Periodically collects information about processes on the server. • Software: Periodically collects software version information on the server.

Function	Function description
Log Retrieve	• Logon history: Logs of successful logons. • Brute force cracking: Logs of brute force cracking attacks. • Process snapshot: Logs of processes on the server at a specific time. • Port snapshot: Logs of listener ports on the server at a specific time. • Account snapshot: Account logon informatio n on the server at a specific time. • Process initiation log: Logs of process initiation on the server. • Network connection log: Logs of outgoing connections from the server.

Key concepts

The client of Server Guard Advanced Edition consists of the webshell feature database, webshell isolation module, patch feature database, and patch-based repair module.

- The webshell feature database is used to check whether a file is consistent with the features in the feature database. If it is consistent, the file is sent to the Server Guard server, which will further analyze whether the file is a Trojan based on additional feature databases.
- The webshell isolation module is used to isolate a file if the Server Guard server determines that the file is a Trojan.
- The patch feature database is used to check whether a file meets features in the feature database. If yes, the file is sent to the Server Guard server, which will further analyzes whether the file is a vulnerability based on additional feature databases.
- The patch-based repair module is used to repair a file if the Server Guard server determines that the file is confirmed as a vulnerability.

The client of Server Guard Advanced Edition is available in two versions: Windows version and Linux version. The client automatically connects to the server for online upgrade.

The server of Server Guard Advanced Edition consists of the Aegis-server, Defender, and Aegis-health-check modules.

- Aegis-server consists of the communication module and client check module. It interacts with Aegis-client to collect information about Trojan and patch files, and reports information about remote logons, attempted brute-force password cracking, and successful brute-force password cracking to Defender.

- Defender analyzes information about the remote logon and brute-force password cracking, and checks whether brute-force password cracking is successful.
- Aegis-health-check performs baseline checks. It sends the baseline check commands to the client through Aegis-server, collects data returned by the client through Aegis-server, and uses the returned data to make modifications to the baseline check status.

Use case

Server Guard Advanced Edition is applicable to host security protection in the following scenarios:

- **Generic software exploitation**

  Server Guard Advanced Edition can detect vulnerabilities that are exploited maliciously by hackers using generic software. Once detected, a vulnerability can be quickly repaired in one click.
- **Web application services**

  Server Guard Advanced Edition can effectively prevent internal and external attacks initiated by hackers regardless of whether the service is internal or external.

Performance indicators

To ensure the ECS host performance, the client of Server Guard Advanced Edition is subject to the following performance constraints:

- **Normal working status**: The CPU usage of the client of Server Guard Advanced Edition is 1%, and the memory usage is 50 MB.
- **Peak working status**: The CPU usage of the client of Server Guard Advanced Edition is 10%, and the memory usage is 80 MB. If the peak value of CPU usage or memory usage is exceeded, the client of Server Guard Advanced Edition automatically stops.

12.3.2.2 DDoS Cleaning

Alibaba Cloud designed and developed the DDoS cleaning module based on the cloud computing architecture to protect the cloud platform against massive DDoS attacks.

Features

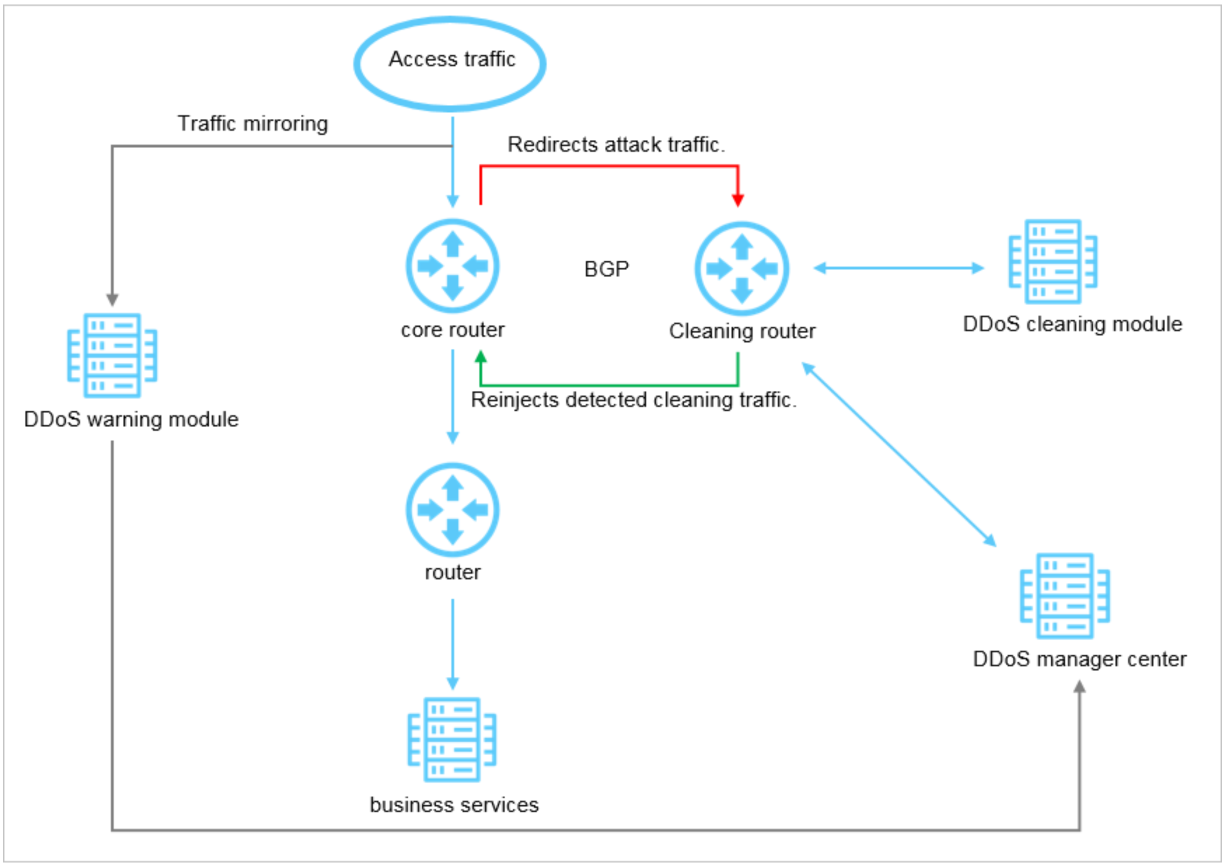
The DDoS cleaning module provides the following functions:

Function	Function description
DDoS attack cleaning	Detects and protects the system against attacks, such as SYN flood, ACK flood, ICMP flood, UDP flood, NTP flood, DNS flood, and HTTP flood.
DDoS attack viewing	Allows you to view DDoS attack events on the GUI and search for DDoS attack events by IP address, status, and event information.
DDoS traffic analysis	Allows you to analyze the traffic of a DDoS attack, view the traffic protocol of the DDoS attack, and display Top 10 IP addresses related to this attack.

Key concepts

After detection and scheduling performed by the network traffic monitoring module, the DDoS cleaning module redirects, cleans, and reinjects detected attack traffic. This provides protection against DDoS attacks and ensures normal service provisioning. The traffic cleaning process is shown in *Figure 12-2: Traffic cleaning process*.

Figure 12-2: Traffic cleaning process



When the DDoS attack threshold is reached on a targeted host, Apsara Stack Security automatically identifies the attack traffic and starts traffic cleaning.

Benefits

The DDoS cleaning module has the following features and advantages:

- **Full coverage of common DDoS attack types**

  The DDoS cleaning module protects your system against various types of DDoS attacks at the network, transmission, or application layer, including HTTP flood, SYN flood, UDP flood, UDP DNS query flood, (M)Stream flood, ICMP flood, HTTP GET flood, and more. It also sends status updates through SMS to notify you of your website defense status.
- **Response, and protection activation within one second**

  The DDoS cleaning module adopts world leading detection and protection technologies and completes attack discovery, traffic redirection, and traffic cleaning within one second. The traffic thresholds are configured to trigger protection. Furthermore, statistics of network behaviors are recorded to accurately identify DDoS attacks, which greatly reduces network jitter and ensures availability of your business in the case of a DDoS attack.
- **Highly elastic and redundant anti-DDoS capabilities**

  The DDoS cleaning module can filter 20 Gbit/s of attack traffic at the minimum. Based on high elasticity and high redundancy of the cloud computing architecture, the DDoS cleaning module can be seamlessly re-sized in the cloud environment to achieve highly elastic anti-DDoS capabilities.
- **Full protection to avoid malicious use of cloud resources**

  The DDoS cleaning module not only protects your system from external DDoS attacks, but also detects malicious use of internal cloud resources. Once an ECS instance is found to be used to initiate DDoS attacks, the network traffic monitoring module will collaborate with the HIDS module to restrict the network access of the hijacked ECS instance and generate an alert, so as to effectively control internal hosts.

Performance indicators

A single DDoS cleaning device can provide a processing capability for 10 Gbit/s traffic.

12.3.2.3 Web Application Firewall

Web Application Firewall (WAF) can protect website applications against attacks of common web vulnerabilities including SQL injection, XSS, and other types of web application attacks, or HTTP flood attacks, and other types of attacks that affect website availability by consuming resources. In addition, WAF allows you to develop precise protection policies, based on characteristics of your website, to filter out the malicious web requests sent to your website.

The WAF module protects the traffic on HTTP and HTTPS websites. On the WAF console, you can import certificates and private keys to realize end-to-end encryption of your services, prevent your data from being intercepted on the links, and meet security protection requirements for HTTPS services.

The WAF module also supports rule sorting in protection scenarios and allows you to adjust the relationship between precise protection and other security protection policies. For example, you can determine whether to enable HTTP flood and general web protection policies after matching requests against the precise protection policy. As a custom protection policy, precise protection is always given the highest priority during request matching.

Features

The WAF module provides the following functions:

Function	Description
Protection against common web attacks	Protects your website against SQL injections, XSS, harmful file upload, files with vulnerabilities, common directory traversal, common CMS vulnerabilities, code execution injection, webshells, scanner attacks, and other types of web attacks. The observation and blocking modes are provided to handle web attacks as follows: • In observation mode, WAF generates alerts for the attacks, but does not immediately block the traffic. This facilitates evaluation and prevents false alerts. • In blocking mode, WAF directly blocks attack-related requests.
HTTP flood attack protection	Takes statistics on the request URL frequency, access address distribution, and abnormal response codes, and intercepts abnormal behaviors. Provides two HTTP flood protection modes for HTTP flood attacks: default and emergency modes. If your website cannot be accessed due to HTTP flood attacks, you can enable the emergency mode to enhance HTTP flood protection and relieve HTTP flood attacks. Supports custom access frequency control for URL access with a single source IP address.
Precise access control	Provides a friendly configuration console interface and supports condition combinations for common HTTP fields, including IP, URL,

Function	Description
	Referer, and User-Agent. This allows you to create powerful precise access control policies that are applicable to scenarios such as anti-leeching and website background protection.
Automatic banning of malicious IP addresses	Automatically bans an IP address for a period of time if the IP address continuously initiates web attacks to the domain name.
Region banning	Provides the region banning capabilities based on geographical locations to ban source IP addresses of a specified province or region other than China at one click.

Use cases

WAF is applicable for the protection of various web applications in fields such as finance, e-commerce, O2O, Internet+, games, government, and insurance. It is designed to solve the following problems:

- Prevents data leaks and avoids intrusions from hacker injections, which may result in leaks of core databases of your website.
- Prevents malicious HTTP flood attacks to safeguard website availability.
- Prevents Trojans from being uploaded to web pages, ensuring credibility of your website.
- Provides virtual patches to address the latest website vulnerabilities exposed, and provides quick fixes as they are released.

Performance indicators

- Traffic that can be carried by a single WAF is as follows:
  - QPS: 10,000 for HTTP or 2,000 for HTTPS
  - Concurrent connection count: 10,000,000
  - New connection count: 30,000
- The attack detection latency of WAF is less than 10ms.

12.3.2.4 Cloud Firewall

The Cloud Firewall module is a firewall service applicable to Apsara Stack. It resolves problems caused by rapid Apsara Stack service changes, such as blurred security boundaries or even failures to define security boundaries. Cloud Firewall implements secure access control on east-west traffic in Apsara Stack. It uses groundbreaking techniques such as business sorting and business isolation based on visualized business results.

- **Zero-configured business visibility**: automatically groups business without affecting business operations, so that you can clearly view the business structure.
- **Business orderliness**: uses visualization technologies and a topological graphical interface to help you keep business in order as cloud services increase rapidly.
- **Micro-segmentation**: enables secure micro-segmentation deployment for specific server applications in Apsara Stack.

Functions

The following table lists functions of Cloud Firewall.

Function	Description
Topology mechanism	Displays all cloud server information and inter-server access relationships within a topology grouped by clicks, float views and highlights. Global business is displayed visually.
Display associations	Associates server information with the access relationship and displays the association through multiple visualization methods during role-based management of cloud servers. It combines access segments into an integrated access relationship.
Role management	Defines roles for cloud servers. Cloud Firewall provides shortcuts to locate server usage and their access relationship when your business goes offline.
Smart search	Provides overlying search conditions and diversified rule actions to help administrators locate the servers or traffic they are looking for in complicated business typologies.
Business grouping	Groups cloud servers based on business.
Traffic curve based policy definition	Verifies the validity of traffic with visualized traffic graphs. A click on a traffic curve enables the delivery of policies.
Access relationship authorization	Access policies may be enabled or changed from time to time during normal business accesses. Cloud Firewall visualizes the traffic to help you easily enable or change your whitelist policies.
Observation mode	Simulates the delivery of policies. If you are unsure about the efficiency of the policies to be delivered, you can use the observation mode to simulate the impact of the policies on the traffic, and optimize the policies as required.
Allow all traffic	Provides a temporary policy for special cases that allow all traffic to the current business. This is used to allow you more time to troubleshoot problems.

Working principles

Cloud Firewall is a security service based on the distributed architecture. It learns and displays business traffic in Apsara Stack, and assists you with rapid business grouping and security policy definition.

- Cloud Firewall restores business traffic, such as server information and access relationship between servers. It performs a series of algorithm-based data analysis to help you rapidly locate what you want to view, such as servers and access connections.
- Cloud Firewall helps you review traffic successfully, group disordered business, and deploy security policies.
- Cloud Firewall proactively learns and globally displays continuous business traffic to help you maintain orderly business.

Use cases

Cloud Firewall can be used to:

- **Implement micro-segmentation**: You can use Cloud Firewall to implement fine-grained micro-segmentation. Cloud Firewall manages ports that must be enabled to prevent business interruptions in a more fine-grained manner through business and role grouping. Therefore, it reduces attacks and security risks.
- **Check whether traffic is secure**: You can check whether traffic is secure in the traffic view of Cloud Firewall. For example, you can check whether HTTP traffic has been switched to HTTPS traffic, or whether the traffic destined for TCP port 3306 (service port of MySQL) includes traffic from the Internet.
- **Determine whether the business is affected by server changes**: When servers need to be migrated or shut down, you can use Cloud Firewall to see the relevant server traffic, and determine whether the business is affected by server changes.
- **Help with rapid business expansion**: Cloud Firewall defines policies in an IP-free manner, to ensure that policies do not frequently change when business increases rapidly. For example, you can assign some newly added servers with the same roles at peak hours to implement expansion, without the need for any modifications to the policies.
- **Implement visualized traffic and rapid operations and maintenance**: Traditionally, the **`tcpdump`** command or packet capture software was the only method to view the connections between inbound and outbound servers. Cloud Firewall uses traffic flow visualization to display the related information visually.

- **Detect port misuse**: With multiple development departments, there is a chance that a service (same applications and processes) provided on a server uses different ports. In cases such as these, port resources are wasted, which in turn complicates operations and maintenance. Based on visualized traffic, Cloud Firewall can clearly identify port misuse.

Benefits

The following table lists characteristics and benefits of Cloud Firewall compared with traditional firewalls.

Traditional firewall	Cloud Firewall
Does not focus on your business.	Provides you with values by visualizing traffic flow. Policies can be created and deployed based on actual business traffic .
Does not focus on the correctnes s of policies.	Ensures correct policies to the maximum extent through visualized traffic.
Does not focus on the complexity of future policy operations and maintenance.	Uses topology mechanisms to display asset information and access relationship between assets. It simplifies operations and maintenance.

12.3.2.5 Situation Awareness

The Situation Awareness module enables the following vulnerability detection requirements:

- **Vulnerability analysis**: Vulnerability analysis is based on the stateless scan technology. Situation Awareness works together with the network traffic monitoring module and uses a combination of dynamic detection and static matching scanning modes to provide you with automated, high-performance, and precise web vulnerability scanning capabilities.
- **Big data security analysis platform**: Situation Awareness uses machine learning and data modeling to find potential infiltration and attack risks. From the attacker's perspective, it effectively captures zero-day vulnerability attacks mounted by advanced attackers and new virus attacks, and displays ongoing security attacks. It also visually presents this information, keeping you aware of business security in an easy to understand way. This solves the problem of data leaks due to cyber attacks and allows you to discover the hacker's identity using the tracing service.

Features

The Situation Awareness module provides the following functions:

Function	Function description
Security situation overview	Provides the overall security information, including the number of emergencies, today's attacks, today's vulnerabilities, security attack trend, latest thread analysis, latest intelligence, and protected assets information.
Access analysis	Analyzes all external access to web services in the protection range , including top 10 accessed services, number of IP addresses that access services normally, number of IP addresses that access services maliciously, number of crawling IP addresses, and some access detailed samples.
Visualization dashboard	Provides a dashboard to display the map-based traffic and host security status.
Intrusion analysis	Supports threat detection in the following models: • MySQL privilege escalation and webshell writing • Download of malicious files or scripts in the Linux system • Minerd process detection on the host side • Windows logon credential interception • VBScript abnormal commands • Download of suspicious files with PowerShell • Scanner attack • Minerd process running • Malicious use of Pingback • Central control of botnets • Shell reflection • Malicious use of Redis • MongoDB ransom • MySQL database ransom
Traffic analysis	Records statistics on traffic in the monitoring range, including today 's traffic, traffic from 30 days, traffic from 90 days, and QPS. Today 's traffic, and the traffic from 30 days and 90 days of a specific IP addresses can be displayed.
Malicious host recognition	Detects external attack behaviors of internal malicious hosts, and identifies the controlled internal hosts, HTTP flood attacks, and DDoS attack commands.
Web attack detection	Checks the use of web vulnerabilities, scan tools, upload, connected webshells, SQL injection, XSS attacks, local and remote files that contain "include", and execution of codes and commands.

Function	Function description
Utilization of server vulnerabilities	Converts feature characters into a binary string for matching based on the packet features, for example, utilizing the vulnerabilities in Redis application.
Application vulnerability analysis	Scans vulnerabilities at the web application layer, provides measures for immediately verifying the scanning results and corresponding repair suggestions, periodically performs automated scanning by integrating NAT assets and host assets, verifies detected application vulnerabilities in a centralized manner, and updates the application vulnerability status after verification.
Host vulnerability analysis	Scans and detects vulnerabilities at the host layer, and provides scanning results and corresponding repair suggestions.
Weak password analysis	Scans weak passwords for logon to common systems, such as web, SSH, and FTP, supports addition of custom weak password, performs automated scanning every night by integrating NAT assets and host assets, verifies detected weak passwords in a centralized manner, and updates the weak password detection time after verification.
Configuration item detection	Scans the access to external service pages, generates alerts for leaks of web page configuration items, verifies the detected leaks of configuration items every night, and updates the detection time.

Key concepts

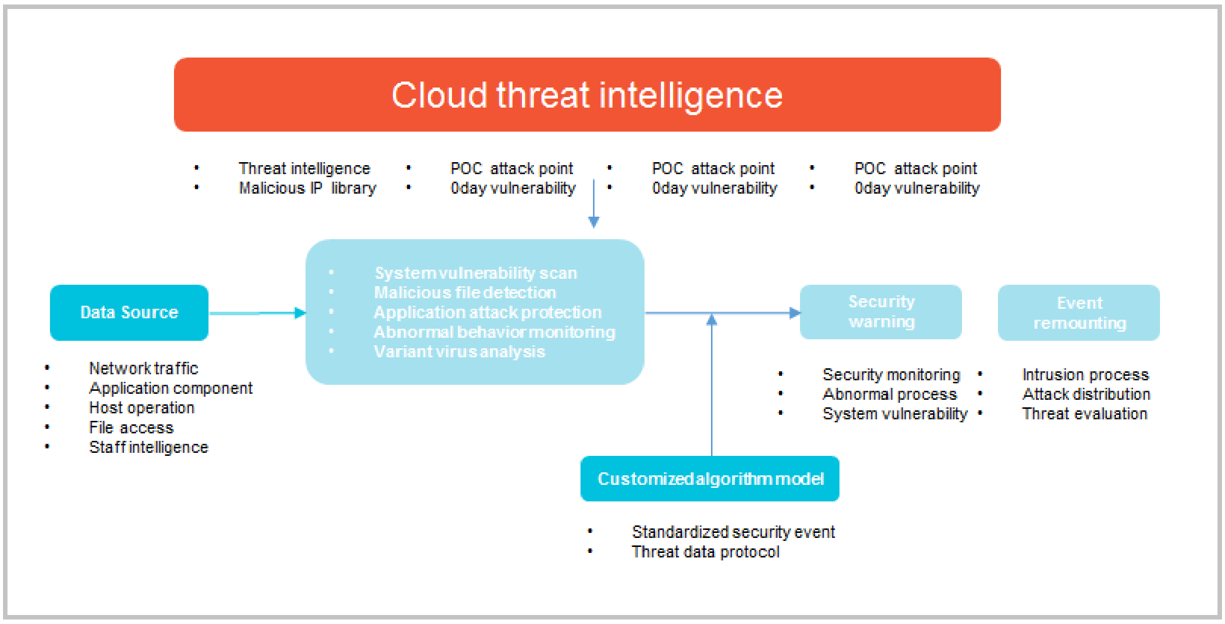
The vulnerability analysis function involves the Cactus-batch and Cactus-keeper modules.

- The Cactus-batch module processes data, and sends the processed URLs that need to be scanned to the Cactus-keeper module using a message queue.
- The Cactus-keeper module is integrated with the scan engine and equipped with Apsara Stack's rich scanning rules and plug-in library. It scans for system vulnerabilities, weak passwords, and configuration items, and detects and reports vulnerabilities in the system. This allows you to observe any defects in the system and take measures to resolve them.

The big data security analysis platform uses two different analysis methods: corresponding rule analysis and machine learning modeling analysis, to perform aggregated analysis on collected logs. Additionally, big data analysis is performed based on the cloud threat intelligence to report security alerts and backtrack events.

Figure 12-3: Situation Awareness module shows the overall concept of the Situation Awareness module.

Figure 12-3: Situation Awareness module



- Logs collected by the Situation Awareness module are classified into the following types:
  - Service data, including the HTTP traffic, 5-tuple data, and host syslogs.
  - Alibaba Cloud product data, including the logon logs and operation command logs of ECS instances, statements executed by RDS instances, and operation logs of MaxCompute instances.
- Threat intelligence of the Situation Awareness module comes from the following sources:
  - Intelligence obtained through data analysis, such as malicious IP address library that are used to initiate pertinent attacks, attacking measures, and hacker information.
  - Data collected from the external entities, such as the known sample information like the IP address credit database, zero-day vulnerability database, virus database, webshell database, and weak password database.

Benefits

Situation Awareness has the following features and advantages:

- **Fast vulnerability scan speed and full coverage in vulnerability detection**

  The vulnerability analysis function adopts the stateless scan technology, and can concurrently scan 10,000 IP addresses per second when the bandwidth is 5 MB. With its patented third-party vulnerability scan technology, the Situation Awareness module can rapidly scan third-party CMSs by means of fingerprint recognition.

The module works with the network traffic monitoring module to monitor URL requests on the network in real time and scan all target URLs and interfaces. It also supports scanning by binding to the proxy, HTTPS, and DNS.

The vulnerability analysis function supports the following aspects:

- The function supports scanning of over 30 common web vulnerabilities and over 150 exclusive web application vulnerabilities, including OWASP, WASC, and CNVD vulnerabil ities.
- The function supports scanning weak passwords in common systems and databases.
- The function supports detecting malicious tampering of script languages (such as Web2.0 and AJAX) and environments (such as PHP, ASP, .NET, and Java).
- The function supports detecting malicious tampering of complex character encoding.
- The function supports detecting malicious tampering of compression methods such as Chunk, Gzip, and Deflate.
- The function supports detecting malicious tampering of a wide range of authentication methods, such as Basic, NTLM, Cookie, and SSL.

Backed by Alibaba Cloud's big data computing capabilities, the Situation Awareness module can perform data mining and analysis on massive attacks detected on the Alibaba Cloud platform. By doing so, Apsara Stack Security obtains samples and generates intelligence about the latest attack behaviors in real time, detects zero-day vulnerabilities, generates a security vulnerability database, and applies the database to the vulnerability analysis system to further ensure that the vulnerability analysis module can provide rapid and comprehensive analysis results.

- **Big data threat analysis**

  Situation Awareness not only provides big data analysis and computing capabilities suitable for petabytes of data, but also uses machine learning to collect security data and threat intelligen ce from across the Internet. This allows Situation Awareness to establish comprehensive and intelligent security threat models for use in the actual business scenarios involving millions of users.

  Specifically, Situation Awareness uses the big data analysis method along with intelligent machine learning and modeling analysis to focus on new threats and security trends faced by cloud computing users of the data center, including attacks on web applications, brute-force password cracking, hacker intrusions, and vulnerabilities at the application and host layers.

- **Dashboard display**

  The Situation Awareness module uses a graphical dashboard to display results of big data threat analysis as graphs to help security decision making of the cloud computing platform.

12.3.2.6 Bastion Host

The bastion host module provides complete audit playback and permission control services for O&M of ECS instances. Based on the AAAA solution that centrally manages accounts, authentication, authorization, and audit, the bastion host module improves security of O&M management through features such as identity management, authorization management, two-factor authentication, monitoring and disconnection of real-time sessions, audit video playback, and risky command query.

The bastion host module meets various laws and regulation requirements, including classified protection, China Banking Regulatory Commission (CBRC), China Securities Regulatory Commission (CSRC), PCI Security Standards Council, and enterprise internal control management requirements.

Features

The bastion host module provides the following functions:

Function	Function description
Logon authentication	Supports local authentication and two-factor authentication, such as the mobile app dynamic password and SMS password.
Credential hosting and single-point logon	Supports hosting of ECS accounts and passwords (or SSH key). To log on to an ECS instance, O&M personnel only need to log on to the bastion host system, and the account and password of this ECS instance are not required.
System O&M	Supports calling local tools on the web side to realize single-point logon. You can perform O&M by logging on to the bastion host system from a local client and then selecting a server.
O&M monitoring and blocking	Monitors operations of O&M personnel in real time and supports blocking of unauthorized operations by interrupting operation sessions.
Log playback and post-event backtracking	• Provides the log playback function that allows you to specify the position for playback using keywords. • Provides the command recording function and supports searching by key command. • Provides the image recording function and supports searching by keyword.

Function	Function description
	• Provides the file audit function and records detailed information such as names of uploaded or downloaded files.

Use cases

When high-profile customers such as governments, financial institutes, and enterprises migrate their services to Apsara Stack, O&M security of the cloud resources of users becomes a top priority. In traditional environments, all servers and applications are hosted over the cloud, and the following two traditional O&M methods are normally provided:

- An ECS instance is configured as the springboard, and only the O&M port (such as TCP port 22 and TCP port 3389) of the springboard ECS instance is opened. O&M personnel then log on to the springboard ECS instance first, and then access other ECS instances for O&M operations through the springboard ECS instance.
- For advanced O&M personnel, they can build their own VPN, and log on to the VPN to interwork with or perform O&M on ECS instances within the cloud platform.

The preceding methods solve the majority security problems of O&M to some extent, however, the following security risks remain:

- When the SSH (TCP port 22) and RDP (TCP port 3389) remote logon modes of ECS instances are opened to the public networks, vulnerabilities of RDP and SSH protocols may be maliciously used, and accounts and passwords may be cracked through brute-force attacks. Once the accounts and passwords are cracked, enterprise confidential data and user data may be leaked, bringing huge losses to enterprises.
- If the O&M, development, and third-party O&M personnel share high-privilege accounts (such as root and administrator), and unauthorized operations cause leak of sensitive data, it is difficult to identify persons who are responsible when security events occur.
- The development and O&M processes are not transparent. If abnormal operations are detected , it is difficult to analyze operations afterwards.
- Users of the financial cloud and e-government cloud must comply with related laws and regulations, for example, those stipulated by CBRC, Ministry of Public Security, and classified protection. The preceding O&M methods cannot meet these regulatory requirements.

Deploying the bastion host system for centralized O&M management can completely eliminate the preceding security risks and enhance security of O&M management.

The bastion host system is applicable to the following scenarios:

- **Audit compliance requirements are strict.**

  Rigid security supervision requirements are inherent to the financial and insurance industries , and a sound audit mechanism must be built. The bastion host system can be used to implement:

  - Isolation of department permissions: Implements effective management and audit of departments based on isolated department permissions.
  - Unified O&M portal: Provides a unified O&M portal for O&M personnel for centralized logon.
  - Compliance with audit requirements: Builds a robust cloud-based O&M audit mechanism that meets the industrial regulatory requirements.
- **O&M management is efficient and stable.**

  Enterprises in the Internet industry develop rapidly, and the number of employees and systems increases. Therefore, these enterprises need an efficient and stable operation audit system. The bastion host system can be used to implement:

  - Highly concurrent sessions: Supports concurrent sessions involving thousands of users.
  - Stable operation: Delivers a highly stable SLA-based assurance.
  - O&M fault backtracking: Establishes O&M redlines by backtracing operations when the O& M personnel have misoperations.

Benefits

The bastion host module has the following features and advantages:

- **Multi-protocol support**: Supports various protocols, such as SSH, Telnet, RDP, VNC, SFTP, FTP, and Rlogin.
- **Authorization rule and policy control**
  - Supports approval, interception, and blocking of commands.
  - Supports file transmission control.
  - Supports access control by factors, such as the source IP address and time.
- **Audit compliance**: Meets audit requirements put forward by financial industry regulatory units and information security classified protection regulations.
- **Stable, reliable, secure and creditable**
  - Based on the robustness of the Apsara Stack, the annual availability of a bastion host instance is as high as 99.95%.

- Two-factor authentication can be implemented using a mobile phone number, static password, and SMS dynamic password (or mobile app password) to ensure security and reliability of each visitor's identity.

12.4 Benefits

Apsara Stack Security operates from the core to the edge of Apsara Stack. It uses the network traffic monitoring module to identify malicious attacks at the network layer and block them in real time. At the host layer, Apsara Stack Security detects and removes Trojans and malicious files in real time, ensuring that your ECS instances are not taken over by hackers. In addition, Apsara Stack Security intercepts brute-force password cracking and warns of remote logons, so that attackers cannot take advantage of weak passwords to log on to your systems and steal or damage business data.

Apsara Stack Security is composed of multiple function modules that provide in-depth defense and multi-point linkage in network outlets, networks, and servers of Apsara Stack. You can use its unified management view to easily manage your cloud platform in a centralized manner. With this control system, you can manage the security policies for all of the security protection modules, and perform association analysis on the relevant logs.

In-depth defense security system architecture

Apsara Stack Security is composed of multilevel security protection modules, covering network security, host security, application security, vulnerability analysis, and other security aspects. This forms an in-depth defense system on cloud borders, within cloud networks, and on ECS instances . Using the centralized management center to coordinate scheduling and integrating the security information from the various modules, this service can make extremely accurate judgments, and detect and block malicious attacks at the most suitable points. Apsara Stack Security effectively protects cloud environments against external intrusion and safeguards users' business systems.

Security solutions fully integrated with the cloud platform

Apsara Stack Security is based on a decade's worth of experience in providing security services to Alibaba Group's internal businesses, and six years of safeguarding the Alibaba Cloud division . Applying knowledge gained from a considerable amount of security research achievements, security data analyses, and security operations and management approaches, Alibaba Cloud built a professional cloud security team of experts and combined the rich experience of these experts to develop an attack protection product specifically designed for cloud computing platforms.

This product, Apsara Stack Security, can effectively protect the security of users' cloud network environments and business systems on public clouds and Apsara Stack.

With all software components virtualized, Apsara Stack Security is compatible with a broad range of hardware and can be quickly deployed, scaled up, and commissioned, enabling the highly elastic feature of cloud computing to be realized. The protection modules situated on cloud borders and in cloud networks use a bypass architecture designed for cloud businesses to minimize the impact on the cloud platform services. The virtualized protection modules deployed on ECS instances suit the flexibility of these virtual machines.

Tenant-based security self-service awareness

Because cloud platforms are tenant-facing, Apsara Stack Security provides a self-service console portal for tenants for them to view their security protection situation and generate simple reports . By applying appropriate external resource configurations, tenants can receive automated SMS and email alerts to gain a better insight into their cloud platform.

Continuation of Alibaba Cloud's security capabilities

Alibaba Cloud analyzes over 10 TB of security data each day, and the results of this analysis are used to strengthen fundamental Apsara Stack Security capabilities, such as the malicious IP library, malicious behavior library, malicious sample library, and security vulnerability library. Then , these capabilities are rapidly deployed in the various Apsara Stack Security protection modules to enhance the protection of this service and provide you with greater security.

13 Key Management Service (KMS)

13.1 Product overview

Key Management Service (KMS) is a secure and easy-to-use management service provided by Alibaba Cloud. The confidentiality, integrity, and availability of keys are guaranteed at a low cost.

With the help of KMS, you can use keys securely and conveniently, and focus on developing encryption/decryption scenarios.

13.2 Product architecture

KMS is deployed in different regions. Each region provides the same functions, but the data is mutually independent. In a single region, KMS adopts a distributed architecture composed of multiple equivalent nodes. All the nodes in a single region provide the same level of availability, allowing you to resize the service based on your actual access needs.

The KMS architecture is shown in *# 13-1: Product architecture*.

图 13-1: Product architecture

KMS is divided into four modules:

- Storage

  This module is for storing exported key tokens (EKTs) and other metadata.
- AAA

  This modole is for authentication, authorization, and auditing.
- KMSHOST

  This module is for processing user API requests.
- HSA (Hardware Security Appliance)

  The Hardware Security Module (HSM) is for processing the cryptographic logic of KMS powered by RAFT protocol distributed storage and trusted computing technology.

13.3 Functions and features

13.3.1 Convenient key management

You can use the APIs provided by KMS or the KMS console to conveniently manage your keys.

- You can disable and enable user keys at any time. After a key is disabled, the data encrypted using this key cannot be decrypted.
- A pre-deletion policy is used to delete keys. You can cancel key pre-deletion at any time, reducing the potential impact of accidental operations.
- You can use Resource Access Management (RAM) to manage key permissions and separate key encryption and decryption permissions.
- You can use EncryptionContext to enhance control over keys and ciphertext data.

13.3.2 Envelope encryption technology

Although KMS provides the Encrypt API, it does not actually encrypt data. KMS provides a Customer Master Key (CMK) management service and data key encryption/decryption service. You have to use the data keys to encrypt data yourself.

You can encrypt data using your own data key and then use the Encrypt interface to protect your data key. Or, you can obtain a data key from the KMS GenerateDataKey API.

Encryption process

For the envelop encryption process, see *# 13-2: Encryption flowchart*.

图 13-2: Encryption flowchart

As you can see in *# 13-2: Encryption flowchart*, the encryption process is as follows:

1. Use the specified CMK to generate a data key and obtain the key and encrypted key.

   Or, you can generate your own data key and use the Encrypt interface to obtain the corresponding encrypted key.
2. Use the data key to encrypt the data and obtain the ciphertext data.
3. Store the ciphertext data together with the encrypted key.

Decryption process

For the envelop decryption process, see *# 13-3: Decryption flowchart*.

图 13-3: Decryption flowchart

As you can see in *# 13-3: Decryption flowchart*, the decryption process is as follows:

1. Use KMS to decrypt the encrypted key.
2. Obtain the plaintext key.
3. Use the plaintext key to decrypt the ciphertext data and obtain the plaintext data.

13.3.3 Secure key storage

KMS guarantees the security of keys during storage in the following ways:

- Customer Master Key (CMK) plaintext only appears in the memory of the HSA module. The KMS storage module only stores the CMK ciphertext.
- CMKs are encrypted using the HSA module's domain key. This domain key is rotated once per day.
- The domain key is encrypted for storage using trusted computing technology and stored based on a distributed storage protocol. This guarantees the high reliability of the domain key.

14 StreamCompute

14.1 Product History

Alibaba Cloud StreamCompute was developed to meet the needs of Alibaba Group's global e-commerce shopping festival, known colloquially as 11.11 or Double Eleven. We expect that Alibaba Cloud StreamCompute can help more enterprises to realtime their own big data business.

In previous years, the use of open-source software Apache Storm was implemented in order to achieve real-time streaming data services of Double Eleven data. The real-time business in this period is in its embryonic stage, and its scale is small. Data developers use native API of Storm to develop streaming jobs. The development threshold is high, and system debugging is difficult. There is also a lot of repetitive manual work. Alibaba Cloud engineers began to consider business encapsulation and abstraction in response to such repetitive work.

First of all, we hope to have an integrated solution of streaming data computing and batch processing. Both Spark and Flink have stream and batch processing capabilities, but their approach is opposite. Spark Streaming transformates streams into small batch. One problem of this scheme is that the lower the delay we need, the greater the proportion of the extra overhead , which causes the Spark Streaming to be difficult to delay for second or even sub second. Flink regards batch as finite stream, which can retain a series of optimizations specific to batch processing while the stream and batch share most of the code. For this reason, if you want to use a set of engines to solve the stream and batch processing, it must be based on stream processing , so we decide to choose an excellent stream processing engine first. Stream processing can be divided into two types according to functions: stateless and stateful. The introduction of state management in the stream processing framework greatly improves the system's expressive ability , allowing you to easily implement complex processing logic, which is a leap in stream processing.

The development of any technology must follow the growth trajectory of the innovation to the popular, and the turning point must lie in the functional maturity and the cost reduction of the technology. Alibaba Cloud engineers began to think about how to reduce the threshold of data analysis. Flink has a lot of innovation in the architecture, which is in a leading position, but there are some shortages in the implementation of the project. Different jobs may run in the same process, which greatly reduces the performance of the system. In order to solve this problem, Alibaba Cloud engineer reimplements the combination of Yarn. Flink is used to ensure consistenc y through the mechanism of checkpoint, but the original mechanism efficiency is low, which leads to the unavailability of the large state (the state of the incremental calculation). Blink greatly optimizes the checkpoint and can efficiently handle a large state. Stability and scalability are essential in production. Through the refining on large clusters, Blink has solved a series of problems and bottlenecks in this area, and has become a computing engine that can support the core business. At the same time, we extend the SQL layer of Flink to Blink so that it can support more complex business in a more complete way, and it has become the core computing engine of Alibaba Cloud StreamCompute.

14.2 Features

Blink, which is the underlying engine of Alibaba Cloud StreamCompute, is developed on the basis of Flink, inheriting all the advantages of Flink and whose Table API is improved to make it more complete, so you can use the same SQL for batch processing and stream processing. The more powerful YARN mode is still 100% compatible with Flink's API and wider ecosystem. The advantages of Alibaba Cloud StreamCompute are as follows:

- **Powerful features**

  Different from other ope-source stream computing middleware, which only provides a rough computing framework, and a lot of stream computing details need to be re-built by engineers . Alibaba Cloud StreamCompute integrates many full link functions to facilitate your full link stream computing development.

  - **A powerful stream computing engine**
    - The StreamCompute engine supports standard StreamSQL for stream computing and automatic recovery from a variety of failures, ensuring data processing accuracy in case of a fault.
    - It provides multiple built-in string processing, time, and statistics functions.
    - Its precise control of computing resources ensures that the jobs of different tenants are isolated.
  - The number of key performance indicators are 3 to 4 times that of the open-source Flink, and the data computation delay is optimized to second level or sub second level, and the single job throughput can be in the million (record per second) level, and the single cluster size can be thousand level.
  - StreamCompute can integrate with data storage systems such as MaxCompute, DataHub, Log Service, ApsaraDB for RDS, Table Store, and AnalyticDB. You do not need to perform any additional data integration work, and StreamCompute directly reads and writes data to these products.

- **Managed real-time computing services**

  Unlike open source or self-built stream processing services, StreamCompute is a fully managed stream computing engine that can run queries for streaming data without presetting or managing any infrastructure. With StreamCompute, you can enable streaming data service capability in one-click. StreamCompute seamlessly integrates services such as data development, data O&M, and monitoring alarms, which facilitate your management and transference of stream computing at low cost.

  StreamCompute supports fully isolated management services with tenant, and provides the most effective isolation and comprehensive protection from the top work space environment to the bottom running machines, allowing users to use StreamCompute securely.

- **Good streaming development experience**

  Supports standard SQL (namely, BlinkSQL) and provides various built-in functions such as string processing, time, and statistics, replacing inefficient and complex Flink development industry-wide. This enables users to complete real-time big data analysis and processing using simple BlinkSQL, greatly reducing the complexity and concerns of real-time big data processing .

  Provides assistance kits for different stages including data development, data O&M, and monitoring alarms for full-link StreamCompute, so that it only takes a minimal amount of steps to completely release stream computing tasks.

- **Low cost**

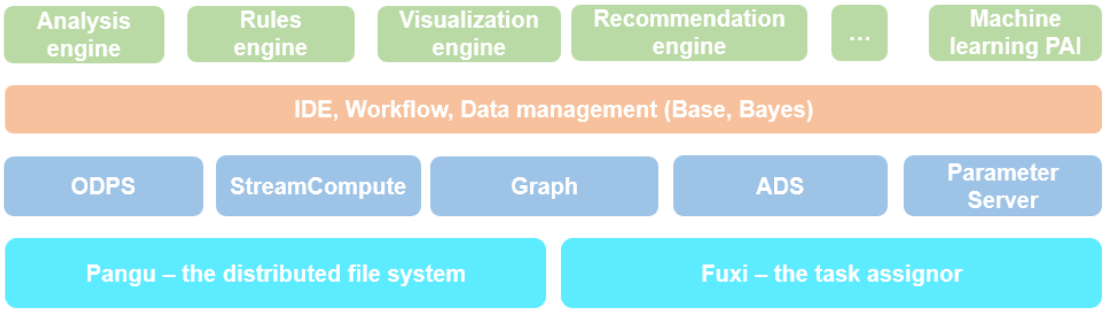
  With improvements to the SQL execution engine, StreamCompute is more competitive than native Flink in terms of both development costs and operating costs.

14.3 Strategic Position and Development Trajectory

StreamCompute runs on a cluster of thousands of nodes in the Alibaba Group. It serves hundreds of real-time applications for over 20 business units, processing hundreds of billions of messages per day and about one petabyte of traffic volumes. StreamCompute has become one of the core distributed computing services of Alibaba Group.

The location of StreamCompute in the Alibaba Cloud platform is shown in *Figure 14-1: StreamCompute strategic position* .

Figure 14-1: StreamCompute strategic position



Alibaba Cloud StreamCompute will be strengthened in the following aspects:

- Computing engine: We will focus on performance enhancement, multiple message processing semantic support, and other aspects.
- Programming interface: Provides richer APIs, supports multiple languages, be compatible with open-source system API, such as Storm API, Beam API and so on.
- Language: Enriches the SQL expression ability of Streaming scenes, and increases the support of syntax and semantics such as Temporal and CEP.
- Product: Improves StreamCompute's debuggability, one key deployment, hot upgrading, training system and so on continuously.

14.4 Architecture

14.4.1 Business Architecture

StreamCompute is defined as a lightweight streaming compute engine that uses SQL expressions

.

- **Data production**

  The source where data is produced. Data production generally takes place in server logs, database logs, sensors, and third-party systems. This streaming data enters the data integration module where it drives StreamCompute.
- **Data integration**

  This module integrates streaming data and acts as a hub for data publishing and subscription. The data can be collected from the DataHub service, the IoTHub service, and ECS's Log Service.
- **Data computing**

StreamCompute subscribes to the streaming data provided by data integration to drive streaming computation of data.

- **Data storage**

  StreamCompute does not provide any storage resources. Instead, it writes the results of streaming processing and compute to other storage resources, including relational databases, NoSQL databases, and OLAP systems.
- **Data consumption**

  Different data storage resources allow you to consume data in various ways. For example, the storage for message queues can be used for alarms, while that for relational databases can be used for online business support.

14.4.2 Technical Architecture

StreamCompute is a real-time incremental computing platform. It provides StreamSQL-like syntax and uses the MapReduceMerge (MRM) model for incremental computing. StreamComp ute provides an excellent failover mechanism to ensure data accuracy in the case of various exceptions.

StreamCompute includes the following components:

- **Data application layer**: This provides a development platform for you to develop new business and submit jobs. A comprehensive monitoring and alarm system is provided to inform the business end of any job delay. You can also use Blink UI or another system to view the operating status of online jobs and performance bottlenecks, allowing you to quickly and effectively optimize your jobs.
- **Data development**: This layer parses Blink SQL statements, generates logical and physical execution plans, and ultimately converts execution plans into executable directed acyclic graphs (DAGs). This layer generates various directed graphs modeled by the DAGs obtained at the SQL layer. It is used to process specific business logic. Generally, a model contains three parts:
  - Map: This performs data filter, distribution (group), Join (MapJoin), and other operations.
  - Reduce: This performs aggregation within a single batch (StreamCompute packs stream data into batches, each of which contains multiple data entries).
  - Merge: This merges the computing results from the batch with the previous results (state) to get a new state. After processing N number (the value N is configurable) of batches, the

checkpoint operation is performed to save the current state to the State system (such as HBase or Tair).

- **Blink Core**: It provides a variety of computing models, Table APIs and Blink SQL. The underlying layer supports DataStream API and DataSet API, and the lowest layer needs Blink Runtime which is responsible for resource scheduling to ensure jobs stable.
- **Distributed resource Scheduling**: The StreamCompute cluster is built on the Gallardo scheduling system, which ensures the effective operation and recovery of StreamCompute.
- **Physical layer**: It refers to the powerful cluster support provided by Alibaba Cloud.

15 E-MapReduce

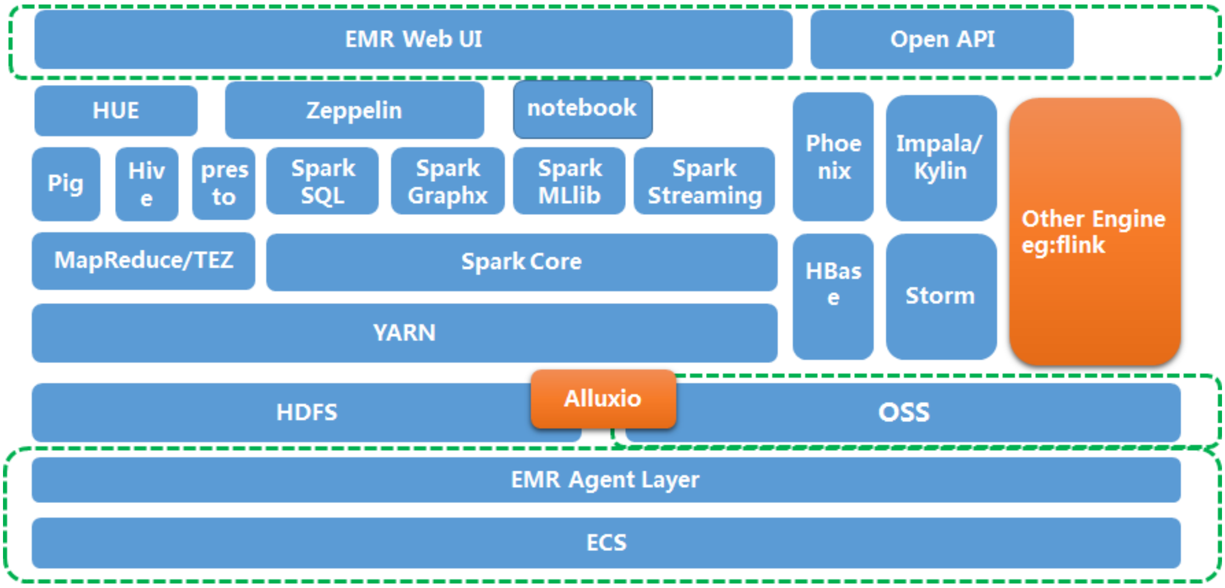
15.1 Overview

Elastic MapReduce (E-MapReduce) is a big data processing and analytic service. It provides cluster, job, and data management for users by leveraging open source big data ecosystem components such as Hadoop, Spark, Kafka, and Strom.

15.2 Architecture

Figure 15-1: EMR architecture shows the architecture of EMR.

Figure 15-1: EMR architecture



EMR clusters are created based on the Hadoop ecosystem. EMR clusters can exchange data seamlessly with Alibaba Cloud services, such as Object Storage Service (OSS) and Relational Database Service (RDS). This enables you to share and transmit data between multiple systems to meet different business demands. E-MapReduce provides a series of OpenAPIs to facilitate you to operate clusters, jobs and execution plans.

For more introductions to the components in the Hadoop ecosystem, see *Glossary*.

15.3 Features

15.3.1 Clusters

An EMR cluster is a Hadoop or Spark cluster that includes one or more ECS instances.

EMR provides an integration solution for cluster management tools, such as host model selection , cluster creation, cluster configuration, cluster operation, job configuration, job execution, cluster management, and performance monitoring. This frees you from cluster creation tasks, such as purchasing, preparing, and maintainance, allowing you to focus on the processing logic of your applications.

EMR also offers service customization, allowing you to select different cluster services to meet your business requirements. For example, to perform daily data statistics and batch computing, you only need to select the Hadoop service for EMR. If you also need streaming computing and real-time computing, then you need to add the Spark service.

15.3.2 Jobs

To run a computing task in EMR, you must create a job.

EMR supports multiple job types such as Spark, Hadoop, Hive, Pig, Sqoop, Spark SQL, and Shell. You can create a job to define the commands that you need to execute and the policy for handling execution failures.

15.3.3 Execution plans

An execution plan is a set of jobs. When you execute an execution plan, you can select an existing cluster for the plan or dynamically create a temporary cluster. You can choose to manually or periodically execute an execution plan by configuring scheduling policies. The biggest benefit of using execution plans is resource-saving. The system resources are used based on plan execution demands.

EMR supports the following execution plan scheduling policies:

- Periodical execution: Specifies the plan execution interval and initial execution time. A plan is executed at the specified intervals.
- Manual execution: Requires you to manually execute an execution plan by clicking.

和

15.3.4 Alarm management

EMR supports alarm managment, which enables you to associate execution plans with alarm contact groups. You can enable alarm feature on the Execution Plan Details page. When EMR finishes executing a execution plan, it sends a SMS to all contacts in alarm contact groups that are associated with the execution plan. The SMS includes the plan name, job execution results (quantities of plans succeeded and failed), correspondinsg execution cluster name, and specific execution time.

15.4 Benefits

Compared with manually creating clusters, EMR offers an easier way for you to comprehensively manage the EMR clusters. Additionally, EMR offers the following benefits:

- Deep integration

  EMR is deeply integrated with other Alibaba Cloud services such as OSS, MNS, RDS, and MaxCompute. This enables these services to act as the input source or output destination of the Hadoop or Spartk compute engine in EMR.
- Security management

  EMR is integrated with Resource Access Management (RAM), using primary and sub-accounts to control service access.

16 Quick BI

16.1 What is Quick BI

Quick BI is a flexible and lightweight self-service platform, and it provides BI tools based on cloud computing.

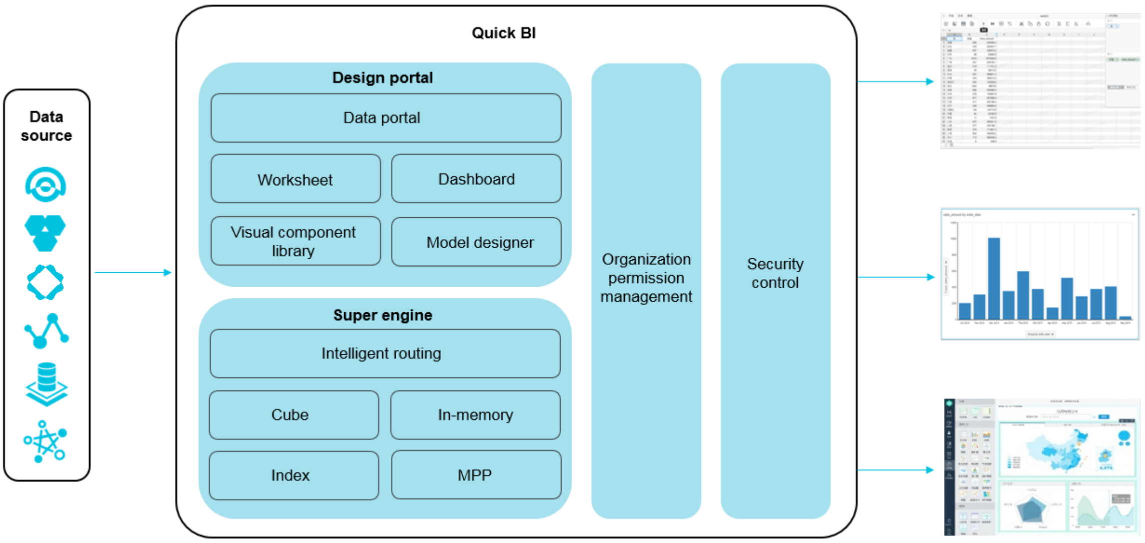
Quick BI can connect to multiple data sources, including cloud data sources such as MaxCompute (ODPS), HybirdDB for MySQL, AnalyticDB, and HybridDB (Greenplum), as well as your MySQL database on ECS, meanwhile, the data source from the VPC is also supported . Quick BI provides a real-time online analysis service tailored to massive data. With an intelligent data modeling tool, Quick BI reduces data acquisition cost by a large margin and makes it much easier to use . Besides, the drag-drop operation and the rich set of visual chart controls allow you to easily complete data perspective analysis, self-service data acquisition, business data profiling, report making, and data portal building.

In addition to a data viewer for business personnel, Quick BI also turns everyone into a data analyst to achieve data-based operation of enterprises.

16.2 Architecture

The architecture of Quick BI is shown in the following figure:

Figure 16-1: Quick BI architecture



Modules and features of Quick BI

- **Data connection module**

  Compatible with various cloud data sources, including but not limited to MaxCompute, RDS (MySQL, PostgreSQL, SQL Server), AnalyticDB, HybridDB (MySQL, PostgreSQL), used to encapsulate standard query APIs of meta data and other data from data sources.
- **Data preprocessing module**

  Provides lightweight ETC processing of data sources and supports custom SQL of MaxCompute. More data sources will be supported in the future.
- **Data modeling**

  Responsible for OLAP modeling of data sources, transforming data sources to multi-dimensional analysis model, supporting standard semantics such as dimensions (including date and geographic position), measurement, and star-type topology model, as well as computing field, and allowing you to process dimensions and measurements again by using SQL syntax of current data source.
- **Worksheet/Workbook**

  Provides operations related to online electronic spreadsheet (webexcel), including data analysis (such as row and column filtering, common/advanced filtering, classified aggregation , AutoSum, conditional formatting), data export, text processing, sheet processing, and other operations.
- **Dashboard**

  Assembles visual chart controls into a dashboard in a drag-drop manner, and supports 17 charts (such as line chart, pie chart, bar chart, funnel chart, tree chart, bubble map, color map , and indicator board), four basic controls (query conditions, TAB, IFRAME, and text box), and inter-chart data linkage.
- **Data portal**

  Assembles dashboards into a data portal in a drag-drop manner, and supports embedded link (dashboard), external link (third-party URL), and settings of template and menu bar.
- **QUERY engine**

  Queries data sources.
- **Organization permission management**

  Manages permissions based on <organization - workspace> architecture and user roles under workspace to achieve permission control and enable different users to view different tables.

- **Row-level permission management**

  Controls row-level permission of data and enables users of different roles to view different data from one report.
- **Share/Publish**

  Shares worksheets, dashboards, and data portals to other logged-in users, and publishes dashboards on the Internet for non-logged-in users to access.

16.2.1 Deployment

Quick BI automation deployment is carried out through the Tianji.

16.2.2 Components and functions

Quick BI is divided into the following categories of service roles:

- base-biz-yunbi-dbinit: Perform metadata initialization
- quickbi-redis-slave: Redis caches slave
- quickbi-redis-master: Redis caches master
- base-biz-yunbi-executor: Quick BI agent services
- base-biz-yunbi: Web services of Quick BI home page
- ServiceTest: Automated test service

16.3 Features

Quick BI offers the following functions:

Seamless integration with cloud-based database

Supports various Alibaba Cloud data sources, including but not limited to MaxCompute, HybirdDB for MySQL (MySQL, PostgreSQL, SQL Server), AnalyticDB, and HybridDB (MySQL, PostgreSQL).

Chart

Diverse options for data visualization. The built-in 17 types of visual charts (such as bar chart, line chart, pie chart, radar chart, and scatter chart) can meet data presentation demands of different scenarios. Besides, it can automatically recognize data features and recommend an appropriate visualization solution.

Analysis

Multi-dimensional data analysis. The web page based environment supports Microsoft Excel-like drag-and-drop operations, data import with one click, and real-time analysis. This allows you to analyze data from different perspectives without having to build a new model.

Quick building of data portal

Drag-and-drop operations, powerful data modeling, and rich visual charts help you build a data portal in a short time.

Real-time

Supports online analysis of massive data without preprocessing, thus greatly improving the analysis efficiency.

Secure management of data permissions

Provides organizational member management, and supports row-level data permissions to enable users of different roles to view different reports as well as to view different data from a same report table.

16.4 Benefits

The benefits of Quick BI can be summarized as follows:

High compatibility

Supports multiple data sources such as HybirdDB for MySQL, MaxCompute, and AnalyticDB.

Fast response

Responds in seconds for hundreds of millions of data.

Powerful capabilities

The built-in complete spreadsheet tools allow you to easily create complex Chinese statements.

Ease of use

Rich data visualization, automatic identification of data features, and automatic intelligence function can help you to generate the most appropriate chart.