Alibaba Cloud
Apsara Stack Enterprise

User Guide

Version: 1808..

Issue: 20180831

Legal disclaimer

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

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

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

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

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 instructio ns, 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 What is the Apsara Stack console

This section describes the definition and operation process of the Apsara Stack console.

Overview

The Apsara Stack console is customized for government and enterprise customers based on the Apsara Stack platform. It intends to improve IT management, solve operation problems, and provide service capabilities of industrial cloud computing. It provides large-scale, cost-effective, and one-stop cloud computing and big data services for customers in many industries, such as governments, education, healthcare, finance, and enterprises.

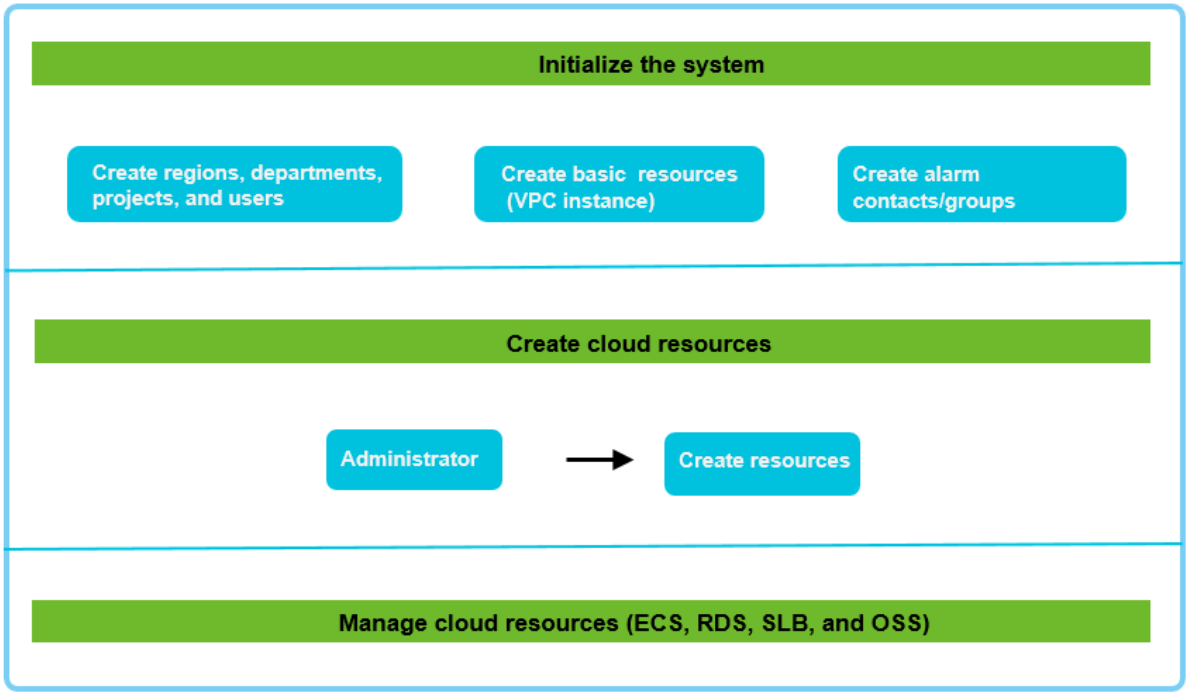
The Apsara Stack console builds the government and enterprise Apsara Stack platform that supports different business types, simplifies management and deployment of physical and virtual resources, and helps you easily and rapidly establish your own business system with higher resource utilizations and lower Operation and Maintenance (O&M) costs. It shifts your attention from operation and O&M to business, brings the Internet economic model to government and enterprise customers, and builds a brand new ecological chain based on cloud computing.

Overall operation process

The main operations available in the Apsara Stack console are as follows:

- Initialize the system: Complete the basic system configurations, such as creating regions, departments, projects, users, basic resources (Virtual Private Cloud (VPC) instances), cloud monitoring contacts, and contact groups.
- Create cloud resources: The administrator directly creates resources as required.
- Manage cloud resources: Manage resources, such as starting, using, and releasing resources, and changing resource configurations.

Figure 1-1: Operation process of the Apsara Stack console



2 Configuration requirements

Before accessing the Apsara Stack console, check if your browser and operating system meet the configuration requirements.

Your local computer must meet the requirements in *Table 2-1: Configuration requirements* to log on to the Apsara Stack console.

Table 2-1: Configuration requirements

Content	Requirement
Browser	• Internet Explorer: 11 or later versions • Chrome (recommended): 42.0.0 or later versions • Firefox: 30 or later versions • Safari: 9.0.2 or later versions
Operating system	• Windows XP, Windows 7, or later versions • Mac OS X

3 Log on to the Apsara Stack console

Take the Chrome browser as an example to describe how to log on to the Apsara Stack console as cloud product users.

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

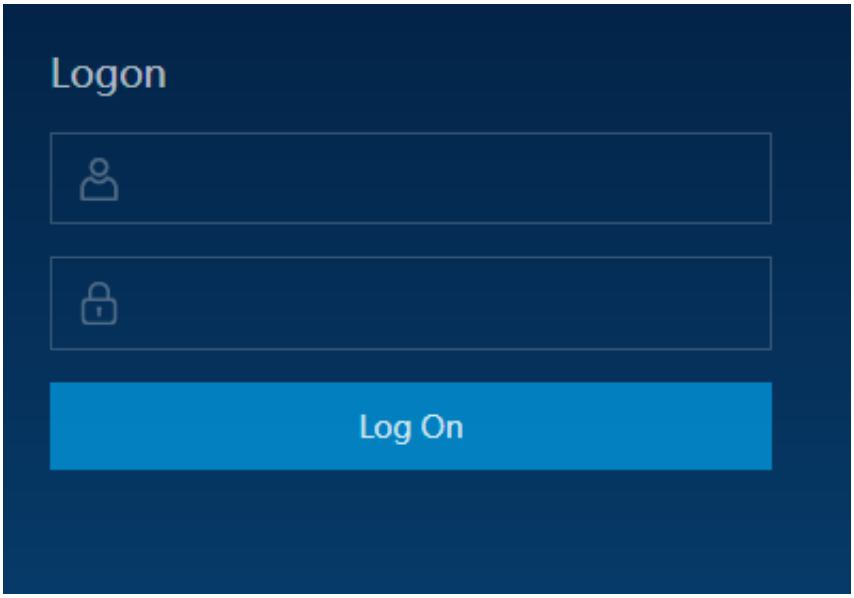
  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 3-1: Log on to the Apsara Stack console*.

   Figure 3-1: Log on to the Apsara Stack console

3. Enter the correct username and password.

- The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
- You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

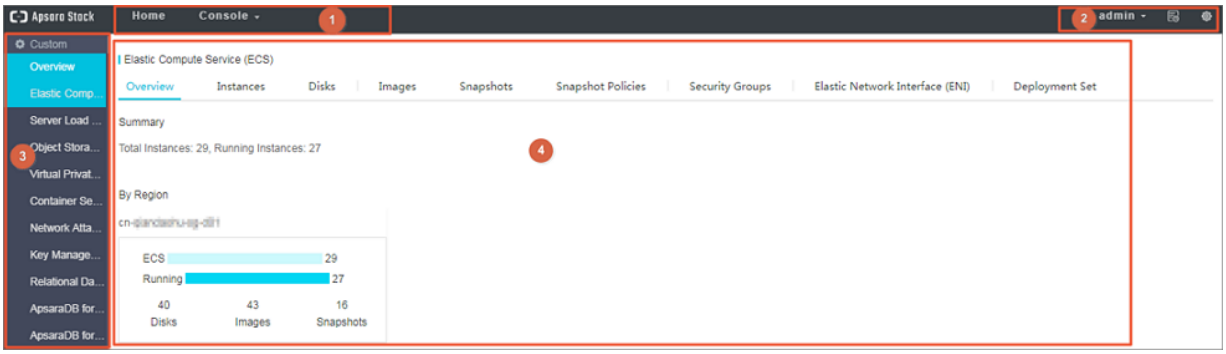
4. Click **Log On** to go to the **Home** page.

4 Familiarize yourself with the Web page

The Web page of the Apsara Stack console is consisted of four areas: main menu bar, information area of the current logon user, custom menu pane, and operation area.

The Web page of the Apsara Stack console is consisted of four areas, as shown in *Figure 4-1: Web page*.

Figure 4-1: Web page



For more information about the functional areas of the Web page, see *Table 4-1: Functional areas of the Web page*.

Table 4-1: Functional areas of the Web page

Area		Description
1	Main menu bar	• Home: Displays the resource overview and monitoring status in the Apsara Stack console. • Console: Manages the overall system and all resources. It contains the following modules: — Compute, Storage & Networking: Manages all types of basic cloud products and resources. — Database: Manages all types of database products and resources. — Big Data: Manages all types of big data products and resources. — Administration: Manages the CloudMonitor Center, System Reports, Operation Log, and Task Center of the system. — Operations Center: Manages resource allocation of the system. — User Center: Manages the departments, projects, roles, users, and logon policies of the system to make sure that the system runs properly.

Area		Description
		Note: The menu bar varies with different roles. See your menu bar for relevant functions.
2	Information area of the current logon user	: Click this to display the **Personal Information** page of the current logon user or log out of the console. On the **Personal Information** page, you can: • View your basic information. • Modify your information. • Change your portrait. • Change your password. • View the AccessKey. • View the AccessKey used for third-party access. : Click this to display the history and most frequently accessed menu items. : Click this to go to the **System Configuration** page.
3	Custom menu pane	Used to configure your common menu items. : Click this to hide or show the menu pane.
4	Operation area	Displays the function configuration interface of the selected menu item .

5 Initial system configurations

5.1 Relationships among departments, projects, users, and roles

The Apsara Stack console follows service principles to uniformly manage the users, roles, organizations, and projects related to cloud data centers, which allows you to grant different resource access permissions to users.

As the core module for centralized permission management, User Center integrates the functions of user management, role management, department management, and project management.

Roles, departments, and projects are described as follows:

- Role

  A collection of access permissions. When creating users, you must assign roles to users to meet their access control requirements on the system.
- Department

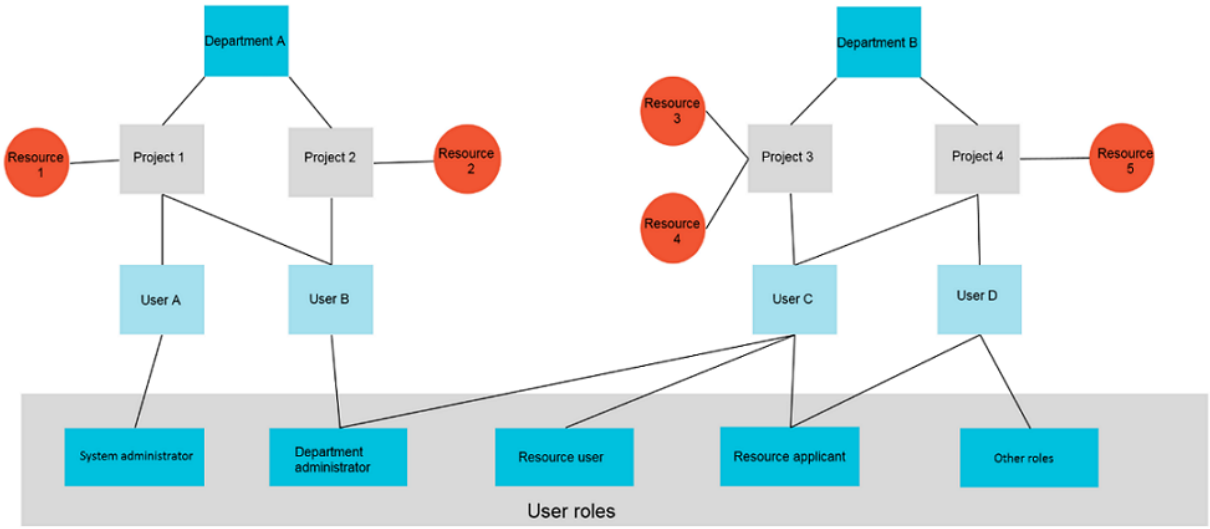
  After the Apsara Stack console is deployed, a root department is created by default. You can create departments under the root department.

  The departments are displayed hierarchically and you can create sub-departments under each department.
- Project

  A container where resources are stored. All resources must be created under a project.

For the relationships among departments, users, projects, roles, and cloud resources, see *Relationships among departments, users, projects, roles, and cloud resources*.

Figure 5-1: Relationships among departments, users, projects, roles, and cloud resources



- A department can have multiple projects, but each project can only belong to one department.
- Each project can have multiple cloud resources and users, but each cloud resource can only belong to one project.
- A user can have multiple projects, that is, a user can participate in multiple projects of the same department.
- Each user can have multiple roles, and each role can be assigned to multiple users.

The quantity relationships among departments, users, projects, roles, and cloud resources are as shown in *Relationship table*.

Table 5-1: Relationship table

Relationship	Relationship type	Description
Department vs. project	One to many	A department can have multiple projects, but each project can only belong to one department.
Department vs. user	One to many	A department can have multiple users, but each user can only belong to one department.
Project vs. user	Many to many	A user can have multiple projects, and a project can be assigned to multiple users.
User vs. role	One to many	A user can have multiple roles, and a role can be assigned to multiple users.

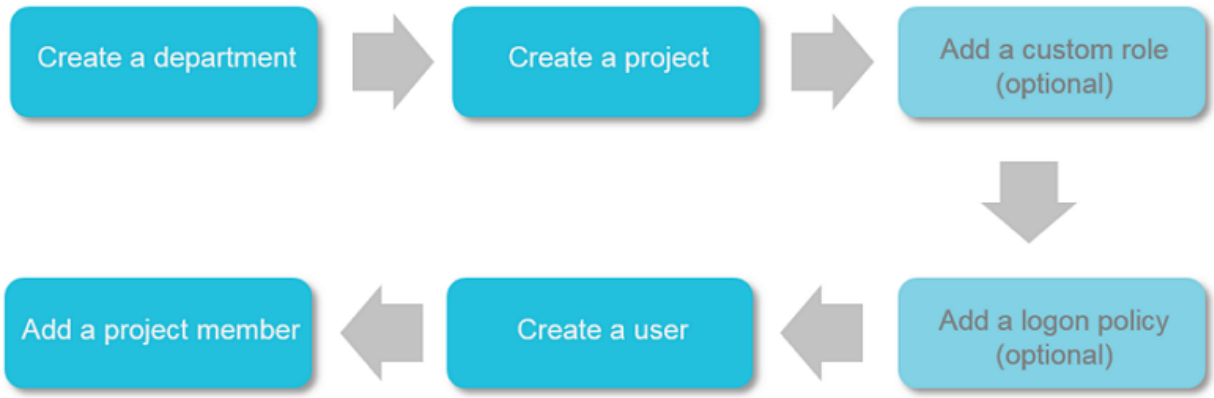
Relationship	Relationship type	Description
Project vs. resource	One to many	A project can have multiple resources, but each cloud resource can only belong to one project.

5.2 Configuration process

This section describes the initial configuration process of the system.

The administrator must complete the initial configuration of the system as shown in *Figure 5-2: Initial system configuration process* before using the Apsara Stack console.

Figure 5-2: Initial system configuration process



5.3 Create a department

Create a department to store projects and the resources within the projects.

Context

After the Apsara Stack console is deployed, a root department is created by default. You can create departments under the root department. The departments are displayed hierarchically and you can create sub-departments under each department.

Departments added under the root department are level-1 departments, departments added under the level-1 departments are level-2 departments, and so on. In the Apsara Stack console, the sub-departments of a department refer to departments of all levels under the department.

Departments reflect the tree structure of an enterprise or business unit. A user can only belong to one department.

You can create a department under an existing department. The created department is a sub-department of the existing department.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Department Management**.
3. Select a department and click **Add Department**.

   The **Add Department** dialog box appears.
4. Enter the department name.

   The name must be 2-20 characters long and can contain English letters, numbers, and Chinese characters.
5. Click **Confirm**.

5.4 Create a project

You must create a project before applying for resources.

Prerequisites

Make sure that you have created a department before creating a project. For more information, see *Create a department*.

Context

You can create at most 20 projects under each level-1 department.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Project Management**.

   The **Project Management** page appears.
3. Click **Add Projects**.
4. The **Add Projects** dialog box appears. Select a department and enter the project name.
5. Click **Confirm**.

5.5 Add a custom role

You can add custom roles in the Apsara Stack console to better assign permissions to users.

Context

The system has 13 roles by default. The super administrator initializes system information and creates system administrators. Both system administrators and department administrators are administrators, and the rest of default roles are users.

Before adding a custom role, note that:

- You must have the permissions to manage users and projects if you want to add or modify users.
- You must have the permissions to view Virtual Private Cloud (VPC) instances, users, and projects, and permissions to manage users and projects if you want to create VPC-related resources.
- You must have the management permissions in CloudMonitor Center if you want to create alarm items.
- The total number of custom and default roles cannot exceed 20.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Role Management**.
3. Click **Add Role**. The **Add Roles** dialog box appears.

   Configure parameters for adding a role, as shown in *Parameter description*.

   Table 5-2: Parameter description

Parameter	Description
Role Name	The name of a role, which must be 1-15 characters long and can contain English letters, numbers, and Chinese characters.
Description	The description of a role, which must be 1-100 characters long and can contain English letters, numbers, Chinese characters, commas (,), semicolons (;), and underscores (_).
Permission Scope	• Department The permissions apply to all departments of the corresponding modules. • Department/Sub-department The permissions apply to the department to which the user belongs and its sub-departments. • Project

Parameter	Description
	The permissions apply to the projects that the user has joined.
Select Permissions	Specify the operation permissions to cloud products. Double-click the modules in the **Available Permissions** section to select the corresponding permissions, or click **Import** to select all permissions.

4. Click **Confirm**.

5.6 Create a logon policy

The administrator can configure logon polices to control the logon address and time for users.

Context

A default logon policy is automatically generated when the Apsara Stack console provides services. This policy does not have any limits on the logon time and address, and cannot be deleted.

With the logon policies configured, users can access the Apsara Stack console at the permitted time and from permitted IP addresses. This improves the security of the console.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Logon Policy Management**.
3. Click **Add Policy**.
4. In the displayed **Configure Authorization Policy** dialog box, enter the policy name, permitted logon time, and permitted logon address.

Table 5-3: Parameter description

Parameter	Description
Policy Name	The name must be 1-15 characters long and can contain English letters, numbers, and Chinese characters, but cannot be the same as any other existing policy name.
Logon/Logout Time	The permitted logon time is a time period. After being configured, you can only log on during the specified time period.
Client IP Addresses	The permitted logon address is an IP address segment. After being configured, you can only log on from the IP addresses within the specified IP address segment.

5. Click **Confirm**.

   You can edit or delete the existing logon policies.

   - Click the icon at the right of the logon policy and select **Edit** to modify the policy.
   - Click the icon at the right of the logon policy and select **Delete** to delete the policy.

   > **Note:**
   > You cannot delete the default logon policy.

6. Optional: Bind a user to a logon policy. For more information, see *Change the logon policy of a user*.

   > **Note:**
   > - After a user is bound to a logon policy, this user can only log on at the permitted time and from permitted IP addresses configured in the policy.
   > - If the user does not want to be limited by the bound logon policy, the user must submit an application to the administrator. After approving the application, the administrator binds the user to a logon policy that meets the user's requirements.

5.7 Create a user

The administrator can create users and assign roles to users to meet their access control requirements on the system.

Prerequisites

Before creating a user, make sure that:

- You have created a department. For more information, see *Create a department*.
- You have created a custom role if you want to customize the role. For more information, see *Add a custom role*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Click **Add**. The **Add User** dialog box appears.
5. Configure parameters for creating a user, as shown in *Table 5-4: Parameter description*.

Table 5-4: Parameter description

Parameter	Description
Username	The cloud platform account name of the user. The name must be 3-30 characters long, start with a letter or number, and can contain letters, numbers , hyphens (-), underscores (_), and at signs (@).
Display Name	The name must be 2-30 characters long and can contain letters, numbers, Chinese characters, hyphens (-), underscores (_), and at signs (@).
Department	Select a department for the user.
Role	Select a role for the user. At most five roles can be assigned to a user.
Logon Policy	Select a logon policy for the user. It restricts the time period and IP addresses for the user to log on. By default, newly created users are automatically bound to the default policy. **Note:** By default, the default policy does not restrict the time and IP addresses for users to log on. To restrict the logon time and IP addresses of a user, you can modify the user's logon policy or create a logon policy for the user. For more information, see *Create a logon policy*.
Cellphone Number	The mobile phone number of the user. It is used to notify the user of resource applications and usage by SMS. Make sure the entered mobile phone number is correct. Update the number in time if it is changed.
Landline	The landline number of the user. It must be 4-20 characters long, and can contain numbers (0-9) and hyphens (-).
Email	The email address of the user. It is used to notify the user of resource applications and usage by email. Make sure the entered email address is correct. Update the email address in time if it is changed.

For the relationships among departments, users, and roles, see *Relationships among departments, projects, users, and roles*.

6. Click **Confirm**.

5.8 Add a project member

Add a member to a project to allow the member to use the resources of the project.

Context

The members of a project have the permissions to use resources of the project.

Deleting resources from a project does not affect the members of the project. Similarly, deleting members from a project does not affect the resources of the project.

You can delete the project members that are no longer in use. A deleted project member cannot access the resources of the project.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Project Management**.
3. Click the  icon at the right of the project and select **View Details**.
4. Click the **Project Members List** tab.
5. Click **Add Members**.
6. In the displayed **Add Project Members** dialog box, select a department and a project member.
7. Click **Confirm**.

   The project member is added successfully. You can view information about this member on the **Project Member List** page.

   To remove one or more members from the project, complete the following steps:

   1. Select one or more members and click **Delete**.
   2. In the displayed **Delete** dialog box, select **Yes**.
   3. Click **Confirm**.

6 Initial resource configurations

6.1 Create cloud resource quotas

The administrator can create quotas for cloud resources.

Context

You can create quotas for a child department. If the parent department (except for a level-1 parent department) has a quota, the result that the quota of the parent department minus the quotas of other child departments is the maximum quota that can be configured for the child department. The result cannot be smaller than the amount of resources already created for the child department.

This section takes the Elastic Compute Service (ECS) quota as an example to describe how to create quotas. You can create quotas for other cloud resources in a similar way.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Operations Center** > **Quota Management**.
3. Select a department on the left.
4. Click the  icon in the ECS section.
5. Configure the total quotas in the displayed **Create ECS Quota** dialog box.

   For more information about the quota parameters, see *Quota parameter description*.
6. Click **Confirm**.

6.2 Create a cloud resource

The administrator can create a cloud product instance in the console of each cloud product based on the project requirements.

For more information about how to create a cloud product instance, see the user guide of each cloud product.

7 CloudMonitor Center

7.1 Overview of CloudMonitor Center

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

Currently, CloudMonitor Center can be used to monitor metrics of Elastic Compute Service (ECS), Server Load Balancer (SLB), Relational Database Service (RDS), and Object Storage Service (OSS). You can use the metrics to set alarm rules and notification polices to keep up with the running status and performance of product instances. Consider a scale-up if you receive an insufficient resource alarm.

CloudMonitor Center has the following functions:

- Automatic monitoring: The monitoring is automatically started based on your created ECS resources or auto scaling groups. You are not required to start it manually or install any plug -ins. You can view the monitoring data of specific instances on the monitoring page after applying for resources.
- Flexible alarm: You can configure alarms flexibly, such as setting alarms and thresholds for monitoring metrics, pausing alarms, and enabling alarms.
- Real-time notification: Set the alarm notification to receive notifications by SMS or email in real time. If the status of an alarm rule changes, such as alarms are triggered, data is insufficient, or alarms are cleared, the system informs you by SMS or email.

7.2 Description of cloud monitoring metrics

CloudMonitor Center tests the service availability based on the monitoring metrics of cloud resources. You can configure alarm rules and notification polices based on the monitoring metrics to keep up with the running status and performance of product instances.

CloudMonitor Center can monitor resources of ECS, SLB, RDS, and OSS. Monitoring metrics supported by each service are described as follows.

Table 7-1: ECS monitoring metrics

Metric	Description	Measured object	Calculation formula	Remarks
CPU Utilization	Used to measure the CPU utilization (%) of a measured object.	ECS instance	CPU utilization of the ECS instance/ Total CPU cores of the ECS instance	None
Memory Utilization	Used to measure the memory utilization (%) of a measured object.	ECS instance	Memory utilization of the ECS instance /Total memory of the ECS instance	The memory utilization calculated by CloudMonitor Center does not include cache utilization. Therefore, when you run the **`free`** or **`top`** command to query the memory utilization of a Linux server, the result may be inconsistent with the memory utilization displayed in the Apsara Stack console.
Disk I/O Read	Used to measure the volume of data read from a measured object per second (KB/s).	ECS instance	Total bytes read from the ECS instance disk /Statistical cycle	For Linux hosts, the disk I/O monitoring data is obtained by using the iostat tool. If your Linux host has no disk I/O data, check if iostat is installed on your host . If not, Redhat or CentOS users can use yum to install the tool, and Ubuntu or Debian users can use apt-get to install the tool.
Disk I/O Write	Used to measure the volume of data written to a measured object per second (KB/s).	ECS instance	Total bytes written to the ECS instance disk /Statistical cycle	None
Disk Utilization	Used to measure the disk utilization (%) of a measured object.	ECS instance	Used capacity of the ECS instance disk/Total capacity of	None

Metric	Description	Measured object	Calculation formula	Remarks
			the ECS instance disk	
Inbound Traffic	Used to measure the inbound network traffic of a measured object per second (Kbit/s).	ECS instance	None	None
Outbound Traffic	Used to measure the outbound network traffic of a measured object per second (Kbit/s).	ECS instance	None	When the bandwidth you bought is used up, access fails or requests slow down. On the monitoring chart, eth0 indicates the intranet NIC of the server, and eth1 indicates the Internet NIC of the server.
TCP Connections	Total number of TCP connections set up by the server.	ECS instance	None	None
Processes	After you set an alarm rule with this monitoring item, the specified running processes are counted and the total number of these processes is displayed.	ECS instance	None	To monitor the running status of processes on the server, set an alarm rule with this monitoring item to trigger the alarm when the number of processes is unequal to the actual number of processes.
Average Load	A concept in Linux, the average load value of the server.	ECS instance		The value cannot be greater than 1. If your server has a multi-core processor, the value must be divided by the number of CPU cores. If the value is greater than 1

Metric	Description	Measured object	Calculation formula	Remarks
				, processes are queued up and the server slows down.

Note:

For ECS instances, monitoring plug-ins must be installed to collect the metric data at the operating system level.

Installation method: Click **Install** next to an ECS instance. Alternatively, select ECS instances and click **Install Plugins**.

The monitoring chart displays monitoring data 5-10 minutes after the monitoring plug-ins are installed.

Table 7-2: RDS monitoring metrics

Metric	Description	Measured object	Calculation formula
CPU Utilization	Used to measure the CPU utilization (%) of a measured object.	RDS instance	CPU utilization of the RDS instance/Total CPU cores of the RDS instance
Memory Utilization	Used to measure the memory utilization (%) of a measured object.	RDS instance	Memory utilization of the RDS instance/Total memory of the RDS instance
Disk Utilization	Used to measure the disk utilization (%) of a measured object.	RDS instance	None
IOPS Utilization	Used to measure the number of I/O requests of a measured object per second.	RDS instance	Number of I/O requests of the RDS instance/Statistical cycle
Connection Utilization	Used to measure the number of applications that can be connected to the measured object per second.	RDS instance	Number of applications that can be connected to the RDS instance per second/Statistical cycle

Table 7-3: SLB monitoring metrics

Metric	Description	Measured object	Remarks
Port Outbound Packets per Second	Number of packets sent by SLB per second.	SLB instance	None
Port Inbound Packets per Second	Number of packets received by SLB per second.	SLB instance	None
Port Inbound Data per Second	Traffic consumed to access SLB from the external.	SLB instance	None
Port Outbound Data per Second	Traffic consumed by SLB to access the external.	SLB instance	None
Active Port Connections	Number of all connections in the ESTABLISHED status.	SLB instance	It can be interpreted as, but cannot be equivalent to the concurrent connections. This is because a persistent connection transmits multiple file requests concurrently.
Inactive Port Connections	Number of all TCP connections except connections in the ESTABLISHED status.	SLB instance	None
New Port Connections	Number of TCP connections in SYN_SENT status in the three-way handshake in a statistical cycle.	SLB instance	Active Port Connections, Inactive Port Connections, and New Port Connections are all used to measure the number of requests for connecting a client to an SLB instance.

Table 7-4: OSS monitoring metrics

Metric	Description	Measured object
Reads	Used to measure the number of times that a measured object is read.	OSS instance
Internal Server Errors	Used to measure the number of errors of a measured object.	OSS instance
Public Network Inbound Traffic	Used to measure the inbound Internet network traffic (bytes) of a measured object per second.	OSS instance
Public Network Outbound Traffic	Used to measure the outbound Internet network traffic (bytes) of a measured object per second.	OSS instance
Classic Network Inbound Traffic	Used to measure the inbound intranet network traffic (bytes) of a measured object per second.	OSS instance
Classic Network Outbound Traffic	Used to measure the outbound intranet network traffic (bytes) of a measured object per second.	OSS instance
Writes	Used to measure the number of times that a measured object is written.	OSS instance
Storage Space Used	Used to measure the used storage space (bytes) of a measured object.	OSS instance

7.3 Manage alarm contacts

7.3.1 Create an alarm contact

You can create an alarm contact to receive alarms.

Context

An alarm contact is a person who receives alarms. Alarms can be sent by SMS or email. When monitoring data meets the conditions specified in alarm rules, the system sends alarm notifications to the relevant alarm contacts.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm Contact**.
3. Click the **Alarm Contact** tab.
4. Click **Create Contact**.

   The **Set Alarm Contact** dialog box appears.
5. Configure parameters for creating an alarm contact, as shown in *Parameter description*.

   Table 7-5: Parameter description

Parameter	Description
Username	The username of the alarm contact.
Cellphone Number	The mobile phone number of the alarm contact. It is used to send alarm notifications to the alarm contact by SMS. Make sure the entered mobile phone number is correct. Update the number in time if it is changed.
Email	The email address of the alarm contact. It is used to send alarm notifications to the alarm contact by email. Make sure the entered email address is correct. Update the email address in time if it is changed.
DingTalk ID	The DingTalk ID of the alarm contact.

6. Click **Confirm**.

7.3.2 Add an alarm contact to alarm groups

You can add a created alarm contact to alarm groups for better management.

Prerequisites

- You have created an alarm contact. For more information, see *Create an alarm contact*.
- You have created an alarm group. For more information, see *Create an alarm group*.

Context

An alarm contact can be added to multiple alarm groups.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm Contact**.
3. Click the **Alarm Contact** tab.

4. Select one or more alarm contacts and click **Add to Alarm Group**.
5. In the displayed **Modify Alarm Group** dialog box, select alarm groups and click **Confirm**.

7.3.3 Query alarm contacts

You can query the information and alarm groups of alarm contacts on the **Alarm Contact** page.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm Contact**.
3. Click the **Alarm Contact** tab.
4. Select **Name**, **Cellphone Number**, **Email**, or **DingTalk ID** as the query condition. Enter the keyword in the search box and then click **Search** to query the information and alarm groups of alarm contacts.

7.3.4 Modify alarm contact information

If the information of an alarm contact changes, you can modify the information on the **Alarm Contact** page.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm Contact**.
3. Click the **Alarm Contact** tab.
4. Click the  icon at the right of the alarm contact and select **Edit**.
5. In the displayed **Set Alarm Contact** dialog box, modify the contact information of the alarm contact, including the cellphone number, email, and DingTalk ID.

7.3.5 Delete an alarm contact

You can delete an alarm contact that is no longer in use based on the business requirements.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm Contact**.
3. Click the **Alarm Contact** tab.
4. Perform the following operations:
   - Delete an alarm contact.

Click the icon at the right of the alarm contact and select **Delete**.

- Delete multiple alarm contacts.

  Select multiple alarm contacts and click **Delete Alarm Contacts** in the upper-right corner.

5. In the displayed **Delete Contact** dialog box, click **Confirm**.

7.4 Manage alarm groups

7.4.1 Create an alarm group

An alarm group is a group of alarm contacts. It contains one or more alarm contacts.

Context

When setting an alarm rule, you must select an alarm group to receive the alarm notifications. For each monitoring item, alarm notifications are sent to the members in the alarm group according to the configured notification method if the alarm threshold is reached.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm Contact**.
3. Click the **Alarm Group** tab.
4. Click **Create Contact Group**.

   The **Alarm Group Management** dialog box appears.
5. Configure parameters for creating an alarm group, as shown in *Parameter description*.

   Table 7-6: Parameter description

Parameter	Description
Group Name	The name of the alarm group. It must be 2-20 characters long and can contain letters, numbers, Chinese characters, and underscores (_).
Remarks	The description of the alarm group. It must be 0-256 characters long and can contain letters, numbers, Chinese characters, hyphens (-), and underscores (_).
Choose Contacts	Add contacts to the alarm group as follows: Select contacts under **Current Contacts** and click **->** to add them to the **Selected Contacts**. To remove a selected contact, click **<-**.

Parameter	Description
	Note: If the contact is not created, create an empty alarm group first. Then, *create an alarm contact* and *add the alarm contact to the alarm group*.

6. Click **Confirm**.
7. Optional: To remove an alarm contact from the alarm group, go to the **Alarm Group** page, and click **Delete** at the right of the alarm contact.

7.4.2 Modify alarm notification methods

Phone notifications, email notifications, and DingTalk notifications are enabled by default. You can disable the unnecessary notification methods.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm Contact**.
3. Click the **Alarm Group** tab.
4. Enable or disable the phone notifications, email notifications, and DingTalk notifications by turning on or off the switches at the right of the alarm contact.

7.5 Manage alarm rules

7.5.1 Create an alarm rule

You can create an alarm rule for an instance to monitor this instance.

Context

We recommend that you create an alarm group before setting an alarm rule. You can also create an alarm group when setting an alarm rule. For more information about how to create an alarm group, see *Create an alarm group*.

Alarm rules configured in CloudMonitor Center are used to monitor the server performance. In this way, you can detect and handle server problems in time, which guarantees the secure, stable, and effective operation of servers.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Monitor**.

3. Select a cloud product.
4. Click the icon at the right of an instance and select **Alarm Rules** to go to the **Alarm Item** page.

> **Note:**
> You can also use the search function to locate a specific instance and create an alarm rule for the instance.

5. Click **Create Alarm Rule**.

   The **Set Alarm Rules** dialog box appears.
6. Configure parameters for creating an alarm rule, as shown in *Table 7-7: Parameter description*.

Table 7-7: Parameter description

Parameter	Description
Monitoring Item	Select a monitoring item from the drop-down list. For more information about monitoring items, see *Description of cloud monitoring metrics*.
Statistical Cycle	Select a statistical cycle from the drop-down list. The statistical cycle is the interval at which data statistics are generated.
Calculation Method	Select a calculation method from the drop-down list. The following calculation methods are available: • Average: An alarm is triggered when the average value of all monitoring data collected in a statistical cycle exceeds the threshold. • Maximum: An alarm is triggered when the maximum value of the monitoring data collected in a statistical cycle exceeds the threshold. • Minimum: An alarm is triggered when the minimum value of the monitoring data collected in a statistical cycle exceeds the threshold. • Original: An alarm is triggered when the original value of the monitoring data collected in a statistical cycle exceeds the threshold.

7. Click **Next**.
8. Configure parameters for the notification object, as shown in *Table 7-8: Parameter description*.

   A notification object is an alarm contact. For more information about how to configure an alarm contact, see *Create an alarm contact*.

Table 7-8: Parameter description

Parameter	Description
Alarm Retries	Select the number of retries before an alarm is triggered from the drop-down list. If the value exceeds the threshold for consecutive statistical cycles, an alarm is triggered. Alarm contacts are notified only after the threshold is exceeded.
Contact Notification Group	Select a contact notification group. After you set an alarm rule for a monitoring item, an alarm notification is sent to the alarm contacts if the monitoring data meets conditions configured in the alarm rule.
Notification Time	Select the notification time, which is a time range during which alarm notifications are sent.

9. Click **Confirm**.

7.5.2 Create multiple alarm rules

You can create the same alarm rule for multiple instances to monitor these instances.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Monitor**.
3. Select a cloud product and then select multiple instances.
4. Click **Create Alarm Rules** in the upper-right corner.

   The **Set Alarm Rules** dialog box appears.
5. Configure parameters for creating an alarm rule, as shown in *Table 7-9: Parameter description*.

Table 7-9: Parameter description

Parameter	Description
Monitoring Item	Select a monitoring item from the drop-down list. For more information about monitoring items, see *Description of cloud monitoring metrics*.
Statistical Cycle	Select a statistical cycle from the drop-down list. The statistical cycle is the interval at which data statistics are generated.

Parameter	Description
Calculation Method	Select a calculation method from the drop-down list. The following calculation methods are available: • Average: An alarm is triggered when the average value of all monitoring data collected in a statistical cycle exceeds the threshold. • Maximum: An alarm is triggered when the maximum value of the monitoring data collected in a statistical cycle exceeds the threshold. • Minimum: An alarm is triggered when the minimum value of the monitoring data collected in a statistical cycle exceeds the threshold. • Original: An alarm is triggered when the original value of the monitoring data collected in a statistical cycle exceeds the threshold.

6. Click **Next**.
7. Configure parameters for the notification object, as shown in *Table 7-10: Parameter description*.

   A notification object is an alarm contact. For more information about how to configure an alarm contact, see *Create an alarm contact*.

   Table 7-10: Parameter description

Parameter	Description
Alarm Retries	Select the number of retries before an alarm is triggered from the drop-down list. If the value exceeds the threshold for consecutive statistical cycles, an alarm is triggered. Alarm contacts are notified only after the threshold is exceeded.
Contact Notification Group	Select a contact notification group. After you set an alarm rule for a monitoring item, an alarm notification is sent to the alarm contacts if the monitoring data meets conditions configured in the alarm rule.
Notification Time	Select the notification time, which is a time range during which alarm notifications are sent.

8. Click **Confirm**.

7.6 Manage alarm items

Alarm items are used to display monitoring items in the alarm rules of CloudMonitor Center.

The system provides the ECS, RDS, SLB, and OSS alarm items. The alarm items enable you to

quickly and conveniently view monitoring items, and guarantee the secure, stable, and effective operation of servers.

Management operations are similar among alarm items of ECS, RDS, SLB, and OSS. Take the ECS alarm items as an example.

7.6.1 View alarm items

You can view your alarm items on the **Alarm Items** page after creating an alarm rule.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm**.
3. Select a cloud product, such as **ECS**. The ECS alarm items are displayed.
4. Enter the ID and name of a monitored resource, select a region, monitoring item, alarm status, and enabled status, and then click **Search** to query alarm items.

7.6.2 View alarm history

After alarms are triggered, you can view the alarm history on the **Alarm Items** page.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm**.
3. Select a cloud product.
4. Click the icon at the right of the alarm item and select **Alarm History** to view the alarm history.

7.6.3 Edit an alarm item

To modify the alarm rule of an alarm item, you can edit the alarm item on the **Alarm Items** page of your cloud product.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm**.
3. Select a cloud product.
4. Click the icon at the right of the alarm item and select **Edit** to modify the alarm rule.

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

7.6.4 Pause an alarm item

You can pause one or more alarm items as per your needs.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm**.
3. Select a cloud product.
4. Perform the following operations:
   - Pause an alarm item.

     Click the  icon at the right of the alarm item and select **Pause**.
   - Pause multiple alarm items.

     Select multiple alarm items, and click **Pause** in the upper-right corner.
5. In the displayed dialog box, click **Confirm**.

   After you pause the alarm items, their alarm notifications are not sent to the alarm contacts.

7.6.5 Start multiple alarm items

You can start multiple paused alarm items.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm**.
3. Select a cloud product.
4. Select the paused alarm items and then click **Start**.

7.6.6 View alarm notification objects

After creating alarm rules, you can view the notification objects of each alarm item on the **Alarm Items** page.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm**.
3. Select a cloud product.

4. Click the alarm contact or alarm group under the **`Notification Object`** column.

   The detailed information of the alarm contacts is displayed in the appeared dialog box.

7.6.7 Delete an alarm item

You can delete an alarm item that is no longer in use.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm**.
3. Select a cloud product.
4. Perform the following operations:
   - Delete an alarm item.

     Click the  icon at the right of the alarm item and select **Delete**.
   - Delete multiple alarm items.

     Select multiple alarm items, and click **Delete** in the upper-right corner.
5. In the displayed dialog box, click **Confirm**.

7.7 View monitoring charts

You can view the monitoring chart to know the running status of each instance.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Monitor**.
3. Select a cloud product.
4. Click the  icon at the right of the instance and select **Monitoring Chart**.

   You can view the monitoring data of each monitoring item on the displayed page.

7.8 View alarm information

You can view alarm information to know the running status of ECS, RDS, SLB, and OSS and obtain the exception information in time.

Context

Alarm information is used to display the information of alarm items that do not meet the requirements of alarm rules.

Note:

This section takes the ECS alarm information as an example. Operations of the other cloud resources are similar.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Select **Console** > **Administration** > **CloudMonitor Center** > **Alarm Information**. The **Alarm Information** page appears.
3. You can filter alarm information based on the region, monitored resource ID, monitored resource name, monitoring item, and date. See the following table for the field description of alarm information.

   Table 7-11: Field description

Field	Description
Region	Region in which the monitored object resides.
Monitored Resource ID/ Name	Instance ID or name of the monitored object.
Monitoring Item	Monitoring item of the monitored object.
Description	Detailed description of the alarm information.
Trigger Status	Alarm trigger status, including **Alarms** and **Insufficient Data**.
Threshold	Threshold of the monitoring item.
Alarm Value	Value of the monitoring item when the alarm is triggered.
Start Time	Time when the alarm is started.
End Time	Time when the alarm is ended.

4. Optional: Click **Export** to export the current alarm information to your local computer as an .xls file.

   The exported file is named `alarm.xls` and stored in `C:\Users\Username\Downloads`.

8 Daily management of resources

8.1 View the Home page

The Apsara Stack console uses charts to display the usage and monitoring metrics of existing system resources in all regions.

Note:

Resource types vary with regions. For resource types available to you, see your **Home** page.

8.1.1 View resource overview of a region

The resource overview summarizes the usage of all resources.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. Click **Home** to view the usage of all resources.

   For more information about the fields in the resource overview, see *Field description*.

Table 8-1: Field description

Cloud product	Content	Description
ECS	Number of ECS instances	Total number of ECS instances of the current user.
	Number of disks	Total number of disks of the current user.
	Number of images	Total number of images of the current user.
	Number of snapshots	Total number of snapshots of the current user.
SLB	Number of SLB instances	Total number of SLB instances of the current user.
OSS	Number of OSS instances	Total number of OSS instances of the current user.
RDS	Number of RDS for MySQL /SQLServer/PostgreSQL instances	Total number of RDS for MySQL/SQLServer/ PostgreSQL instances of the current user.

3. In each cloud resource overview, click the total number of resources and then you are redirected to the corresponding resource page. You can view the detailed resource information on the resource page.

8.1.2 View alarm information of a region

You can view the number of alarms triggered for each resource on the alarm distribution diagram to quickly know the current health status of each resource.

Procedure

1. *Log on to the Apsara Stack console* as an administrator or user.
2. The alarm information of each resource is displayed on the **Home** page.
3. Click the tab of a resource to view alarm information of that resource.

   For more information about alarm items of each resource, see *Description of cloud monitoring metrics*.

8.2 Manage cloud resource reports

The Apsara Stack console provides the following types of reports: resource reports, alarm reports, resource usage evaluation reports, and resource monitoring reports. The administrator can view and download these reports.

8.2.1 Create a report download task

Create a report download task on the **System Report** page before previewing or downloading reports.

Context

You can create a download task for the following reports:

- Resource report

  A resource report summarizes the current number of ECS, SLB, RDS, OSS, VPC, and MongoDB instances in the Apsara Stack console and the details of each instance, including the region, department, project, and status of the instance.
- Alarm report

  An alarm report summarizes the alarm information generated by ECS, SLB, RDS, and OSS.
- Resource usage evaluation report

A resource usage evaluation report summarizes the usage of ECS, SLB, RDS, and OSS resources. You can view resource usage evaluation reports to know the usage of each resource and prevent waste or overload use of resources.

- Resource monitoring report

  A resource monitoring report summarizes the monitoring information generated by ECS, SLB, RDS, and OSS.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Administration** > **System Reports**.
3. On the **System Report** page, select a tab.
4. Configure the filter conditions or evaluation rules based on business requirements and click **Create Download Task**.
5. In the displayed **Create Download Task** dialog box, enter a report name and click **Create**.

   After the download task is created, you can view the status of this task under the **Download Center** tab.

8.2.2 Modify the report name

The administrator can modify the report name under the **Download Center** tab after a download task is created.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Administration** > **System Reports**.
3. Click the **Download Center** tab.
4. Click the icon at the right of the download task, and select **Modify Report Name**.

   Note:
   You can also filter the download tasks based on the task status or start date to modify report names.

5. In the displayed dialog box, enter the report name and click **Confirm**.

8.2.3 Preview and download reports

The administrator can preview and download reports based on report names and types.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Administration** > **System Reports**.
3. Click the **Download Center** tab.
4. Select a report based on the report name and type. Click the icon, and then select **Preview**.

   The **Preview Report** page appears.
5. Select a report based on the report name and type. Click the icon, and then select **Download**.
6. In the displayed dialog box, click **Confirm**.

   The downloaded file is stored in `C:\Users\Username\Downloads` by default.

   Where:

   - The downloaded resource reports are named `resource.xls`.
   - The downloaded alarm reports are named `alarm.xls`.
   - The downloaded resource usage evaluation reports are named `Evaluation.xls`.
   - The downloaded resource monitoring reports are named `ResourceMonitor.xls`.

8.2.4 Delete a report download task

The administrator can delete a report download task that is no longer in use.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Administration** > **System Reports**.
3. Click the **Download Center** tab.
4. Click the icon at the right of the download task, and select **Delete**.
5. In the displayed dialog box, click **Confirm**.

8.3 Manage quotas

The administrator can configure quotas of resources for departments in the Apsara Stack console to reasonably distribute resources among departments. The department administrator can create resources for the department within the quotas. If the quotas of the department are used up, the system does not allow the department administrator to create resources for the department. To continue to create resources, increase quotas for the department first.

8.3.1 Quota parameter description

ECS

Parameter	Description
Total CPU Quota (Cores)	Total number of CPU cores that can be configured for ECS.
Total Memory Quota (GB)	Total memory size that can be configured for ECS.
Total Disk Quota (GB)	Total number of cloud disks that can be configured for each ECS instance.

RDS (including primary instances and read-only instances)

Parameter	Description
Total CPU Quota (Cores)	Total number of CPU cores that can be configured for RDS (MySQL/SQLServer/ PostgreSQL).
Total Memory Quota (GB)	Total memory size that can be configured for RDS (MySQL/SQLServer/PostgreSQL).
Total Storage Quota (GB)	Total storage size that can be configured for RDS (MySQL/SQLServer/PostgreSQL).

SLB

Parameter	Description
Total Public IP Addresses	Total number of Internet IP addresses that can be configured for SLB.

Parameter	Description
Total Internal IP Addresses Quota	Total number of intranet IP addresses that can be configured for SLB.

OSS

Parameter	Description
Total OSS Quota	Total number of buckets that can be configured for OSS if **Unlimited Size** is selected.
Total OSS Capacity Quota (GB)	Total size of buckets that can be configured for OSS if **Fixed Size** is selected.

8.3.2 View quotas

The administrator can view the total and remaining quotas of cloud resources for different departments and regions.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Operations Center** > **Quota Management**.
3. Select a department. View the total and remaining quotas of cloud resources in this department.

   For more information about the quota parameters of different resources, see *Quota parameter description*.

8.3.3 Modify a quota

The administrator can modify cloud resource quotas based on the department requirements.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Operations Center** > **Quota Management**.
3. Select a department.
4. Click the  icon of the cloud product.

   The quota modification dialog box of the cloud product appears.

> **Note:**
> For OSS, select a product type first.

5. Click **Confirm**.

8.3.4 Delete quotas

The administrator can delete quotas as required.

Prerequisites

Before deleting quotas, make sure that all sub-departments of the selected department do not have any quotas.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Operations Center** > **Quota Management**.
3. Select a department.
4. Click the icon of the cloud product.

   The **Delete Quota** dialog box appears.

   > **Note:**
   > For OSS, select a product type first.

5. Click **Confirm**.

9 Manage RAM roles

Resource Access Management (RAM) is a resource access control service provided by Apsara Stack, which allows you to control the permissions of RAM roles to access the cloud resources of your department.

9.1 Create a RAM role

To host your cloud resources with a cloud service, you must create a RAM role for this cloud service.

Context

A RAM role is used to grant cloud services permissions to perform operations on resources.

To limit the operations that a service can perform on behalf of you, you must configure a RAM role for the service. Then, the service performs operations in the configured RAM role.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Compute, Storage & Networking** > **Resource Access Management**. The **RAM Role** page appears.
3. In the upper-right corner, click **New RAM Role**.

   The **Default Service Role** page appears.
4. Select the department, region, and cloud service you want to host. Then, click **Create** to complete the authorization.

   The default service role is displayed on the **RAM Role** page after being created.

   For more information about the default service roles that can be created and their role names, see *Table 9-1: Default service role description*.

   Table 9-1: Default service role description

Role name	Service name	Role description
AliyunCloudFirewallAccessingECSRole	Cloud Firewall	Used to grant Cloud Firewall to use this role to access ECS.
AliyunECSImageExportDefaultRole	ECS	Used to grant ECS to use this role to export images.

Role name	Service name	Role description
AliyunECSImageImport DefaultRole		Used to grant ECS to use this role to import images.
AliyunEMRDefaultRole	E-MapReduce	Used to grant E-MapReduce to use this role to access resources of other cloud products that belong to the same department.
AliyunEMRECSDefaultRole		Used to grant E-MapReduce jobs to use this role to access your cloud resources.

For example, select A as the **Region**, B as the **Department**, and Cloud Firewall as the **Service** to create a RAM role. After the RAM role is created successfully, Cloud Firewall in region A can use the created RAM role to access resources of other cloud products that belong to department B in region A.

9.2 View role details

The administrator can view the role details to know the name, description, created time, and global resource descriptor of the role.

Prerequisites

A RAM role is created.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Compute, Storage & Networking** > **Resource Access Management**. The **RAM Role** page appears.
3. Click the  icon at the right of the role, and select **View Details**.

   The **Role Details** page appears.

   View the role details, including the name, description, created time, and global resource descriptor of the role.

9.3 View a role policy

The administrator can know the detailed permissions of a role by viewing a role policy.

Prerequisites

A RAM role is created.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Compute, Storage & Networking** > **Resource Access Management**. The **RAM Role** page appears.
3. Click the icon at the right of the role, and select **View Details**.

   The **Role Details** page appears.
4. Click the **Role Policy** tab.
5. Click the icon at the right of the policy, and select **View Details** to view the policy details, including the name, description, type, and contents of the policy.

10 System maintenance

10.1 Manage departments

10.1.1 Modify the department name

The administrator can modify the department name.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Department Management**.
3. Select a department and click **Modify Department**.
4. In the displayed **Modify Department** dialog box, modify the department name and click **Confirm**.

10.1.2 View projects of a department

The administrator can view projects of a department to view the project information.

Context

Departments reflect the tree structure of an enterprise or business unit. A department can have multiple projects.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Department Management**.
3. Select a department.

   Projects of this department are displayed on the right.

10.1.3 Obtain the AccessKey of a department

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Department Management**.
3. Select a department and click **Create AccessKey** to obtain the account name, AccessKey, and PrimaryKey of the department.

   Note:

The account name, AccessKey, and PrimaryKey are automatically allocated to a level-1 department. The sub-departments use the same account name, AccessKey, and PrimaryKey as their level-1 department.

10.1.4 Delete a department

The administrator can delete a department that is no longer in use.

Prerequisites

Note:
Make sure the department to be deleted does not contain any users, projects, or sub-departments. Otherwise, the department cannot be deleted.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Department Management**.
3. Select a department and click **Delete Department**.

10.2 Manage projects

A project is a container where resources are stored. All resources must be applied for and created under a project.

10.2.1 Modify the project name

The administrator can modify the project name.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Project Management**.
3. Click the icon at the right of the project and select **Modify Project Name**.
4. In the displayed **Modify Project Name** dialog box, modify the project name and click **Confirm**.

10.2.2 View project details

The administrator can view the basic information of a project, including the name, ID, department, created time, and headcount, by viewing the project details.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.

2. Select **Console** > **User Center** > **Project Management**.
3. Click the icon at the right of the project and select **View Details**.

   The **Project Details** page appears.

10.2.3 View project members

To use resources of a project, you must be a member of the project. Check if you are in the member list of the project.

Context

The members of a project have the permissions to use resources of the project.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Project Management**.
3. Click the icon at the right of the project and select **View Details**.

   The **Project Details** page appears.
4. Click the **Project Members List** tab. You can view all the members of the project and their contact information.

10.2.4 View resource information of a project

You can view the resource information of a project in the project resource list if you want to use that cloud resource.

Context

All the cloud resources of a project are displayed in the project resource list.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Project Management**.
3. Click the icon at the right of the project and select **View Details**.

   The **Project Details** page appears.
4. Click the **Project Resource List** tab.
5. In the **Project Resource List**, view all cloud resources of the project.

6. Select a cloud product.
7. Click the icon at the right of the resource and select **View Details** to view the resource details.

10.2.5 Release resources

The administrator can release the resources that are no longer in use in a project.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Project Management**.
3. Click the icon at the right of the project and select **View Details**.

   The **Project Details** page appears.
4. Click the **Project Resource List** tab and then:
   - Release a single resource.

     Select a cloud product. Click the icon at the right of the resource, and select **Release Resources**. In the displayed dialog box, click **Confirm**.
   - Release multiple resources.

     Select a cloud product. Select multiple resources, and then click **Delete** in the upper-right corner.

10.2.6 Delete a project

The administrator can delete a project that is no longer in use when the project is complete or changed.

Prerequisites

> **Note:**
> Make sure the project to be deleted does not contain any resources or project members. Otherwise, the project cannot be deleted.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Project Management**.

3. Click the icon at the right of the project and select **Delete Project**.

> **Note:**
> Alternatively, you can click the icon at the right of the project and select **View Details**. On the **Project Details** page, click **Delete Project** to delete the project.

4. Click **Confirm**.

10.3 Manage roles

A role is a collection of access permissions. Each role corresponds to a range of permissions. A user can have multiple roles, which means that this user has all the permissions defined in these roles. You can use a role to grant the same permissions to a group of users.

10.3.1 View role details

The administrator can view permissions of a role on the **Role Management** page.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Role Management**.
3. Click the icon at the right of the role. View the permissions of this role in the displayed dialog box.

10.3.2 Modify a custom role

The administrator can modify the description and permissions of a custom role.

Context

> **Note:**
> System default roles cannot be modified.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Role Management**.
3. Click the icon at the right of the role, and select **Edit**.
4. In the displayed **Modify Role** dialog box, modify the description, permission scope, and permission list of the role.

5. Click **Confirm**.

10.3.3 Delete a custom role

The administrator can delete a custom role that is no longer in use.

Context

> **Note:**
> System default roles cannot be deleted.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Role Management**.
3. Click the icon at the right of the role, and select **Delete**.

   The **Confirm Deletion** dialog box appears.
4. Click **Confirm**.

10.4 Manage users

10.4.1 View basic information of a user

The administrator can view the basic information of a user to know the department, role, and contact information of the user.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Click the icon at the right of the user and select **User Information** to view the basic information of the user.

10.4.2 Modify user information

The administrator can modify the display name and contact information of a user if the user information is changed.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.

2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Click the icon at the right of the user and select **Edit**.
5. In the displayed **Edit User** dialog box, modify the display name and contact information of the user.

10.4.3 Change the logon policy of a user

For better management, the administrator can change the logon policy of a user to modify the permitted logon time and IP addresses for the user.

Prerequisites

A logon policy has been created. For more information, see *Create a logon policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Click the icon at the right of the user and select **Manage Logon Policy**.
5. In the displayed **Assign Logon Policy** dialog box, select the logon policy.
6. Click **Confirm**.

   After the logon policy of the user is changed, the user is limited by the new policy.

   If the user does not want to be limited by the bound logon policy, the user must submit an application to the administrator. After approving the application, the administrator binds the user to a logon policy that meets the user's requirements.

10.4.4 Modify user roles

The administrator can modify user roles by adding, changing, or deleting roles for a user.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Click the icon at the right of the user, and select **Authorize**.

The **Modify Authorization** dialog box appears.

5. In the **Role** field, add, change, or delete roles for the user.
6. Click **Confirm**.

10.4.5 Obtain the AccessKey of a personal account

If you have activated Object Storage Service (OSS) and want to access it, you must obtain the AccessKey ID and AccessKey Secret of a personal account for logon authorization.

Procedure

1. *Log on to the Apsara Stack console*.
2. In the upper-right corner of the Web page, click the icon and select **Personal Information**.
3. Click **AccessKey**.

   The **Get AccessKey** dialog box appears.
4. Click **Confirm**.

   The region, department, AccessKey ID, and AccessKey Secret of the current user are displayed on the right.

10.4.6 Authorize third-party access

To call APIs of the Apsara Stack console, the administrator must authorize third-party access to obtain the AccessKey used for third-party access.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Click the icon at the right of the user and select **Authorize Third-Party Access**.
5. In the displayed **Authorize Third-Party Access** dialog box, click **Authorize**.

> **Note:**
> **Authorize Third-Party Access** is enabled by default. You can click **Recreate an AccessKey** or **Remove the AccessKey** in the displayed dialog box.

10.4.7 Reset logon password

The administrator can reset the logon passwords for users if they forget their logon passwords.

Prerequisites

Only users who have the permissions to manage users and projects can reset the logon passwords.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Click the  icon at the right of the user and select **User Information**.
5. On the **User Information** page, click **Reset Password**. The system automatically generates a new password and sends the new password to the user by SMS.

10.4.8 Export initial user password

If the user does not receive any SMS notification after the password is reset, the administrator can export the initial user password and notify the user of the password orally.

Prerequisites

The password has been reset. For more information, see *Reset logon password*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Select a user and click **Export Initial User Password**.

   The password file `UserInitPassword.txt` is generated.

10.4.9 Enable or disable a user

To prevent a user from logging on to the Apsara Stack console, the administrator can disable the user. A disabled user must be enabled before logging on to the Apsara Stack console.

Context

A user is activated by default after being created.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Perform the following operations:
   - Click the icon at the right of an **Activated** user, and select **Disable** to disable this user.
   - Click the icon at the right of a **Disabled** user, and select **Activate** to enable this user.

10.4.10 Delete a user

The administrator can delete a user based on business requirements.

Prerequisites

The user has been removed from all projects. For more information, see *Add a project member*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Click the icon at the right of the user and select **Delete**.
5. In the displayed **Confirm Deletion** dialog box, click **Confirm**.

   The deleted user still exists in the database, but does not belong to any department or have any role, and cannot log on to the Apsara Stack console.

10.4.11 Restore a user

After a user is deleted, the administrator can locate and restore the user in the **Deleted Users** list.

Context

Except for the department and role, the other basic information and the logon password of a restored user are the same as those before the user was deleted.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Deleted Users** tab.

4. Click the  icon at the right of the user, and select **Recover**.

   The **Restore User** dialog box appears.

5. Select a department and a role for the restored user and click **Confirm**.

10.5 Manage logon policies

10.5.1 View a logon policy

The administrator can view a logon policy to know the permitted logon time and IP addresses of a user.

Context

A default logon policy is automatically generated when the Apsara Stack console provides services. This policy does not have any limits on the logon time and IP addresses.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **Logon Policy Management**.
3. Optional: Enter the policy name in the search box and click **Search**.

   The search result appears.

4. View the logon policy, which includes the permitted logon time and IP addresses of users.

10.5.2 Bind a logon policy to multiple users

Prerequisites

- You have created users. For more information about how to create users, see *Create a user*.
- You have created a logon policy. For more information about how to create a logon policy, see *Create a logon policy*.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **User Center** > **User Management**.
3. Click the **Users** tab.
4. Select multiple users and click **Assign Logon Policy** to bind a logon policy to multiple users.

10.6 Manage operation logs

Logs record a series of operations performed by all users in the Apsara Stack console.

10.6.1 View logs

The administrator can view logs to know the usage status of resources, such as ECS, RDS, and SLB, and the running status of all function modules in the Apsara Stack console in real time.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Administration** > **Operation Log**. The **Operation Log** page appears.
3. You can filter operation logs by username, module, level, instance ID, start date, and end date.

   For more information about the fields in the search results, see *Field description*.

   Table 10-1: Field description

Field	Description
Time	Operation time.
Username	Name of the operator.
Module	• ECS: Records all operations related to ECS instances, including creating, modifying, deleting, and querying ECS instances. • RDS: Records all operations related to RDS instances, including creating, modifying, deleting, and querying RDS instances. • OSS: Records all operations related to OSS instances, including creating, modifying, deleting, and querying OSS instances. • OTS: Records all operations related to Table Store instances, including operations of Table Store instances and tables. • SLB: Records all operations related to SLB instances, including creating, modifying, deleting, and querying SLB instances. • VPC: Records all operations related to VPC instances, including creating, modifying, deleting, and querying VPC instances, and managing VSwitches and VRouters. • AUTH: Records all operations related to user roles, including adding and deleting user roles. • USER: Records user activities, including logon time and logout time. • PROJECT: Records all operations related to projects, including creating, updating, querying, and deleting projects, and adding and deleting project members.

Field	Description
	• DEPARTMENT: Records all operations related to departments, including creating, modifying, and deleting departments. • LOGINPOLICY: Records all operations related to logon polices, including creating, modifying, and deleting logon policies.
Operation Object	The instance ID/name of the operation object.
Region	The region in which the operation object resides.
Level	The operation level, including Information, Notification, Warnings, Error, Important, Emergency, Alarms, and Debug.
Action	The action type, including logon, logout, and display.
Details	Brief introduction to the operation objectives.

4. Optional: Click **Export** to export the current logs to your local computer as an .xls file.

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

10.6.2 Delete logs

The administrator can delete logs within a specific time period if they are no longer in use.

Context

Note:

Logs cannot be recovered after being deleted, so proceed with caution.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Administration** > **Operation Log**. The **Operation Log** page appears.
3. Configure the query conditions and then click **Search**.
4. Click **Delete Log** to delete logs within a specific time period.

10.7 System configurations

10.7.1 Configure the system OSS

Configure the system OSS to specify the storage path for uploaded attachments.

Prerequisites

Before configuring the system OSS, select an OSS bucket as the system OSS and obtain the AccessKey ID and AccessKey Secret. AccessKey ID and AccessKey Secret are used to identify a visitor. The system uses AccessKey ID and AccessKey Secret to access OSS.

Obtain AccessKey ID and AccessKey Secret as follows:

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Compute, Storage & Networking** > **Object Storage Service**. View the region and department to which the bucket belongs.
3. Select **Console** > **User Center** > **Department Management**. Locate the region and department of the bucket. Select the department and then click **Create AccessKey**.

Context

By default, the storage path for attachments is not configured in the Apsara Stack console, and no attachment upload function is available. Configure the system OSS to specify the storage path for uploaded attachments to implement the high-reliability storage of large numbers of attachments.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the upper-right corner of the Web page, click the  icon. The **System Configuration** page appears.
3. Click the **OSS Configuration** tab.
4. Set **`Storage Mode`** to **`OSS`**.
5. Configure parameters for the system OSS.

   For more information about the parameters, see *Parameter description*.

   Table 10-2: Parameter description

Parameter	Description
OSS Endpoint	The endpoint address of OSS. Obtain the endpoint by viewing the bucket details.
Bucket Name	The name of the bucket.
AK ID and AK Secret	The AccessKey used to access OSS. AccessKey ID is used to identify a user, and AccessKey Secret is a key used to authenticate the user.

6. Click **Save**.
7. Click **Test Connection**.

To modify the OSS configuration, click **Reset** and configure the system OSS again.

10.7.2 Manage resource notification objects

You can add users as notification objects or remove users from notification objects by configuring the resource notifications. When resources are created or deleted in the Apsara Stack console, emails and SMS notifications are sent to the configured notification objects. Users who are added as notification objects can keep up with the resource usage.

10.7.2.1 Configure resource notification objects

Configure the resource notification objects to receive emails and SMS notifications from the Apsara Stack console when resources are created or deleted.

Context

Note:
The Apsara Stack console can send resource notifications for ECS, OSS, RDS, and SLB resources.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the upper-right corner of the Web page, click the icon. The **System Configuration** page appears.
3. Click the **Resource Notification Configuration** tab.
4. Click **Configuration Details** in the **Resource Notification Configuration** section.

   The **Add User** dialog box appears.
5. Select users under **Available Users** and click -> to add them to **Selected Users**. Click **Add** to complete the configurations.

10.7.2.2 View resource notification objects

View the information about resource notification objects.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the upper-right corner of the Web page, click the icon. The **System Configuration** page appears.

3. Click the **Resource Notification Configuration** tab.
4. View the resource notification objects.
5. Optional: Alternatively, you can click the username of a resource notification object to view the details of this notification object.

10.7.2.3 Delete a resource notification object

The administrator can remove users who are no longer required to be notified of new changes from resource notification objects because of business change or other reasons.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the upper-right corner of the Web page, click the  icon. The **System Configuration** page appears.
3. Click the **Resource Notification Configuration** tab.
4. Click the  icon at the right of the user and select **Delete**.
5. In the displayed **Delete User** dialog box, click **Confirm**.

10.7.3 Set ECS startup configuration

In **Resource Notification Configuration**, configure whether or not the ECS instance is automatically started after it is created.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the upper-right corner of the Web page, click the  icon. The **System Configuration** page appears.
3. Click the **Resource Notification Configuration** tab.
4. In the **ECS Startup Configuration** section, select the **Automatically start the ECS instance after it is created** check box and then click **Configuration Details**.

   A system prompt appears, indicating the instance has been configured.

10.7.4 Set alarm gateway configurations

10.7.4.1 Configure email notifications

Prerequisites

Make sure the SMTP server URL is obtained before configuring email notifications.

To obtain the SMTP server URL and port, view the official description about the mailbox system to be configured. Generally, the SMTP server URL is in the format of smtp.xxxx.com. For example, the SMTP server URL of the 163 mailbox is smtp.163.com.

The system sends email notifications by using the configured email address and email password.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the upper-right corner of the Web page, click the  icon. The **System Configuration** page appears.
3. Click the **Alarm Gateway Configuration** tab.
4. In the **Email Notification Settings** section, click **Configuration Details**.

   The **Email Notification Settings** dialog box appears.
5. Enter the SMTP server URL, email address, and email password, and then select the SMTP server port.
6. Click **Confirm**.

   To modify the configurations, click **Clear Configurations** and re-configure the parameters.

10.7.4.2 Configure DingTalk alarm notifications

Context

To send alarm notifications by using DingTalk, you must obtain the CorpID, CorpSecret, and AgentID.

Procedure

1. Obtain the AgentID.
   a) Log on to oa.dingtalk.com as a DingTalk administrator.
   b) Click **Applications** and locate the **Application Base** section.
   c) Click the icon on an application and then select **Set**.
   d) In the displayed dialog box, obtain the AgentID.
2. Obtain the CorpID and CorpSecret.
   a) Log on to oa.dingtalk.com as a DingTalk administrator.
   b) Click **Applications** and locate the **Create your app** section.

c) Click **Open Application** to go to the DingTalk Developer console.

d) In the left-side navigation pane, click **Account Management**. In the **Account Information** section, obtain the CorpID and CorpSecret.

3. *Log on to the Apsara Stack console* as an administrator.
4. In the upper-right corner of the Web page, click the icon. The **System Configuration** page appears.
5. Click the **Alarm Gateway Configuration** tab.
6. In the **DingTalk Alarm Notification Settings** section, click **Configuration Details**.

   The **DingTalk Notification Settings** dialog box appears.
7. Enter the CorpID, CorpSecret, and AgentID.
8. Click **Confirm**.

   To modify the configurations, click **Clear Configurations** and re-configure the parameters.

10.7.4.3 Configure SMS alarm notifications

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. In the upper-right corner of the Web page, click the icon. The **System Configuration** page appears.
3. Click the **Alarm Gateway Configuration** tab.
4. In the **Configure SMS Alarm Notifications** section, click **Configuration Details**.

   The **SMS Notification Configuration** dialog box appears.
5. Enter the notification URL, AccessKey ID, and AccessKey Secret.
6. Click **Confirm**.

   To modify the configurations, click **Clear Configurations** and re-configure the parameters.

10.8 Task Center

After creating a task (for example, an ECS instance) in the Apsara Stack console, you can view the task status in Task Center.

10.8.1 View running tasks

The administrator can view the details of running tasks.

Context

You can query tasks by department, task ID, task name, task type, and created time.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Administration** > **Task Center**.
3. Click the **Execute Task** tab.
4. Configure the query conditions and then click **Search**.

   In the search result, view the task details.

10.8.2 View previous tasks

The administrator can view the details of completed tasks in **Previous Tasks**.

Context

A previous task is a task that has been completed.

You can query tasks by department, task ID, task name, task type, and created time.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Select **Console** > **Administration** > **Task Center**.
3. Click the **Previous Tasks** tab.
4. Configure the query conditions and then click **Search**.
5. Optional: If a task failed, click **Error** under the **Status** column.

   View the failure details of the task.

10.9 Typical scenarios

10.9.1 Use departments, projects, roles, and user accounts

This section uses a simple scenario to introduce how to use departments, projects, roles, and user accounts.

Prerequisites

The basic information, including name and contact information, of all users is obtained.

Context

A company establishes a new department (department A) with 10 employees to be in charge of project B. The administrator must create the accounts for these 10 users and grant them the corresponding permissions.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Create department A. For more information, see *Create a department*.
3. Create project B. For more information, see *Create a project*.

   Select department A as the **Department** when creating project B.
4. Optional: Create one or more custom roles based on the business permissions to be granted to users. For more information, see *Add a custom role*.

   If custom roles are not required, use the system default roles.
5. Optional: Create a logon policy. For more information, see *Create a logon policy*.

   To not limit the logon time and IP addresses of users, use the default logon policy.
6. Create an account for each of the 10 users based on their own business conditions. For more information, see *Create a user*.

   When creating a user:

   - Select department A as the **Department**.
   - Select the created custom role or system default role as the **Role**.
   - Select the created logon policy or default logon policy as the **Logon Policy**.
7. Add these 10 users to the member list of project B. Then, they can use the cloud resources in the project. For more information, see *Add a project member*.

10.9.2 Modify the department of a user

Context

Scenario:

User A is transferred to another department because of the department structure adjustment inside an enterprise. The administrator must modify the department to which user A belongs.

Procedure

1. *Log on to the Apsara Stack console* as an administrator.
2. Remove user A from the project member list.

   For more information about how to remove a project member, see *Add a project member*.
3. Delete user A.

   For more information, see *Delete a user*.
4. Restore user A from the **Deleted Users** list.

   Select the new department and role of user A in the **Restore User** dialog box.

   For more information, see *Restore a user*.

   The username, password, cellphone number, and email address after user A is restored are the same as those before user A was deleted.

11 Personalized settings

You can change your logon password and portrait after logging on to the Apsara Stack console.

11.1 Configure custom menu items

You can customize menu items in the left-side navigation pane as per your needs for convenience.

Procedure

1. In the Apsara Stack console, click **Console** and then select a menu item.

   The default menu items are displayed in the left-side navigation pane.

2. Click the icon next to **Custom Menu** in the left-side navigation pane.

   The **Configure Quick Access Menu** dialog box appears.

3. Click the + icon to add main menu or submenu items to the left-side navigation pane.

   If the main menu item is added, its submenu items are also added. If a submenu item is not required, locate the submenu item under **Add to the Left Navigation Pane** and click the icon to remove it from the left-side navigation pane.

   Note:
   Click the icon to view submenu items of a main menu item.

4. Click **Confirm**.

   The added menu items are displayed in the left-side navigation pane.

11.2 Change your logon password

To improve security, change your logon password in time.

Procedure

1. In the upper-right corner of the Web page, click the icon and select **Personal Information**.

   The **Personal Information** page appears.

2. Click **Change Password**. Enter the **Current Password**, **New Password**, and **Confirm Password**.

3. Click **Submit Changes**.

11.3 Change your portrait

You can upload your custom portrait for better identification.

Procedure

1. In the upper-right corner of the Web page, click the  icon and select **Personal Information**.

   The **Personal Information** page appears.

2. Click **Change Picture**.

   The **Change Picture** section appears.

3. Click **Upload Files**.

4. Select the picture you want to upload and click **Open**. The picture to be uploaded appears in the preview area.

   To change the picture, click **Reset** and select another picture.

5. Click **Confirm** to change your portrait.

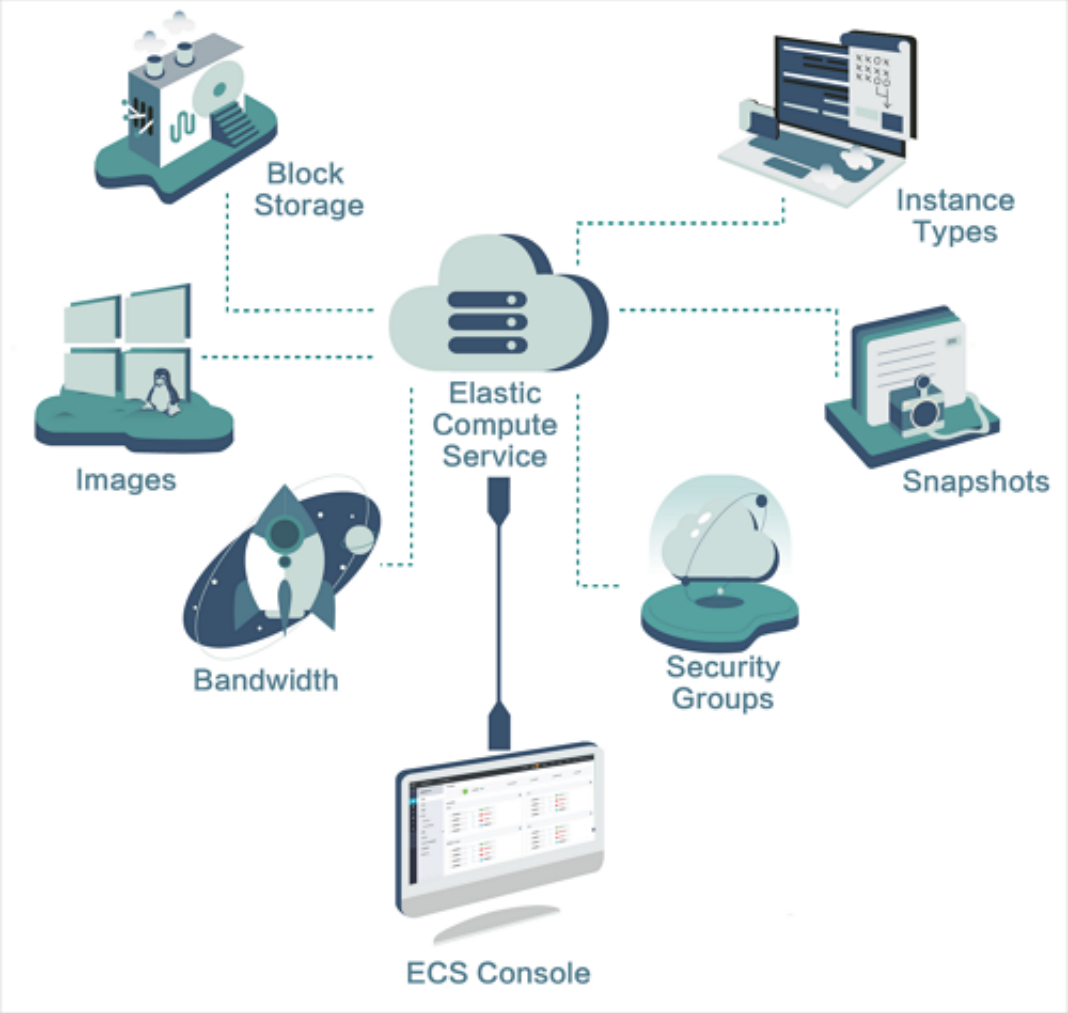
12 Elastic Compute Service (ECS)

12.1 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 12-1: Concept of an ECS instance* illustrates the services supported by an ECS instance.

Figure 12-1: Concept of an ECS instance



12.1.1 Instance types

For an ECS instance, its type specifies two attributes, its CPU (such as model and clock speed) and memory. To definitely determine the application scenario, however, you must select the image, disk, and network service at the same time. *Table 12-1: Instance types* details the instances of various attributes within each ECS instance type family.

Table 12-1: Instance types

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
N4	ecs.n4.small	N/A	1	2.0	1
	ecs.n4.large	N/A	2	4.0	1

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.n4.xlarge	N/A	4	8.0	2
	ecs.n4.2xlarge	N/A	8	16.0	2
	ecs.n4.4xlarge	N/A	16	32.0	2
	ecs.n4.8xlarge	N/A	32	64.0	2
MN4	ecs.mn4.small	N/A	1	4.0	1
	ecs.mn4.large	N/A	2	8.0	1
	ecs.mn4.xlarge	N/A	4	16.0	2
	ecs.mn4. 2xlarge	N/A	8	32.0	3
	ecs.mn4. 4xlarge	N/A	16	64.0	8
	ecs.mn4. 8xlarge	N/A	32	128.0	8
E4	ecs.e4.small	N/A	1	8.0	1
	ecs.e4.large	N/A	2	16.0	1
	ecs.e4.xlarge	N/A	4	32.0	2
	ecs.e4.2xlarge	N/A	8	64.0	3
	ecs.e4.4xlarge	N/A	16	128.0	8
XN4	ecs.xn4.small	N/A	1	1.0	1
gn5	ecs.gn5-c4g1. xlarge	440	4	30.0	2
	ecs.gn5-c8g1. 2xlarge	440	8	60.0	3
	ecs.gn5-c4g1. 2xlarge	880	8	60.0	3
	ecs.gn5-c8g1. 4xlarge	880	16	120.0	8
	ecs.gn5-c28g1 .7xlarge	440	28	112.0	8

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.gn5-c8g1.8xlarge	1,760	32	240.0	8
	ecs.gn5-c28g1.14xlarge	880	56	224.0	8
	ecs.gn5-c8g1.14xlarge	3,520	56	480.0	8
d1	ecs.d1.2xlarge	4 * 5,500	8	32.0	3
	ecs.d1.4xlarge	8 * 5,500	16	64.0	8
	ecs.d1.6xlarge	12 * 5,500	24	96.0	8
	ecs.d1-c8d3.8xlarge	12 * 5,500	32	128.0	8
	ecs.d1.8xlarge	16 * 5,500	32	128.0	8
	ecs.d1-c14d3.14xlarge	12 * 5,500	56	160.0	8
	ecs.d1.14xlarge	28 * 5,500	56	224.0	8
gn4	ecs.gn4-c4g1.xlarge	N/A	4	30.0	2
	ecs.gn4-c8g1.2xlarge	N/A	8	60.0	3
	ecs.gn4.8xlarge	N/A	32	48.0	8
	ecs.gn4-c4g1.2xlarge	N/A	8	60.0	3
	ecs.gn4-c8g1.4xlarge	N/A	16	60.0	8
	ecs.gn4.14xlarge	N/A	56	96.0	8
ga1	ecs.ga1.xlarge	1*87	4	10.0	2
	ecs.ga1.2xlarge	1*175	8	20.0	3

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.ga1. 4xlarge	1*350	16	40.0	8
	ecs.ga1. 8xlarge	1*700	32	80.0	8
	ecs.ga1. 14xlarge	1*1,400	56	160.0	8
se1ne	ecs.se1ne. large	N/A	2	16.0	1
	ecs.se1ne. xlarge	N/A	4	32.0	2
	ecs.se1ne. 2xlarge	N/A	8	64.0	3
	ecs.se1ne. 4xlarge	N/A	16	128.0	8
	ecs.se1ne. 8xlarge	N/A	32	256.0	8
	ecs.se1ne. 14xlarge	N/A	56	480.0	8
sn2ne	ecs.sn2ne. large	N/A	2	8.0	1
	ecs.sn2ne. xlarge	N/A	4	16.0	2
	ecs.sn2ne. 2xlarge	N/A	8	32.0	3
	ecs.sn2ne. 4xlarge	N/A	16	64.0	8
	ecs.sn2ne. 8xlarge	N/A	32	128.0	8
	ecs.sn2ne. 14xlarge	N/A	56	224.0	8
sn1ne	ecs.sn1ne. large	N/A	2	4.0	1

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.sn1ne. xlarge	N/A	4	8.0	2
	ecs.sn1ne. 2xlarge	N/A	8	16.0	3
	ecs.sn1ne. 4xlarge	N/A	16	32.0	8
	ecs.sn1ne. 8xlarge	N/A	32	64.0	8
gn5i	ecs.gn5i-c2g1. large	N/A	2	8.0	1
	ecs.gn5i-c4g1. xlarge	N/A	4	16.0	2
	ecs.gn5i-c8g1. 2xlarge	N/A	8	32.0	2
	ecs.gn5i-c16g1 .4xlarge	N/A	16	64.0	2
	ecs.gn5i-c28g1 .14xlarge	N/A	56	224.0	2
g5	ecs.g5.large	N/A	2	8.0	2
	ecs.g5.xlarge	N/A	4	16.0	3
	ecs.g5.2xlarge	N/A	8	32.0	4
	ecs.g5.4xlarge	N/A	16	64.0	8
	ecs.g5.6xlarge	N/A	24	96.0	8
	ecs.g5.8xlarge	N/A	32	128.0	8
	ecs.g5. 16xlarge	N/A	64	256.0	8
	ecs.g5. 22xlarge	N/A	88	352.0	15
c5	ecs.c5.large	N/A	2	4.0	2
	ecs.c5.xlarge	N/A	4	8.0	3
	ecs.c5.2xlarge	N/A	8	16.0	4

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.c5.4xlarge	N/A	16	32.0	8
	ecs.c5.6xlarge	N/A	24	48.0	8
	ecs.c5.8xlarge	N/A	32	64.0	8
	ecs.c5. 16xlarge	N/A	64	128.0	8
r5	ecs.r5.large	N/A	2	16.0	2
	ecs.r5.xlarge	N/A	4	32.0	3
	ecs.r5.2xlarge	N/A	8	64.0	4
	ecs.r5.4xlarge	N/A	16	128.0	8
	ecs.r5.6xlarge	N/A	24	192.0	8
	ecs.r5.8xlarge	N/A	32	256.0	8
	ecs.r5.16xlarge	N/A	64	512.0	8
	ecs.r5.22xlarge	N/A	88	704.0	15
se1	ecs.se1.large	N/A	2	16.0	2
	ecs.se1.xlarge	N/A	4	32.0	3
	ecs.se1. 2xlarge	N/A	8	64.0	4
	ecs.se1. 4xlarge	N/A	16	128.0	8
	ecs.se1. 8xlarge	N/A	32	256.0	8
	ecs.se1. 14xlarge	N/A	56	480.0	8
d1ne	ecs.d1ne. 2xlarge	4 * 5,500	8	32.0	4
	ecs.d1ne. 4xlarge	8 * 5,500	16	64.0	8
	ecs.d1ne. 6xlarge	12 * 5,500	24	96.0	8

Instance type family	Instance type	Local storage (GiB)	vCPU (Core)	Memory (GiB)	ENI (including 1 primary elastic NIC)
	ecs.d1ne.8xlarge	16 * 5,500	32	128.0	8
	ecs.d1ne.14xlarge	28 * 5,500	56	224.0	8
f3	ecs.f3-c16f1.4xlarge	N/A	16	64.0	8
	ecs.f3-c16f1.8xlarge	N/A	32	128.0	8
	ecs.f3-c16f1.16xlarge	N/A	64	256.0	16
ebmg5	ecs.ebmg5.24xlarge	N/A	96	384.0	32
i2	ecs.i2.xlarge	1 * 894	4	32.0	3
	ecs.i2.2xlarge	1 * 1,788	8	64.0	4
	ecs.i2.4xlarge	2 * 1,788	16	128.0	8
	ecs.i2.8xlarge	4 * 1,788	32	256.0	8
	ecs.i2.16xlarge	8 * 1,788	64	512.0	8
re5	ecs.re5.15xlarge	N/A	60	990.0	8
	ecs.re5.30xlarge	N/A	120	1,980.0	15
	ecs.re5.45xlarge	N/A	180	2,970.0	15

12.1.2 Instance lifecycle

The lifecycle of an instance begins with creation and ends with release. *Table 12-2: Lifecycle description* shows the different states of an instance during its entire lifecycle.

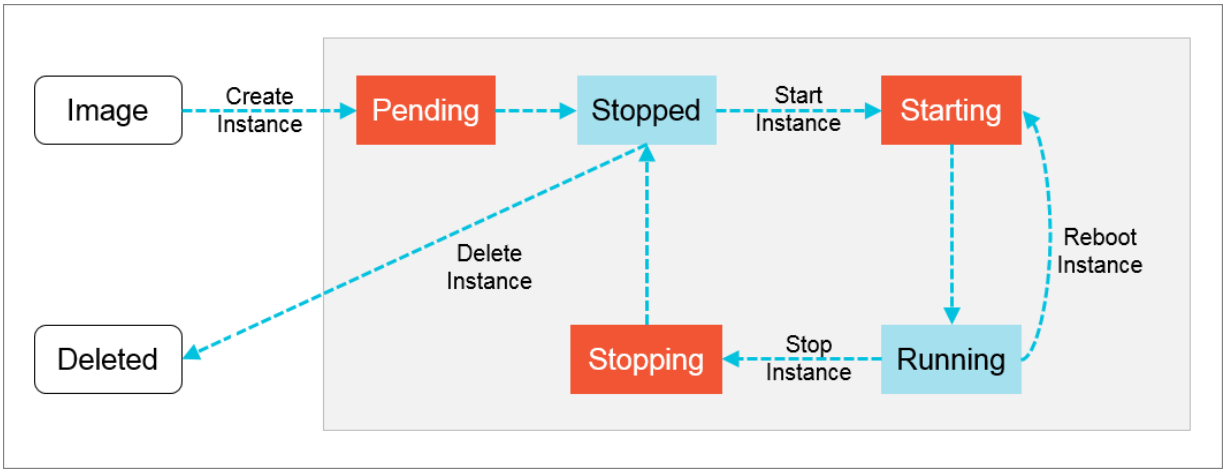
Table 12-2: Lifecycle description

Status	Status attribute	Description	Corresponding API status
Creating task	Intermediate status	Instance creation in progress. Waiting for start. If an instance is in this status for a long time, an exception occurs.	**Pending**
Starting	Intermediate status	It is the state entered by an instance prior to the **Running** state before you perform a restart or start operation for that instance on the console or via an API. If an instance is in this status for a long time, an exception occurs.	**Starting**
Running	Stable status	Indicates that the instance is running smoothly. The instance in this state can accommodate your business needs.	**Running**
Stopping	Intermediate status	An instance is in this status after the Stop operation is performed on the console or using an API but before the instance enters the **Stop** state. If an instance is in this status for a long time, an exception occurs.	**Stopping**
Stop	Stable status	Indicates the instance has been stopped normally. In this status, the instance cannot accommodate external services.	**Stopped**
Re-initializing	Intermediate status	An instance is in this status after the system disk and/or data disk is re-initialized in the console or using an API until it is **Running**. If an instance is in this status for a long time, an exception occurs.	**Stopped**
Replacing System Disk	Intermediate status	An instance is in this status after the operating system is replaced or another such operation is performed on the console using an	**Stopped**

Status	Status attribute	Description	Corresponding API status
		API until, and before the instance enters the **`Running`** state. If an instance is in this status for a long time, an exception occurs.	

Table 12-2: Lifecycle description describes mappings between console statuses and API statuses. The API status chart is shown in Figure *Figure 12-2: API status chart*.

Figure 12-2: API status chart



12.2 Instructions before use

12.2.1 Prohibitions

- Do not upgrade the ECS kernel or operating system without prior authorization.
- Do not start SELinux for the other Linux systems except CentOS and RedHat.
- Do not detach PVDriver.
- Do not arbitrarily modify the MAC address of the network adapter.

12.2.2 Suggestions

- For ECS instances with memory above 4 GB, use a 64-bit operating system (a 32-bit operating system has a 4 GB limitation in memory addressing).
- A 32-bit Windows operating system supports CPUs with up to four cores.
- To guarantee service continuity and avoid service unavailability due to downtime migration, we recommend that you configure service applications to automatically start at system startup.

12.2.3 Restrictions

- Windows does not support instance specifications higher than 64vCPU.
- Virtual application installation and subsequent virtualization (for example, using VMware) are not yet supported.
- Currently, ECS does not support sound card applications (only a GPU instance supports analog audio cards) and cannot connect to external hardware devices (such as hardware dongles, USB drives, external hard disks, and USB keys of banks).
- Currently, ECS does not support multicast protocols. If you want to use multicast services, we recommend that you use the unicast point-to-point method.

12.2.4 Precautions for using ECS instances in Windows

- If an instance uses a local disk as its data disk, there will be a risk of data loss. If you are unable to ensure the reliability of the data architecture, we strongly recommend you to use a cloud disk to establish your instance.
- Do not close the built-in shutdownmon.exe process. Otherwise, it may take longer to restart your Windows server.
- For normal use of the server, do not rename, delete, or disable administrator accounts.
- The use of virtual memory is not recommended.
- If you have changed your computer name, you must synchronize the related key values in the registry; otherwise, the computer name cannot be modified, causing a failure to install certain third-party programs. The following key values must be modified in the registry:

```
HKEY_LOCAL_MACHINE\SYSTEM\ControlSet001\Control\ComputerName\
ActiveComputerName
```

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\ComputerName\
ComputerName
```

12.2.5 Precautions for using ECS instances in Linux

- Do not modify the default /etc/issue file of a Linux instance. Otherwise, the system release version of the custom image created on the basis of the instance will be unidentifiable, and the instance created by using this image cannot be started.
- Do not change the permission of any directory in the partition in which the root directory resides , particularly, the permissions of /etc, /sbin, /bin, /boot, /dev, /usr, and /lib. Improper modification of permissions may cause exceptions.
- Do not rename, delete, or disable Linux root accounts.

- Do not compile or perform any other operations on the Linux kernel.
- The use of swap for partitioning is not recommended.
- Do not enable the NetWorkManager service. This service conflicts with the system's internal network service, causing network exceptions.

12.2.6 Restrictions on instance type families

Some instance type families have requirements on images, bandwidths, and related drivers.

Instance type family ga1

To create instance type family ga1, you need to use the following images pre-installed with drivers :

- Ubuntu16.04 pre-installed with the AMD GPU driver
- Windows Server 2016 Chinese Edition pre-installed with the AMD GPU driver
- Windows 2008 R2 Chinese Edition pre-installed with the AMD GPU driver

Notes:

- The ga1 instance uses a driver optimized by Alibaba Cloud and AMD. The driver is included in the image provided by Alibaba Cloud. Driver download link is not provided and driver installation by the client is not supported.
- If the GPU driver fails to work properly because its related components are detached or removed, you need to restore GPU-related functions by *Change System Disks*.

  Note:
  Changing system disks may cause data loss.
- If the driver of a visual compute ga1 instance with GPUs fails to work properly because an incorrect image is selected, you need to reselect an image pre-installed with the AMD GPU driver through *Change System Disks*.
- If you use an image of Windows 2008 or earlier versions, the function of **Connect to Management Terminal** on the Alibaba Cloud console is unavailable after the installed GPU driver takes effect. The management terminal is unresponsive with a black screen or stuck at the startup interface. You can use other protocols to access the system, such as the remote desktop protocol (RDP) of Windows.
- The RDP of Windows does not support DirectX, OpenGL, and other related applications. You need to install VNC and a client or configure other supported protocols, such as PCOIP and XenDeskop HDX 3D.

Instance type family gn4

- Bandwidth: Select a bandwidth.

  Note:
  If you use an image of Windows 2008 R2, the function of **Connect to Management Terminal** on the Alibaba Cloud console cannot be used to connect to the gn4 instance after the installed GPU driver takes effect. Therefore, you need to set the bandwidth to a non-zero value or bind the created instance to an elastic public IP.

- Image: Select an image.

  If pre-installation of the NVIDIA GPU driver is unnecessary, you can select any image, and *Download and install the GPU driver*.

Instance type families gn5i and gn5

- Bandwidth: Select a bandwidth.

  Note:
  If you use an image of Windows 2008 R2, the function of **Connect to Management Terminal** on the Alibaba Cloud console cannot be used to connect to the gn5i or gn5 instance after the installed GPU driver takes effect.

- Image: Select an image.

  If pre-installation of the NVIDIA GPU driver is unnecessary, you can select any image, and *Download and install the GPU driver*.

12.2.7 DDoS protection

To benefit from the DDoS protection capability, you need to purchase the Alibaba Cloud Security Advanced Edition. For details, see *Cite LeftAlibaba Cloud Security OverviewCite Right*.

12.3 Quick start

12.3.1 Log on to the ECS console

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.

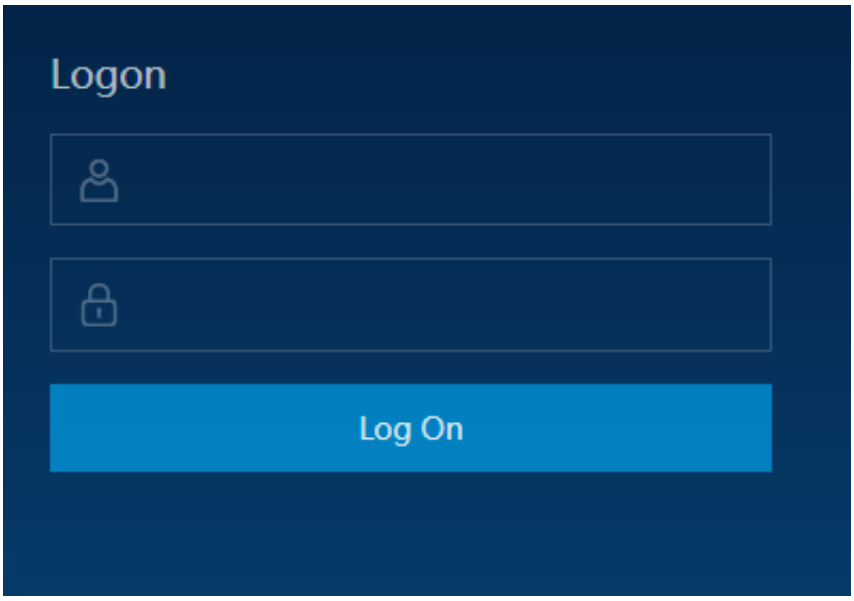
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 12-3: Log on to the Apsara Stack console*.

   Figure 12-3: Log on to the Apsara Stack console

3. Enter the correct username and password.

   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click **Log On** to go to the **Home** page.
5. In the menu bar at the top of the page, select **Console** > **Compute, Storage&Networking** > **Elastic Compute Service**.

12.3.2 Create a security group

Before creating an ECS instance in a VPC, you need to create a security group. The security group can be used to set network access control for a single or multiple ECS instances and is an important means of network security isolation.

Context

To create a security group, do the following:

Procedure

1. *Log on to the ECS console*.
2. Click the **Security Groups** tab and then click **Create Security Group**.
3. On the **Create Security Group** page, provide the following information and click **Confirm**.

Table 12-3: Security group configuration

Item	Description
Security Group Name	Required. Enter the name of a security group.
Region	Required. Select the region to which the security group belongs. It must be the same region to which the VPC belongs.
Department	Required. Select the department to which the security group belongs. It must be the same department to which the VPC belongs.
Description	Optional. Enter a description for the security group.
Project	Optional. Select a project to which you want to add the security group.
Network Type	Required. The parameter value is **VPC** by default.
VPC	Required. Select a VPC to which the security group belongs.

12.3.3 Create an instance

Prerequisites

The following should be noted before creating an insance:

- Before creating an instance, complete the creation of a VPC and switch.

- Before you create an instance, check that a security group is available. If not, *Create a security group* first.
- Before creating a GPU instance, refer to *Restrictions on instance type families*.

To create an instance, do the following:

Procedure

1. *Log on to the ECS console* and go to the **Instance** page.
2. Click **Create Instance**.
3. On the **Create Cloud Server (ECS)** page, complete the following configurations:

Table 12-4: Instance configuration

Item	Description
Region	• **`Region`**: select a region for the ECS instance to be created. • **`Zone`**: Zones are physical areas with independent power grids and networks within a region. Intranet communication and fault isolation are both enabled between different zones. If you want to improve application availability, we recommend that you create instances in different zones.
Configurations	• **`Department`**: select a department for the ECS instance. • **`Project`**: select a project for the ECS instance.
Network	• **`Network Type`**: It is a required parameter and set to VPC by default. Select a specific VPC and switch name. • **`Security Groups`**: It is a required parameter. **Note:** Before creating an ECS instance, you must create a security group. • **`Configure Private IP`**: It is an optional parameter. Determine an IP address segment based on the CIDR block where the switch is located. **Note:** • If it is null, the system will designate a private IP Address automatically. • If a private IP address is configured, instances cannot be created in batch.

Item	Description
Instances	• **`Instance Series`**: select **`Series 3`** or **`Series 4`** . • **`I/O Optimized`**: It is an optional parameter. By default, it is an I/O optimized instance. • **`Instance Specifications`**: It is an optional parameter. Select CPU and memory based on application requirements. Windows-based images are not applicable to some CPU and memory combinations. See *Suggestions*.
Images	**`Image Type`**: It is a required parameter. Select **`Public Image`** or **`Custom Image`** as the image type for your operating system.
Storage	• **`System Disk`**: It is a required parameter. Select an SSD cloud disk or ultra cloud disk as the system disk for installing the operating system. • **`Data Disk`**: It is an optional parameter. You can choose between SSD cloud disks and ultra cloud disks. Up to 16 data disks can be added. The maximum capacity of each data disk is 32 TB. You can select **Release with Instance** or **Encrypt**. **Note:** • If you check **`Release with Instance`**, this disk will be released together with the instance and the data is not recoverable; if Release with Instance is not checked, the system will detach this disk from the instance when this instance is released, and the disk will not be released. • If you check **`Encrypt`**, the disk created is an encrypted disk; if it is not checked, the disk created is a non-encrypted disk. • If you do not add data disks here, you can add them later by following the procedure described in *Create disks*.
Password	It is a required parameter. You can select **Configure Now** or **Configure After** to set your password on the console by using the password re-setting function. **Note:** **`Login Password`** can be 8 to 30 characters in length. It must contain at least three of the following character types: uppercase letters, lowercase letters, numbers, and special characters.
Deployment Set	Select the name of the deployment set to which the instance will be added.

Item	Description
Instance Name	It is an optional parameter. We recommend that you set a recognizable name for your instance.  **Note:** The instance name can be 2 to 114 characters in length and can contain letters, Chinese characters, numbers, underscores (_), and hyphens (-). It must start with a letter or Chinese character.
Custom Data	You can enter the corresponding custom data encoding schemes. If the data to be entered adopt Base64 encoding schemes, check **Enter Base64 encoded finformation**.  **Note:** Windows supports two formats: bat and powershell. Before Base64 codes, the first line is [bat] or [powershell]. Linux supports shell scripts.
Quantity	The default value is 1. If you apply for more than one instance, these instances are created in batch based on your current settings.  **Note:** You can create up to 50 ECS instances in batch. If a private IP address is configured, batch creation is not supported.

4. Click **Create**.

Result

You can view the created instance in the instance list. Check whether the instance created is in the **`Running`** state. If so, the instance is successfully created.

12.3.4 Connect to an instance

After creating an ECS instance, you can connect to the instance and install application software in the instance.

 **Note:**
The username is "root" in Linux and "administrator" in Windows.

You can connect to and manage your ECS instance in either of the following ways:

- Use a remote connection tool to connect to the ECS instance.

> **Note:**
> Check that your ECS instance has an elastic Internet IP address.

- Connect to the ECS instance by using the function of **Connect to Management Terminal** on the cloud console.

12.3.4.1 Connect to a Linux instance using the SSH command in the Linux or Mac OS X environment

Procedure

1. Enter the following command: `ssh root@instance IP`.
2. Enter the password of the *root* user for this instance upon logon.

12.3.4.2 Connect to a Linux instance using a remote connection tool in the Windows environment

Prerequisites

The usage of different remote connection tools is similar. This article describes how to connect to a remote instance through PuTTY. Download PuTTY at *http://www.chiark.greenend.org.uk/~sgtatham/putty/* .

To connect to an instance, do the following:

Procedure

1. Download and install PuTTY for Windows.
2. Start the PuTTY client and complete the following settings:
   - **`Host Name (or IP address)`**: enter the EIP address for the instance.
   - **`Port`**: Set it to the default port number 22.
   - **`Connection Type`**: select **`SSH`**.
   - **`Saved Session`**: the name of the session. Click **Save**. After the settings are saved, PuTTY remembers the name and IP address of the instance. You do not have to enter the IP address every time you connect to the instance.
3. Click **Open** to connect to the instance.

   When you connect to the instance for the first time, a **PuTTY Security Alert** dialog box appears. Click **Yes**.
4. Enter the username **`root`** and press the **Enter** key.

5. Enter the password of your instance and press the **Enter** key.

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

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

12.3.4.3 Connect to a Windows instance using the remote desktop connection function in the Windows environment

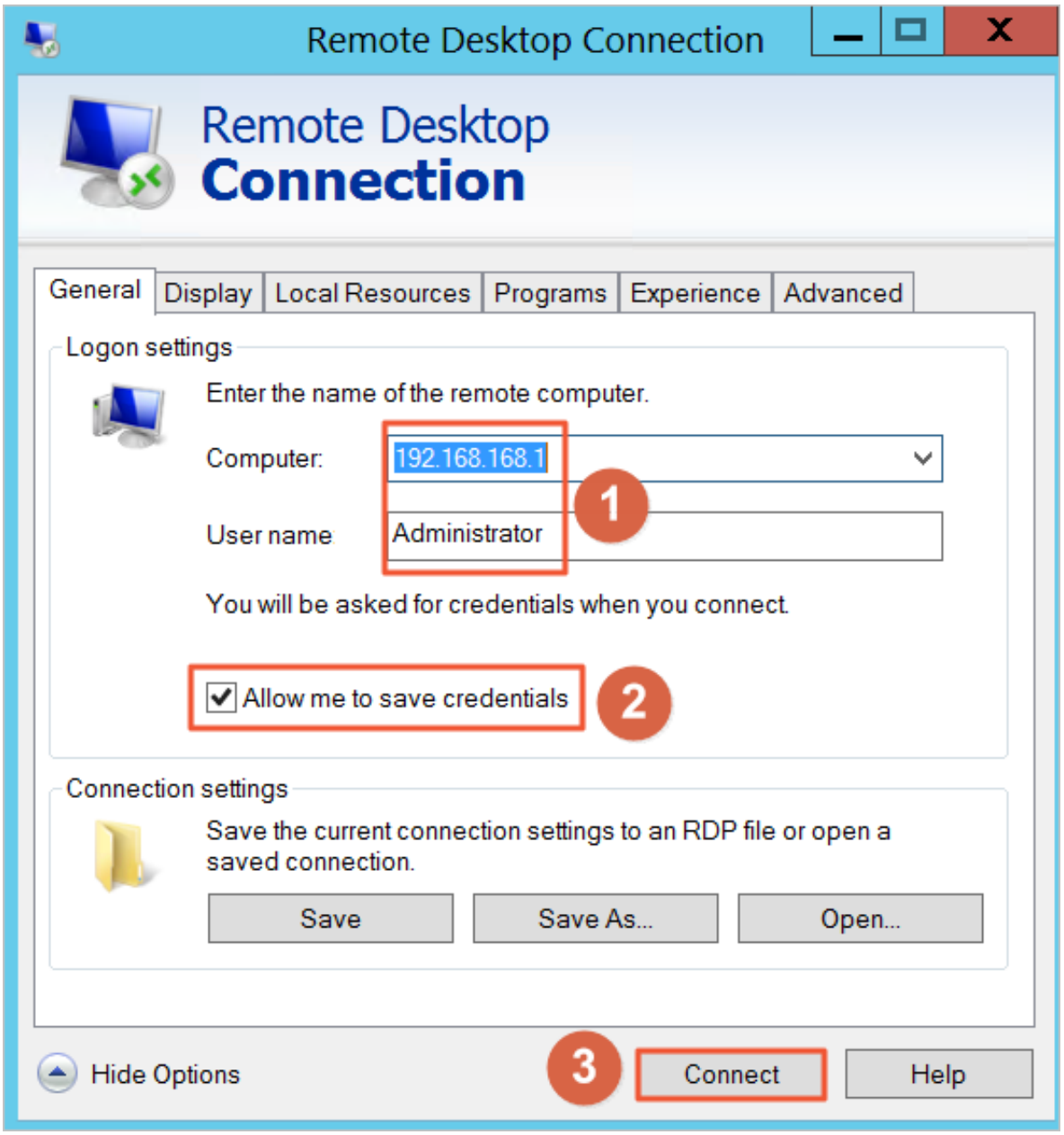
This article describes how to connect to a Windows instance through the remote desktop connection function of Windows.

Context

To connect to an instance, do the following:

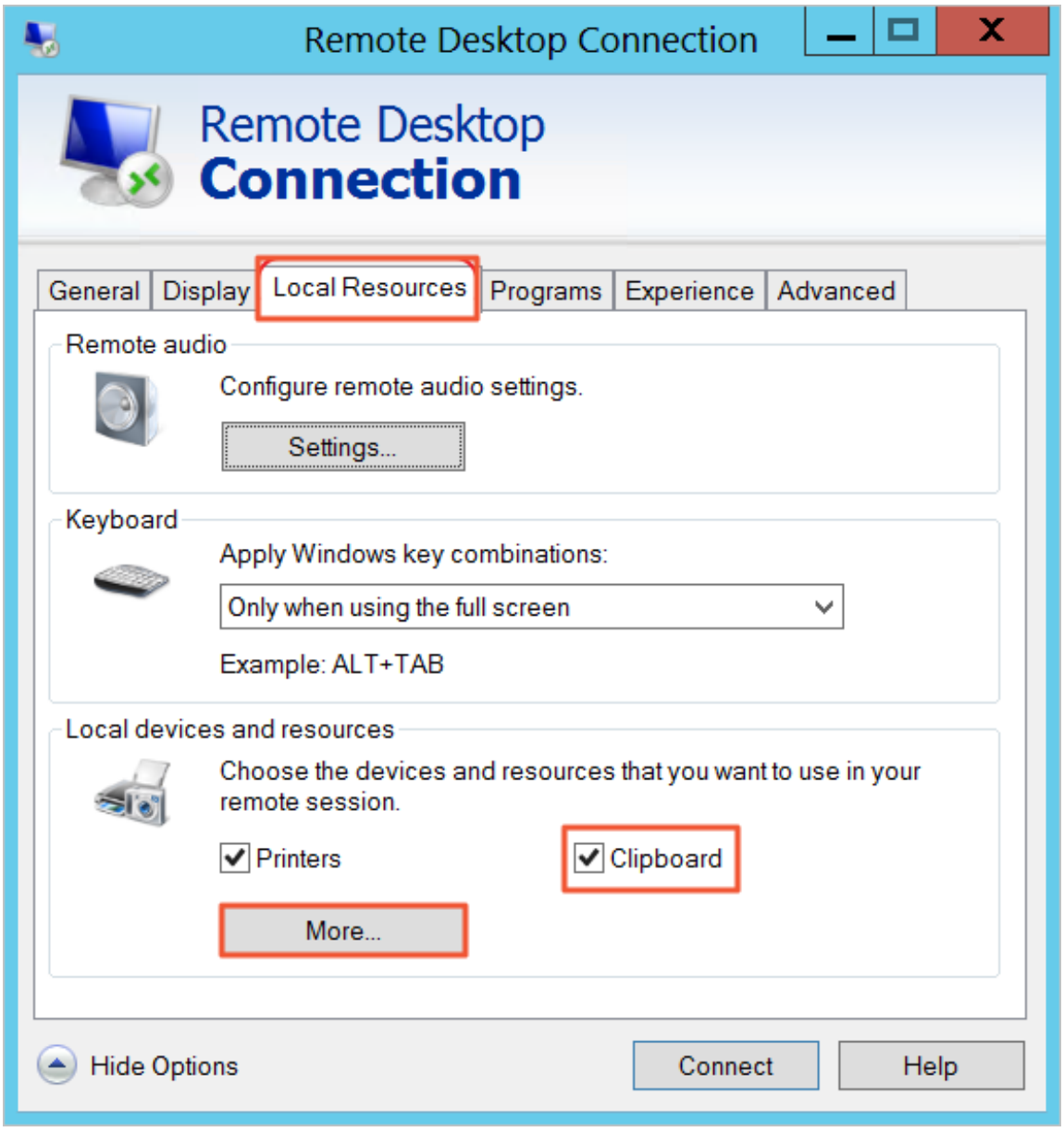
Procedure

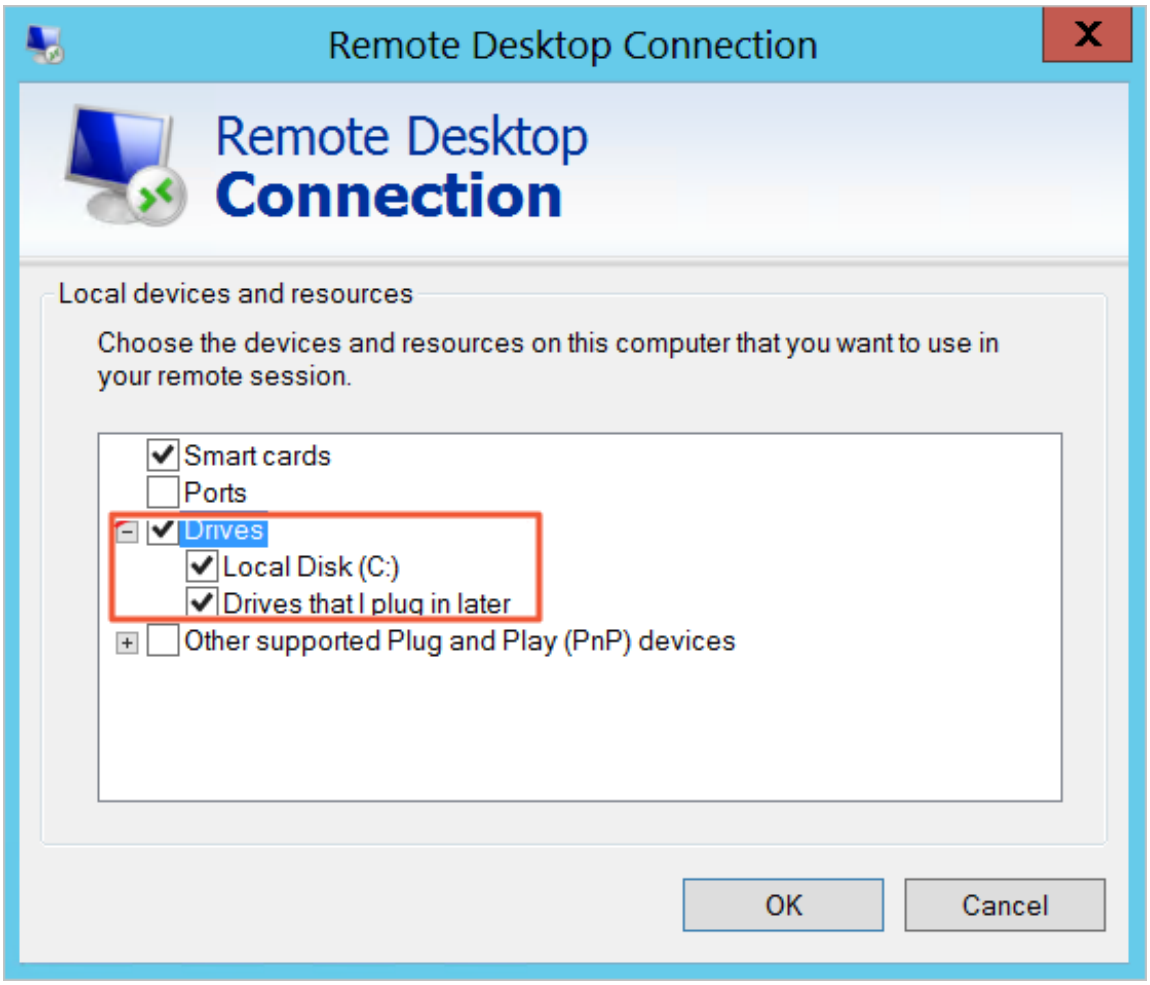
1. Start the remote desktop connection function in any of the following ways:
   - Click **Start**, enter **Remote Desktop Connection** in the search box, and click **Remote Desktop Connection** in the list that appears.
   - Enter **`mstsc`** in the search box and click **mstsc** in the list that appears.
   - Press the shortcut key **`Windows key+R`** to bring up the **Start** dialog box, enter **`mstsc`**, and press the Enter key to start the remote desktop connection function.
2. In the **Remote Desktop Connection** dialog box, enter the EIP address of the instance and click **Show Options (O)**.
3. Enter the username. The default value is **`Administrator`**. If you do not want to enter the password upon subsequent logon, select **Allow me to save credentials (R)**. After completing the settings, click **Connect** to connect to the instance.



You can also complete the following settings before connecting to the instance.

- If you want to copy local text to the instance, click the **Local Resource** tab and select **Clipboard**.
- If you want to copy a local file to the instance, click the **Local Resource** tab and then **Details**. Select **Drivers** and then the drive letter of the data disk where the file is stored. After completing the settings, click **OK**.


Remote Desktop Connection
Remote Desktop
Connection
General
Display
Local Resources
Programs
Experience
Advanced
Remote audio
Configure remote audio settings.
Settings...
Keyboard
Apply Windows key combinations:
Only when using the full screen
Example: ALT+TAB
Local devices and resources
Choose the devices and resources that you want to use in your remote session.
Printers
Clipboard
More...
Hide Options
Connect
Help



- You can click the **Display** tab to adjust the size of the remote desktop. Normally you can use full screen mode.

4. In the dialog box that appears, enter the password of the **`Administrator`** account of the Windows instance and click **OK** to connect to the instance.

Result

If the **Remote Desktop Connection** window displays a Windows desktop, a connection is successfully established to the instance.

12.3.4.4 Connect to an instance by logging on to VNC on the cloud console

When an ordinary remote connection tool (for example, PuTTY, Xshell, and SecureCRT) is not usable, you can use the **Connect to Management Terminal** function on the cloud console to connect to an ECS instance.

Prerequisites

- To use VNC, you must import the root certificate to the web browser. For more information, see *Install a certificate*.
- Before logging on to the VNC, complete the operation of *Change the VNC password*.

> **Note:**
> The VNC password is used to connect the management terminal of the console, and the password of the instance is used to log on to the instance. Pay attention to the difference between them.

Context

The function of **Connect to Management Terminal** can be used in the following scenarios:

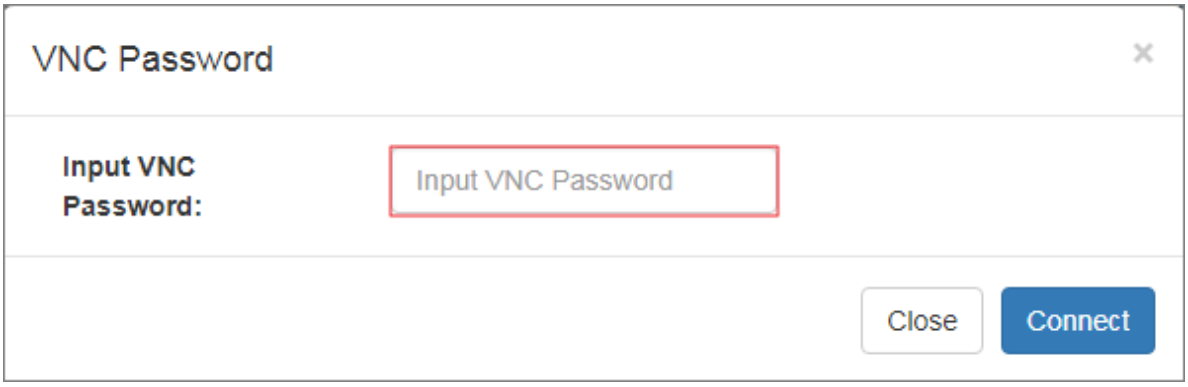
- Check the progress when the instance boot speed is slow (for example, when self-check is initiated).
- The connection to your instance with remote connection software fails due to incorrect instance settings. For example, if the firewall has been enabled due to a misoperation, you can use the **Connect to Management Terminal** function to connect to your instance and then disable the firewall.
- The connection to your instance fails due to high CPU or bandwidth consumption by applications (such as CPU or bandwidth fully occupied by processes due to a botnet attack). In this case, you can use the **Connect to Management Terminal** function to connect to your ECS instance and end exceptional processes.

To connect to an instance, do the following:

Procedure

1. *Log on to the ECS console* and go to the **Instance** page.
2. In the **Action** column of the target instance, click the icon and select **View Details** from the drop-down list.
3. On the **Instance Details** page, click **Connect to Management Terminal**.
4. In the dialog box that appears, enter your VNC password and click **Connect**.

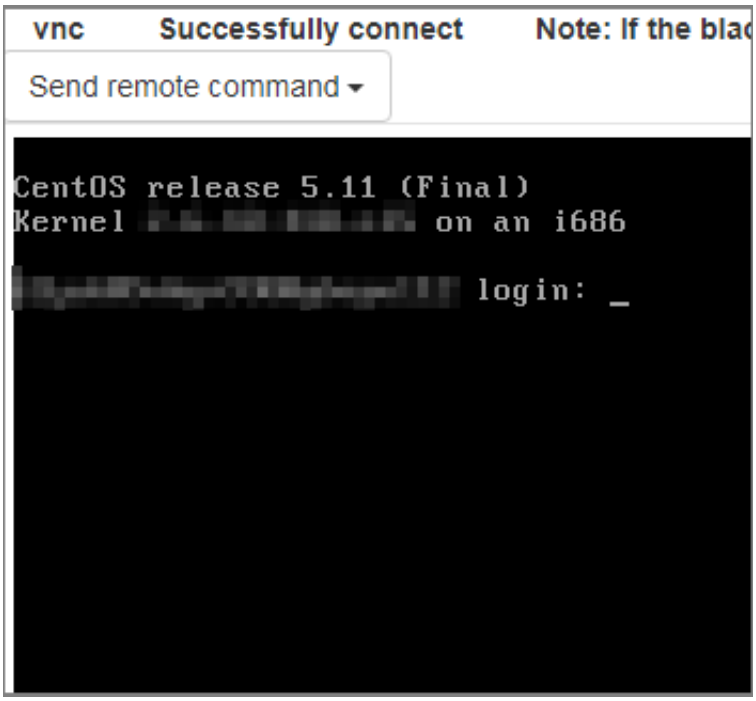
Figure 12-4: Connect to VNC



5. After the connection is established, the system will display the logon page.

   Take the Linux system as an example. When the instance is connected, you can view the message below: *Figure 12-5: Log on to a Linux instance*.

Figure 12-5: Log on to a Linux instance



6. Enter the username and password and then log on to the instance.

   - If your instance runs on a Linux operating system, you must enter the username *root* and logon password for the instance.
   - If your instance runs on a Windows operating system, you must enter the username *administrator* and logon password for the instance.

Note:

For a Linux operating system, the password entering process will not be displayed. In this case, enter the password and press **Enter**.

12.4 Instance management

An ECS instance is the minimal unit that can provide compute services for your business. It provides computing capabilities of specific specifications.

12.4.1 View an instance

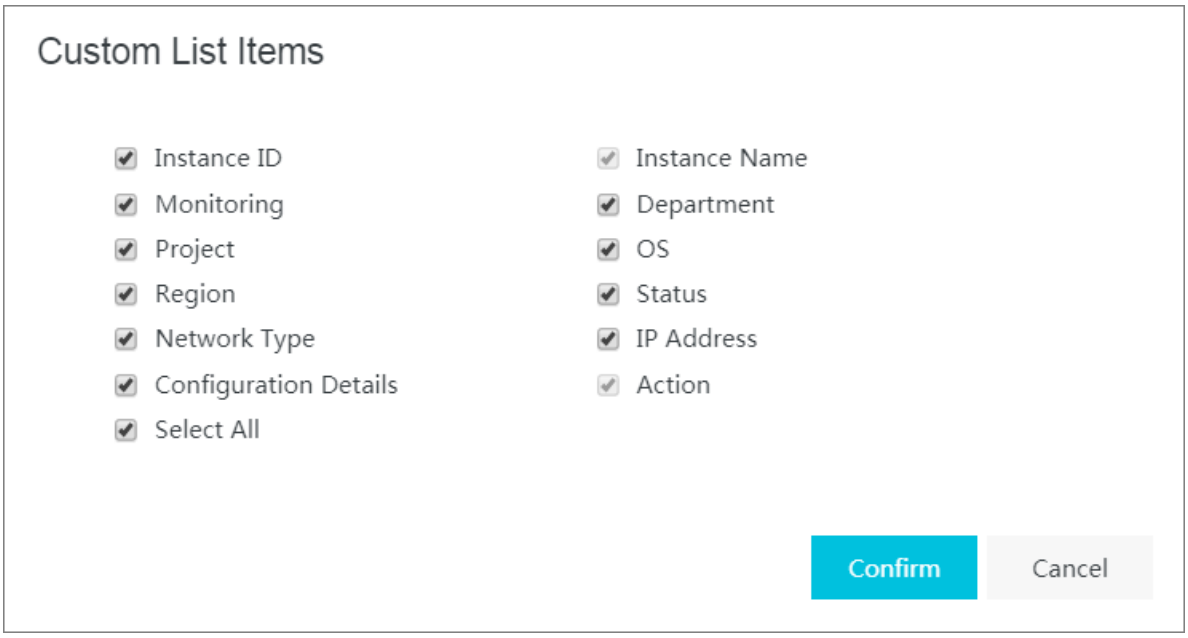
You can log on to the ECS console to view all your instances and their details.

Procedure

1. *Log on to the ECS console*.

   On the upper-right corner of the **Instances** page, click **Set**, select the items to be displayed in the **Custom List Items** box, and click **Confirm**. As shown in *Figure 12-6: Set the custom list*.

   Figure 12-6: Set the custom list



2. On the **Instances** page, select a **`Department`** and a **`Region`** or enter an **`Instance Name`** and click **Search** to find the target instance.

Note:

Click **Instance Name** and you can choose other filtering conditions from the drop-down menu: **Instance ID**, **Instance Status**, **VPC ID**、**IP Address**, and **Project Name**.

3. In the **Action** column of the instance, click the icon and choose **View Details** to enter the **Instance Details** page and view the details of the instance.

> **Note:**
> On the **Instances** page, you can also click an instance ID to go to the **Instance Details** page.

12.4.2 Edit an instance

You can modify the name and description of an existing instance on the ECS console.

Procedure

1. *Log on to the ECS console* and find the ECS instance you want to edit.
2. In the **Action** column of the target instance, click the icon and select **Edit** from the drop-down list.
3. In the dialog box that appears, edit the **Name**, **Description**, and **Custom Data** of the instance and click **Confirm**.

> **Note:**
> For the custom data, Windows supports two formats: bat and powershell. Before a Base64 code, the first line is [bat] or [powershell]. Linux supports shell scripts.

12.4.3 Start, stop, or reboot an instance

On the ECS Console, you can start, stop, or reboot an instance just like on a real server.

Procedure

1. *Log on to the ECS console*.
2. On the **Instances** page, in the **Action** column of the corresponding instance, click and select **Reboot**, **Stop**, or **Start** from the drop-down list.
   - You can stop or restart an instance only when the instance is in the **Running** state. You can start an instance only when the instance is in the **Stopped** state.

> **Note:**
> The Stop and Restart operations will stop your instance and interrupt your business. Therefore, exercise caution when performing these operations.

- You can click the icon behind a corresponding instance and choose **View Details** on the drop-down menu. On the **Instance Details** page, click **Reboot**, **Stop** or **Start** on the upper-right corner of the page to restart, stop, or start that instance.

12.4.4 Delete an instance

You can manually delete unnecessary instances.

Prerequisites

The instances to be deleted are in the **Stopped** state.

Procedure

1. *Log on to the ECS console* and go to the **Instances** page.
2. Identify the ECS instance to be deleted. In the **Action** column of the instance, click the icon and select **Delete** from the drop-down list and click **OK** on the dialog box popped up.

> **Note:**
> Alternatively, in the **Action** column of the instance, click the icon and choose **View Details**. On the **Instance Details** page, click **Delete** and then click **OK** in the dialog box popped up to delete the instance.

12.4.5 Modify configurations

You can modify the CPU and memory specifications of an instance by changing its configurations.

Procedure

1. *Log on to the ECS console*, go to the **Instances** page, and find the instance whose configurations need to be changed.
2. In the **Action** column of the instance, click the icon, choose **Change Configurations** from the drop-down menu, select new CPU and memory specifications, and then click **OK**.

> **Note:**
> After the configurations of the instance are modified, *Restart the instance* on the console for the new configurations to take effect.

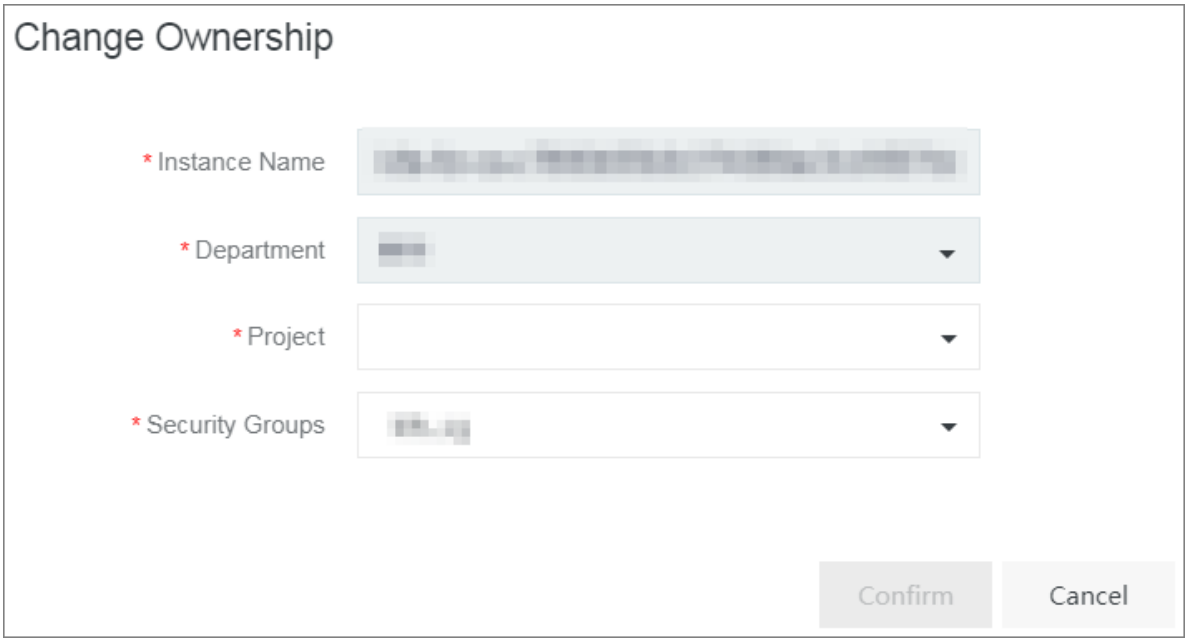
12.4.6 Change ownership

The ownership change function allows you to change the department and project for an instance.

Procedure

1. *Log on to the ECS console* and find the ECS instance whose ownership you want to change.
2. In the **Action** column of the instance, click the icon and select **Change Ownership** from the drop-down list.
3. In the dialog box for **Change Ownership**, select new **Department** and **Project** and then click **Confirm**. See *Figure 12-7: Change ownership*.

Figure 12-7: Change ownership

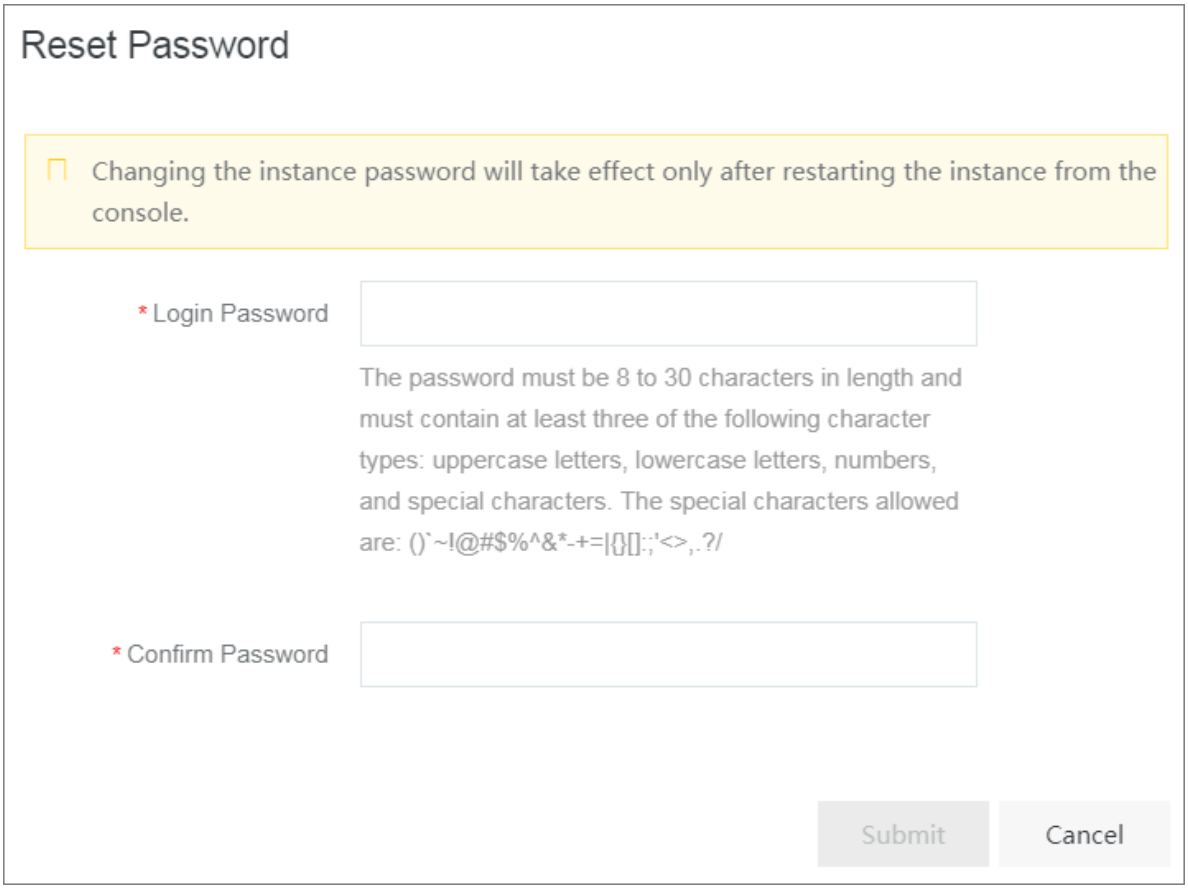


12.4.7 Change the ECS instance log-on password

Procedure

1. *Log on to the ECS console* and find the ECS instance whose password you want to change. .
2. In the **Action** column of the instance, click the icon and select **View Details** from the drop-down list.
3. On the **Instance Details** page, click **Change Password**.
4. In the dialog box that appears, enter a new **Login Password** and **Confirm Password**. Then click **Submit**. See *Figure 12-8: Reset your password*.

Figure 12-8: Reset your password



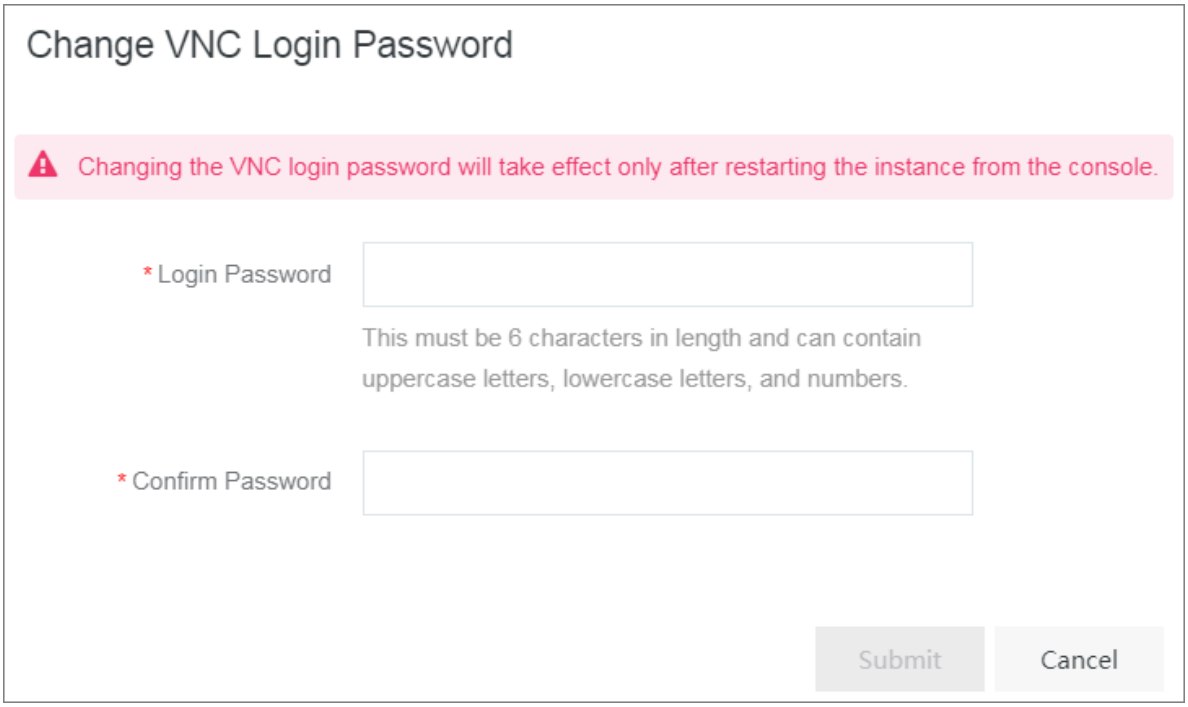
5. *Reboot the instance* on the console for the new instance password to take effect.

12.4.8 Change the VNC password

Procedure

1. *Log on to the ECS console* and find the ECS instance whose VNC password you want to change.
2. In the **Action** column of the instance, click the icon and choose **View Details** from the drop-down menu.
3. On the **Instance Details** page, click **Change Management Terminal Password**.
4. In the dialog box that appears, enter a new **`Login Password`** and **`Confirm Password`**. Then click **Submit**. As shown in *Figure 12-9: Change the VNC password*.

Figure 12-9: Change the VNC password



Note:

Reboot the instance on the console for the new password of the VNC to take effect.

12.4.9 Join a security group

Method 1: Join a security group on the instance page

1. *Log on to the ECS console*.
2. Enter the **Instances** page. In the **Action** column of an instance, click the icon and select **View Details** from the drop-down list.
3. In the top navigation bar, click the **Instance Security Group** tab, and click **Join Security Group**.
4. In the **Move to Security Group**dialog box that appears, select the target security group and click **Confirm**.

Method 2: Add an instance to a security group on the security group page

You can also add an instance to a security group in the following procedure:

1. *Log on to the ECS console* and go to the **Security Groups** page.
2. Click the security group ID to go to the **ECS Instances** page.

> **Note:**
> Alternatively, in the **Action** column of a security group, click the icon and select **View Details** to enter the **ECS Instances** page.

3. In the upper-right corner of the **ECS Instances** page, click **Import ECS Instances**.
4. In the **Move to Security Group** dialog box, select an instance and click **Confirm**.

12.4.10 UserData

Provided by Alibaba Cloud, the UserData function allows you to customize the startup behavior of an ECS instance and to pass data into an ECS instance.

Context

This function works on an ECS instance running in either Windows or Linux. UserData can be categorized into:

- User-defined scripts, which can be passed into an instance for running when the instance is started.
- General-purpose data, which can be passed into an instance that may reference such data.

Usage instructions of user-defined data

- **Operating environment**

  Only instances that meet all the following conditions can use user-defined data:

  — Network type: VPC.
  — Image: a system image or a custom image inherited from a system image.
  — Operating system: One operating system listed in *Table 12-5: Supported types* is used.

Table 12-5: Supported types

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

— To pass user-defined scripts, enter them according to the operating system and related script type.

> **Note:**
> Only English letters are allowed, and no unnecessary characters are allowed.

■ If the scripts have been Base64 encoded, you must select **The input has been base64 encoded** when entering the user-defined scripts for an instance.

> **Note:**
> The size of the scripts must not exceed 16 KB before Base64 encoding.

■ For Linux instances: The scripts must be entered in the format of *Script types for Linux instances*.

■ For Windows instances: The scripts must use `[bat]` or `[powershell]` as the first line.

- After instances are started, run commands to check:

  — Execution result of the user-defined scripts

  — Passed general-purpose data

- **Console**: You can edit the user-defined data in the console. If you are editing user-defined scripts, the script type determines whether the scripts should be run again after modification. For example, for a Linux instance, if you modified scripts of `bootcmd` type in the **`Cloud Config`**, the scripts are automatically run every time the instance is restarted.
- **OpenAPI**: You can use user-defined data by using OpenAPI. For more information, see **CreateInstance** and **ModifyInstanceAttribute** in *Cite LeftECS Developer GuideCite Right*.

User-defined scripts for Linux instances

The open-source cloud-init package is used by Alibaba Cloud to implement the user-defined scripts for Linux instances and configure the instances automatically. Alibaba Cloud Linux instances adopt the script types that come with the cloud-init package.

Instructions of user-defined scripts for Linux instances

- The user-defined scripts for Linux instances are run when the instance is running and before `/etc/init` is run.
- By default, user-defined scripts for Linux instances are run with the **root** privilege.

Script types for Linux instances

- **User-Data Script**

- Description: The script is used as the means for customized configuration, such as the shell script.
- Format: Start the first line with `#!`, such as `#!/bin/sh`.
- Limits: Before Base64 encoding, the scripts, including the first line, must not exceed 16 KB.
- Execution frequency: The scripts are run only when the instance is started for the first time.
- Example:

```
#!/bin/sh
```

```
echo "Hello World.  The time is now $(date -R)!" | tee /root/
output10.txt
```

- **Cloud Config Data**
  - Description: a method pre-defined in the cloud-init package to configure some services of the instance, such as the yum repository and SSH key pair.
  - Format: The first line must be `#cloud-config`.
  - Limits: Before Base64 encoding, the scripts, including the first line, must not exceed 16 KB.
  - Execution frequency: The script is run at a frequency that is determined by the services to be configured.
  - Example:

```
#cloud-config
```

```
apt:
```

```
primary:
```

```
- arches: [default]
```

```
uri: http://us.archive.ubuntu.com/ubuntu/
```

- **Include**
  - Description: Save the specific configuration in the form of text file and pass the file into cloud -init in the form of URL for processing.
  - Format: The first line must be `#include`.
  - Limits: Before Base64 encoding, the scripts, including the first line, must not exceed 16 KB.
  - Execution frequency: The script is run at a frequency that is determined by the script type in the actual text file.

- Example:

```
#include
```

```
http://ecs-image-test.oss-cn-hangzhou.aliyuncs.com/userdata/
cloudconfig
```

- **Gzip compressed content**
  - Description: The cloud-init has a limit of 16 KB on the contents of various user-defined scripts, so you may need to compress a script file before passing it into an instance.
  - Format: In the format of .gz file. It is passed into the instance in the form of `#include` URL.
  - Execution frequency: The script is run at a frequency that is determined by the compressed content.
  - Example:

```
#include
```

```
http://ecs-image-test.oss-cn-hangzhou.aliyuncs.com/userdata/config
.gz
```

Check user-defined data of Linux instances

To check the user-defined data of a Linux instance, run the following command within the instance :

```
curl http://100.100.100.200/latest/user-data
```

User-defined scripts for Windows instances

Developed by Alibaba Cloud, user-defined scripts for Windows instances allow you to customize the startup behavior of Windows ECS instances.

Two types of user-defined scripts are available for Windows instances.

- Batch script: Use `[bat]` as the first line. Before Base64 encoding, the content of the script must not exceed 16 KB.
- PowerShell script: Use `[powershell]` as the first line. Before Base64 encoding, the content of the script must not exceed 16 KB.

Check user-defined data of Windows instances

To check the user-defined data of a Windows instance, run the following PowerShell command within the instance:

```
Invoke-RestMethod http://100.100.100.200/latest/user-data/
```

12.4.11 Change private IP

Each instance is assigned with a private network adaptor and bound to a private IP address. The private IP address is on the IP address segment of the switch. Before you change the private IP address of an instance, make sure that the instance is in the **Stopped** state.

Procedure

1. *Log on to the ECS console* and find the ECS instance whose private IP address you want to change.
2. *Stop the instance*.
3. In the **`Action`** column of the instance, click the icon and select **Change Private IP** from the drop-down list.
4. In the **Change Private IP**dialog box that appears, enter a new **`Private IP`** and click **Confirm**.

12.4.12 Install a certificate

Before you log on to a VNC, export the certificate from your site and install it in your local web browser.

Procedure

1. Access the target Apsara Stack domain and press **F12** or **Fn+F12** to view the certificate on the displayed page, as shown in the following figures.

Figure 12-10: View a certificate

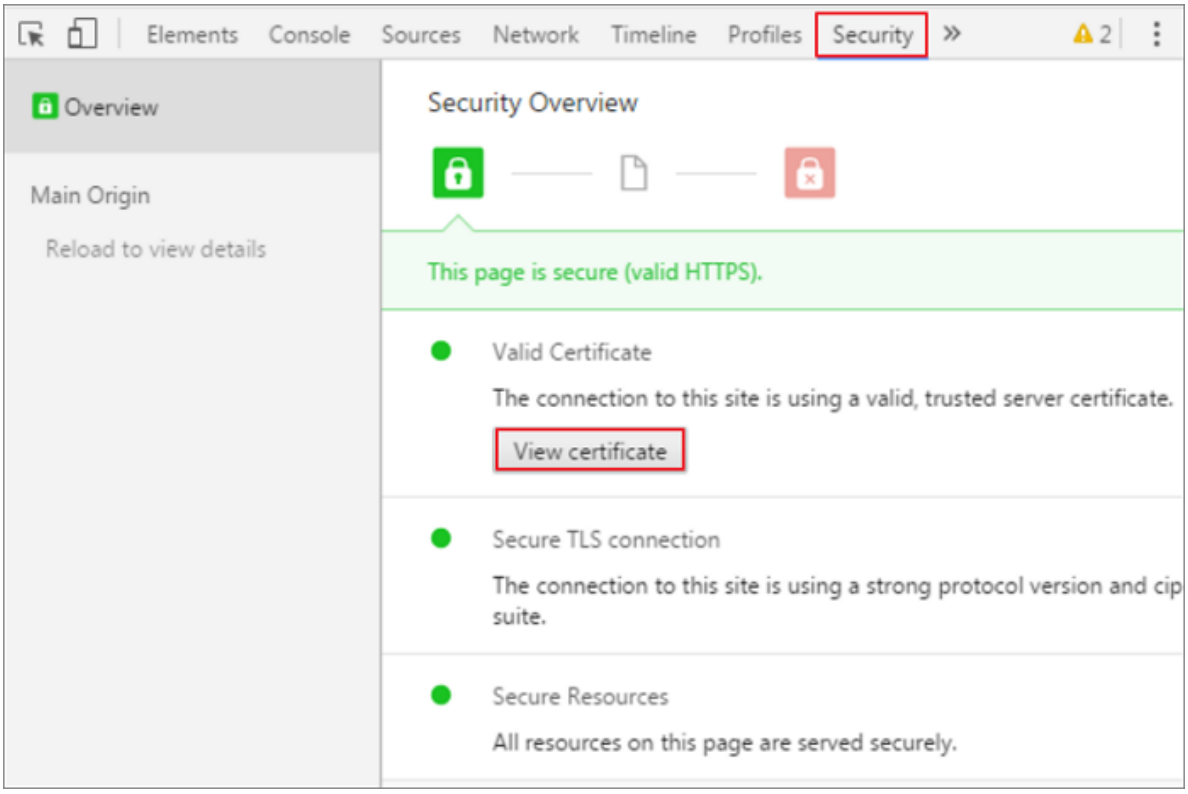
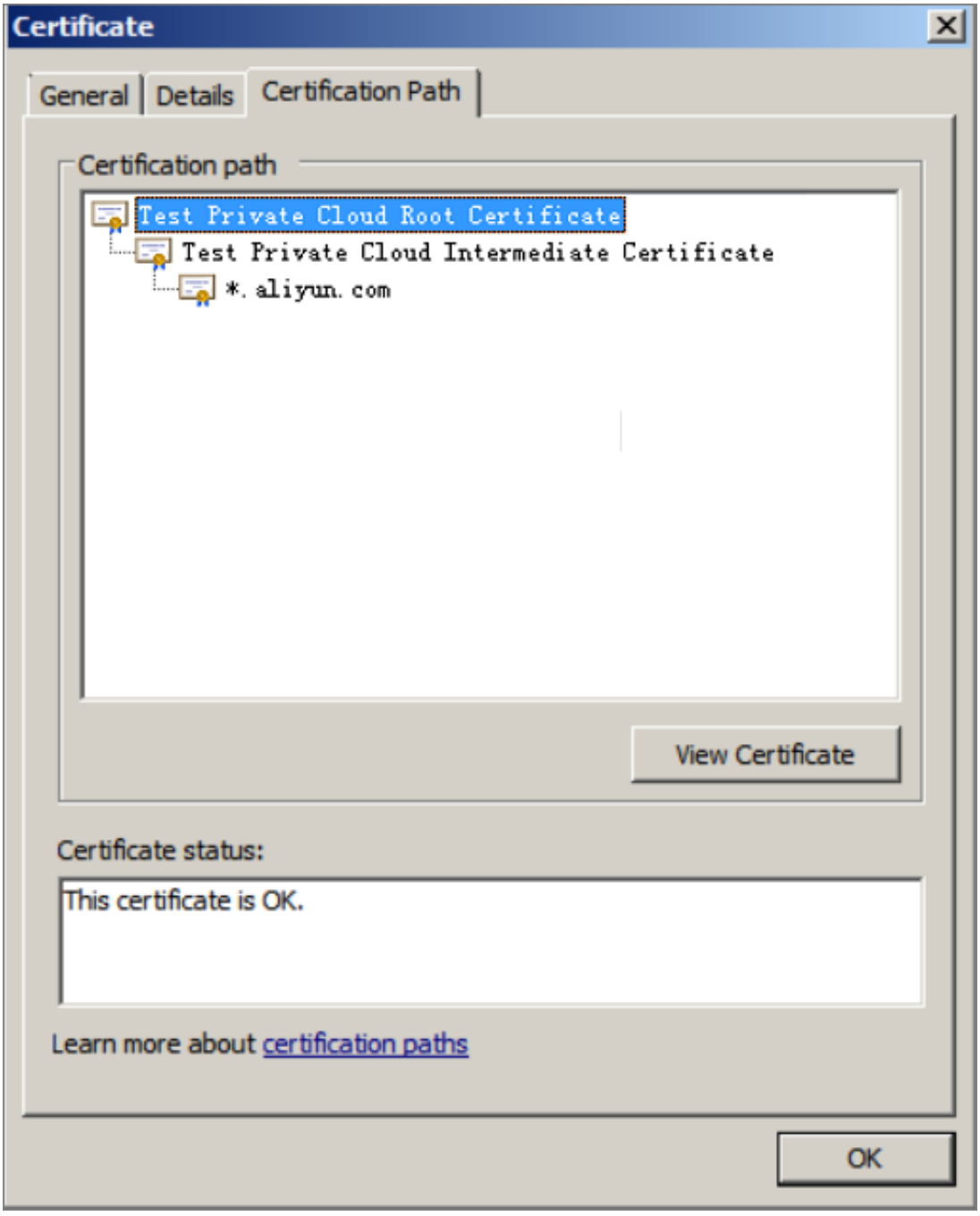


Figure 12-11: Select a certificate



2. On the certificate page, click **Copy to File**. Enter a name for the certificate and save the certificate to your local computer, as shown in the following figures.

Figure 12-12: Copy the certificate to a file

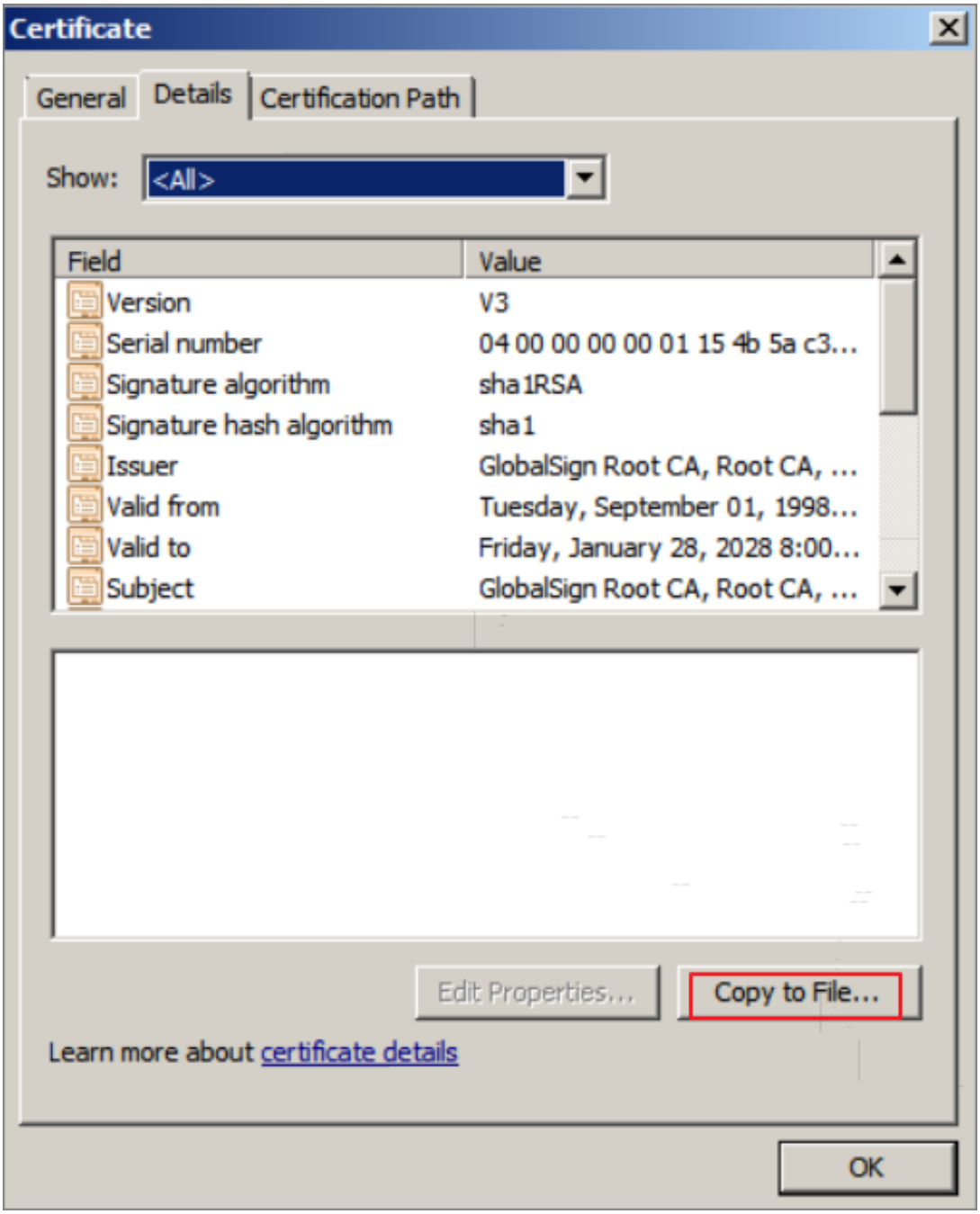


Figure 12-13: Certificate Export Wizard

Figure 12-14: Select the format of the exported file

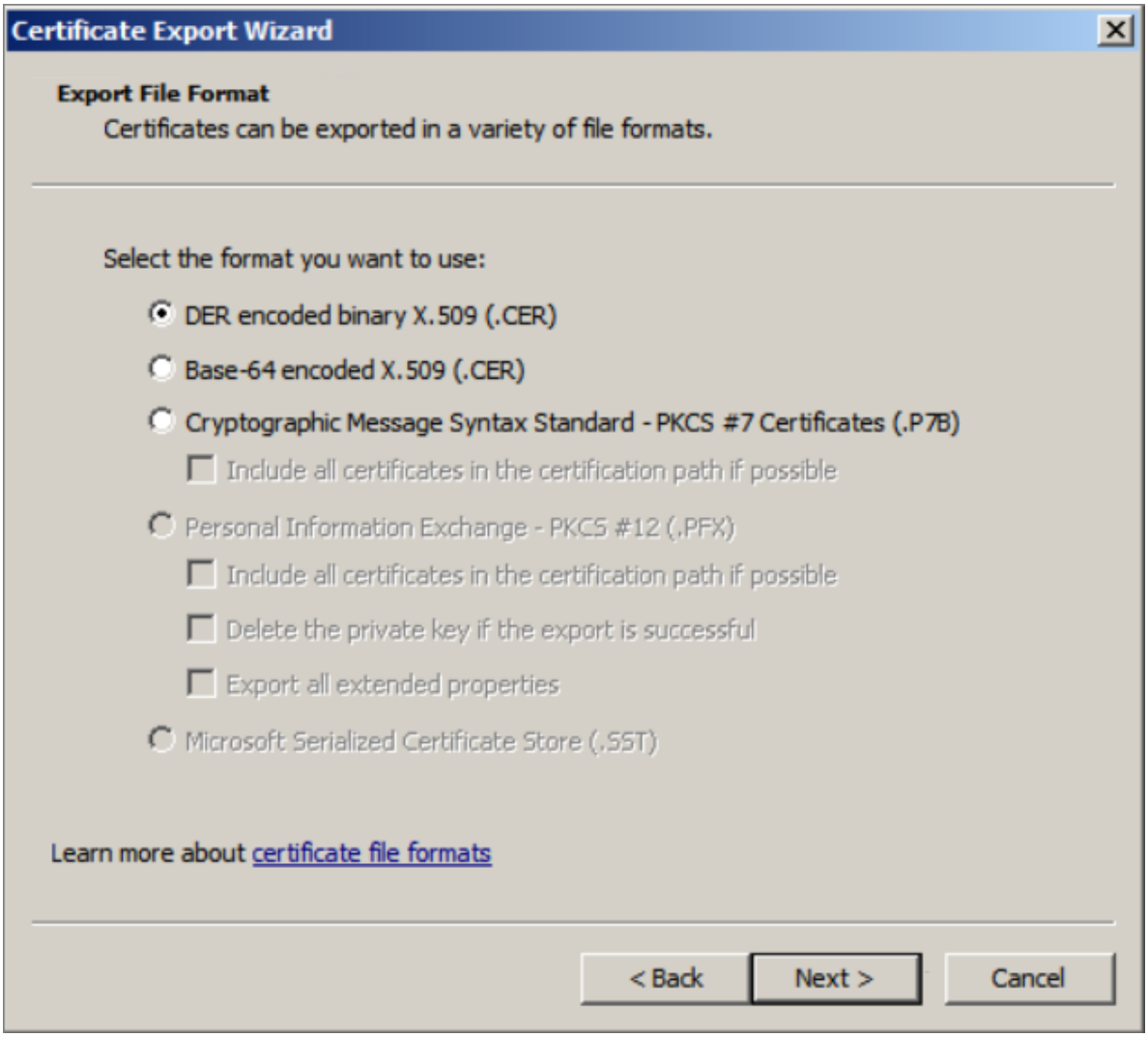


Figure 12-15: Select the file to be exported

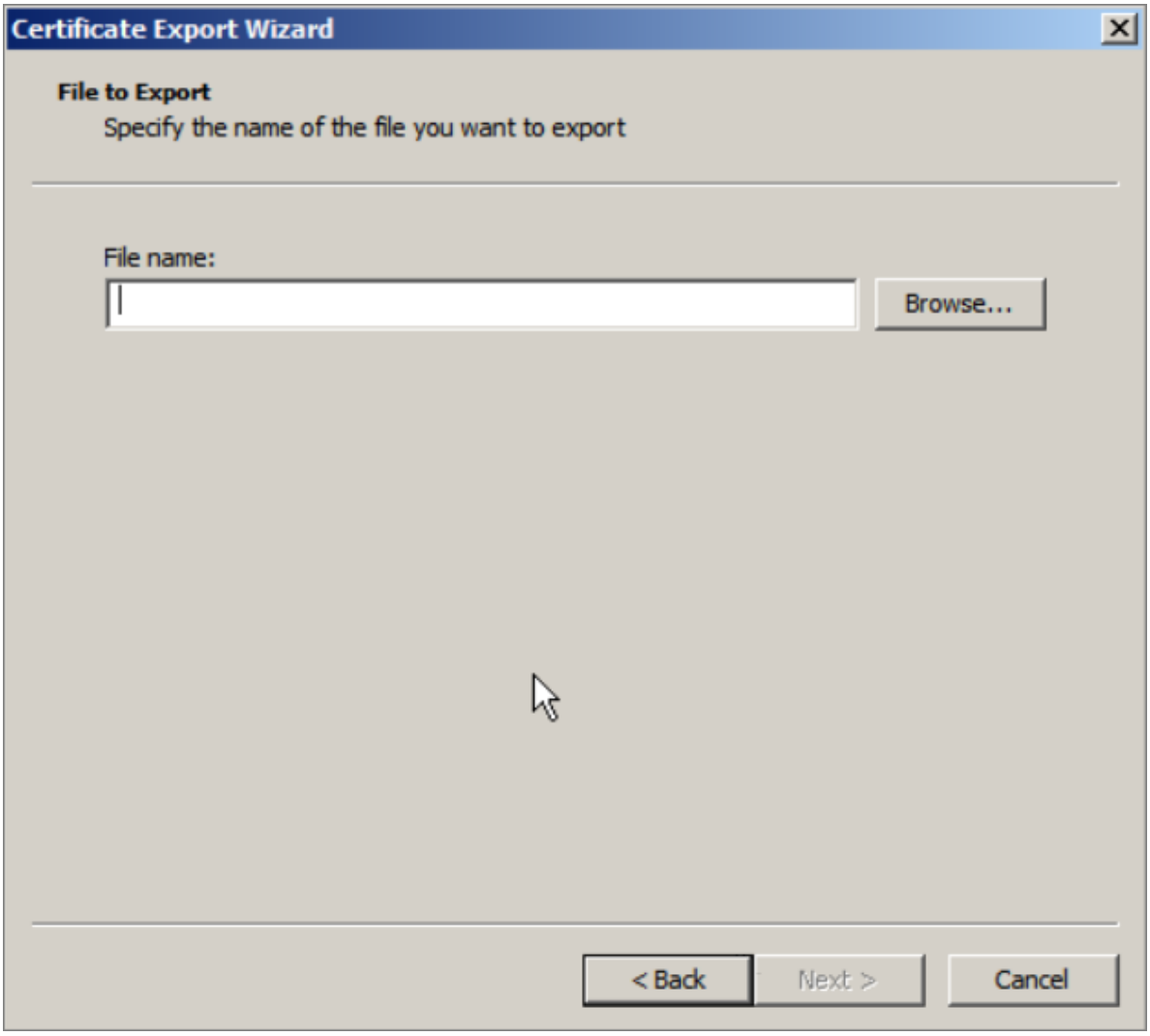
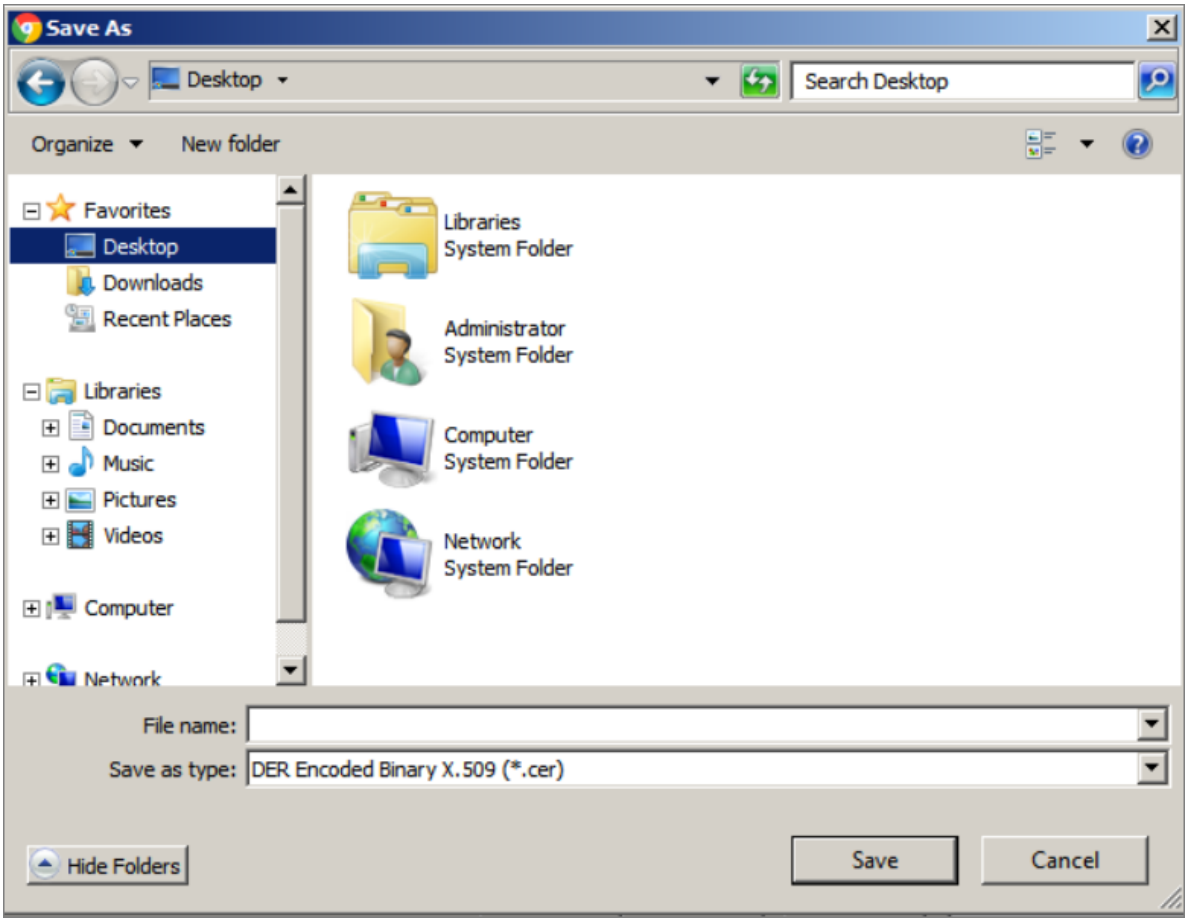


Figure 12-16: Enter a name for the certificate

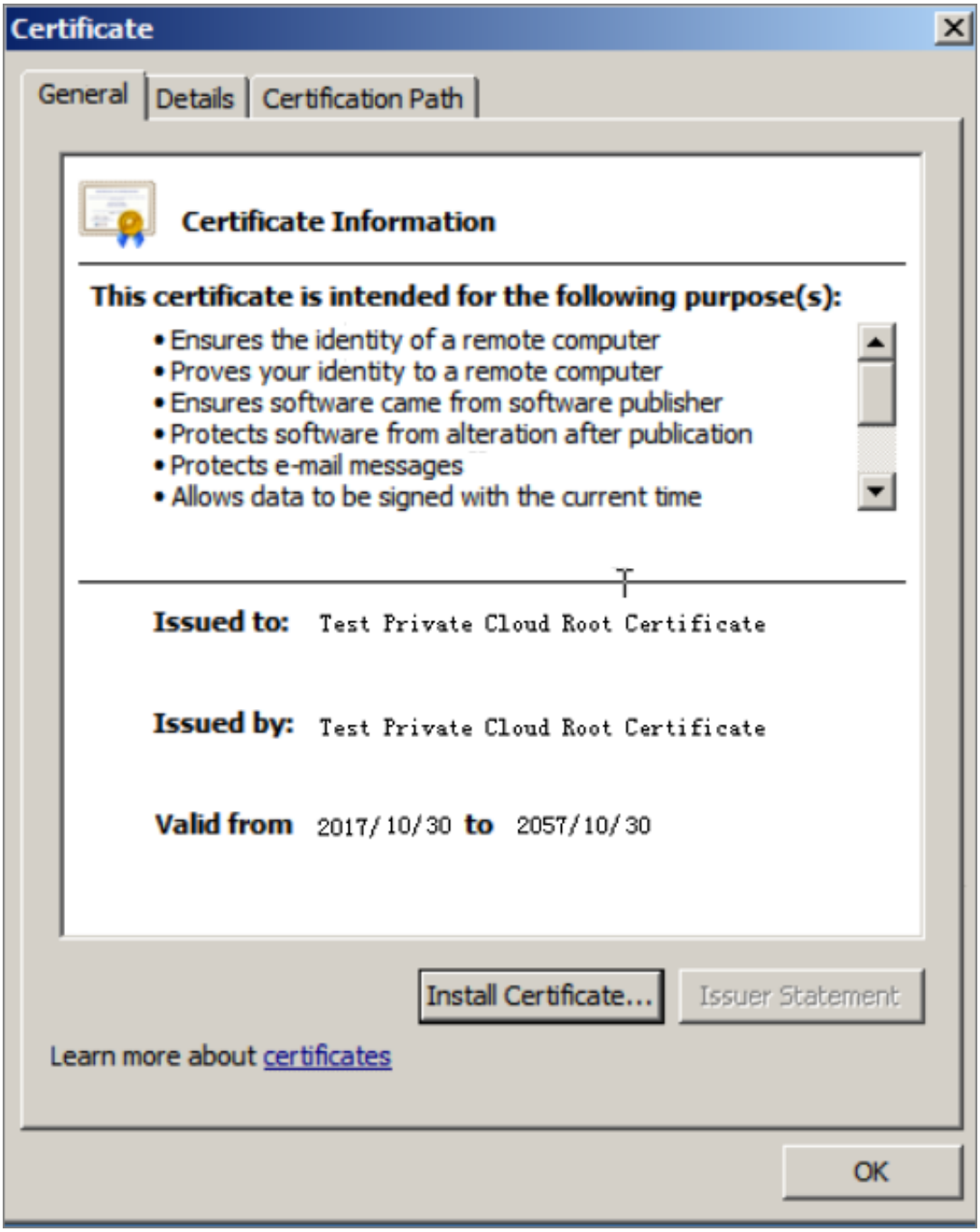


3. Click **Finish**. A prompt box appears, indicating that the certificate is successfully exported.

certificate

4. Double-click the locally stored certificate. On the page shown in the following figure, click **Install Certificate**.

Figure 12-17: Install a certificate



5. On the page that appears, select **Put All Certificates to Following Storage** and click **Browse**. Then select **Trusted Root Certificate Authority** and click **OK**.

Figure 12-18: Store certificates

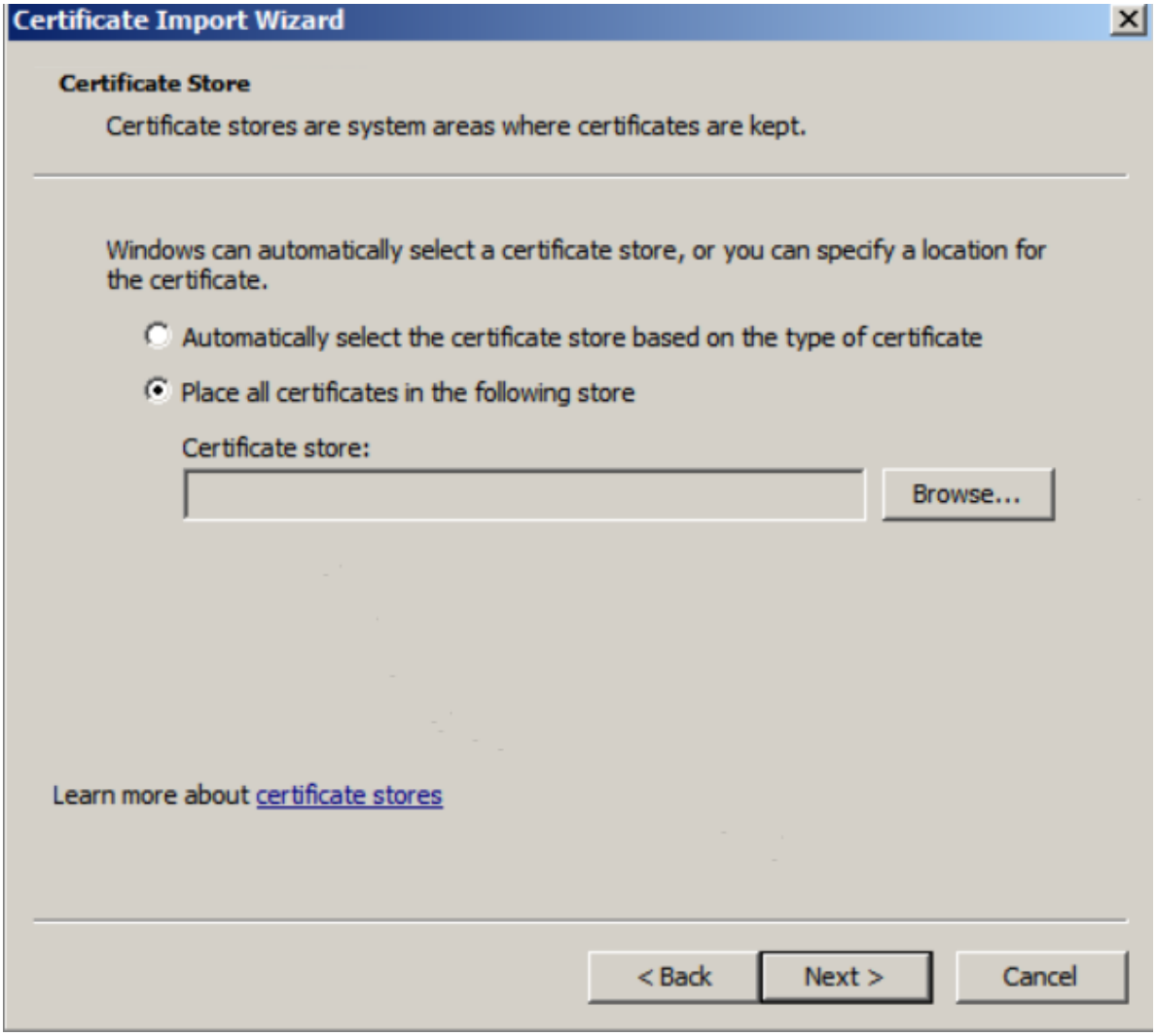
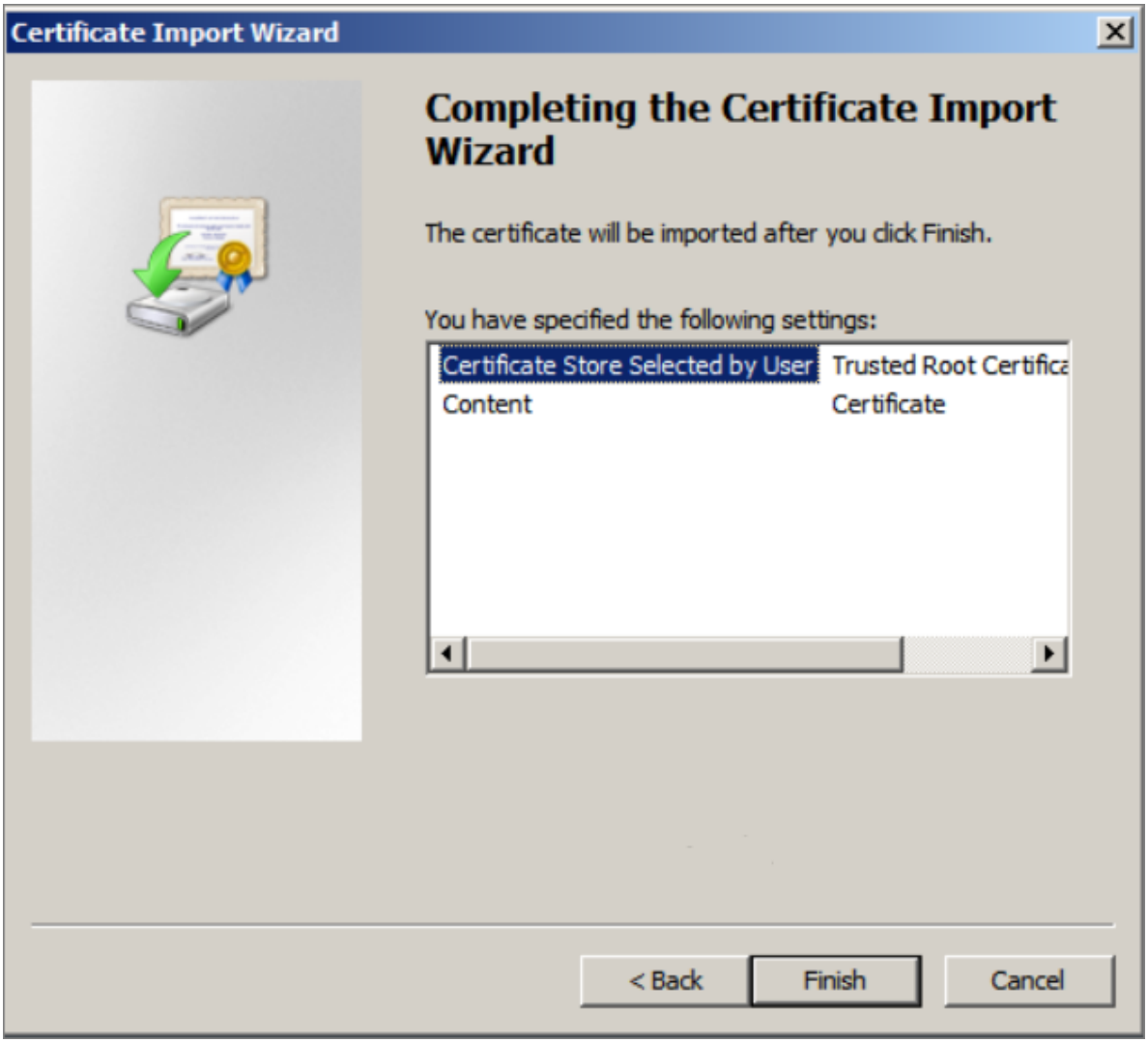


Figure 12-19: Certificate storage location

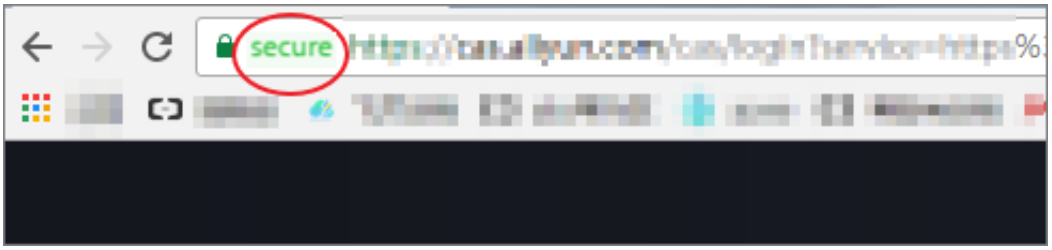
6. After the import, click **Finish**, as shown in *Figure 12-20: Import completed*. A prompt box appears, indicating a successful import.

Figure 12-20: Import completed



7. Restart the web browser and log on to Apsara Stack Console. If the certificate is successfully installed, the "secure" sign on the left of the URL is in green, as shown in *Figure 12-21: Restart the web browser*.

Figure 12-21: Restart the web browser



12.4.13 Download and install the GPU driver

To install a GPU driver for your instance instead of using an image pre-installed with a GPU driver , do as follows:

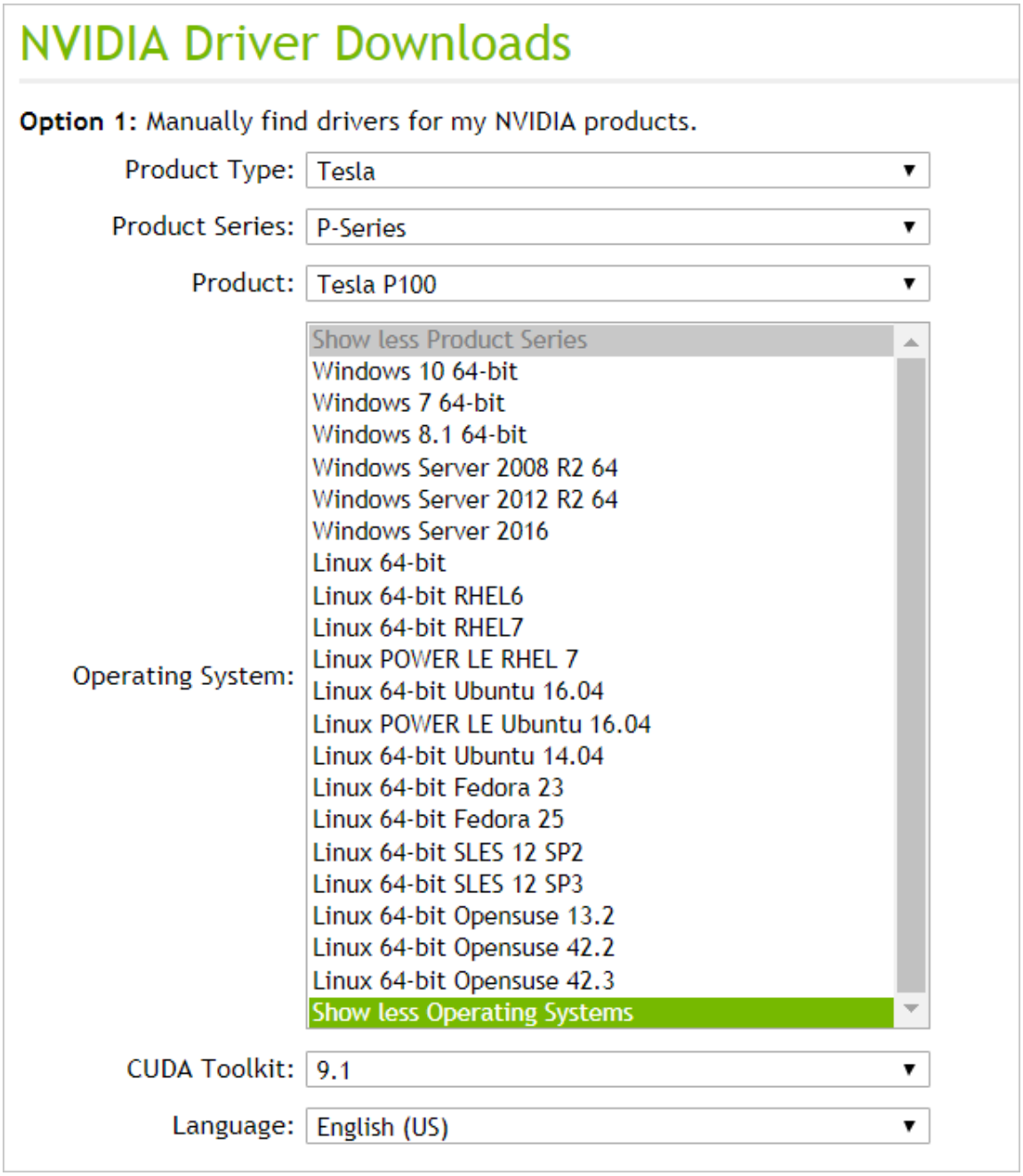
Procedure

1. Obtain the installation package.
   a) Visit the *NVIDIA website*.
   b) Search for the driver for your instance. The filter information is as follows:
      - Product type: Tesla
      - Product series: P-Series
      - Product family: Tesla P100
      - Operating system: Select a version compatible with your instance image.

        Note:

        - If the server operating system is not displayed in the drop-down list, click **Select All Operating Systems** at the bottom of the drop-down list.
        - If your instance uses a Linux image that is not in the list, you can select **Linux 64-bit**.

Figure 12-22: Download the driver



c) Click **Download** after the search result is displayed.

2. Download and install the GPU driver.

   - If your instance runs on a Windows operating system, double-click the installation package to install the GPU driver.
   - If your instance runs on a Linux operating system, install the GPU driver as follows:

1. Download and install the kernel-devel and kernel-header packages of the corresponding kernel.

   Note:
   If your image runs on CentOS 7.3, install the kernel-devel and kernel-header packages of kernel 3.10.0-514.26.2.el7.x86_64. If you use other images, you can search for and download suitable packages.

2. Run the following command to verify that the kernel-devel and kernel-header packages are successfully downloaded and installed:

   ```
   sudo rpm -qa | grep $(uname -r)
   ```

   For example, if your image runs on CentOS 7.3, the following command output indicates that the kernel-devel and kernel-header packages are successfully installed:

   ```
   kernel-3.10.0-514.26.2.el7.x86_64
   ```

   ```
   kernel-headers-3.10.0-514.26.2.el7.x86_64
   ```

   ```
   kernel-tools-libs-3.10.0-514.26.2.el7.x86_64
   ```

   ```
   python-perf-3.10.0-514.26.2.el7.x86_64
   ```

   ```
   kernel-tools-3.10.0-514.26.2.el7.x86_64
   ```

3. Install the GPU driver as instructed in the **Other Information** tab on the GPU driver download page of the *NVIDIA website*.

Result

If you use an image of Windows 2008 R2 or earlier versions, when you use the function of **Connect to Management Terminal** on the Alibaba Cloud console to access the instance after the installed GPU driver takes effect, the instance interface is unresponsive with a black screen or stuck at the startup interface. In this case, you need to connect to your ECS instance remotely through other protocols, such as the RDP of Windows.

12.5 Disk management

ECS disks can be classified into basic cloud disks, SSD cloud disks, and ultra cloud disks.

A mount point is the position of an ECS disk on the disk controller bus.

The selected mount point corresponds to the disk device number in Linux, and is consistent with the disk sequence in the disk manager in Windows.

12.5.1 Cloud disk

ECS instances are provided with cloud disks for data storage. You can directly identify the cloud disks in the operating system, and perform read or write operations.

You can use cloud disks like physical disks in ECS instances. Each cloud disk must be attached to an ECS instance and formatted before becoming usable.

Cloud disks are provided in three types: basic cloud disks, SSD cloud disks, and ultra cloud disks.

Cloud disks have the following basic features:

- Robust data security.
- High IOPS for random and sequential read/write.
- An ECS instance can have one or more (up to 17) cloud disks, including system disks and data disks.
- In the case of downtime migration, data prior to the downtime is saved.

Snapshots and images are stored in the OSS instance in the same region as your ECS instance. With the distributed storage feature, the system does not copy entire data to a disk that is created from a snapshot or image. Instead, the data on the disk is loaded by block per your needs, that is , only the required data is copied to the disk, which enhances the service flexibility and resource utilization.

A typical data block is several megabytes. It is read directly from the disk.

With the distributed storage feature, snapshots and images are copied to the disks block by block in the background when there are less I/O operations, which optimizes your overall I/O experience .

Therefore, when you access a cloud disk for the first time, you may experience a significant decrease in the disk's I/O performance. However, the I/O performance restores to the normal level during subsequent access. When you perform a high-load operation, such as reading data, we recommend that you do it through full-disk access.

12.5.2 Create disks

Context

You can create a block storage on the ECS console to resize the system storage space.

- One instance can have up to 16 data disks (this quota includes both cloud disks and shared block storages).
- A shared block storage can be attached to more than two ECS instances simultaneously. In the current version, one shared block storage can be attached to four ECS instances.
- Each ultra cloud disk (or ultra shared block storage) or SSD cloud disk (or SSD shared block storage) supports up to 32 TB capacity.

Note:

- Currently, ECS instances do not support combining multiple cloud disks. After creation, each cloud disk is an independent entity. The space of multiple cloud disks cannot be combined through formatting. We recommend that you plan the disk quantity and capacity in advance.
- A snapshot is intended for an independent disk, so data may be different after you perform snapshot rollback under the Logical Volume Management (LVM). Thus, if you have created multiple disks, we do not recommend that you configure LVM for them.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.
2. Click **Create Disk**.
3. You can view the following configuration on the **Create Disk** page:

Table 12-6: Disk configuration

Item	Description
Region	• **`Region`**: It is a required parameter. Select the region in which the disk resides. • **`Zone`**: It is a required parameter. Select the zone in which the disk resides.
Configurations	• **`Name`**: It is a required parameter. Enter the name of the disk. • **`Department`**: It is a required parameter. Select the department in which the disk resides. • **`Project`**: It is a required parameter. Select the project in which the disk resides. • **`Storage`**: It is a required parameter. Select the specific storage type of the disk, which can be **Disk** or **Shared block storage.**

Item	Description
	• **Type**: It is a required parameter. After selecting the storage type, you can specify the disk type as **Ultra cloud disk** (ultra block storage) or **SSD cloud disk** (SSD block storage) as needed. • **Encrypted**: It is an optional parameter. It specifies whether the created disk will be encrypted. • **Use Snapshots**: It is an optional parameter. After checking **Use Snapshots**, you still need to select the corresponding snapshot.  **Note:** • If you have checked **Encryption** for the disk in the previous option, this option does not appear. • If the disk size specified by the user is smaller than the selected snapshot size, the disk size actually generated will be in line with the snapshot size; otherwise, the disk size will be the value specified by the user.

4. Click **Confirm**.

Result

In the disk list, check whether the disk is in the **Available** state. If so, the disk is successfully created.

What's next

The procedure varies depending on the operating system of the instance.

- If the Linux operating system is selected for the instance, you must *Attach a disk* and then *Partition, format, and attach data disks in Linux*.
- If a Windows operating system is selected for the instance, you must *Attach a disk* and then *Partition and format data disks in Windows*.

12.5.3 View disks

You can log on to the ECS console to view all your disks and their details.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.

2. Select a **Department** and **Region** or enter a **Disk Name** and click **Search** to find the target disk.

   Note:
   Click **Disk Name**, and you can also choose the following filtering conditions from the drop-down menu: **Disk ID**, **Instance Name**, **Disk Status**, **Disk Usage Type**, and **Snapshot Policy ID**.

3. In the **Action** column of the disk, click the icon and select **View Details** from the drop-down list to go to the **Disk Details** page and view the details of the disk.

12.5.4 Roll back a disk

When you want to roll back the data on a disk to a previous time point, you can do so through disk rollback.

Prerequisites

Make sure that the instance of the target disk is in the **Stopped** state.

Note:
Snapshot rollback is irreversible. After rollback is finished, the original data cannot be restored. Exercise caution when performing this operation.

Procedure

1. *Log on to the ECS console* and go to the **Snapshots** page.
2. On the **Snapshots** page and in the **Action** column of the snapshot, click and select **Rollback Disk** from the drop-down list.
3. In the confirmation box that appears, click **Confirm**.

   Note:
   If you select **Start instance right after rollback**, the instance will start automatically after the cloud disk is rolled back successfully.

12.5.5 Edit disk attributes

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.

2. On the **Disks** page, select the disk for which you want to modify the attributes. In the **Action** column of the disk, click the icon and select **View Details** from the drop-down list to go to the **Disk Details** page.

   Note:
   You can also click the disk ID to go to the **Disk Details** page.

3. Click **Modify Properties**.

   You can set the following disk attributes:

   - **Disk Name**: It is a string of 2 to 128 characters, including numbers, periods (.), underscores (_), and hyphens (-). It must begin with an uppercase or lowercase English letter or a Chinese character.
   - **Disk Description**: It is a string of 2 to 256 characters. It cannot start with 'http://' or 'https://'.

4. Click **Confirm**.

12.5.6 Attach a disk

You can only attach independent cloud disks to ECS instances.

Note:

- When you attach a cloud disk to an ECS instance, you must check that the ECS instance is in the **Running** or **Stopped** state and the instance's security control indicator is not in the **Locked** state.
- When you attach an independent cloud disk, the cloud disk must be in the **Available** state.
- You can attach up to 16 data disks (including cloud disks and shared block storages) to one ECS instance.
- You must attach an independent cloud disk to an instance in the same zone.
- You can attach an independent cloud disk only as a data disk, but not a system disk.

You can attach a disk in either of the following ways:

- *Attach a disk on the Instance Details page*.
- *Attach a disk on the Disk List page*.

12.5.6.1 Attach a disk on the Instance Details page

Prerequisites

- Before you attach a disk, complete the following operation: *Create disks*.
- When you attach a data disk, ensure the cloud disk is in the **`Available`** state.

Context

You do not need to perform the attach operation for a data disk created together with an instance.

Procedure

1. *Log on to the ECS console* and go to the **Instances** page.
2. On the **Instances** list page, click the ID of the ECS instance to which you want to attach a disk.
3. Enter the **Instance Details** page and click the **Disks** tab.
4. Click **Attach**.
5. In the displayed dialog box, provide the following information:
   - **`Target Disk`**: It is a required parameter. Select an existing cloud disk that is in the **`Available`** state.
   - Select **`Deleted with Instance`**.

     Note:
     By default, the value of this option is **`No`**. If you select **`Yes`**, when the instance is deleted, the disk will be deleted together.
6. Click **Submit**.

12.5.6.2 Attach a disk on the Disk List page

Prerequisites

- Before you attach a disk, complete the following operation: *Create disks*.
- When you attach a data disk, ensure the cloud disk is in the **`Available`** state.

Context

You do not need to perform the attach operation for a data disk created together with an instance.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.
2. In the **Action** column of the disk to be attached, click the icon and choose **Attach**.

3. In the **Attach** dialog box, complete the following settings:
   - **Destination Instance**: Select the ECS instance to which you want to attach the selected cloud disk.
   - Select **Are you sure you want to configure the disk to delete along with the instance?**. The default value is **No**. Set it to **Yes** if you want to release the disk when the instance is deleted.
4. Click **Submit**.

12.5.7 Partition and format disks

ECS supports only secondary partitioning of **Data Disks**, but does not support secondary partitioning of **System Disks** (either in Windows or Linux operating systems). If you use a third-party tool to perform secondary partitioning of a system disk, you may encounter unknown risks, such as system crash and data loss.

Note:
Before you attach a data disk, *Create disks*.

12.5.7.1 Partition, format, and attach data disks in Linux

Prerequisites

- Complete *Connect to an instance*.
- Complete *Create disks* and *Attach a disk*.

Procedure

The data disks of Linux ECS instance are not partitioned or formatted. You can follow these steps to partition and format the data disks:

1. View the data disks. Before you partition and format the data disks, run the `fdisk -l` command (instead of `df –h`) to view the data disks.

   The output of the `fdisk -l` command shows information about the data disks, such as */dev/vdb* in the following figure. If */dev/vdb* is not displayed, the ECS instance has no data disks and you do not have to attach data disks.

```
[root@iZ********eZ ~]# fdisk -l

Disk /dev/vda: 42.9 GB, 42949672960 bytes
255 heads, 63 sectors/track, 5221 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
```

```
Disk identifier: 0x00078f9c

Device Boot      Start         End      Blocks   Id  System
/dev/vda1   *           1        5222    41940992   83  Linux

Disk /dev/vdb: 21.5 GB, 21474836480 bytes
16 heads, 63 sectors/track, 41610 cylinders
Units = cylinders of 1008 * 512 = 516096 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x00000000
```

2. Partition the data disks. Run the `fdisk /dev/vdb` command to partition the data disks, as shown in the preceding figure. Enter the following commands in sequence as prompted:

   a) `n` command: Creates a partition.

   b) `p`: Creates a primary partition.

   c) **`Partition number (1 to 4)`**: Number of the new partition, an integer in the range from 1 to 4. You can create up to four partitions. In this example, 1 is entered to indicate Partition 1.

   d) **`First cylinder`**: Start position of the partition. You can select the default value by pressing the **Enter** key. Also, you can enter a number in the range from 1 to 41610 and then press the **Enter** key. In this example, the default value 1 is used.

   e) **`Last cylinder`**: End position of the partition. You can select the default value by pressing **Enter**. Also, you can enter a number in the range from 1 to 11748 and then press the Enter key. In this example, the default value is used.

   f) Optional: If you want to create multiple partitions, you can repeat Steps a through e until all the four partitions are configured.

   g) Run the `wq` command to start partitioning.

```
[root@iZ********eZ ~]# fdisk /dev/vdb
Device contains neither a valid DOS partition table, nor Sun, SGI or
 OSF disklabel
Building a new DOS disklabel with disk identifier 0x01ac58fe.
Changes will remain in memory only, until you decide to write them.
After that, of course, the previous content won't be recoverable.

Warning: invalid flag 0x0000 of partition table 4 will be corrected
 by w(rite)

WARNING: DOS-compatible mode is deprecated. It's strongly
 recommended to
        switch off the mode (command 'c') and change display units
 to
        sectors (command 'u').

Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
```

```
p
Partition number (1-4): 1
First cylinder (1-41610, default 1):
Using default value 1
Last cylinder, +cylinders or +size{K,M,G} (1-41610, default 41610):
Using default value 41610

Command (m for help): wq
The partition table has been altered!
```

3. View the new partition. Run the `fdisk -l` command to list all the partitions, as shown in *code*. If the command output shows */dev/vdb1*, the partition *vdb1* is successfully created.

```
[root@iZ********eZ ~]# fdisk -l

Disk /dev/vda: 42.9 GB, 42949672960 bytes
255 heads, 63 sectors/track, 5221 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x00078f9c

Device Boot      Start         End      Blocks   Id  System
/dev/vda1   *           1        5222    41940992   83  Linux

Disk /dev/vdb: 21.5 GB, 21474836480 bytes
16 heads, 63 sectors/track, 41610 cylinders
Units = cylinders of 1008 * 512 = 516096 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x01ac58fe
Device Boot      Start         End      Blocks   Id  System
/dev/vdb1               1       41610    20971408+  83  Linux
```

4. Format the new partition. For example, you can run the `mkfs.ext3 /dev/vdb1` command to format the new partition as ext3. The time required for formatting varies depending on the hard disk size. You can also format the new partition as another file system type. For example, you can run the `mkfs.ext4 /dev/vdb1` command to format it as ext4.

> **Note:**
> Compared with ext2, ext3 only adds the log function. Compared with ext3, ext4 improves some important data structures. ext4 provides better performance and reliability, and more diverse functions.

```
[root@iZ********leZ ~]# mkfs.ext3 /dev/vdb1
mke2fs 1.41.12 (17-May-2010)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
Stride=0 blocks, Stripe width=0 blocks
1310720 inodes, 5242852 blocks
262142 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=4294967296
```

```
160 block groups
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
 32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632,
 2654208,
 4096000

Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 25 mounts or
180 days, whichever comes first.  Use tune2fs -c or -i to override.
```

5. Add partition information. Run the `echo '/dev/vdb1 /mnt ext3 defaults 0 0' >> /etc/fstab` command to add information about the new partition and then run the `cat /etc/fstab` command to view the partition information.

Note:

- This example adds partition information to the ext3 file system. You can add partition information to another file system type, such as ext4.
- Ubuntu 12.04 does not support barriers. Therefore, in Ubuntu 12.04, the `echo '/dev/vdb1 /mnt ext3 barrier=0 0 0' >> /etc/fstab` command is used to add partition information.

```
[root@iZ********eZ ~]# echo '/dev/vdb1 /mnt ext3 defaults 0 0' >> /etc
/fstab
[root@iZbp19cdhgdj0aw5r2izleZ ~]# cat /etc/fstab

#
# /etc/fstab
# Created by anaconda on Thu Aug 14 21:16:42 2014
#
# Accessible filesystems, by reference, are maintained under '/dev/
disk'
# See man pages fstab(5), findfs(8), mount(8) and/or blkid(8) for more
 info
#
UUID=94e4e384-0ace-437f-bc96-057dd64f42ee / ext4 defaults,barrier=0 1
 1
tmpfs                   /dev/shm                tmpfs   defaults
  0 0
devpts                  /dev/pts                devpts  gid=5,mode=620
  0 0
sysfs                   /sys                    sysfs   defaults
  0 0
proc                    /proc                   proc    defaults
  0 0
/dev/vdb1 /mnt ext3 defaults 0 0
```

To attach the data disk to a folder separately, for example, to store webpages separately, modify */mnt* of the preceding command.

6. Attach the new partition. Run the `Run mount -a` command to attach all the partitions listed in */etc/fstab* and then run the `df -h` command to check the attachment. If the following information is displayed, the partitions are successfully attached and the new partitions are available for use.

```
[root@iZ********eZ ~]# mount -a
[root@iZ********eZ ~]# df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/vda1        40G  5.6G   32G  15% /
tmpfs           499M     0  499M   0% /dev/shm
/dev/vdb1        20G  173M   19G   1% /mnt
```

12.5.7.2 Partition and format data disks in Windows

This section describes how to partition and format the data disks of a Windows instance.

Prerequisites

- Complete *Connect to an instance*.
- Complete *Create disks* and *Attach a disk*.

Context

The operations mentioned in this section only apply to Windows 2008.

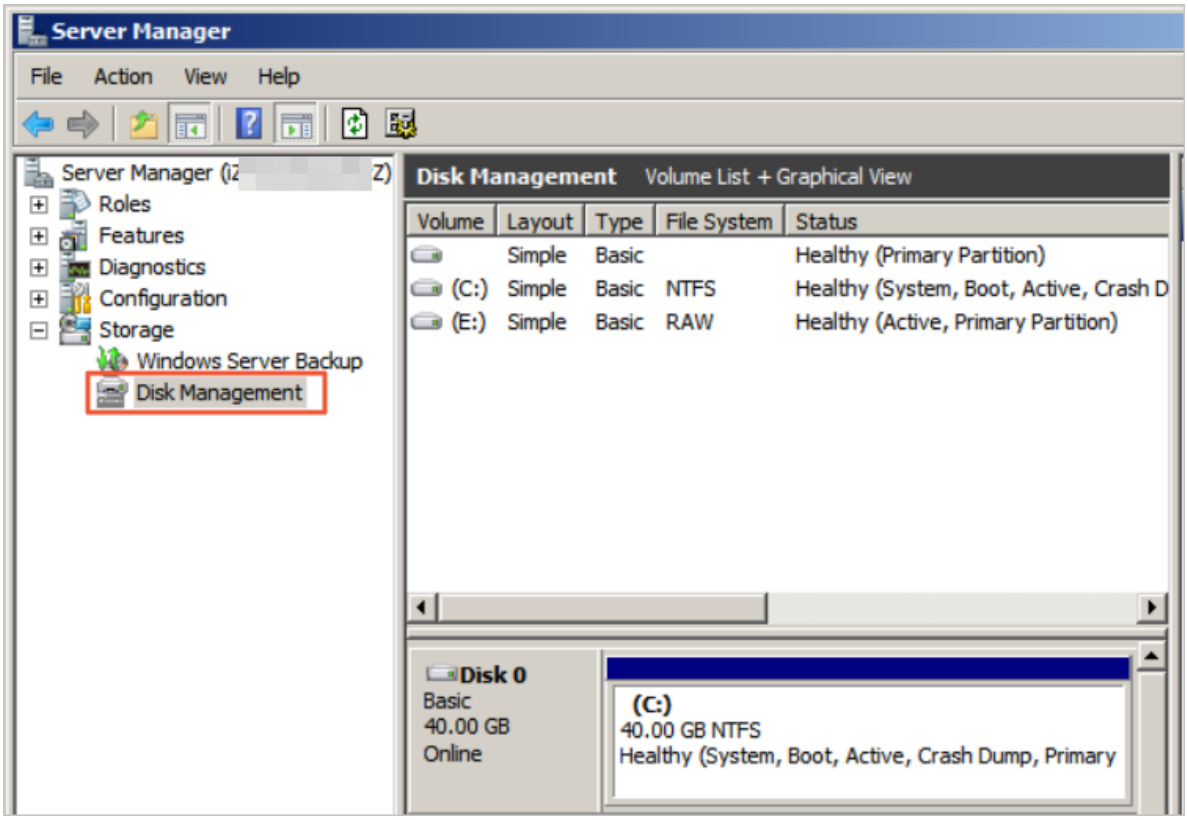
Procedure

Note:

If your data disks are in the **Offline** state, change them to the **Online** state before you allocate volume numbers and capacities to the data disks.

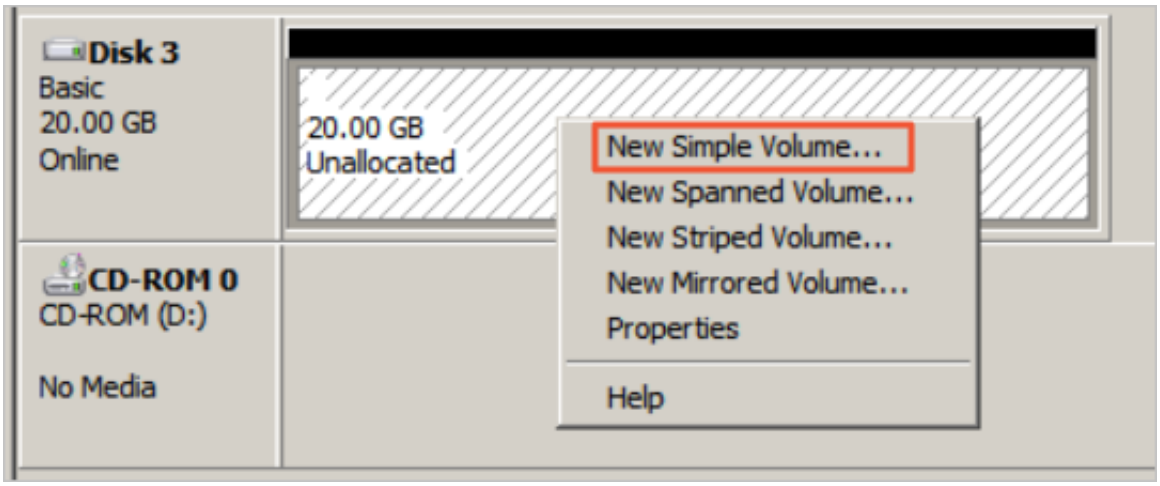
1. Click **Server Manager** on the toolbar in the lower-left corner to start the server manager.
2. On the left-side navigation bar of the **Server Manager** window, choose **Storage** > **Disk Management**.

Figure 12-23: Manage disks



3. Right-click an empty partition and choose **New Simple Volume** from the shortcut menu.

Figure 12-24: Choose the "New Simple Volume" option

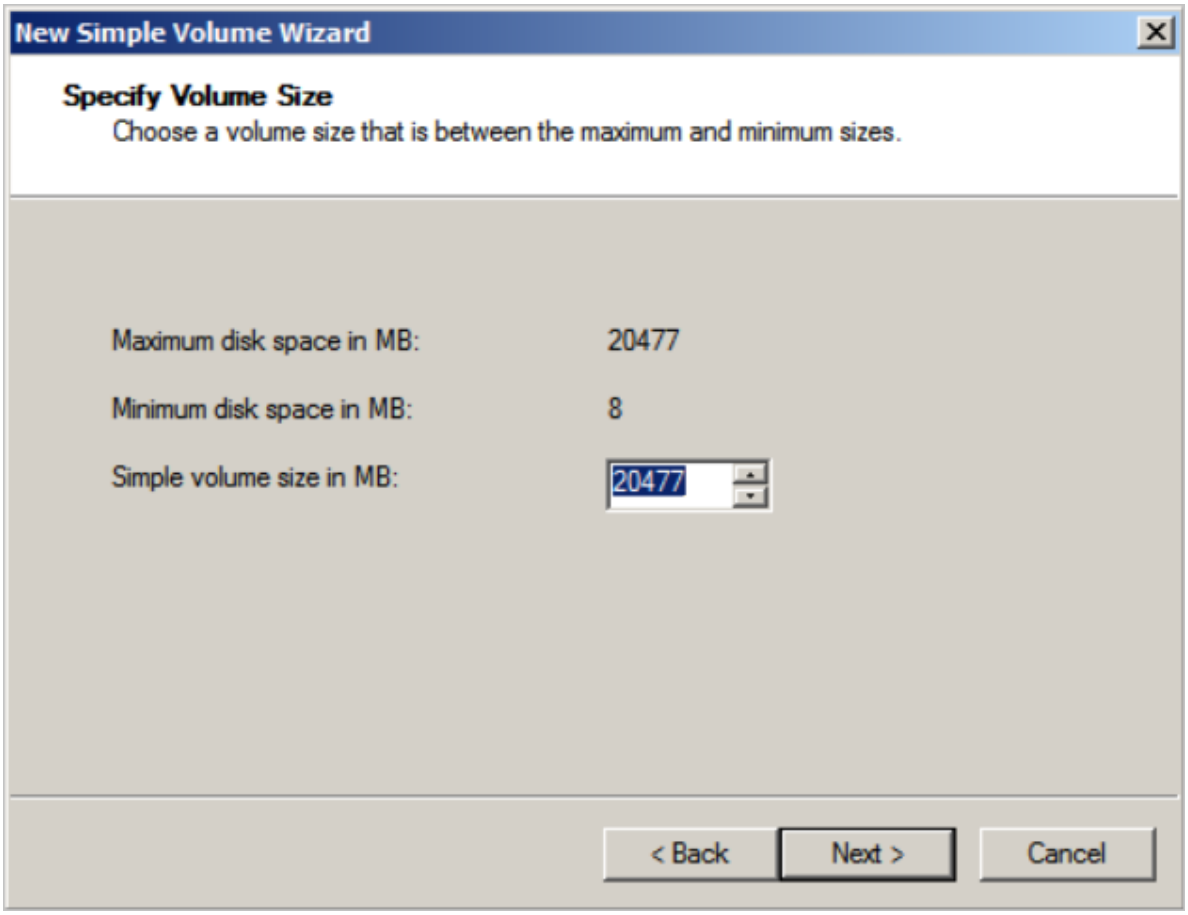


4. The "New Simple Volume" wizard appears. Click **Next**.

Figure 12-25: "New Simple Volume" wizard

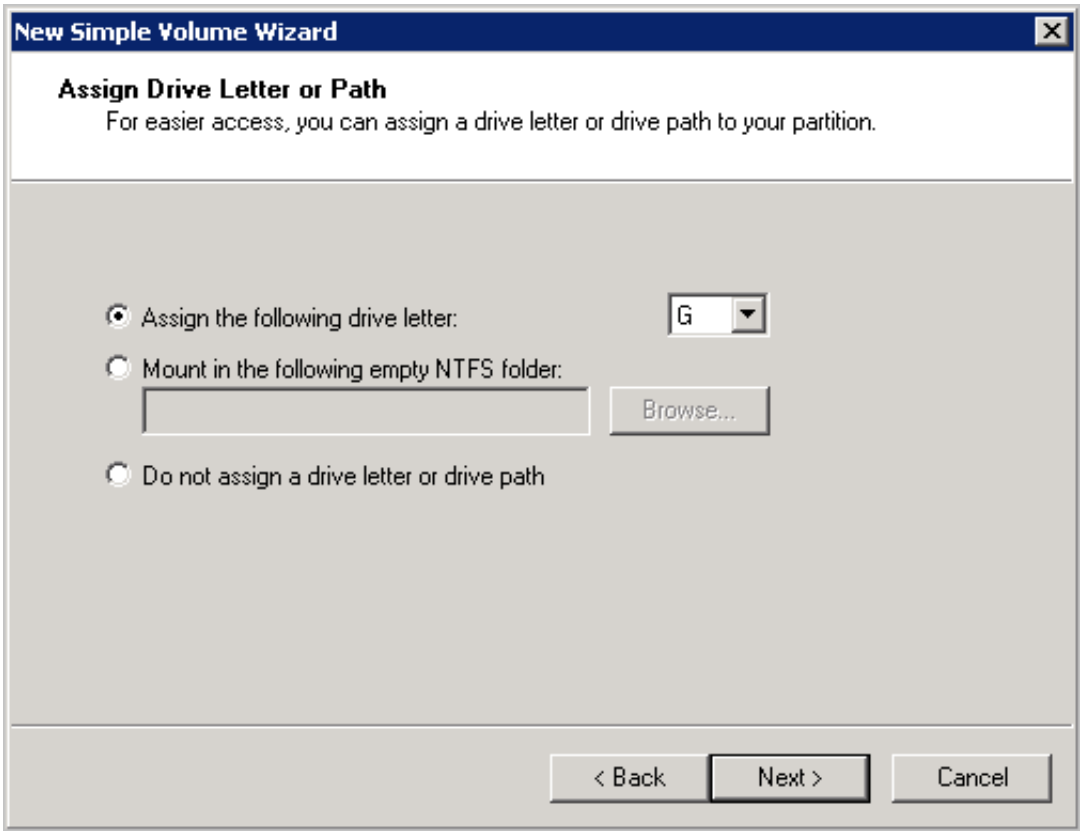
5. Set the size of the simple volume, that is, the partition size. The default value is **`Maximum Disk Space`**. You can specify the partition size as needed. After you complete the settings, click **Next**.

Figure 12-26: Set the partition size



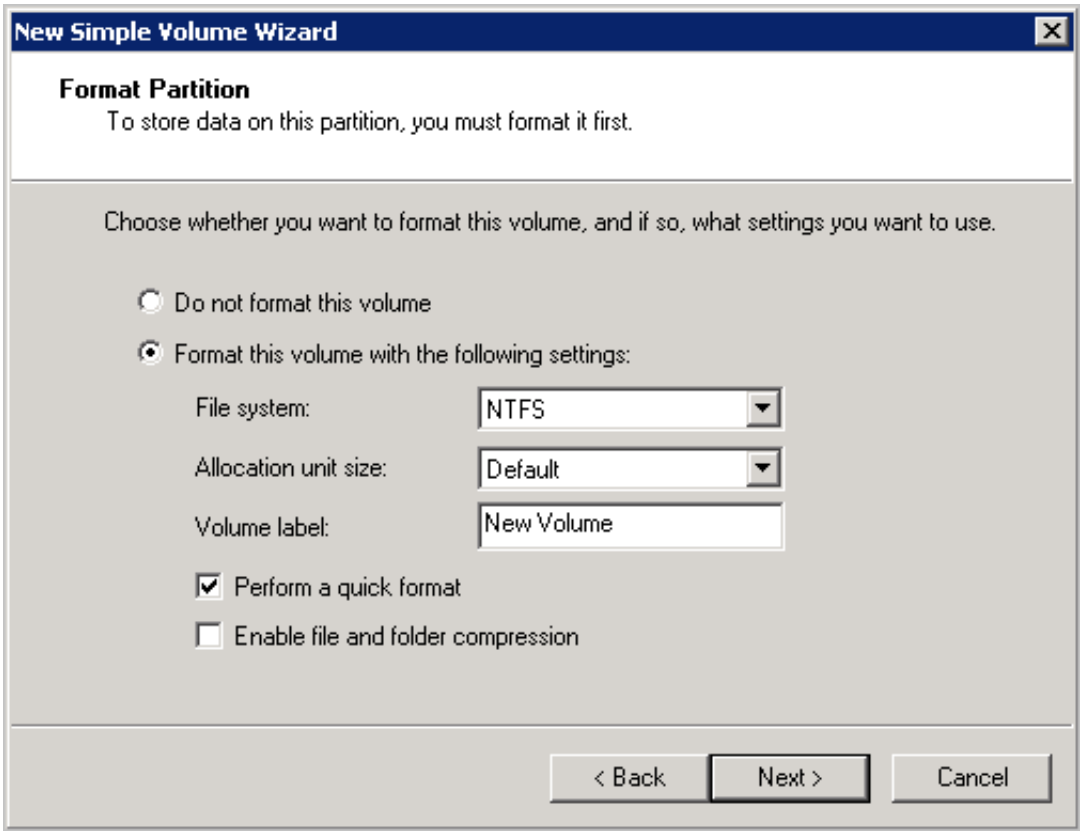
6. Specify the drive letter, which is listed in the alphabetic order by default. Click **Next**.

Figure 12-27: Allocate the drive letter



7. Format the partition. We recommend that you format the partition using the default settings of the wizard. After you complete the settings, click **Next** to start formatting.

Figure 12-28: Format the partition



8. When the wizard prompts that partitioning is complete, click **Finish** to close the wizard. The partition is successfully created.

12.5.8 Resize a system disk

Note:
Read the following Precautions before resizing the system disk.

Risks

- This operation requires you to stop your instance, which means interruption of your business.
- After replacement, you must redeploy the business runtime environment on the new system disk. There is a possibility of long interruption of your business. Extreme caution should be exercised when performing this operation.
- Your manually created snapshots are retained after the system disk is resized. However, because the disk ID is changed, you can no longer use the manually created snapshots on the original system disk to roll back the new system disk. The retained snapshots can still be used to create custom images.

- To retain enough snapshot quota for the automatic snapshot policy of the new disk, you can *delete unnecessary snapshots*.
- The original system disk is released after resizing.

Notes

- When resizing a system disk, you cannot reduce the disk capacity, but can only increase or keep the disk capacity.
- Resizing the system disk will not change the IP address and MAC address of your instance.
- The system disk type cannot be changed.
- Windows 2003 does not support system disk resizing.

Procedure of system disk resizing

If you are sure to resize a system disk, follow these steps:

1. *Create a snapshot for a system disk*.
2. *Create an image from a snapshot*.
3. *Replace a system disk*.
4. *Set a snapshot policy for a system disk*.

12.5.8.1 Create a snapshot for a system disk

Prerequisites

Make sure that the instance of the target disk is in the **`Stopped`** state.

Context

If you do not want to retain the data on the system disk, skip this step and proceed to *Replace a system disk*. To avoid impact on your business, do not create snapshots during traffic peak periods. It takes about 40 minutes to create a 40 GB snapshot for the first time. Reserve enough time. When you create a snapshot, check that the system disk has sufficient space. We recommend that you reserve **1 GB** space. Otherwise, the system may not start properly after the system disk is resized.

Procedure

1. *Log on to the ECS console* and go to the **Instances** page.
2. Select a **`Department`** and **`Region`** or enter an **`Instance Name`** and click "Search" to find the target instance.

3. Click the instance whose system disk you want to replace, or in the **Action** column of the instance, click and choose **View Details** from the drop-down menu to go to the **Instance Details**page.
4. Click the **Disks** tab.
5. Find the system disk. In the **Action** column of the system disk, click the icon and select **Create Snapshot** from the drop-down list.
6. In the **Create Snapshot** dialog box that appears, enter a name for the snapshot and click **Confirm** to create the snapshot.

   Note:
   The name of a snapshot cannot start with **auto** because **auto** has been reserved as the name prefix for the snapshot that the system automatically creates for you.
7. Click the **Instance Snapshot** tab, and you can view the snapshot creation progress and status. When the **Progress** is 100%, the snapshot is successfully created.

12.5.8.2 Create an image from a snapshot

Context

- If you do not want to continue using the current operating system or retaining its data, skip this procedure and proceed to *Replace a system disk*.
- If you want to continue using the current system disk, you need to make an image based on the current system disk. After the system disk is resized, you can completely copy all the data to a new environment.
- You can perform an alternative operation to create an image of the system disk. For details, see *Create custom images from snapshots*.

Note:
When you create an image, check that the system disk has sufficient space. We recommend that you reserve 1 GB space. Otherwise, the system may not start properly after the system disk is resized.

Procedure

1. *Log on to the ECS console* and go to the **Instances** page.
2. On the **Instances** page, select a **Department** and **Region** or enter an **Instance Name** and click **Search** to find the target instance.

3. In the **Action** column of the instance, click and select **View Details** to go to the **Instance Details** page.

> **Note:**
> On the **Instances** page, you can also click an instance ID to go to the **Instance Details** page.

4. Click the **Instance Snapshot** tab. In the **Action** column of the target snapshot, click and choose **Create Custom Image** from the drop-down menu.
5. In the **Create Custom Image** dialog box, enter a name and description for the custom image and click **Confirm**.

> **Note:**
> - Remember the image name. You must select the custom image when replacing the system disk.
> - Do not select "Add Data Disk Snapshot". Selection of data disks is not supported during the system disk replacement.

Result

After the image is successfully created, it is displayed on the **Image** page.

12.5.8.3 Replace a system disk

You can replace the system disk of an instance to increase the disk capacity, for example, from 40 GB to 100 GB.

Prerequisites

- To prevent data loss, back up the data on the system disk properly.
- The user snapshot of the system disk will be retained, but the auto snapshot policy will become invalid and you need to reset the policy.
- After the system disk is replaced, the system disk ID will be changed and the former system disk will be deleted.
- Before you replace a system disk, the instance should be in the **Stopped** state.

Procedure

1. *Log on to the ECS console* and go to the **Instances** page.
2. In the **Action** column of the instance, click and select **View details** from the drop-down list.
3. On the **Instance Details** page, click **Change System Disk**.

4. On the **Change System Disk** page, complete the following operation:

> **Note:**
> Before replacing a system disk, read the following instructions and precautions carefully:

- **Image type**: If you want to retain the data of the original system disk, select the custom image you created before. If you do not want to retain the data of the original system disk, you can select a public image.
- **System disk**: The type of the system disk is unchangeable, but you can set the capacity of the new disk. The new capacity cannot be smaller than that of the original one. The maximum capacity of the new disk is 500 GB.

5. After confirming the information, click **Confirm** and the capacity of the system disk is scaled up.

> **Note:**
> You can monitor the system status on the console. Typically, it takes about 10 minutes to finish the replacement. When it is done, the instance will start automatically.

12.5.8.4 Set a snapshot policy for a system disk

After you replace a system disk, you must set a snapshot policy for the new system disk if automatic snapshotting is needed. For more information, see *Configure an automatic snapshot policy*.

12.5.9 Detaching a disk

You can detach a data disk rather than a system disk. Local disks cannot be detached.

Prerequisites

Pay attention to the following before detaching a data disk:

- In Windows, you need to log on to the instance and perform **Offline** operation for the disk via disk management. After the command is executed successfully, you can enter the console to detach the disk.

> **Note:**
> To ensure data integrity, we recommend that you pause read/write operations for all the file systems in this disk. Otherwise, the data that is not read or written completely may be lost.

- In Linux, you need to log on to the instance and run the `unmount` command for the disk. After the command is executed successfully, you can enter the console to perform the detach operation for the disk.

  > **Note:**
  > If you have enabled automatically attaching data disk partitions during instance startup, before detaching the data disk, you must delete the attaching information of the data disk partitions from the /etc/fstab file first. Otherwise, you cannot connect the instance after the instance is restarted.

- The data disk to be detached must be in the **`Running`** state.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.
2. On the **Disks** page, select the disk you want to detach. In the **`Action`** column, click the management icon and select **Uninstall** from the drop-down list.
3. In the **Uninstall Disk** dialog box that appears, select the instance to which the disk is attached, confirm the information, and then click **Submit**.

Result

In the disk list, check whether the disk is in the **`Available`** state. If so, the disk is successfully detached from the instance.

12.6 Image management

An ECS image is a template that contains the software configurations such as the operating system, application server, and application programs of the ECS instance. When creating an instance, you must specify an ECS image. The operating system and software provided by the ECS image are installed in the created instance. You can create a custom image based on a created instance and then create more instances based on the custom image.

12.6.1 Select a suitable image

To create an instance, you must select a suitable image. When selecting an image, consider the following factors:

- Region and zone.
- Select the Linux or Windows operating system.

The 512 MB memory specifications do not support the Windows operating system, while the 4 GB memory specifications and above do not support the 32-bit operating system.

- Select the 32-bit or 64-bit operating system.

When creating an instance, you can select a custom image or public image.

12.6.2 Create a custom image

You can create a custom image and use it to replace the system disks while creating ECS instances.

12.6.2.1 Create custom images from snapshots

You can create custom images from snapshots on the system disk to fully load the operating system and data environment information in the snapshots to the images.

Prerequisites

- The disk attribute of the snapshot must be system disk, and data disks cannot be used to create a custom image.
- Make sure the system disk in your instance has available snapshots.

Procedure

1. *Log on to the ECS console* and go to the **Snapshots** page.
2. Select the system disk snapshot from which you want to create an image. In the **`Action`** column, click  and select **Create Custom Image**.
3. In the **Create Custom Image** dialog box, enter the following configuration information:
   - **`Custom Image Name`**: It is a required parameter. It is a string of 2 to 128 characters, including special characters such as periods (.), hyphens (-), and underscores (_). It must start with an uppercase or lowercase English letter.
   - **`Custom Image Description`**: It is a required parameter. It is a string of 2 to 256 characters. It cannot start with http:// or https://.
4. After completing the configuration information, click **Confirm**.

12.6.2.2 Create a custom image from an instance

By creating a custom image based on an instance, you can fully replicate all disks of the instance, including the data on the system disk and data disks, to the custom image (full image). When you

create a full image from an instance, each disk of the instance creates a snapshot automatically, and all the snapshots constitute a complete custom image.

Prerequisites

To avoid data security risks, delete sensitive data before creating a custom image.

Procedure

1. *Log on to the ECS console* and go to the **Instances** page.
2. Find the instance from which you want to create a custom image. In the **Action** column of the instance, click the  icon and select **Create Custom Image** from the drop-down list.
3. In the **Create Custom Image** dialog box that pops up, provide the following configuration information:
   - **Custom Image Name**: It is a required parameter. It is a string of 2 to 128 characters, including special characters such as periods (.), hyphens (-), and underscores (_). It must start with an uppercase or lowercase English letter or a Chinese character.
   - **Custom Image Description**: It is an optional parameter. It is a string of 2 to 256 characters and can start with http:// or https://.
4. After you complete the settings, click **Confirm**.

12.6.3 View images

You can check the running status of images on the ECS console.

Procedure

1. *Log on to the ECS console* and go to the **Images** page.
2. Select a **Department** and **Region** or enter an **Image Name** and click **Search** to find a particular image and view its details.

> **Note:**
> Click **Image Name** and you can select below filtering conditions for your query: **Image ID** and **Image Type**.

12.6.4 Copy images

To copy a custom image from one region to another, you can use the function of copying images, which is suitable for deploying an application across regions or running the same image environment on ECS instances in different regions. The time consumed for copying an image depends on the network status and concurrent tasks in the queue.

Context

Procedure

1. *Log on to the ECS console* and go to the **Images** page.
2. In the **Action** column of the image to be copied, click the icon , then select **Copy Image** in the drop-down menu.
3. In the **Copy Image** pop-up, you can see the ID of the custom image to be copied. Now configure the following items:
   - **Name**: mandatory. Specify the name of the custom image shown in the target region.
   - **Description**: optional. Provide the description of the custom image shown in the target region..

     Note:
     The description is 2 ~ 256 characters long and cannot start with "http://" or "https://".
   - Click **Confirm**.
4. Switch to the target region and you can see that the custom image is in the status of **Creating**. When the status is **Available**, the image is copied successfully.

12.6.5 Share images

You can share your custom images with other departments.

Prerequisites

Only custom images can be shared.

Procedure

1. *Log on to the ECS console* and go to the **Images** page.
2. In the **Action** column of the image to be shared, click the icon and select **Share Image** from the drop-down list.
3. In the dialog box that appears, select the target **Department** and click **Confirm**.

12.6.6 Import images

12.6.6.1 Notes for importing images

To guarantee the usability of an imported image and to improve the import efficiency, pay attention to the following aspects before importing an image.

The notes vary accounting to the operating system of your instance:

- *Linux*
- *Windows*

Linux

Limits

When importing a Linux image, pay attention to the following:

- Multiple network interfaces are not supported.
- IPv6 addresses are not supported.
- The password can be 8 to 30 characters in length and must contain uppercase/lowercase letters, numbers, and special characters.
- The firewall is disabled, and port 22 is enabled by default.
- The size of the Linux system disk is between 40 to 500 GB.
- DHCP is enabled in the image.
- SELinux is not activated.
- The KVM virtualization platform drivers must be installed.
- We recommend that you *Install cloud-int* to guarantee the successful configuration of host name, NTP source, and yum source.
- Imported Red Hat Enterprise Linux (RHEL) images must have a BYOL license.

Table 12-7: Notes

Item	Images of standard operating systems	Images of non-standard operating systems
Definition	The operating system editions (32-bit and 64-bit) supported by Alibaba Cloud includes: • CentOS • Ubuntu • FreeBSD • SUSE • OpenSUSE • RedHat • Debian • CoreOS • Aliyun Linux	The **non-standard operating system** refers to either of the followings: • The operating system that is not included in the list of operating systems currently supported by Alibaba Cloud. • A standard operating system that fails to comply with the requirements for a standard operating system in terms of critical system configuration files, basic

Item	Images of standard operating systems	Images of non-standard operating systems
		system environments, and applications. To use an image of a non-standard operating system, you are only allowed to choose Others Linux. If you import an image of such an operating system, Alibaba Cloud does not process any of the created instance. After you create the instance, you must connect to the instance by using the function of **Connect to Management Terminal** on the ECS console, and manually configure the IP address, route, and password.
Critical system configuration files	• Do not modify `/etc/issue*`. Otherwise, the system release cannot be properly recognized and the system creation fails. • Do not modify `/boot/grub/menu.lst`. Otherwise, the system may fail to start. • Do not modify `/etc/fstab`. Otherwise, an exception may occur and partitions fail to be loaded, leading to system startup failure. • Do not change`/etc/shadow` to **read-only**. Otherwise, you may be unable to modify the password file, leading to system creation failure. • Do not enable SELinux by modifying `/etc/selinux`	Fail to comply with the requirements of a standard operating system.

Item	Images of standard operating systems	Images of non-standard operating systems
	`/config`. Otherwise, the system may fail to start.	
Requirements for basic system environments	• Do not adjust the partition of the system disk. Currently only a single root partition is supported. • Make sure that the system disk has sufficient free space. • Do not modify critical system files, such as `/sbin`, `/bin`, and `/lib*`. • Before importing an image, confirm the integrity of the file system. • File systems ext3 and ext4 for Linux images are supported.	
Applications	Do not install `qemu-ga` in a custom image. Otherwise, some of the services that Alibaba Cloud needs may become unavailable.	
File format	Currently, only images in RAW and VHD formats are supported. To import images in other formats, use a tool to convert the format before importing the images. We recommend that you import images in the VHD format that has a smaller transmission capacity.	
File size	Setting the system disk size when importing an image: we recommend that you configure the system disk size based on the virtual disk size of the	

Item	Images of standard operating systems	Images of non-standard operating systems
	image (not the image file size). The disk size must be larger than or equal to 40 GB.	

Windows

Limits

When importing a Linux image, pay attention to the following:

- The password can be 8 to 30 characters in length and must contain uppercase/lowercase letters, numbers, and special characters.
- Imported Windows images do not provide the Windows activation service.
- The firewall must be disabled. You cannot perform remote logon if you do not disable the firewall. Port 3389 must be enabled.
- The size of Windows system disk is between 40 to 500 GB.

Operating system editions

You are allowed to import the following operating system editions (32-bit and 64-bit):

- Microsoft Windows Server 2012 R2 (standard edition)
- Microsoft Windows Server 2012 (standard edition and data center edition)
- Microsoft Windows Server 2008 R2 (standard edition, data center edition, and enterprise edition)
- Microsoft Windows Server 2008 (standard edition, data center edition, and enterprise edition)
- Microsoft Windows Server 2003 R2 (Standard Edition, Data Center Edition, and Enterprise Edition)
- Microsoft Windows Server 2003 with Service Pack 1 (SP1) or later (Standard Edition, Data Center Edition, and Enterprise Edition)
- Windows 7 (Professional Edition and Enterprise Edition)

Requirements for basic system environments

- System disks with multiple partitions are supported.
- Make sure that the system disk has sufficient free space.
- Do not modify critical system files.
- Before importing an image, confirm the integrity of the file system.

- The NTFS file system and MBR partition are supported.

Applications

Do not install qemu-ga in a custom image. Otherwise, some of the services that Alibaba Cloud needs may become unavailable.

Supported image formats

- RAW
- VHD

We recommend that you configure the system disk size based on the virtual disk size of the image (not the image file size). The disk size ranges from 40 to 500 GB.

Note:
We recommend that you import images in the VHD format that has a smaller transmission capacity.

12.6.6.2 Install cloud-int

To configure an ECS instance by using an existing image, you can create an instance from an image in the console. To guarantee the successful configuration of the host name, NTP source, and yum source of the imported Linux image, we recommend that you install cloud-init to your source server, VM, or instance before importing an image.

Context

Currently, the following Linux releases support the installation of cloud-init:

- CentOS
- Debian
- Fedora
- FreeBSD
- Gentoo
- RHEL (Red Hat Enterprise Linux)
- SLES (SUSE Linux Enterprise Server)
- Ubuntu

Prerequisites

Make sure that you have installed the following programs to your source server, VM, or instance.

- **git**: Downloads the source code package of cloud-init.

  Command: `yum install git`
- **Python 2.7**: the basis of running and installing cloud-init.

  Command: `yum install python`
- **pip**: Installs some libraries on which cloud-init depends but not included in Python 2.7.

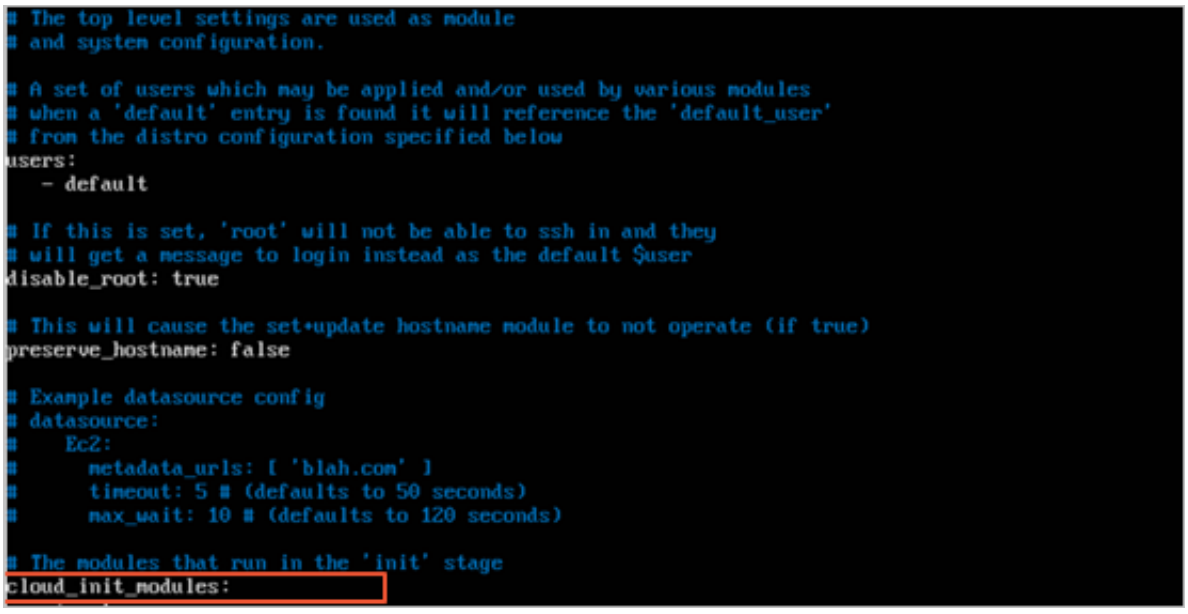
  Command: `yum install python-pip`

The `yum` command is used as an installation example. The use of `zypper` or `apt-get` commands is similar to that of the `yum` command.

Procedure

1. Make sure that you have installed the following programs to your source server, VM, or instance. If your instance is an ECS instance, see *Connect to an instance*.
2. Run `git clone https://git.launchpad.net/cloud-init` to download the cloud-init package.
3. Run `cd cloud-init` to enter the **cloud-init** directory.
4. Run `python setup.py install` to install the installation file **setup.py**.
5. Run `vi /etc/cloud/cloud.cfg` to modify the configuration file **cloud.cfg**.

   Figure 12-29: Modify the configuration file



   Change the preceding content of `cloud_init_modules` to the following:

```
# Example datasource config
```

```
 # The top level settings are used as module
 # and system configuration.
 # A set of users which may be applied and/or used by various
modules
 # when a 'default' entry is found it will reference the 'default_us
er'
 # from the distro configuration specified below
 users:
    - default
 user:
     name: root
     lock_passwd: False
 # If this is set, 'root' will not be able to ssh in and they
 # will get a message to login instead as the above $user
 disable_root: false
 # This will cause the set+update hostname module to not operate (if
 true)
 preserve_hostname: false
 syslog_fix_perms: root:root
 datasource_list: [ AliYun ]
 # Example datasource config
 datasource:
     AliYun:
         support_xen: false
         timeout: 5 # (defaults to 50 seconds)
         max_wait: 60 # (defaults to 120 seconds)
 #      metadata_urls: [ 'blah.com' ]
 # The modules that run in the 'init' stage
 cloud_init_modules:
```

Note:

The missing libraries may vary depending on the system. You can install the missing libraries using pip and run `python setup.py install` again after the installation.

Troubleshooting

- **Library six or library oauthlib is missing**

- If the following message appears, it indicates that the six library is missing from Python. Run `pip install six` to install the six library.

```
File "/root/cloud-init/cloudinit/log.py", line 19, in <module>
import six ImportError: No module named six )
```

- If the following message appears, it indicates that the oauthlib library is missing from Python. Run `pip install oauthlib` to install the oauthlib library.

```
File "/root/cloud-init/cloudinit/url_helper.py", line 20, in <
module>
```

```
import oauthlib.oauth1 as oauth1
```

```
ImportError: No module named oauthlib.oauth1
```

```
)
```

- **No dependency library is specified when an error occurs during installation**

  If no dependency library is specified in the error output, you may run `pip install -r requirements.txt` to install all the dependency libraries listed in the **requirements.txt** file of cloud-init.

What's next

You can *Import images to the console*.

12.6.6.3 Convert image file format

Only image files in RAW or VHD format can be imported. To import images in other formats, use a tool to convert the format before importing the images. You can use the qemu-img tool to convert image files into VHD or RAW from other formats, such as RAW, Qcow2, VMDK, VDI, VHD (vpc), VHDX, qcow1, or QED. You can also use qemu-img to convert image files between RAW and VHD formats.

Install qemu-img and convert image file format

You can use different methods to install qemu-img and convert the image file format based on operating system of your local computer:

- *Windows*
- *Linux*

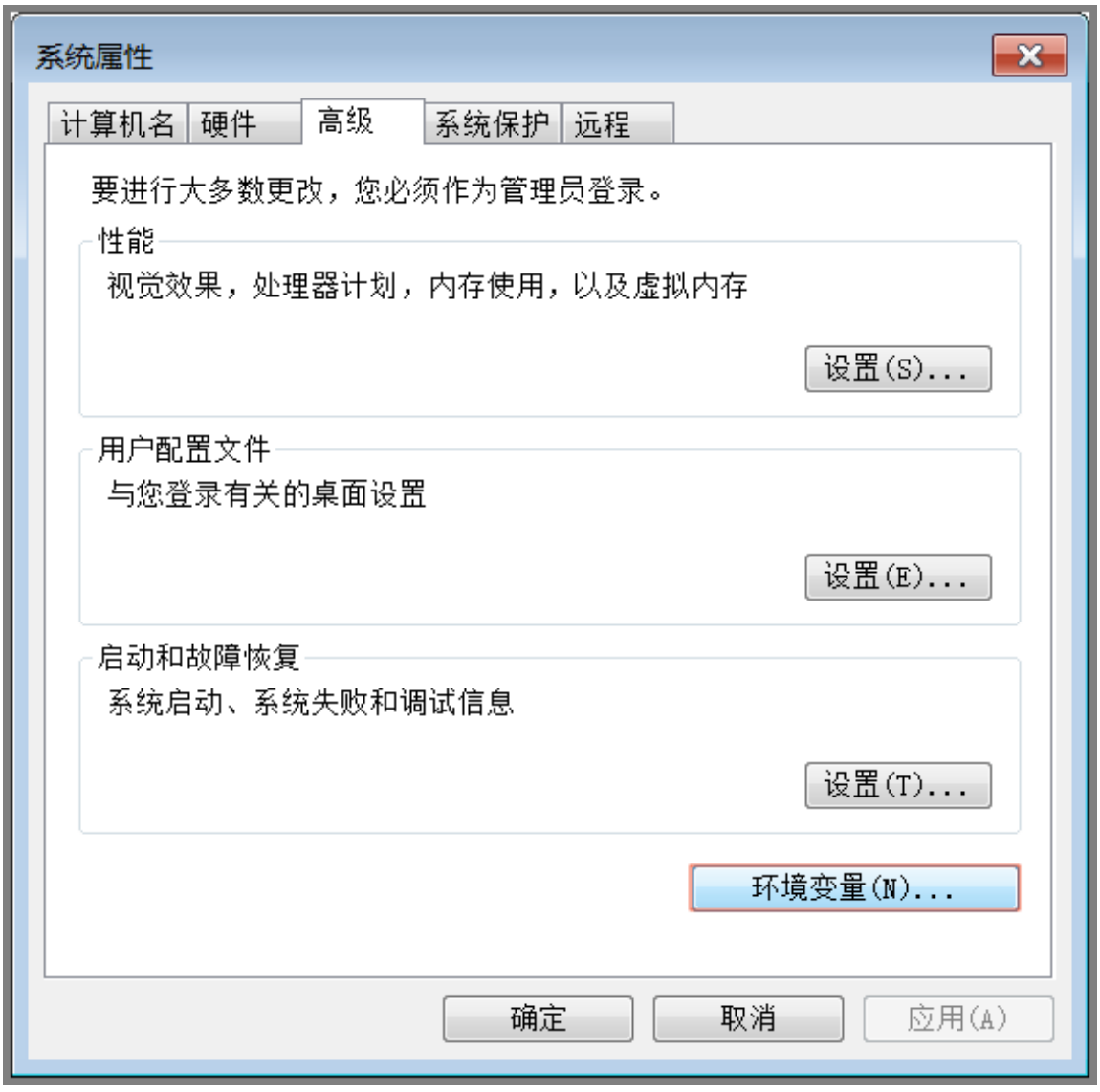
Windows

To install qemu-img and convert the image file format, follow these steps:

Procedure

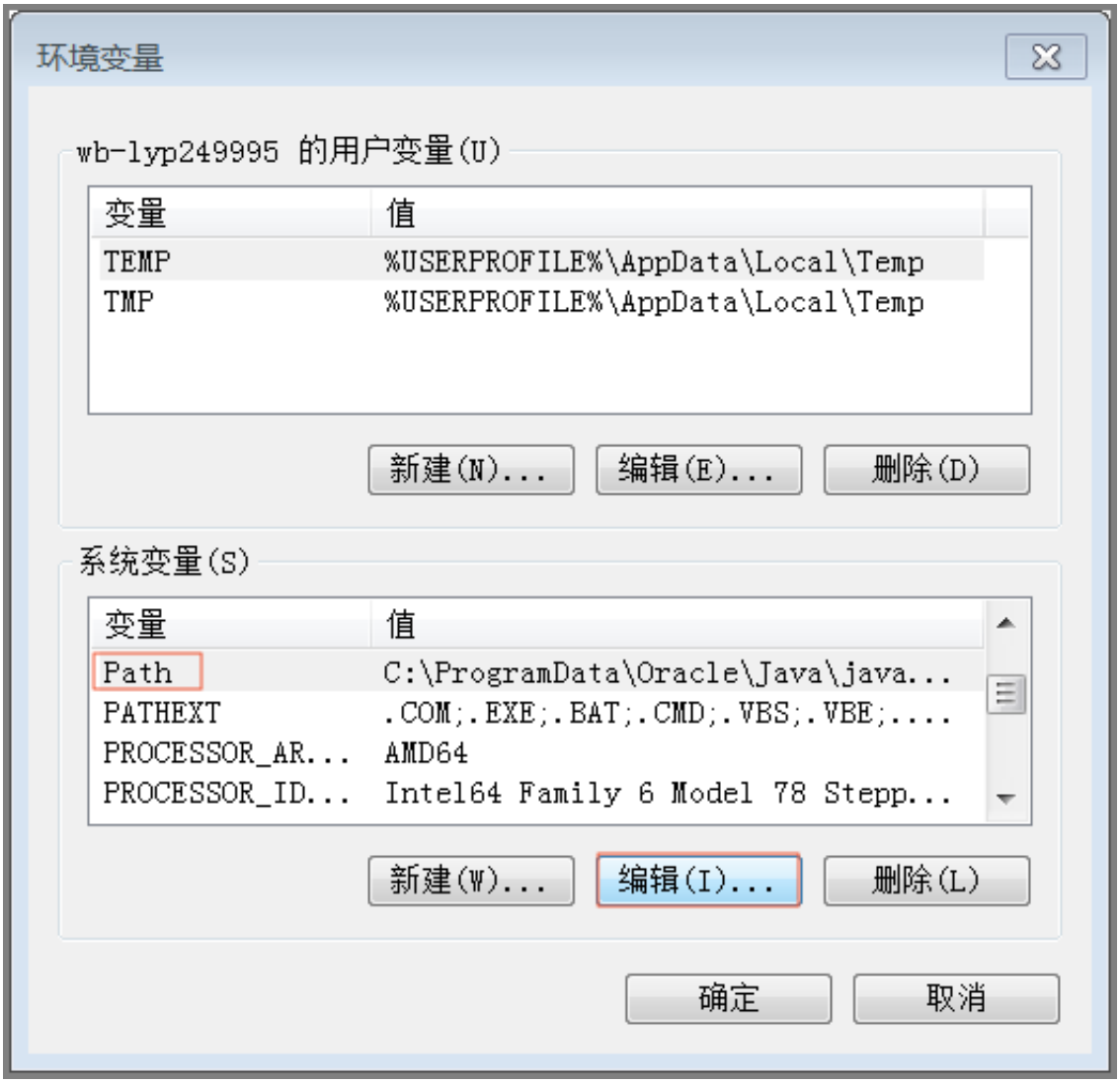
1. Download and*install qemu*. Installation path: *C:\Program Files\qemu*.
2. Do as follows to configure environment variables:
   a) Select **StartComputer**, and right-click **Properties**.
   b) In the left-side navigation pane, click **Advanced System Settings**.
   c) In the **System Properties** dialog box, click the **Advanced** tab and click **Environment Variables**.

   Figure 12-30: System properties



   d) In the **Environment Variables** dialog box, find the *Path* variable in **System Variables**, and click **Edit**. If the *Path* variable does not exist, click **New**.

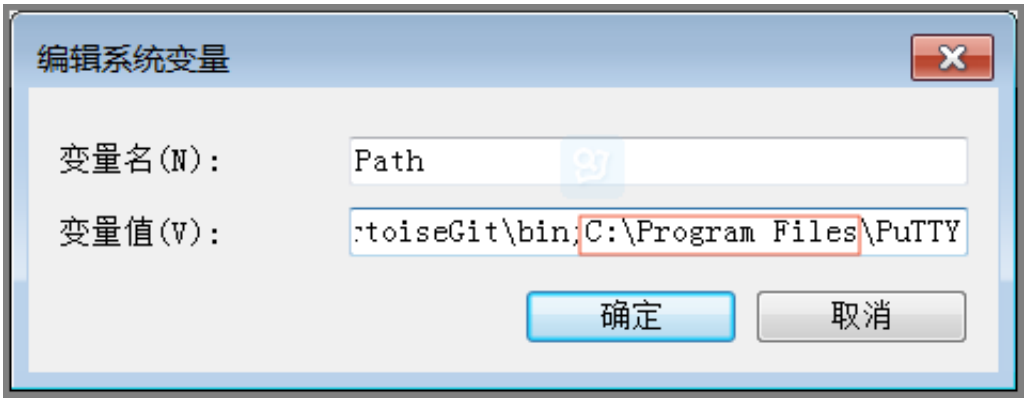
Figure 12-31: Edit variables



e) Add a variable value.

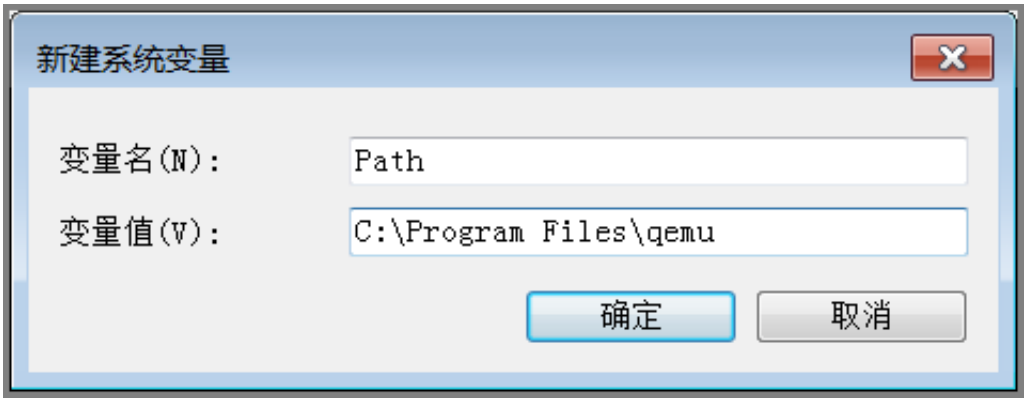
- In the **Edit System Variables** dialog box, add *C:\Program Files\qemu* to **Variable Value**. Different variable values are separated with semicolon (**;**).

Figure 12-32: Edit system variable



- In the **New System Variable** dialog box, enter *Path* in **Variable Name**, and *C:\Program Files\qemu* in **Variable Value**.

Figure 12-33: New system variable



3. Open the **command prompt** in Windows and run the `qemu-img --help` command. If it is displayed successfully, the installation succeeds.
4. In the **command prompt**, run the `cd [directory of the source image file]` command to change the file directory, for example, `cd D:\ConvertImage`.
5. Run the following command in the **command prompt** to convert the image file format: `qemu-img convert -f raw -O qcow2 centos.raw centos.qcow2`.

   The command parameters are described as follows:

   - `-f` is followed by the source image format.
   - `-O` (uppercase is required) is followed by the converted image format, the source file name, and the target file name.

   After the file format is converted, the target file appears in the directory of the source image file.

Linux

To install qemu-img and convert the image file format, follow these steps:

Procedure

1. Install qemu-img, for example:
   - For Ubuntu, run the `apt install qemu-img` command.
   - For CentOS, run the `yum install qemu-img` command.
2. Run the following command to convert the image file format.

   ```
   qemu-img convert -f raw -O qcow2 centos.raw centos.qcow2
   ```

   The command parameters are described as follows:
   - `-f` is followed by the source image format.
   - `-O` (uppercase is required) is followed by the converted image format, the source file name, and the target file name.

12.6.7 Export images

Prerequisites

You have been authorized to export images. For the authorization procedure, see **RAM Management** in *Cite LeftApsara Stack Console User GuideCite Right*.

Procedure

1. *Log on to the ECS console* and go to the **Images** page.
2. In the **Action** column of the image to be exported, click the icon and select **Export Image** from the drop-down list.
3. In the dialog box that appears, select an **OSS Bucket**, set an **OSS Prefix**, and then click **Confirm**.

   Note:

   OSS Prefix: It is an optional parameter, the value of which ranges from 1 to 30 characters and is composed of letters and numbers.

12.6.8 Delete images

Prerequisites

Currently, public images only support the batch delete function.

Procedure

1. *Log on to the ECS console* and go to the **Images** page.
2. Select a **Department** and **Region** or enter an **Image Name** and click **Search** to find the target image.
3. In the **Action** column of the image, click the icon and select **Delete** from the drop-down list.

   Note:
   When you select multiple images, you can click **Delete** in the upper-right corner of the page.
4. In the confirmation box that appears, click **Confirm**.

12.7 Snapshot management

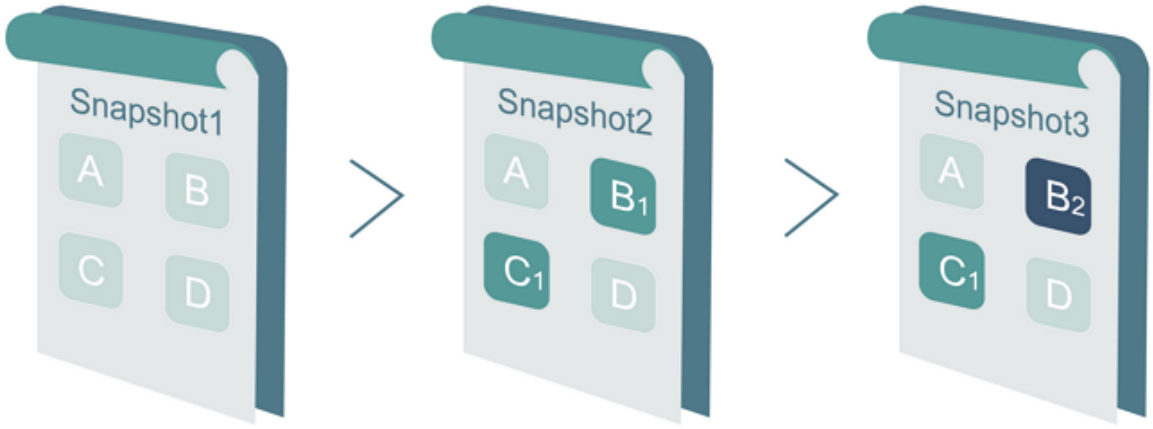
You may have the following needs when working with disks:

- Use the data on a disk as the basic data of another disk when writing or saving data to that disk .
- Restore your data to the expected status. Cloud disks provide a secure storage for your data. However, if the data stored on the disk is incorrect, for example, due to an application error or malicious tampering as a result of an application vulnerability, you may need to restore your data.

Alibaba Cloud allows you to create a snapshot to retain a copy of the data on a disk at a certain time point. You can create disk snapshots on a scheduled basis to guarantee your business continuity.

Snapshots use an incremental backup scheme. Two snapshots are compared so that only the changed data is copied, as shown in the following figure.

Figure 12-34: Operating principles of snapshots



Snapshot 1, Snapshot 2, and Snapshot 3 are the first, second, and third disk snapshots of the disk . The file system checks the disk data block by block. When a snapshot is created, only the blocks with changed data are copied to the snapshot.

Because Snapshot 1 is the first disk snapshot, it copies all of the data on the disk. Snapshot 2 only copies the changed data blocks B1 and C1 and references blocks A and D in Snapshot 1 as its data blocks A and D. Similarly, Snapshot 3 copies the changed data block B2, references blocks A and D in Snapshot 1 as its data blocks A and D, and references data block C1 in Snapshot 2 as its data block C1.

If you restore the disk to the status at the time of Snapshot 3, you can perform snapshot rollback to copy data blocks A, B2, C1, and D in Snapshot 3 to the disk.

If Snapshot 2 is deleted, data block B1 in the snapshot is deleted but data block C1 is not. In this way, when you restore the disk to the status at the time of Snapshot 3, data block C1 can also be restored.

Snapshots are stored in your Object Storage Service (OSS) instance, but the OSS console does not allow you to query, manage, or calculate the bucket space used by the snapshot files. You can operate snapshots only by using the ECS console or APIs.

12.7.1 Create a snapshot

You must create snapshots for your disks manually, because Alibaba Cloud does not automatically create snapshots for you.

Context

Alibaba Cloud provides the snapshot function for each user and imposes a quota on the number of snapshots that can be created. Currently, you can create up to 64 snapshots for each disk.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.
2. Find the disk for which you want to create a snapshot. In the **Action** column of the disk, click  and select **Create Snapshot** from the drop-down list.
3. Enter a **Snapshot Name** and **Snapshot Description**, and click **Confirm** .
4. Click the **Snapshots** tab, and you can view the snapshot creation progress and status. When the **Progress** is 100%, the snapshot is successfully created.

 **Note:**

- The first time you create a snapshot for a disk, it takes a relatively long time because this is a full snapshot.
- When you create a snapshot for a disk with existing snapshots, it takes a relatively short time. The specific duration depends on the volume of data changed between the snapshot to be created and the previous snapshot. The greater the changed volume, the longer the duration.
- Avoid creating snapshots during peak business hours.

12.7.2 View snapshots

You can view all your snapshots on the ECS console.

Procedure

1. *Log on to the ECS console* and go to the **Snapshots** page.
2. Select a **Department** and **Region** or enter a **Snapshot Name** and click **Search** to find the target snapshot.

 **Note:**

Click **Snapshot Name** and you can select other filtering conditions: **Snapshot ID**, **Disk Name**, and **Project ID**.

12.7.3 Delete a snapshot

When you no longer need a snapshot or if you exceed the snapshot quota, you must delete some snapshots.

Prerequisites

- The snapshot is not recoverable after deletion. Please operate with caution.
- If a system disk snapshot has been used to create a custom image, the snapshot cannot be deleted.

Procedure

1. *Log on to the ECS console* and go to the **Snapshots** page.
2. Select a **Department** and **Region** or enter a **Snapshot Name** and click **Search** to find the target snapshot.
3. Find the snapshot you want to delete. In the **Action** column of the snapshot, click the icon and choose **Delete**.

   Note:
   To delete multiple snapshots at the same time, select these snapshots and click **Delete** in the upper part of the snapshot list.
4. In the confirmation box that appears, click **Confirm**.

12.7.4 Application scenarios

In addition to rolling back the source disks, you can also use snapshots in the following situations:

- Create custom images.
- Create data disks when you create an instance.

Create custom images

If you want to use one of the instances as a template, you can quickly create a custom image. For the procedure, see *Create custom images from snapshots*.

Note:
Data disk snapshots cannot be used in the creation of custom images.

Use snapshots to create data disks for instances

You can use a snapshot to create a data disk for an instance so that the new data disk includes the data on the data disk of another instance.

Operation procedure

When you *Create disks*, select **Use Snapshots** to create a disk using the snapshot of another data disk in the same region. The capacity of the new data disk is determined by the snapshot capacity and cannot be changed.

Note:
If you reset a data disk that was created from a snapshot, the data disk restores the data in the snapshot.

12.8 Manage automatic snapshot policies

12.8.1 Create an automatic snapshot policy

You can easily create an automatic snapshot policy for disks by defining such parameters as the time of creation, days for repeated snapshot creation, and snapshot retention days.

Procedure

1. *Log on to the ECS console* and go to the **Snapshot Policies** page.
2. Click **Create**.
3. Define configuration information for the automatic snapshot policy:

Table 12-8: Automatic snapshot policy configuration

Item	Description
Name	Name of the automatic snapshot policy. It is a string of 2 to 128 characters, including numbers, underscores (_), and hyphens (-). It must start with an uppercase/ lowercase letter or a Chinese character.
Region	Sets the region to which the automatic snapshot policy is applied.
Department	Sets the department to which the automatic snapshot policy is applied.
Created At	Time of the day for starting automatic snapshot creation . The value must be on the hour and ranges from 00:00

Item	Description
	to 23:00 (24 time points in total). You can select multiple time points.
Repeat Date	Days of the week for automatic snapshot creation, ranging from Monday to Sunday. You can select multiple days.
Snapshot Retention Period(Days)	By default, the snapshot will be retained permanently. You can enter: 1 to 65535.

4. After completing the settings, click **Confirm**.

What's next

After the automatic snapshot policy is successfully created, you must *Configure an automatic snapshot policy*.

12.8.2 View automatic snapshot policies

You can view all your automatic snapshot policies on the ECS console.

Procedure

1. *Log on to the ECS console* and go to the **Snapshot Policies** page.
2. Select a **Region** or enter a **Snapshot Policy ID** and click **Search** to view the automatic snapshot policy.

> **Note:**
> Click **Snapshot Policy ID** and you can select filtering conditions from the drop-down menu: **Automatic Snapshot Policy Name**.

12.8.3 Edit an automatic snapshot policy

You can modify automatic snapshot policies, including the policy name and snapshot creation time.

Procedure

1. *Log on to the ECS console* and go to the **Snapshot Policies** page.
2. In the **Action** column for the target policy, click and select **Edit**.
3. In the dialog box that appears, you can modify **Snapshot Policy Name**, **Created At**, **Repeat Date**, and **Snapshot Retention Period**. Then click **Confirm**.

12.8.4 Configure an automatic snapshot policy

Prerequisites

- Make sure that the disk to which you want to apply an automatic snapshot policy is in the **Running** state.
- The automatic snapshot command takes the following format: auto_yyyyMMdd_1, for example , auto_20140418_1.

> **Note:**
> - We recommend that you select periods with a low service load for automatic snapshot policy execution.
> - The snapshots you have manually created do not conflict with automatic snapshots. However, if a disk is taking an automatic snapshot, you must wait for it to finish before you can manually create a snapshot.

Procedure

1. *Log on to the ECS console* and go to the **Disks** page.
2. Find the disk for which you want to set an automatic snapshot policy. In the **Action** column of the disk, click the icon and select **Run Automatic Snapshot** from the drop-down list.
3. In the **Implement Automated Snapshot Policy** dialog box that appears, select **Snapshot Policies**, and click **Confirm**.

12.8.5 Configure an automatic snapshot policy for multiple disks

Procedure

1. *Log on to the ECS console* and go to the **Snapshot Policies** page.
2. Click the icon to the right of the target snapshot policies and select **Configure**.
3. Select one or more disks to which you want to configure automatic snapshot policies.
   - In **Optional Disk**, specify multiple disks and click → to choose the disks.
   - In **Selected Disk**, specify multiple disks and click ← to remove the disks.
   - On the top of **Optional Disk**, click **Select All** and click → to choose all disks.
   - On the top of **Selected Disk**, click **Select All** and click ← to remove all disks.
4. Click **Confirm**.

12.8.6 Delete an automatic snapshot policy

You can delete unnecessary automatic snapshot policies.

Procedure

1. *Log on to the ECS console* and go to the **Snapshot Policies** page.
2. Select a **Region**, enter a **Snapshot Policy ID** or **Automatic Snapshot Policy Name**, and click **Search** to find the target snapshot policy.
3. In the **Action** column of the automatic snapshot policy to be deleted, click the icon and select **Delete** from the drop-down list.
4. Click **Confirm**.

12.9 Manage security groups

12.9.1 Security group restrictions

- A single security group cannot contain more than 1,000 instances. If you require Intranet mutual access among more than 1,000 instances, you can allocate them to different security groups and permit mutual access through mutual authorization.
- Each instance can join a maximum of five security groups.
- Each security group can have a maximum of 100 rules.
- Each user can have a maximum of 100 security groups.
- Adjusting security groups does not affect the continuity of your services.
- Security groups are stateful. If packets are permitted in the outbound direction, packets transmitted over this connection are also permitted in the inbound direction.

12.9.2 View security groups

You can view security group information on the ECS console.

Procedure

1. *Log on to the ECS console* and go to the **Security Groups** page.
2. Select a **Department** and **Region** or enter a **Security Group Name** and click **Search** to find the target security group.

> **Note:**
> From the **Security Group Name** drop-down list, you can select **Security Group ID** or **VPC ID**.

3. In the **Action** column of the security group, click the icon and select **View Details** to view detailed information about instances and rules of the security group.

   Note:
   You can also click the security group ID to view its information.

12.9.3 Remove an instance from a security group

Procedure

1. *Log on to the ECS console* and go to the **Security Groups** page.
2. Select a **Department** and **Region**, or enter a **Security Group Name**, and click **Search** to find the target security group.
3. Click the security group ID to go to the **ECS Instances** page.
4. Click the icon next to the target instance, select **Remove from Security Group**, and click **Confirm**.

12.9.4 Delete a security group

Prerequisites

Ensure all the instances are removed from the security group; otherwise, this security group cannot be deleted.

Procedure

1. *Log on to the ECS console* and go to the **Security Groups** page.
2. Select a **Department** and **Region** or enter a **Security Group Name** and click **Search** to find the target security group.
3. In the **Action** column of the security group, click the icon and select **Delete** from the drop-down list.

   Note:
   On the **Security Groups** page, you can select multiple security groups and click **Delete**. In the confirmation box that appears, select **Yes** and click **Confirm**.
4. In the confirmation box that appears, click **Confirm**.

12.9.5 Add security group rules

Procedure

1. *Log on to the ECS console* and go to the **Security Groups** page.
2. Select a **Department** and **Region** or enter a **Security Group Name** and click **Search** to find the target security group.
3. Click the security group ID to go to the **ECS Instances** page.
4. Click the **Rules** tab. In the upper-right corner of the page, click **Add Security Group Rules**.
5. In the dialog box that appears, complete settings and click **Confirm**.

   Parameters are described as follows:

   - **Authorization Policy**: The authorization policies include **Allow** and **Block**.

     Block: Directly discards data packets without any response. If two security groups have the same rules and different authorization policies, **Block** indicates that authorization takes effect and **Allow** indicates that authorization does not take effect.
   - **Rule Direction**:
     - **Outbound**: access by an ECS instance to other ECS instances in the Intranet or to the resources in the Internet.
     - **Inbound**: access to an ECS instance by other ECS instances in the Intranet or by resources in the Internet.
   - **Protocol Type** and **Port Range**: The port range varies with the protocol type. The following table describes the relationship between protocol types and port ranges.

     Table 12-9: Parameter description

Protocol Type	Port Range	Application scenario
All	The port range is displayed as -1/-1, indicating that all ports can be used. It is not configurable.	Used for mutually trusted application scenarios.
TCP	The port range can be customized from 1 to 65535. The valid format is **start port/end port**. Even if there is only one port, the valid format must be used. For example, 80/80 indicates port 80.	Used to permit or deny one or more consecutive ports.
UDP		
ICMP	The port range is displayed as -1/-1, indicating that all ports can be used. It is not configurable.	Run the `ping` command to check the communication status between instances.

Protocol Type	Port Range	Application scenario
GRE	The port range is displayed as -1/-1, indicating that all ports can be used. It is not configurable.	Used for VPN services.

- **`Priority`**: The default value is 1, indicating the highest priority. It is not configurable currently.
- **`Authorization Type`** and **`Authorized IPs`**: The authorization IPs vary with the authorization type. The following table describes the relationship between authorization IPs and authorization types.

Table 12-10: Authorization description

Authorization type	Authorized IPs
IP Range Access	Enter a single IP address or CIDR network segment, for example, 12.1.1.1 or 13.1.1.1/25. Only IPv4 addresses are supported. The address segment 0.0.0.0/0 indicates that access by all IPs is either allowed or denied. Exercise caution when setting authorization IPs.
Security Group Access	Security group access is valid for the Intranet only. Therefore, the security group access rules apply to Intranet access only, but not Internet access. Internet access can only be authorized through IP range access.

12.10 Manage an ENI

ENIs are divided into primary ones and secondary ones. The primary ENI is created by default when you create a VPC instance. The lifecycle of the primary ENI is consistent with that of the instance. You cannot separate the primary ENI from the instance. The following contents are about the secondary ENI. You can create a secondary ENI and attach it to or detach it from instances.

12.10.1 Create an ENI

Procedure

1. *Log on to the ECS console* and go to the **Elastic NIC** page.
2. Click **Create NIC** in the upper-right corner of the page.
3. On the **Create NIC** page, configure the following ENI information and click **Confirm**.

Table 12-11: Configure an NIC

Item	Description
Region	• **Region**: Required. Select a region where the target ENI resides. • **Zone**: Required. Select a physical zone with independent power grids and networks within a region. A zone can communicate with other zones using the Intranet, and is not affected by faults in other zones. If you want to improve application availability, we recommend that you create instances in different zones.
Configurations	• **Department**: Required. Select a department to which the ENI belongs. • **Project**: Required. Select a project to which the ENI belongs. • **VPC**: Required. Select a VPC where your instance resides. The ENI can only be attached to an instance on the same VPC. **Note:** You cannot change the VPC where the created ENI resides. • **Security Groups**: Required. Select a security group of the current VPC. • **ENI Name**: Optional. Set the name of the ENI. • **IP Address**: Optional. Enter the primary Intranet IPv4 address of the ENI. The IPv4 address must be an idle address in the CIDR network segment of VSwitches. If you do not specify an IPv4 address when creating an ENI, the system automatically allocates an idle private IPv4 address for you. • **Description**: Optional. Enter a description for the ENI for future management.

12.10.2 Edit an ENI

You can modify the name, security group, and description of an existing ENI on the ECS console.

Procedure

1. *Log on to the ECS console* and go to the **Elastic NIC** page.

2. Find the target secondary ENI, click the  icon in the **Action** column, and select **Edit** from the drop-down list.
3. In the dialog box that appears, modify the **NIC Name**, **Security Groups**, and **Description** of the ENI, and click **Confirm**.

12.10.3 Attach an ENI to an instance

You can attach a secondary ENI to an instance.

Prerequisites

Pay attention to the following items when attaching an ENI to an ECS instance:

- You can only attach a secondary ENI to an instance.
- An ENI has been created and must be in the **Available** status. For details, see *Create an ENI*.
- The ECS instance must be in the **Stopped** status. For more information about how to stop an instance, see *Start, stop, or reboot an instance*.
- The instance must be in the same VPC as the ENI.
- The VSwitch where the ENI resides must be in the same zone as the target ECS instance.
- An ENI can be attached to only one ECS instance at a time. However, an instance can be associated with multiple ENIs. For more information about the maximum number of ENIs supported by each instance type, see **Instance types** in *Cite LeftECS Product IntroductionCite Right*.

Procedure

1. *Log on to the ECS console* and go to the **Elastic NIC** page.
2. Find the target secondary ENI, click the  icon in the **Action** column, and select **Attach to ECS Instance** from the drop-down list.
3. In the dialog box that appears, select the **Destination Instance** and click **Confirm**.

12.10.4 Detach an ENI from an instance

You can detach a secondary ENI, but not a primary ENI, from an instance.

Prerequisites

- The secondary ENI must be in the **Bound** status.
- The instance must be in the **Stopped** status. For more information about how to stop an instance, see *Start, stop, or reboot an instance*.

Procedure

1. *Log on to the ECS console* and go to the **Elastic NIC** page.
2. Find the target secondary ENI, click the icon in the **Action** column, and select **Detach from ECS Instance** from the drop-down list.
3. In the prompt box that appears, click **Confirm**.

12.10.5 Delete an ENI

Context

You can only delete ENIs one by one, but not multiple ENIs at a time.

Prerequisites

An ENI has been detached from an instance and must be in the **Available** status.

Procedure

1. *Log on to the ECS console* and go to the **Elastic NIC** page.
2. Select a **Department** or **Region**, or enter an **NIC ID**, and click **Search** to find the target ENI.
3. In the **Action** column of the secondary ENI, click the icon and select **Delete** from the drop-down list.
4. In the prompt box that appears, click **Confirm**.

12.11 Manage a deployment set

12.11.1 Create a deployment set

You can create a deployment set.

Procedure

1. *Log on to the ECS console* and go to the **Deployment Set** page.
2. Click **Create Deployment Set**.
3. On the **Create Deployment Set** page, make configuration for a deployment set.

Table 12-12: Configure a deployment set

Item	Description
Region	• **Region**: Required. Select a region where the target deployment set resides.

Item	Description
	• **Zone**: Required. Select a physical zone with independent power grids and networks within a region. A zone can communicate with other zones using the Intranet, and is not affected by faults in other zones. If you want to improve application availability, we recommend that you create instances in different zones.
Configurations	• **Department**: Required. Select a department to which the target deployment set belongs. • **Project**: Required. Select a project to which the target deployment set belongs. • **Deployment Domain**: Required. Set the deployment domain to **Default** or **Switch**. • **Deployment Granularity**: Required. Set the deployment granularity to **Host Machine**, **Rack**, or **Switch**. • **Deployment Strategy**: Required. Set the deployment policy to **Loose Dispersion** or **Strict Dispersion**. • **Deployment Set Name**: Optional. Enter the name of the deployment set. • **Description**: Optional. Enter a description for the deployment set.

4. After completing the settings, click **Confirm**.

Result

You can view the created deployment set with the preceding attributes in the list of deployment sets.

12.11.2 View a deployment set

You can log on to the ECS console to view all your deployment sets and their details.

Procedure

1. *Log on to the ECS console* and go to the **Deployment Set** page.
2. Select a **Department**, **Region**, or enter a **Deployment Set Name**, and click **Search** to find the target deployment set. View detailed information about the deployment set in the list of deployment sets.

> **Note:**
> Click **Deployment Set Name**, and you can select **Deployment Set ID** or **Project ID** from the drop-down list.

12.11.3 Edit a deployment set

You can modify the name and description of an existing deployment set on the ECS console.

Procedure

1. *Log on to the ECS console* and find the deployment set you want to edit.
2. In the **Action** column of the deployment set, click the icon and select **Edit** from the drop-down list.
3. In the dialog box that appears, edit the **Name** or **Description** of the deployment set and click **Confirm**.

12.11.4 Delete a deployment set

You can delete a deployment set.

Prerequisites

ECS instances have been completely removed from the deployment set.

Procedure

1. *Log on to the ECS console* and find the deployment set you want to delete.
2. In the **Action** column of the deployment set, click the icon and select **Delete** from the drop-down list.
3. In the prompt box that appears, click **Confirm**.

12.12 Install FTP software

12.12.1 Install VSFTP in CentOS

Procedure

1. Install VSFTP. Run the `yum install vsftpd -y` command to install VSFTP.
2. Add an FTP account and directory.
   1. Check the location of *nologin*. It is usually in */usr/sbin/nologin* or */sbin/nologin*.
   2. Create an account. Run the following command to create an account with */alidata/www/wwwroot* as your PWFTP home directory. To customize your account name and directory, run `useradd -d /alidata/www/wwwroot -s /sbin/nologin pwftp`.

3. Run the following command to change the account password:

```
passwd pwftp
```

4. Run the following command to modify the permissions on the specified directory:

```
chown -R pwftp.pwftp /alidata/www/wwwroot
```

3. Configure VSFTP.

1. Run the `vi /etc/vsftpd/vsftpd.conf` command to edit the VSFTP configuration file.
2. Change **anonymous_enable=YES** to **anonymous_enable=NO** in the configuration file.
3. Remove the comment tag # before the following configuration:

```
local_enable=YES
        write_enable=YES
        chroot_local_user=YES
```

4. To save the changes, press the **ESC** key and enter the following command: `wq`.

4. Modify the shell configuration by editing */etc/shells* in the `vi` editor. If the file does not contain */usr/sbin/nologin* or */sbin/nologin* (depending on the current system configuration), add either one to the file.
5. Start VSFTP and test logon.

1. Run the `service vsftpd start` command to start VSFTP.
2. Use the account pwftp to test FTP logon. In this example, the directory is */alidata/www/wwwroot*.

12.12.2 Install VSFTP in Ubuntu and Debian

Procedure

1. Run the `apt-get install vsftpd -y` command to install VSFTP.
2. Add an FTP account and directory.

1. Check the location of nologin. It is usually in */usr/sbin/nologin* or */sbin/nologin*.
2. Create an account. Run the following command to create an account with */alidata/www/wwwroot* as your PWFTP directory. To customize your account name and directory, run `useradd -d /alidata/www/wwwroot -s /sbin/nologin pwftp`.
3. Run the `passwd pwftp` command to change the account password.
4. Run the `chown -R pwftp.pwftp /alidata/www/wwwroot` command to modify the permissions on the specified directory.

3. Configure VSFTP.

1. Run the `vi /etc/vsftpd.conf` command to edit the VSFTP configuration file.
2. Change **anonymous_enable=YES** to **anonymous_enable=NO** in the configuration file.
3. Remove the comment tag # before the following configuration:

```
local_enable=YES
        write_enable=YES
        chroot_local_user=YES
        chroot_list_enable=YES
        chroot_list_file=/etc/vsftpd.chroot_list
```

4. Save and exit.
5. Edit the */etc/vsftpd.chroot_list* file to add the FTP account to this file. Save and exit.

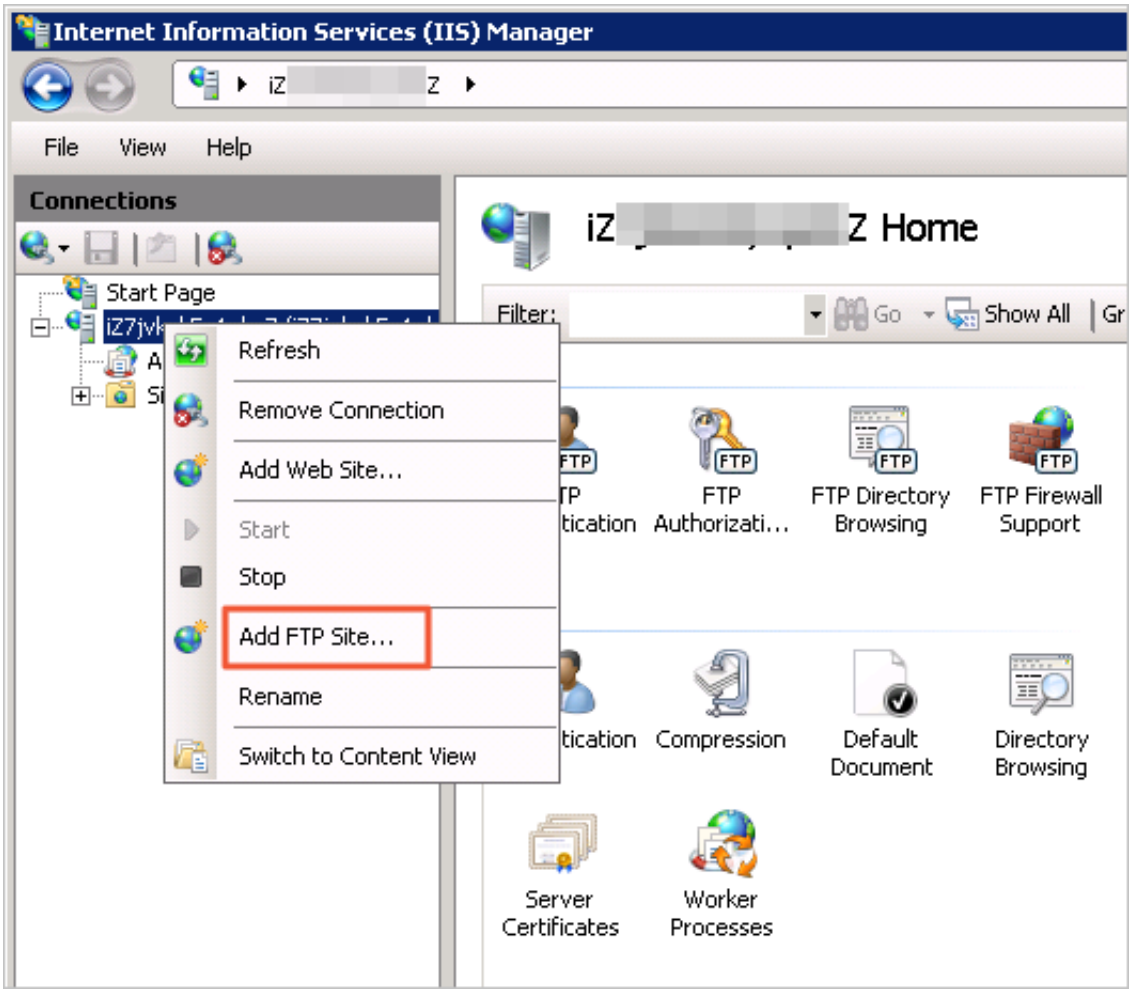
4. Modify the shell configuration by editing */etc/shells* in the `vi` editor. If the file does not contain */usr/sbin/nologin* or */sbin/nologin* (depending on the current system configuration), add either one to the file.
5. Restart VSFTP and test logon.
   1. Run the `service vsftpd restart` command to restart VSFTP.
   2. Use the account pwftp to test FTP logon. The directory is */alidata/www/wwwroot*.

12.12.3 Install and configure FTP in Windows 2008

Procedure

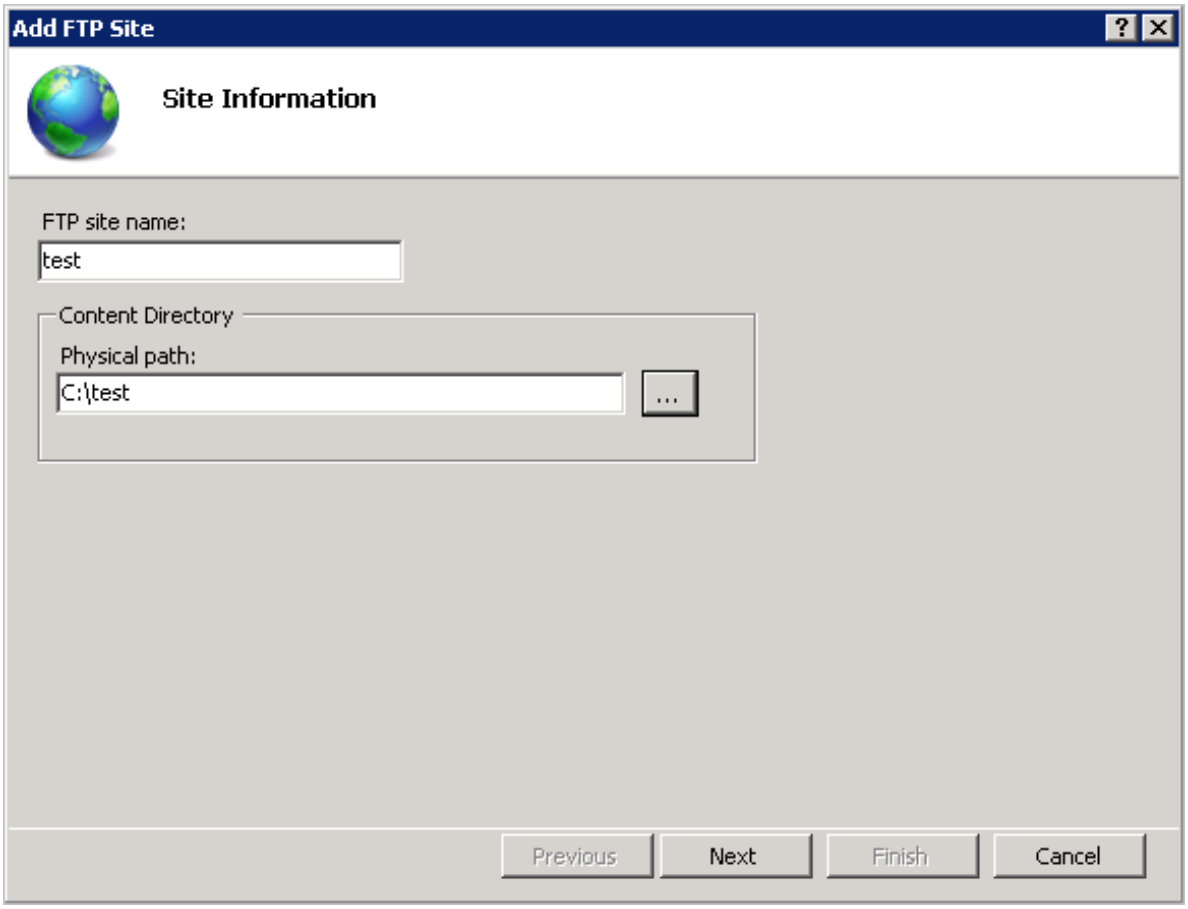
1. After you connect to your ECS instance remotely, choose **Start** > **Management Tools** > **Internet Information Services (IIS) Manager**, right-click the server name, and choose **Add FTP Site** from the shortcut menu.

Figure 12-35: Add an FTP site



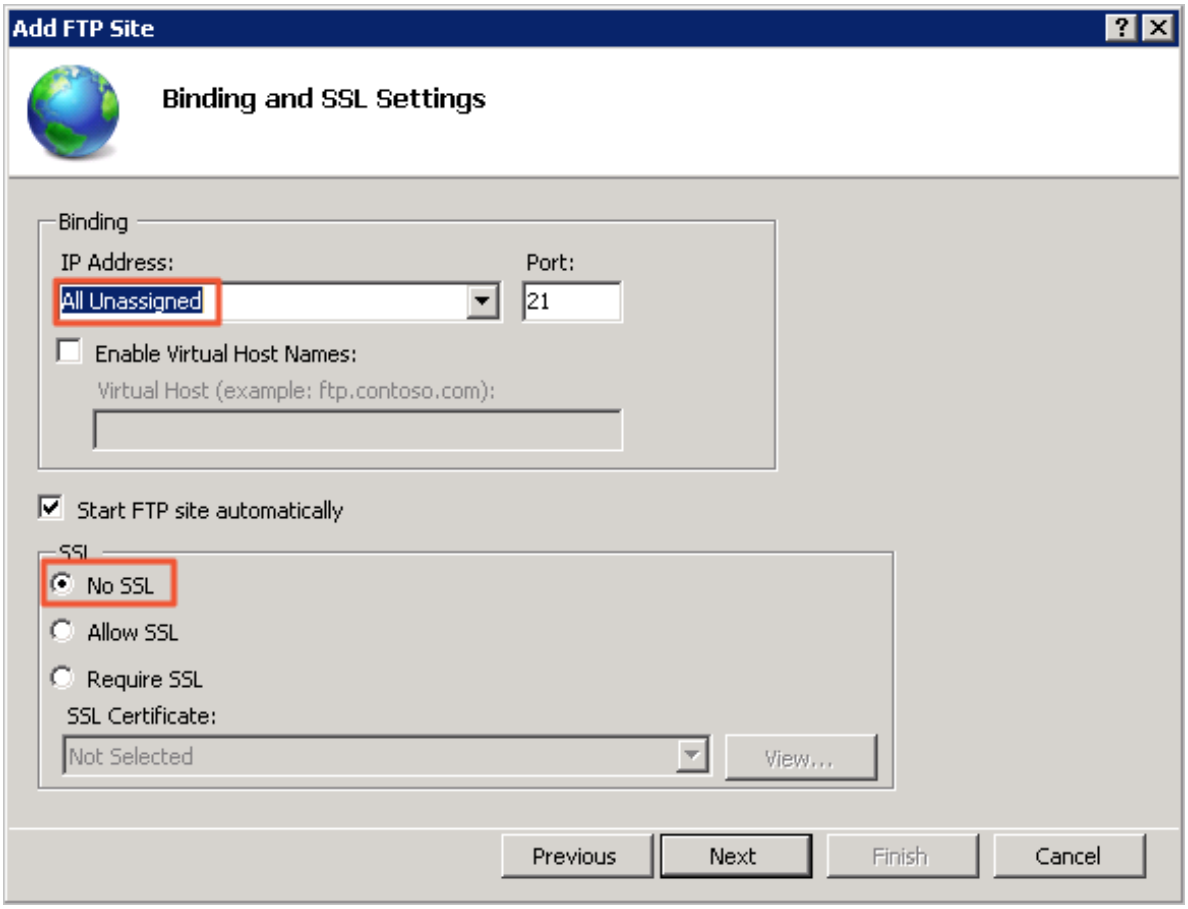
2. Enter the FTP site name and specified path, and click **Next**.

Figure 12-36: Add an FTP site



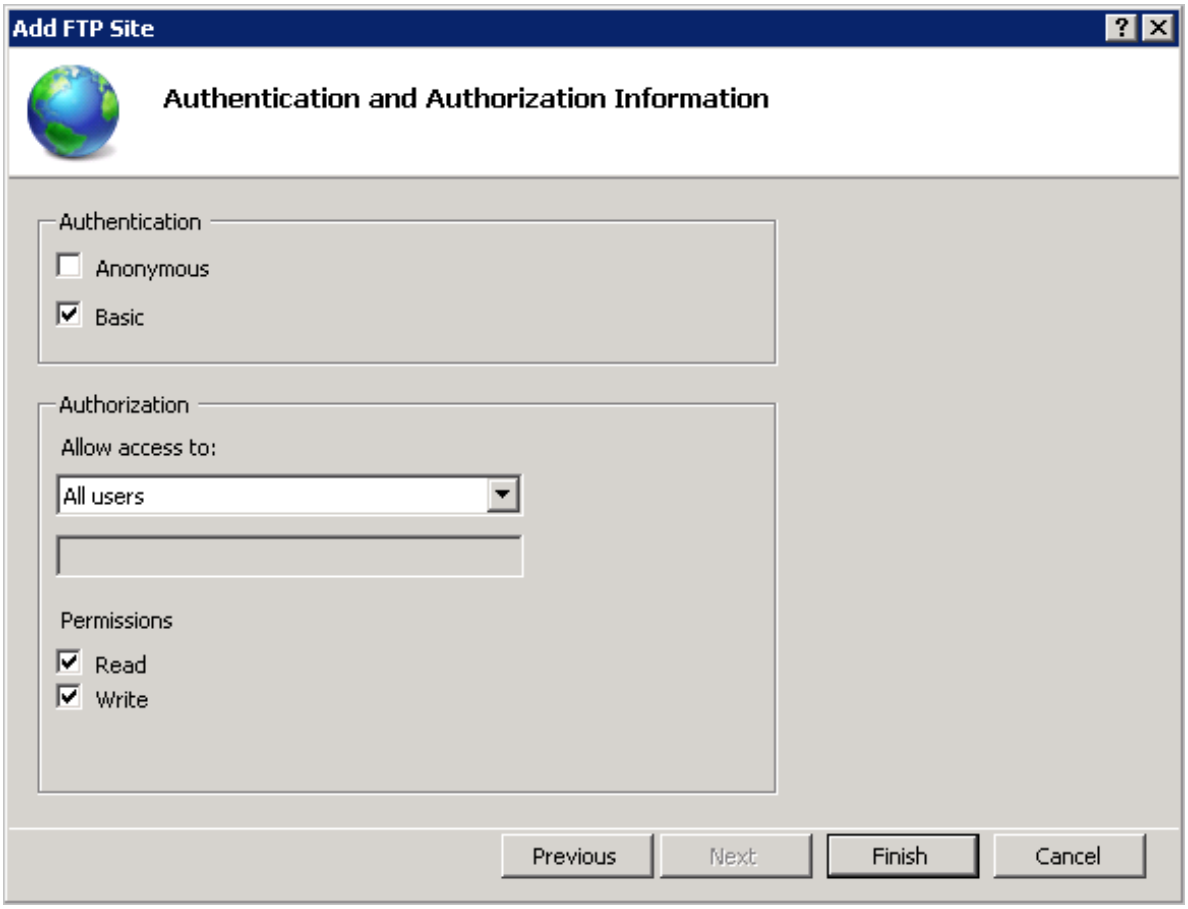
3. As shown in Figure *Figure 12-37: Bind an IP address*, set **IP Address** to **All Unassigned** and **SSL** to **None**.

Figure 12-37: Bind an IP address



4. Set **Authentication** to **Basic**, **Authorization** to **All Users**, and **Permissions** to **Read** and **Write**.

Figure 12-38: Set authentication and authorization



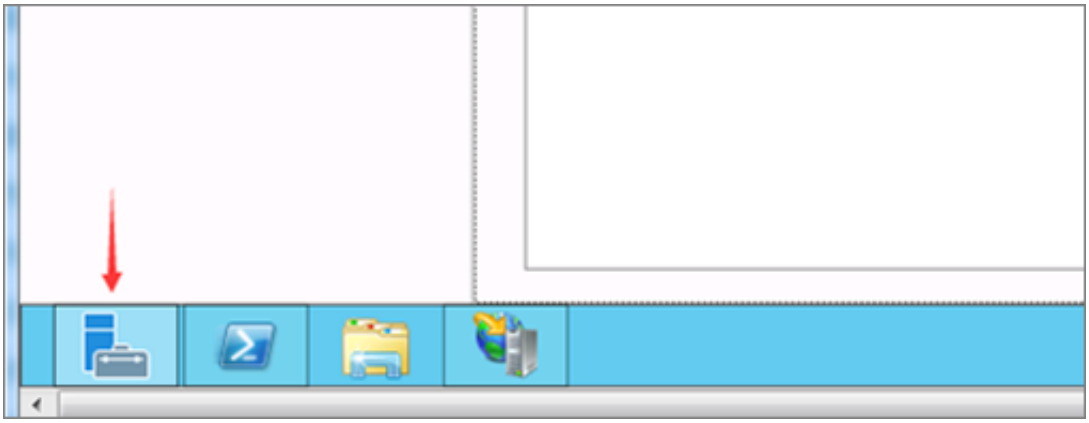
5. Click **Finish** after completing FTP settings. You can use the administrator account and password to upload and download files through FTP.

12.12.4 Install and configure IIS and FTP in ECS Windows 2012

Procedure

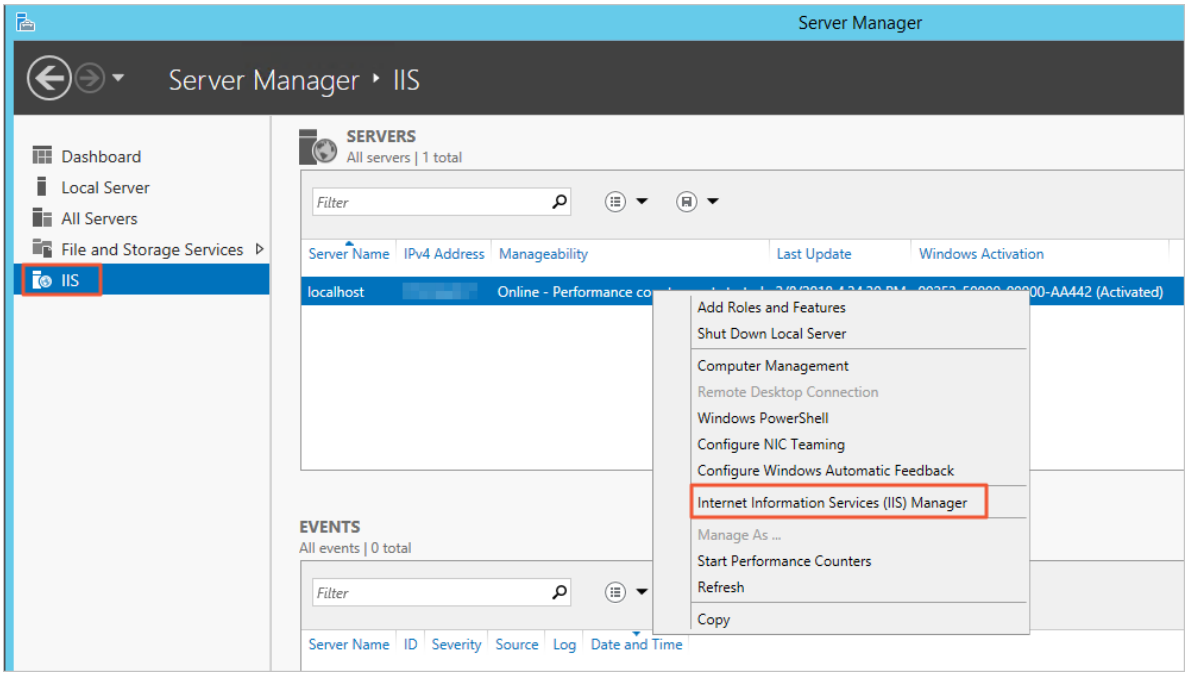
1. In the lower-left corner of the server interface, click **Server Manager** to start the **Server Manager**.

Figure 12-39: Start the server manager



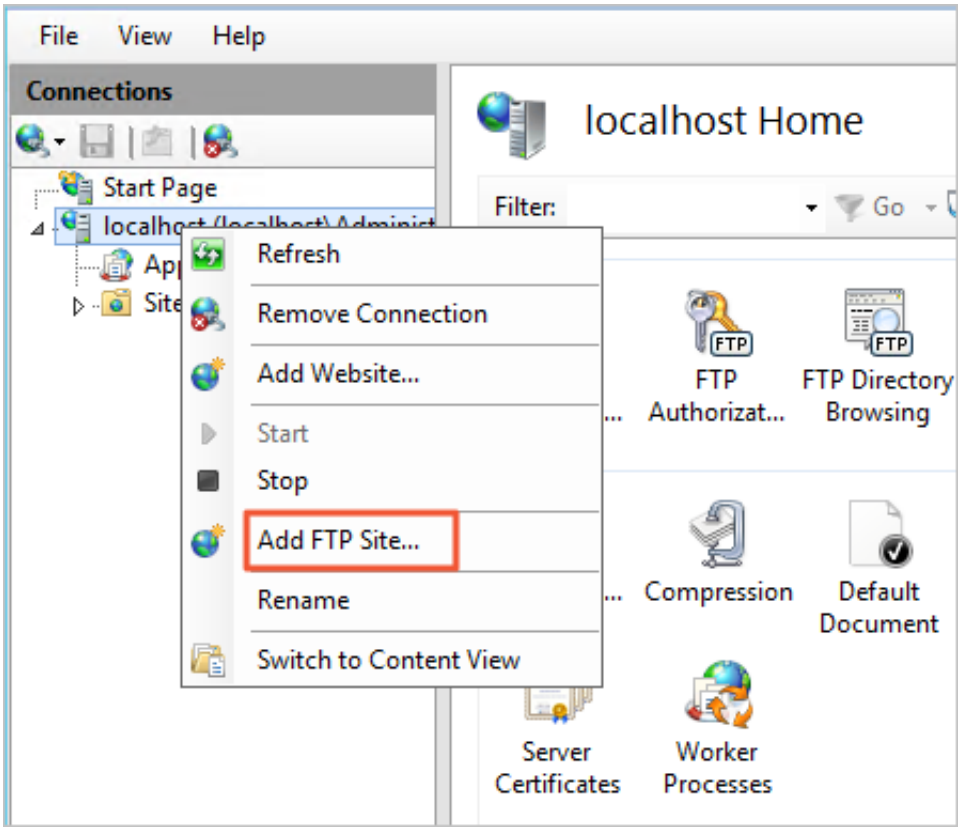
2. Start the IIS manager, as shown in the following figure.

Figure 12-40: Start the IIS manager



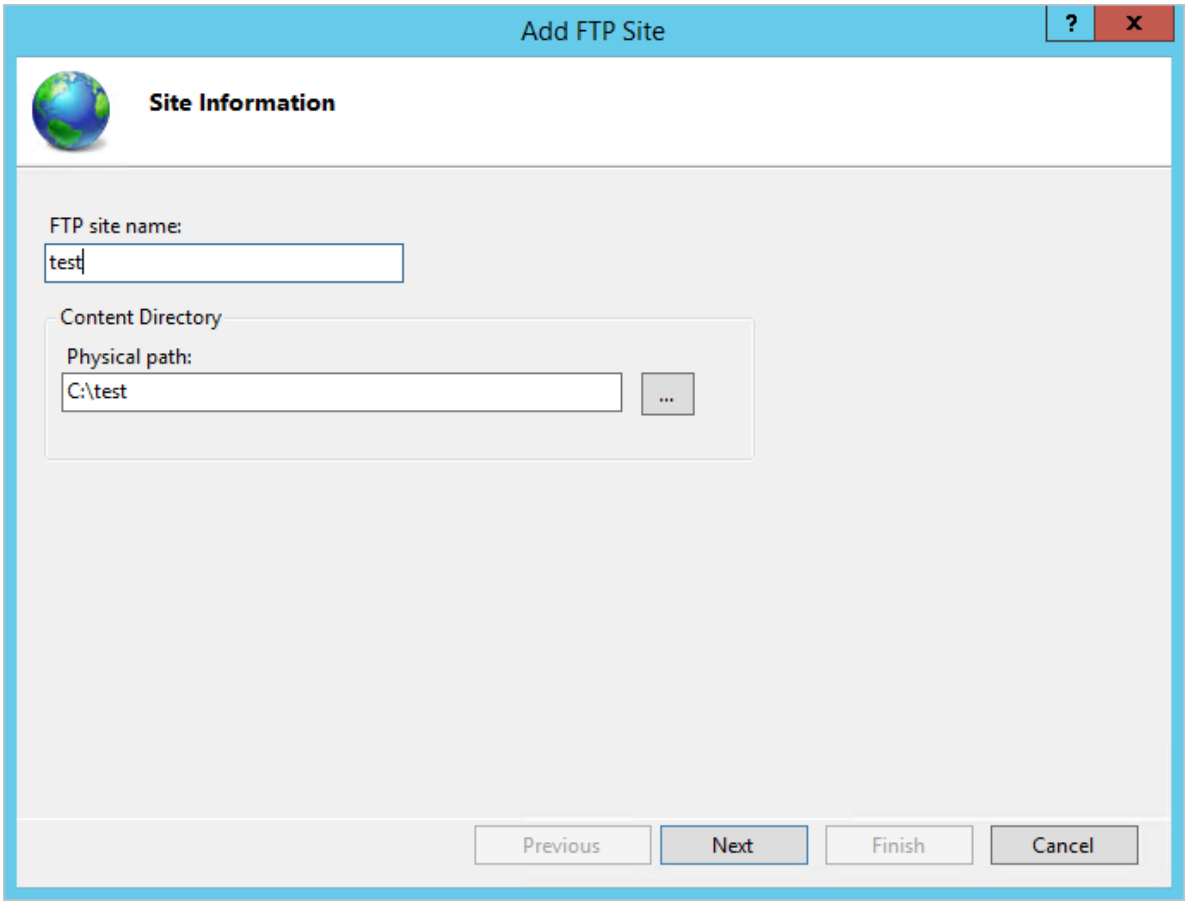
3. Add an FTP site to the IIS manager, as shown in the following figure.

Figure 12-41: Add an FTP site



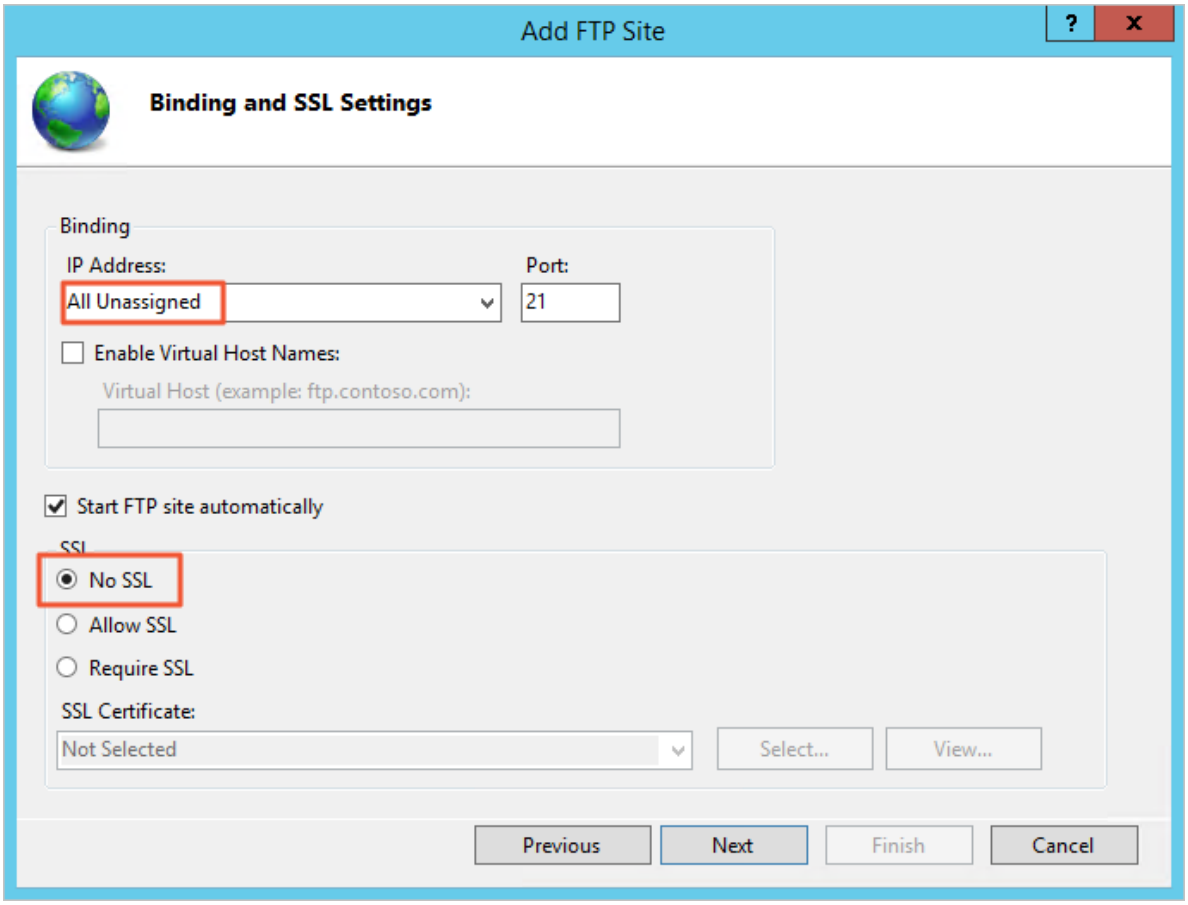
4. Enter the FTP site name and specify the FTP path.

Figure 12-42: Enter site information



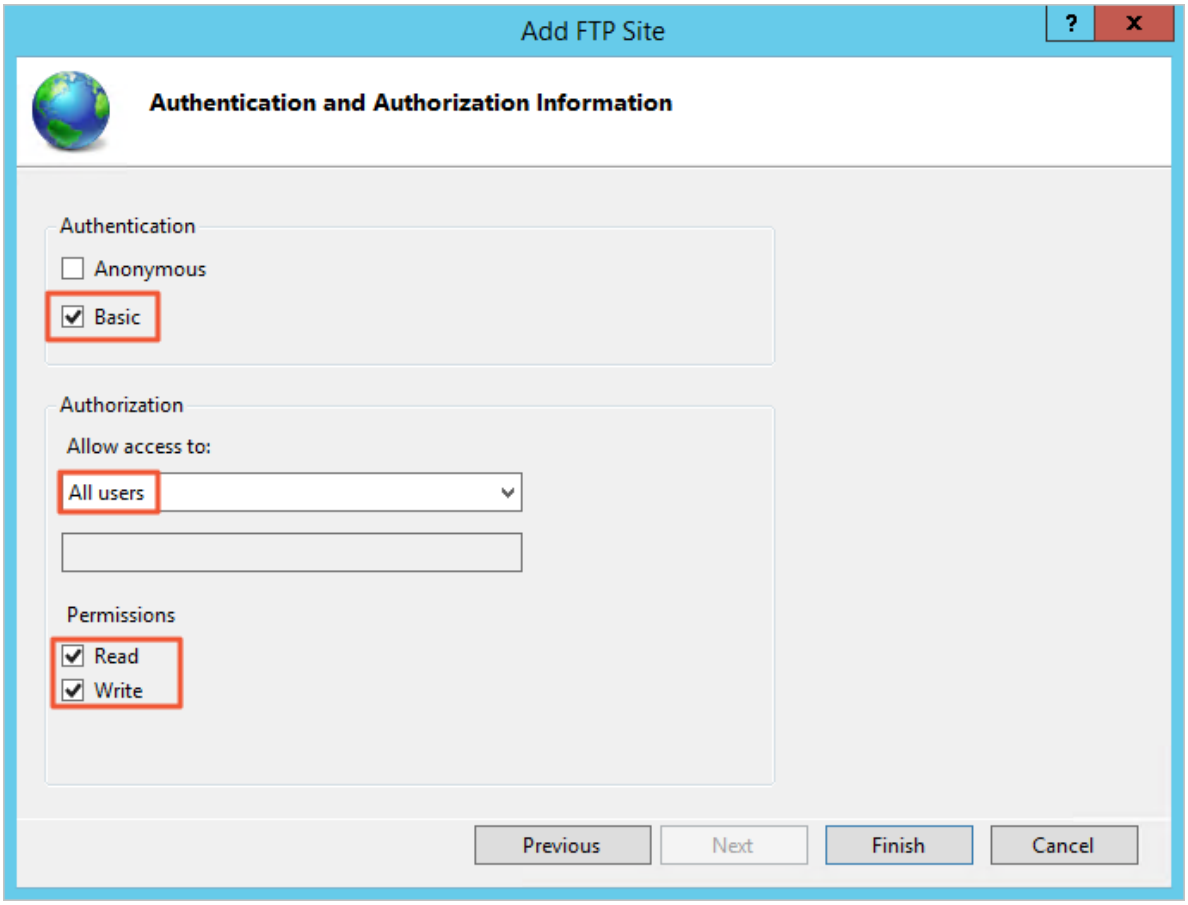
5. As shown in the following figure, set **IP Address** to **All Unassigned** and **SSL** to **None**.

Figure 12-43: Bind an IP address and set SSL



6. Set **Authentication** to **Basic**, **Authorization** to **All Users**, and **Permissions** to **Read** and **Write**.

Figure 12-44: Set authentication and authorization



7. After completing FTP settings, use the default administrator account and password to test logon. Then you can upload and download files.

13 Object Storage Service (OSS)

13.1 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.

13.2 Basic concepts

Object

In OSS, the basic data unit for user operations is an object. The maximum size of a single object is 48.8 TB. An infinite number of objects are allowed in a bucket.

You must have the write permission for a bucket in an OSS instance before uploading an object to the bucket. On the OSS console, the uploaded objects are displayed as files or folders.

Bucket

All files of OSS are stored in buckets. A bucket is a unit for managing stored files. All objects must be stored in a bucket. You can configure the attributes of a bucket to control region, file access, and file lifecycle. These attributes apply to all files stored in the bucket. Therefore, you can create buckets with different attributes to manage multiple files as required.

The storage space in a bucket is non-hierarchical, indicating that it does not use system directories. All files are directly affiliated with their corresponding buckets. However, you can group, classify, and manage relevant files using folders.

AccessKey

AccessKey (AK) is a pair of values (AccessKeyId and AccessKeySecret) used for access identity authentication. OSS verifies the identity of a request sender by using the symmetric

encryption methods of AccessKeyId and AccessKeySecret. The AccessKeyId identifies a user. The AccessKeySecret allows you to encrypt the signature string and enables OSS to verify the AccessKey of the signature string. Do not leak the AccessKeySecret to anyone.

OSS allows a bucket owner to obtain an AccessKey through the following methods:

- Directly apply for an AccessKey on the DTCenter console.
- Apply for an AccessKey for a third-party request sender through the DTCenter console.
- Apply for a third-party AccessKey on Security Token Service (STS).

13.3 Quick start

This user guide introduces how to complete basic tasks, such as how to create buckets, upload files and share files.

13.3.1 Log on to the OSS console

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

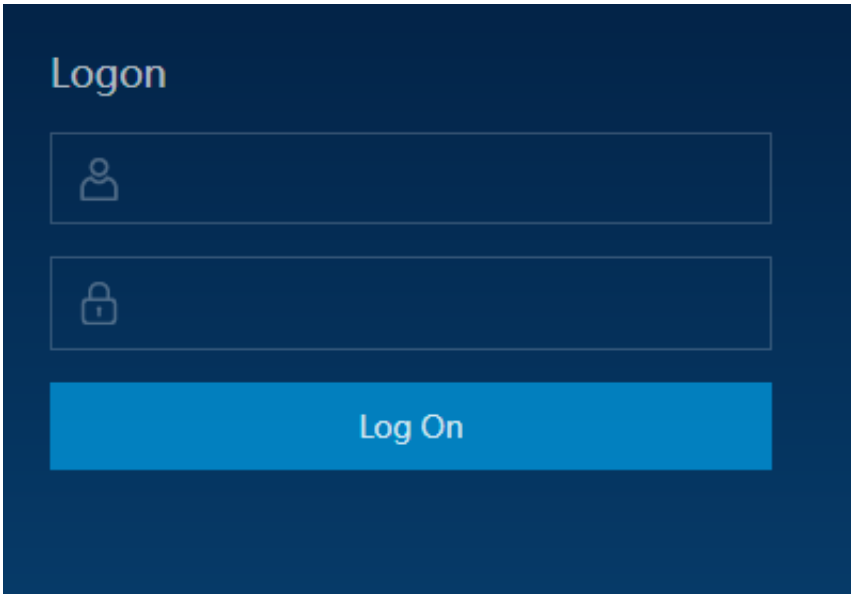
  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 13-1: Log on to the Apsara Stack console*.

Figure 13-1: Log on to the Apsara Stack console



3. Enter the correct username and password.
   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **Log On** to go to the **Home** page.
5. On the menu on the top of the page, select **Console** > **Object Storage Service**.

13.3.2 Create a bucket

Context

Before you upload any files to OSS, you must create a bucket to store files. Attributes of a bucket include its region, access permission, and other basic attributes.

Procedure

1. *Log on to the OSS console*.

2. Click **Create Bucket** to open the **Add Bucket** dialog box. Select related options.

   The parameters are described as follows:

   - **Department**: Select a department from the drop-down list.
   - **Project**: Select a project from the drop-down list.
   - **Region**: Select the region to which the bucket belongs from the drop-down list.

     Note:

     - You cannot change the region for a created bucket.
     - If access to OSS through the ECS intranet is required, you can select the same region in which your ECS is located.
   - **Permissions**: Select a permission for the OSS.
     - **Private**: Only the owner of the bucket can perform read and write operations on the files in the bucket. Other users are not allowed to access the files.
     - **Public (Read Only)**: Only the owner of the bucket can perform write operations on the files in the bucket, while anyone (including anonymous access) can perform read operations on the files.
     - **Public**: Anyone (including anonymous users) can perform read and write operations on the files in the bucket. Exercise caution when using this permission because the fees incurred by these operations will be borne by the owner of the bucket.

     Note:

     You can modify read and write permissions for a created bucket. For more information, see **Edit read/write permissions** in *Cite LeftAlibaba Cloud OSS User GuideCite Right*.
   - **Bucket Name**: Specify the name of the bucket.

     Note:

     - The bucket's name must comply with the naming rules.
     - The bucket's name must be unique in OSS.
     - You cannot change the bucket's name after it is created.
   - **Bucket Capacity**: Specify the capacity of the bucket.
   - **Instances**: Specify the number of bucket instances to apply for. You can create a maximum of 10 bucket instances at a time.

3. Click **Create**. The bucket is successfully created.

13.3.3 Upload a file

Context

After you create a bucket, you can upload all types of files to the bucket. The OSS console allows you to upload files smaller than 500 MB. To upload a file larger than 500 MB, use the software development kit (SDK) or application programming interface (API).

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket management page.
3. Click **Object Management** to go to the file management page.
4. Click **Upload Files** to open the file selection dialog box.
5. Select the file to be uploaded and click **Open**.
6. After the file is successfully uploaded, refresh the page to view the uploaded file.

   You can click **Task Management** to view the upload progress and result.

13.3.4 Obtain a file URL

Context

After you upload a file to a bucket, you can obtain a URL to share and download the file.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket management page.
3. Click **Object Management** to go to the file management page.
4. Click the icon corresponding to the file and select **Get URL** to open the **Get Object URL** dialog box.

   Note:
   If your bucket permission is set to **Private**, you must set a link validity period when you obtain the URL. Click **Get URL** to obtain the URL. The validity period of a URL link is calculated based on the Network Time Protocol (NTP). You can send this link to any user, who can then use it to access the file within the validity period. The URL obtained from a private bucket is a signed URL.

5. Copy the file URL to users so that they can browse or download the file.

13.4 Manage a bucket

Functions described in this section may differ from those in an actual project. Contact the customer manager to confirm available functions. Relevant procedures are for reference only. For more information about actual procedures, see the interface of the OSS console.

13.4.1 View a bucket

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket or the  icon corresponding to the bucket, and select **Details** to go to the bucket information page.
3. On the **Bucket Information** page, you can view detailed information about the bucket, such as the domain used to access the bucket and creation time.

13.4.2 Modify read and write permissions

Context

OSS provides Access Control List (ACL) to control permissions. After a bucket is created, ACL is set to **Private** by default. You can modify ACL after you create a bucket.

OSS ACL provides access control for all buckets. Currently, three access permissions are available for a bucket:

- **Private**: Only the owner or authorized user of a bucket can perform read and write operations on objects in this bucket. Other users are not allowed to access the object in the bucket without authorization.
- **Public-read**: Only the owner or authorized user of a bucket can perform write operations on objects in this bucket, and anyone (including anonymous users) can perform read operations on the object in this bucket.
- **Public-read-write**: Anyone (including anonymous users) can perform read and write operations on objects in a bucket. The fees incurred by these operations are borne by the owner of the buckets. Therefore, exercise caution when using this permission

Procedure

1. *Log on to the OSS console*.

2. Click the name of a bucket to go to the bucket information page.
3. Choose **Bucket Properties** > **Read/Write Permissions**.
4. Select the ACL permissions for the bucket.
5. Click **Set** to save your changes.

13.4.3 Configure static website hosting

Context

You can use the OSS console to configure your bucket to the static website hosting mode. The domain of the bucket can be used to access this static website.

If the default page is blank, static website hosting is disabled.

The default home page is displayed if you directly access the static website through the root domain or any URL (belong to the root domain) ending with a forward slash (/).

Procedure

1. *Log on to the OSS console*.
2. Click the name of the bucket to go to the bucket information page.
3. Choose **Bucket Properties** > **Website Settings**.
4. The **Default Homepage** parameter is used to set the index page (equivalent to **index.html** home page of a website). Only an HTML file that has been stored in the bucket can be used as the default home page.
5. The **Default 404 Page** parameter is set for the default page displayed when an incorrect path is accessed. Only an HTML file that has been stored in the current bucket can be used as the home page. If the **Default 404 Page** parameter is not set, the default page is disabled.
6. Click **Set** to save the static website settings.

13.4.4 Configure logging

Context

You can use the OSS console to enable or disable logging for a bucket. Logs can be stored in a new bucket or in a bucket that has enabled logging.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket information page.

3. Choose **Bucket Properties** > **Logging Settings**.
4. From the **Log Storage Location** drop-down list, select the name of a bucket that is used to store logs. Only buckets that belong to the same user and in the same region can be selected. Select **Not Stored** to directly disable logging.
5. In the **Log Prefix** box, enter the prefix of the log file names, namely, *<TargetPrefix>* in the following naming rules. Logs are recorded in the root directory. You can also add a folder path in front of *<TargetPrefix>*, such as *log/<TargetPrefix>*. Logs are recorded in the *log/* directory.

   Log naming rules

   An example of the naming rules for objects that store the access logs is as follows:

   <TargetPrefix><SourceBucket>YYYY-MM-DD-HH-MM-SS-<UniqueString>

   - *<TargetPrefix>*: log prefix specified by the user.
   - *<SourceBucket>*: name of the source bucket.
   - *YYYY-MM-DD-HH-MM-SS*: China Standard Time (UTC+8) when the log is created. *YYYY* indicates the year, *MM* (the first one) indicates the month, *DD* indicates the day, *HH* indicates the hour, *MM* (the second one) indicates the minute, and *SS* indicates the second.
   - *<UniqueString>*: a string generated by OSS.

   Naming example:

   MyLog-OSS-example2015-09-10-04-00-00-0000

   "MyLog-" is the log prefix specified by the user, "OSS-example" is the name of the source bucket, "2015-09-10-04-00-00" is the log creation time (China Standard Time), and "0000" is the string generated by OSS.
6. Click **Set** to save the logging settings.

13.4.5 Configure anti-leeching

Context

To prevent your data on OSS from being leeched, OSS supports anti-leech through settings of the referer field in the HTTP header. On the OSS console, you can configure a whitelist including the referer field for a bucket or configure whether to accept access requests with the referer field left empty.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket information page.
3. Click **Bucket Properties** > **Anti-Leech Settings**.
4. In the **Referer** box, add a whitelist URL.
5. Configure whether to accept access requests where the referer field is empty.
6. Click **Submit** to save the anti-leech settings.

For example, for a bucket named oss-example, set its whitelist of the referer field to http://www.aliyun.com. Only requests with the referer field of http://www.aliyun.com can access objects in the bucket.

13.4.6 Configure CORS

Context

You can configure Cross-origin resource sharing (CORS) based on the HTML5 protocol to help you achieve cross-origin access. When OSS receives a cross-origin request (or an OPTIONS request), it reads the CORS rules of the bucket and then checks relevant permissions. OSS checks each rule sequentially, and uses the first rule that matches to allow the request, and then returns the corresponding header. If none of the rules match, OSS does not attach any CORS header.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket information page.
3. Choose **Bucket Properties** > **CORS Rules** .
4. Click **Add Rules** to open the **Add CORS Settings** dialog box.
5. Configure the rules in the dialog box.
   - **Source**: Specify the source of the cross-origin request that is allowed. Multiple matching rules are allowed, which are separated by a carriage return (¶). Each matching rule allows only one asterisk (*) for wildcard use.
   - **Method**: Specify the allowed cross-origin request method.
   - **Allowed Header**: Specify the header for an allowed cross-origin request. Multiple matching rules are allowed, which are separated by a carriage return (¶). Each matching rule allows a maximum of one asterisk (*) for wildcard use.

- **Expose Header**: Specify the header of the response that allows the user to access from the application.
- **Cache Time**: Specify the cache time for the returned result of the browser's prefetch (OPTIONS) request to a specific resource.

> **Note:**
> A maximum of 10 rules can be configured for each bucket.

6. Click **OK** to save the rule. You can also modify or delete configured rules.

13.4.7 Manage lifecycle rules

Context

You can use the OSS console to define and manage lifecycle configuration rules for your buckets . A rule can be defined for all objects or a subset (specify the prefix of the object name as the keyword) of objects in a bucket. The configured rule automatically applies to all objects that match the rule. Therefore, you can use lifecycle management to perform various operations, such as simultaneous management of multiple files and automatic deletion of fragments.

- For objects that match the rule, the system ensures that the data is cleared within two days from the effective date.
- Data that is deleted simultaneously based on a lifecycle rule cannot be recovered. Therefore, exercise caution when configuring such a rule.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket information page.
3. Choose **Bucket Properties** > **Lifecycle Settings**.
4. Click **Add Rules** to open the **Add Lifecycle Rules** dialog box.
5. Configure lifecycle rules.
   - **Status**: Specify the status of a rule, indicating whether the rule is enabled or disabled.
   - **Policy**: Specify an object matching policy. You can select **Apply to entire bucket** or **Configure by prefix**.
   - **Prefix**: Assume that you have stored image objects in the bucket and these objects are prefixed with "img/". To manage the lifecycle of these objects, enter `img/`.
   - **Expired**: Specify the date or days for expired objects.

- **Set by Date**: Specify the date for an image object that is deleted after the specified absolute time. Exercise caution when configuring this rule because all files that are created before the date will be deleted.
- **Set by Number of Days**: Specify the days to save the image object after the specified number of days are completed. When the number of days from the last modification time of the object exceeds the specified number of days, this rule is executed to delete the object. For example, if this parameter is set to 30 days, the objects that are last modified on January 1, 2016 are scanned and deleted by OSS on January 31, 2016.

6. Click **OK** to save the rule. After the rule is successfully saved, you can view the configured lifecycle rules in the policy list and perform corresponding **Edit** or **Delete** operations.

13.4.8 Copy cross-cloud server settings

Context

You can synchronize data stored in an OSS bucket to other clouds on the OSS console. You can specify files with certain prefixes to synchronize and configure synchronization policies.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket information page.
3. Click **Bucket Properties** > **Copy Cross-Cloud Server Settings**.
4. Click **Enable Data Synchronization** to open the **Enable Data Synchronization** box.
5. Configure the following parameters in cross-cloud server setting copy, that is, data synchronization.
   - **Synchronization Target Cloud**: Select the target cloud that you want to synchronize the data to.
   - **Synchronization Target Cloud Address**: Enter the address of the target cloud.
   - **Synchronization Target Bucket**: Enter the name of the bucket that you want to synchronize the data to in the target cloud.

   Note:

   The synchronization target cloud must already have a bucket of the same name.

- **Data Synchronization Object**: Select the object that you want to synchronize. You can select **Synchronize All Files** or select **Synchronize Files with** > **Add**, and enter prefixes to synchronize files with them.
- **Data Synchronization Policy**: Select the data synchronization policy. You can select **Write Synchronization** or **Add/Delete/Modify**。
- **Synchronize Historical Data**: Select whehter to synchronize historical data.

6. Click **Confirm** to save the settings. After the setting is saved, you can check the set cross-cloud copy ruls in the setting list. You can also **Edit** or **Delete** a setting in the **Action** column.

13.4.9 Delete a bucket

Prerequisites

Before you delete a bucket, ensure that all files in this bucket are cleared, including file fragments caused by incomplete multipart upload. Otherwise, the bucket cannot be deleted.

Procedure

1. *Log on to the OSS console*.
2. Click the  icon next to the target OSS, and select **Delete** to open the **Delete Bucket** dialog box.
3. Click **OK** to delete the bucket.
4. To delete multiple buckets simultaneously, select multiple buckets and click **Delete**.

13.4.10 Change the capacity

Context

After you create a bucket, you can change the capacity of the bucket as required.

Procedure

1. *Log on to the OSS console*.
2. Click the  icon corresponding to the bucket and select **Change Capacity** to open the **Change Capacity** dialog box.
3. Change the capacity of the bucket, and click **OK**.

13.4.11 Change ownership

Context

After you create a bucket, you can change the ownership (department and project) of a bucket as required.

Procedure

1. *Log on to the OSS console*.
2. Click the icon corresponding to the bucket and select **Change Ownership** to open the **Change Ownership** dialog box.
3. Change the department and project for the bucket, and click **OK**.

13.5 Manage an object

13.5.1 Create a folder

Context

OSS does not use folders. All elements are stored as objects. To use a folder on the OSS console, you actually create an object sized **0** MB and ends with a forward slash (/). This is to sort the same type of files and process them simultaneously. By default, the console displays an object that ends with a forward slash (/) as a folder. This object can be uploaded and downloaded. You can use OSS folders on the OSS console like folders in the Windows operating system.

Note:
The console displays any object that ends with a forward slash (/) as a folder, no matter whether the object contains any data. You can only use the API or SDK to download this object.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket information page.
3. Click **Object Management** to go to the file management page.
4. Click **Create Folder** to open the **Create Folder** dialog box.
5. In the **Folder Name** box, enter the name of the folder.
6. Click **OK** to save the created folder.

13.5.2 Search for a file

Context

This topic describes how to use the OSS console to search for objects whose name begin with the same prefix in a bucket or folder.

When you search for an object by the name prefix, the search string is case-sensitive and cannot contain forward slashes (/). The search range is limited to the root buckets displayed on the file management page or the objects in the current folder (excluding sub-folders and objects in them).

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket information page.
3. Click **Object Management** to go to the file management page.
4. Enter the prefix, for example, "alibaba", in the search box, and press **Enter** or click **Search**. The system lists the names of the objects and folders prefixed with "alibaba" in the root directory of the bucket.

   To search within a folder, open the folder and enter a prefix in the search box. The system lists the names of the objects and folders matching the search prefix in the root directory of the folder.

13.5.3 Configure ACL

Context

You can configure ACL for a single file on the OSS console.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket information page.
3. Click **Object Management** to go to the file management page.
4. Click the  icon corresponding to the file and select **Set File ACL** to open the **Set File ACL** dialog box.
5. Select appropriate read and write permissions from the **Set ACL** drop-down list.
6. Click **OK**.

13.5.4 Delete a file

Context

If you do not need to store uploaded files any longer, delete them to release the occupied space. You can delete one or more files simultaneously on the OSS console.

You can delete up to 50 files simultaneously on the console. To delete only selected files or delete more files simultaneously, use the SDK or API.

Note:
Deleted files cannot be recovered. Therefore, exercise caution when deleting files.

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket information page.
3. Click **Object Management** to go to file management page.
4. Click the icon corresponding to the file and select **Delete** to open the **Delete Object** dialog box.

   Note:
   A folder may fail to be deleted if it contains too many files.
5. Click **OK** to delete the folder.

13.6 Process an image

Image Processing (IMG) is a massive, safe, cost-effective and highly-reliable image processing service provided to external users by OSS.

After you upload source images to OSS, you can process images anytime, anywhere, and on any Internet device through a simple Representational State Transfer (RESTful) API.

IMG provides an image processing API. You must use the OSS upload API to upload images. You can set up image-related services based on IMG.

13.6.1 Create a style

Procedure

1. *Log on to the OSS console*.
2. Click the name of a bucket to go to the bucket management page.
3. Click **Image Processing**.
4. Click **Create Style** to open the **Create Style** dialog box.

The parameters are described as follows:

- **Rule Name**: Specify the name of an image based on the naming rules.
- **Edit Type**: Select **Basic editing** to edit the image style through graphical operations. You can also select **Advanced editing** to use an SDK or a parameter to edit the image style.
- **Preview**: Select the image preview mode.
- **Thumbnail Style**: Set the scaling mode for the image.

  Note:
  The "long side" refers to the side with a larger ratio between the source size and the target size. The same applies to the "short side". For example, for an original image that is scaled from 400x200 to 800x100, the original-to-target ratios are 0.5 (400/800) and 2 (200/100). The side with 200 is the longer side, and the side with 400 is the shorter one because 0.5 is less than 2.
- **Thumbnail Width**: Set the scaling size for the image.
- **Thumbnail Limit**: Set whether to restrict scaling of the image.
- **Fit Direction**: Set the adaptive direction for the image.
- **Image Processing**: Set whether special processing is required for the image.
- **Picture Quality**: Set the image quality.
- **Save Format**: Set the format in which you want to save your image.
- **Add Watermark**: Set the image watermark mode.

5. After you edit the image style, click **Submit** to save the style.
6. After you submit the style, click **Export Style** to download the style to your local device.

13.6.2 Protect a source image

Context

To avoid image piracy risks, image URLs must be restricted so that only thumbnails or watermarked images can be obtained. On the OSS console, you can enable source image protection. After source image protection is enabled, protected image files can only be accessed through a URL with the style name or a signed URL. Direct access to the OSS source file or access by passing in image parameters and modifying the image style is not allowed.

Procedure

1. *Log on to the OSS console*.

2. Click the name of a bucket to go to the bucket information page.
3. Choose **Image Processing** > **Service Management**.
4. Click **Edit** and select whether to enable source image protection.
5. To enable source image protection, you must also set **`File Extensions for Source Image Protection`** to restrict access to source images with one or more suffixes.

   Source image protection is intended for image files. Therefore, you must configure file extensions of the images to be protected. For example, if `.jpg` is configured as the protected image format, `.png` source images can still be accessed.
6. Click **Save** to save the settings.

13.7 Create a single tunnel

Context

You can create a single tunnel between OSS and VPC to access resources in OSS through VPC.

Procedure

1. *Log on to the OSS console*.
2. Click the **OSS VPC Access Control** tab page.
3. Click **Create Single Tunnel** to open the **Create Single Tunnel** dialog box.
4. Set the following attributes for the single tunnel:
   - **Region**: Select a region.
   - **Department**: Select a department or all departments.
   - **Description**: Enter description of the single tunnel.
   - **VPC**: Select a VPC. For more information about the creation of a VPC, see **Create VPC and VSwitch** in *Cite LeftVPC User GuideCite Right*.
   - **VSwitch**: Select a VSwitch. For more information about the creation of a VPC, see **Create VSwitch** in *Cite LeftVPC User GuideCite Right*.
5. Click **OK**.

14 Table Store

14.1 What is Table Store

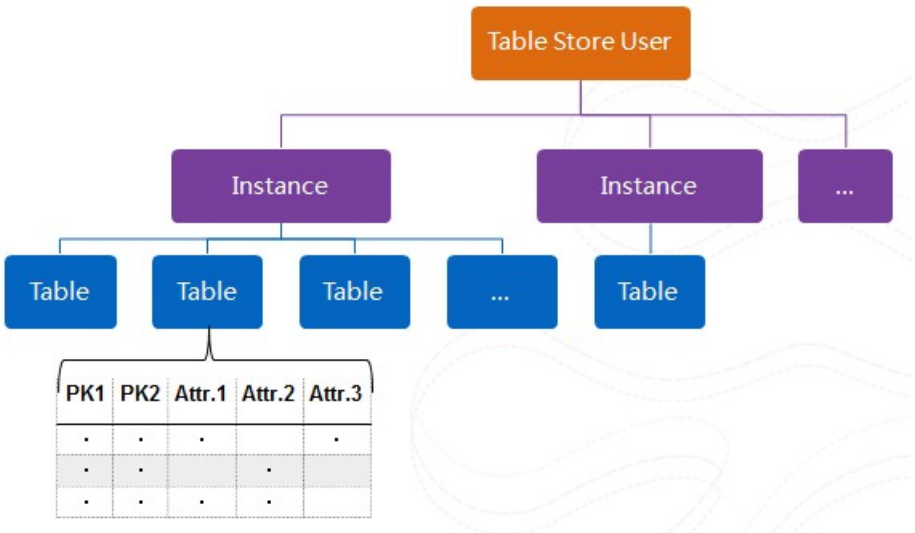
Table Store is a NoSQL database service built on Alibaba Cloud's Apsara distributed file system that can store and access massive structured data in real time.

Table Store allows users to:

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

14.2 Introduction to instances

Figure 14-1: Instances



Instances are entities designed for you to use and manage Table Store. After activating Table Store, you can create instances on the cloud console. You can create and manage tables in the instances.

Instances are the basic units for Table Store resource management. Table Store provides application access control and resource measurement at the instance level.

You can create different instances for different businesses to manage related tables or create different instances for development tests and production environments of the same business. By default, Table Store supports a maximum of 1,024 instances under a cloud account and a maximum of 1,024 tables in each instance.

The following table describes the naming rules of Table Store.

Resource Name	Naming Rules	Example
Instance Name	3 to 16 bytes in length. The character set includes [a-z, A-Z, 0-9] and hyphen (-). The instance name must start with a letter and cannot end with a hyphen (-).	test-instance

14.3 Quick start

This operation instruction describes how to quickly complete some basic tasks, including creating instances and tables.

14.3.1 Log on to the Table Store console

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

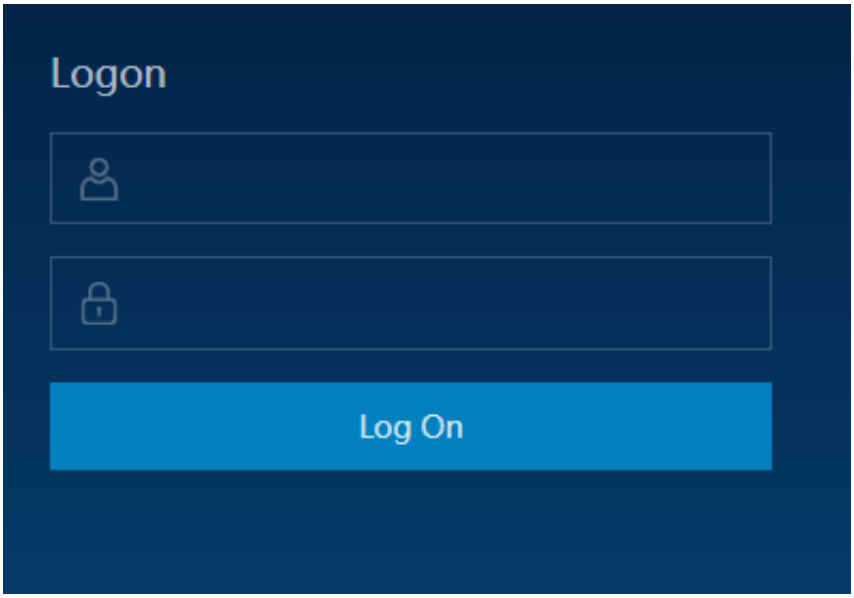
  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 14-2: Log on to the Apsara Stack console*.

Figure 14-2: Log on to the Apsara Stack console



3. Enter the correct username and password.
   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **Log On** to go to the **Home** page.
5. On the menu on the top of the page, select **Console** > **Table Store**.

14.3.2 Create an instance

Procedure

1. *Log on to the Table Store console*.
2. On the Table Store page, click **Create Instance**.
3. Enter the instance name and set the department, project, region, and instance specifications.

Note:

The instance specifications depend on the cluster you customized.

4. Click **Confirm**.

   The new instance is displayed in the list.

14.3.3 Create a table

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance to be managed and click the instance name to go to the **Instance Details** page.
3. Click the **Data Table List** tab.
4. On the **Data Table List** tab, click **Create Data Table**.

   Note:
   You can create a maximum of 64 data tables in an instance.

5. Enter the data table information.

   Table 14-1: Data table parameters

Parameter	Description
Data Table Name	A data table name can contain uppercase/ lowercase letters, numbers, and underlines (_). It must start with a letter or underline (_). The data table name must be unique at the instance level.
Reserved Read Throughput	The reserved read/write throughput can be set to 0. When the reserved read/ write throughput is larger than 0, Table Store allocates and reserves correspond ing resources for the table based on the configuration. The value ranges from 0 to 5000 and must be an integer. Capacity-type instances do not support this parameter.
Reserved Write Throughput	
Data Life Cycle	The minimum data life cycle is 86,400s (one day) or –1 (never expiring).
Maximum Data Version	A non-zero value.

Parameter	Description
	Maximum Data Version indicates the maximum number of data versions that can be stored in each attribute column of a data table. When the number of versions in an attribute column exceeds the parameter value, the earliest version will be deleted asynchronously.
Valid Data Version Margin	The offset of the version of all written data columns from the data write time must be within the range of the valid data version offset. Otherwise, data write may fail. The valid version range of an attribute column is: `[data write time - valid data version margin, data write time + valid data version margin)`.
Table Primary Key	A maximum of four primary keys can be set. The first primary key is the partition key by default. Click **Add Primary Key** to add a new primary key. The primary key type can be **`Integer`** or **`String`**. Once set, the primary key configuration and the key order cannot be modified. The primary key name can contain uppercase /lowercase letters, numbers, and underlines (_). It must start with a letter or underline (_).

6. Click **Confirm**.

   The system automatically returns to the **Data Table List** page and displays the table creation result. After the table is created, it is displayed in the data table list.

14.4 Manage instances

14.4.1 View an instance

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance to be viewed and click the instance name to go to the **Instance Details** page.

The following information is displayed: The status, region, creation time, Internet and internal network access addresses of the instance, as well as whether the instance is bound to VPC.

14.4.2 Release an instance

Prerequisites

Before releasing an instance, delete all tables from the instance. Otherwise, the instance cannot be released.

Procedure

1. *Log on to the Table Store console*.
2. Click the  icon next to the instance to be released and select **Release**.
3. In the **Delete** window, click **Confirm**.

14.5 Manage data tables

14.5.1 Update a table

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance to be managed and click the instance name to go to the **Instance Details** page.
3. Click the **Data Table List** tab.
4. In the data table list, locate the table to be updated, click the  icon on the right, and select **Adjust Read/Write Throughput**.
5. Enter the parameters to be updated.
6. Click **Confirm**. The system returns to the **Data Table List** page and displays the parameter values that take effect immediately.

14.5.2 View details of a data table

Context

You can view the basic information and actual usage of a table on the table management page. The information includes:

- Data table name
- Reserved read/write throughput

- Last modification time
- Primary keys (sorted in the sequence specified during table creation)

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance to be viewed and click the instance name to go to the **Instance Details** page.
3. Click the **Data Table List** tab.
4. Click the name of the table to be viewed to go to the **Basic Information** page of the data table. The basic information about the table is displayed.

14.5.3 Delete a table

Context

> **Note:**
> After a data table is deleted, the data in the table cannot be restored.

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance to be managed and click the instance name to go to the **Instance Details** page.
3. Click the **Data Table List** tab.
4. In the data table list, locate the table to be updated, click the icon on the right, and select **Release**.
5. In the displayed dialog box, click **Confirm**.
   After the deletion is confirmed, the table and the data in the table will be deleted permanently.

14.6 Manage VPC instances

Virtual Private Cloud (VPC) is an isolated network environment built on Apsara Stack. You can take full control of your virtual network instance. For example, you can select a private IP address range, allocate network segments, or configure a route table and gateway. You can also connect a VPC instance to a traditional data center through a leased line or VPN to build an on-demand network environment, achieving smooth cloud migration.

Prerequisites

- You must create a VPC instance first. Select an appropriate node when creating a VPC instance and ensure that the VPC and Table Store instances are in the same node. For more information about how to create a VPC instance, see **Create a VPC Instance** in *Cite LeftVPC User GuideCite Right*.
- After the VPC instance is created, create an ECS instance in the VPC instance. For more information about how to create an ECS instance, see **Create an ECS Instance** in *Cite LeftVPC User GuideCite Right*.

Procedure

1. *Log on to the Table Store console*.
2. Locate the instance to be managed and click the instance name to go to the **Instance Details** page.
3. Click **Bind VPC** to go to the instance and VPC instance binding page.
4. Enter the information and click **Confirm**.
5. After the instance is bound to the VPC instance, the system automatically returns to the instance details page. Information about the bound VPC instance is displayed in the VPC instance list. Click the link in the VPC instance ID column. The Table Store instances bound to the VPC instance and the VPC information list are displayed.

   You can use the ECS instance in the VPC instance to access a Table Store endpoint through the VPC instance access address.

What's next

After use, you can click the  icon next to the VPC instance in the VPC instance list and select **Unbind** to delete binding between the ECS instance and the VPC instance.

After the instance is unbound from the VPC instance, the ECS instance in the VPC instance cannot access Table Store through the preceding address. To access Table Store, you need to bind the instance to the VPC instance again.

14.7 Appendix: Restrictions

The following table describes the restrictions for Table Store. Some of the limit ranges indicate the maximum available values instead of the suggested values. For better performance, set the table structure and data size in a single row properly based on actual conditions and adjust the following configurations.

Item	Limit Range	Description
Number of instances under an Alibaba Cloud account	1024	To raise the limit, contact the technical support staff.
Number of tables in an instance	1024	To raise the limit, contact the technical support staff.
Instance Name Length	3-16 bytes	The character set includes [a-z, A-Z, 0-9] and hyphen (-). It must start with a letter and cannot end with a hyphen (-).
Table Name Length	1-255 bytes	The character set includes [a-z, A-Z, 0-9] and underline (_). The table name must start with a letter or underline (_).
Column Name Length	1-255 bytes	The character set includes [a-z, A-Z, 0-9] and underline (_). The name must start with a letter or underline (_).
Number of Columns Contained in a Primary Key	1-4	A primary key can contain one to four columns.
Size of String Type Primary Key Column Values	1 KB	The values of the String type columns in a single primary key column cannot exceed 1 KB.
Size of String Type Attribute Column Values	2 MB	The values of the String type columns in a single attribute column cannot exceed 2 MB.
Size of Binary Primary Key Column Values	1 KB	The values of the Binary columns in a single primary key column cannot exceed 1 KB.
Size of Binary Attribute Column Values	2 MB	The values of the Binary columns in a single attribute column cannot exceed 2 MB.
Number of Attribute Columns in a Row	Unlimited	The number of attribute columns in a single row is unlimited.
Number of attribute columns written in a single request.	1024	During the PutRow, UpdateRow, or BatchWriteRow operation, the number of attribute columns written in a single row cannot exceed 1024.
Data Size of a Single Row	Unlimited	The total size of all column names and column values for a single row are unlimited.

15 Network Attached Storage (NAS)

15.1 What is NAS

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

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

15.2 Restrictions

- Currently, NAS supports the NFSv3 and NFSv4 protocols.
- Attributes not supported by NFSv4.0 include FATTR4_MIMETYPE, FATTR4_QUOTA_AVAIL_HARD, FATTR4_QUOTA_AVAIL_SOFT, FATTR4_QUOTA_USED, FATTR4_TIME_BACKUP, and FATTR4_TIME_CREATE. The client displays an NFS4ERR_ATTRNOTSUPP error.
- Attributes not supported by NFSv4.1 include FATTR4_DIR_NOTIF_DELAY, FATTR4_DIRENT_NOTIF_DELAY, FATTR4_DACL, FATTR4_SACL, FATTR4_CHANGE_POLICY, FATTR4_FS_STATUS, FATTR4_LAYOUT_HINT, FATTR4_LAYOUT_TYPES, FATTR4_LAYOUT_ALIGNMENT, FATTR4_FS_LOCATIONS_INFO, FATTR4_MDSTHRESHOLD, FATTR4_RETENTION_GET, FATTR4_RETENTION_SET, FATTR4_RETENTEVT_GET, FATTR4_RETENTEVT_SET, FATTR4_RETENTION_HOLD, FATTR4_MODE_SET_MASKED, and FATTR4_FS_CHARSET_CAP. The client displays an NFS4ERR_ATTRNOTSUPP error.
- OPs not supported by NFSv4 include OP_DELEGPURGE, OP_DELEGRETURN, and NFS4_OP_OPENATTR. The client displays an NFS4ERR_NOTSUPP error.
- NFSv4 currently does not support Delegation.
- About UID and GID:
  - For the NFSv3 protocol, if the UID or GID of the file exists in a Linux local account, the corresponding user name and group name are displayed based on the mapping between

the local UID and GID. If the UID or GID of the file does not exist in the local account, the UID and GID are displayed.

- For the NFSv4 protocol, if the version of the local Linux kernel is earlier than 3.0, the UIDs and GIDs of all files are displayed as "nobody". If the version is later than 3.0, the display rule is the same as that of NFSv3.

> **Note:**
> If you use NFSv4 to mount the file system and the version of your Linux kernel is earlier than 3.0, we recommend that you keep the owner or group of the file or directory unchanged. Otherwise, the UID and GID of the file or directory will change to "nobody".

- You can mount a file system on a maximum of 10,000 computing nodes.

15.3 Quick start

This user guide describes how to quickly complete basic tasks, including creating a file system, adding a mount point to the file system, and mounting the file system.

15.3.1 Log on to the NAS console

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

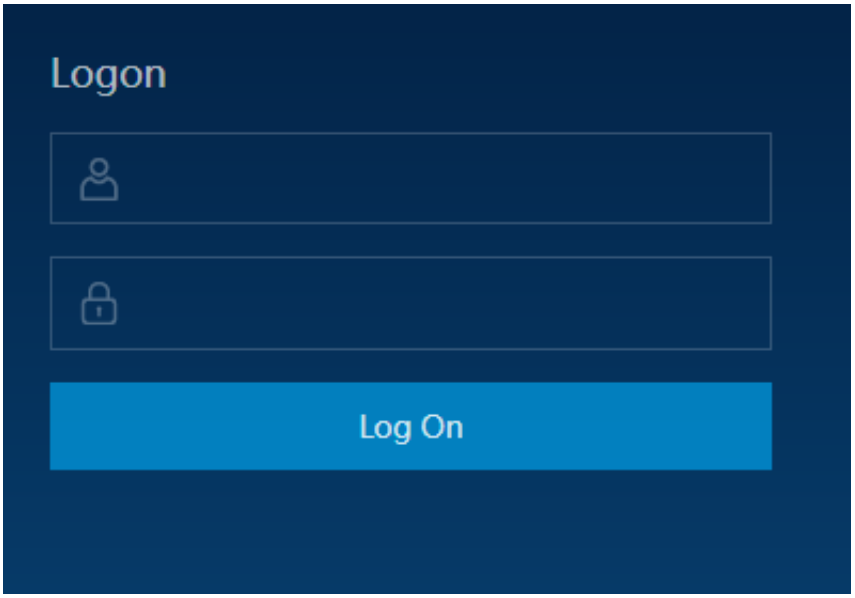
  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 15-1: Log on to the Apsara Stack console*.

Figure 15-1: Log on to the Apsara Stack console



3. Enter the correct username and password.
   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **Log On** to go to the **Home** page.
5. On the menu on the top of the page, select **Console** > **Network Attached Storage (NAS)**.

15.3.2 Create a file system

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS**page, click **Create File System**.

Note:

- You can create up to 1,000 file systems.

- The upper limit of the file system capacity is 10 PB.
- To raise the limit, contact the administrator.

3. On the **Create NAS File System** page, configure the parameters.

   The parameters are shown in *Table 15-1: Description of parameters*.

   Table 15-1: Description of parameters

Parameter	Description
Region	Select a region from the drop-down list.
Department	Select a department from the drop-down list.
Project	Select a project from the drop-down list.
File System Name	Enter a name for the file system.
Storage Type	Select **`Capacity Type`**.
Protocol Type	Select **`NFS`**.

4. Click **Confirm** to create the file system.

15.3.3 Create a permission group

Context

In NAS, each permission group has an IP address whitelist. You can add rules to the permission group to allow specified IP addresses or IP address segments to access the file system, and assign different levels of access permissions to different IP addresses or IP address segments.

Warning:
To guarantee the security of your data, we recommend that you add permission group rules carefully and only authorize necessary IP addresses to access data.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the **Permission Group** tab.
3. Click **Create Permission Group**.

   Note:
   You can create up to 100 permission groups. To raise the limit, contact the administrator.

4. In the **Create Permission Group** dialog box, set the parameters.

The parameters are described in *Table 15-2: Description of parameters*.

Table 15-2: Description of parameters

Parameter	Description
Region	Select a region from the drop-down list.
Department	Select a department from the drop-down list.
Project	Select a project from the drop-down list.
Permission Group Name	Enter a name for the file system.
Network Type	Select **VPC** or **Classic Network**.

5. Click **Confirm** to create the permsiion group.

15.3.4 Create a permission group rule

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, click the **Permission Group** tab, select a permission group and click its name to enter the **Rule list** page of this permission group.
3. Click **Create Rule**.

> **Note:**
> You can create up to 1,000 permission group rules. To raise the limit, contact the administrator.

4. In the **Add Rule** dialog box, set the parameters.

   The parameters are listed in *Table 15-3: Description of permission group rules*.

Table 15-3: Description of permission group rules

Attribute	Value	Description
Authorized IP Address	IP address or IP address segment (You can only enter one IP address in a classic network.)	IP address or IP address segment of the object authorized by the rule.
Read/Write Permissions	Read-Only; Read and Write	Allows the authorized object to perform read-only or read and write operations on the file system.

Attribute	Value	Description
User Permissions	Do Not Limit root User (no_squash); Limit root User (root_squash); Limit All Users (all_squash)	Whether to limit the permission of the authorized objects Linux system users in the file system. When the access permission of a file or directory is determined: • **`Do Not Limit root User ( no_squash)`** allows the root users to access the file system. • **`Limit root User (root_squash)`** considers the root users as nobody. • **`Limit All Users (all_squash )`** considers all users including the root users as nobody.
Priority	1 to 100, with 1 as the highest priority	When the same authorized object matches multiple rules, the rule with the highest priority overwrites the rest of the rules.

5. Click **Confirm** to create the permission group rule.

15.3.5 Add a mount point

After creating a file system and a permission group, you must add a mount point to the file system to mount the file system on computing nodes (ECS, E-HPC, or Container Service instance). NAS supports the mount points of the VPC and classic network types.

Procedure

1. *Log on to the NAS console*.
2. On the **File Storage NAS** page, select and click a file system ID to go to the **System Details** page.
3. Click the **Mount Point** tab.
4. Click **Add Mount Point**.

   Note:
   You can create up to 100 mount points. To raise the limit, contact the administrator.

5. In the **Add Mount Point** dialog box, set the parameters.
   - If **`Mount Point Type`** is set to **`Classic Network`**, select a permission group from the **`permission group`** drop-down list to bind the permission group to the mount point.

- If **`Mount Point Type`** is set to **`VPC`**, select **`VPC`** and **`VSwitch`** corresponding to the mount point. Then select a permission group from the **Permission Group** drop-down list to bind the permission group to the mount point.

> **Note:**
>
> - If Mount Point Type is set to VPC, make sure that the corresponding VPC instance and VSwitch are available.
> - Currently, a mount point in a classic network can be accessed only by ECS instances under the same account.
> - You can mount a mount point on multiple computing nodes (ECS, E-HPC, or Container Service instance) for shared access.

6. Click **Confirm** to add the mount point.

15.3.6 Mount a file system

NAS currently supports the NFSv3 and NFSv4.0 protocols. You can choose a protocol version for mounting a file system according to your needs.

Prerequisites

The following conditions determine whether an ECS instance can access a file system through a mount point:

- If the mount point is the VPC type, you can mount the file system only on the ECS instance in the same VPC instance as the mount point. In addition, ensure that the authorization address of a rule in the permission group bound to the mount point matches the VPC IP address of the ECS instance.
- If the mount point is the classic network type, you can mount the file system only on the ECS instance under the same account as the mount point. In addition, ensure that the authorization address of a rule in the permission group bound to the mount point matches the internal network IP address of the ECS instance.

Before mounting a file system through NFS, check that `nfs-utils` or `nfs-common` is installed. The installation method is as follows:

- CentOS: `sudo yum install nfs-utils`
- Ubuntu or Debian: `sudo apt-get install nfs-common`

Mount a file system through NFSv4.0

Format

```
sudo mount -t nfs -o vers=4.0 <mount point domain name>:<file system
 directory> <target directory to be mounted on the current server>
```

Parameter description

- Mount point domain name: It is automatically generated when you create a file system and a mount point.
- File system directory: A directory of the NAS file system, which may be the root directory "/" or any subdirectory.
- Target directory to be mounted on the current server: Directory that will be mounted on the current server.

Example

- Mount the root directory of the NAS file system:

  ```
  mount -t nfs -o vers=4.0 014544bbf6-wdt41.regionid.nas.example.com
  :/ /local/mntdir
  ```

- Mount the subdirectory sub1 of the NAS file system:

  ```
  mount -t nfs -o vers=4.0 014544bbf6-wdt41.regionid.nas.example.com:/
  sub1 /local/mntdir
  ```

Mount a file system through NFSv3

Format

```
sudo mount -t nfs -o vers=3,nolock,proto=tcp <mount point domain
 name>:<file system directory> <target directory to be mounted on the
 current server>
```

Example

- Mount the root directory of the NAS file system:

```
mount -t nfs -o vers=3,nolock,proto=tcp 014544bbf6-wdt41.regionid.
nas.example.com:/ /local/mntdir
```

- Mount the subdirectory sub1 of the NAS file system:

```
mount -t nfs -o vers=3,nolock,proto=tcp 014544bbf6-wdt41.regionid.
nas.example.com:/sub1 /local/mntdir
```

View the mount point information (NFS)

After mounting, run the following command to check the mounted file system:

```
mount -l
```

Run the following command to check the current capacity of the mounted file system:

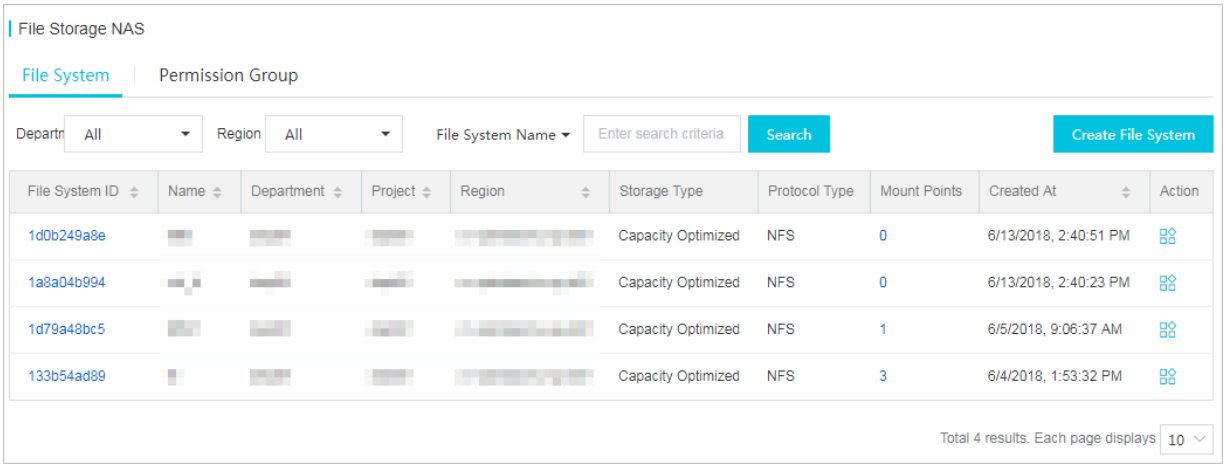
```
df -h
```

15.4 Manage file systems

View the file system instance list

Log on to the NAS console. On the **File Storage NAS** page, you can view the file system instance list, as shown in *Figure 15-2: Instance list*.

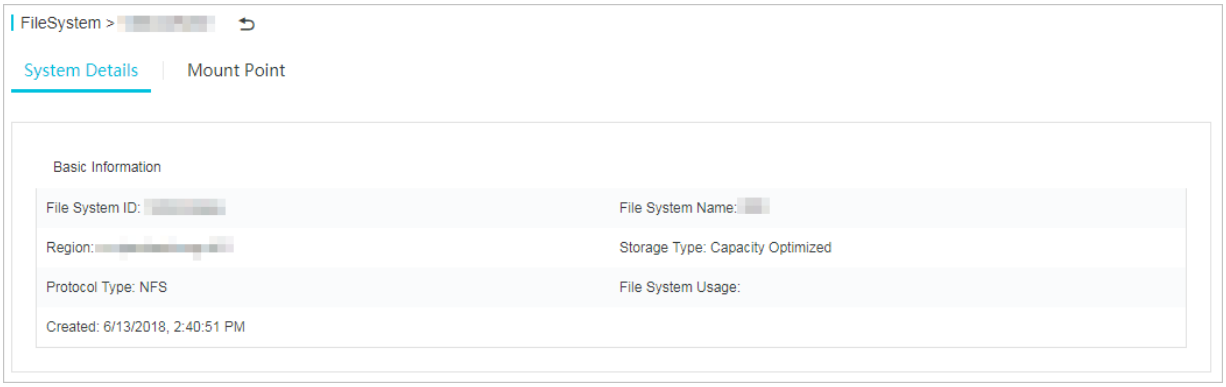
Figure 15-2: Instance list



View details of a file system instance

Log on to the NAS console. Click the file system ID in the file system instance list or choose **Action** > **Details** on the right to go to the file system details page, as shown in *Figure 15-3: Instance details*.

Figure 15-3: Instance details



The file system details page contains the following tabs:

- **System Details** tab: Displays basic information about the file system, including the file system ID, region, and file system storage capacity.
- **Mount Point** tab: Lists the mount points of the file system. You can manage the mount points on this tab.

Delete a file system

Warning:
Before deleting a file system instance, check that the file system has no mount point.

Log on to the NAS console. In the file system instance list, select the file system to delete and click **Action** > **Delete** on the right to delete the file system.

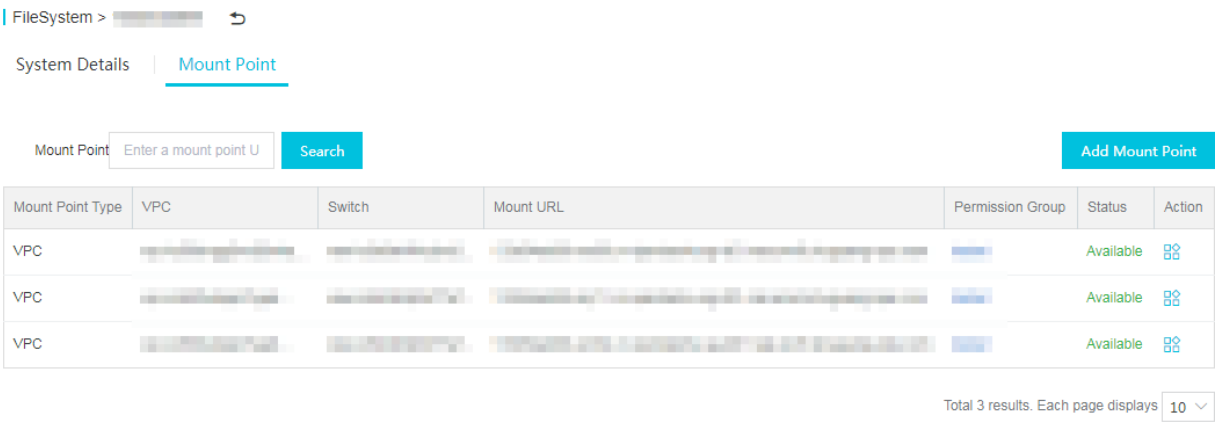
15.5 Manage mount points

View the mount point list

1. *Log on to the NAS console*. On the **File Storage NAS** page, click the file system ID in the file system instance list to go to the **System Details** page.
2. Click the **Mount Point** tab.

You can manage the mount points on this tab. For example, you can add, delete, and modify access groups, and enable and disable mount points.

Figure 15-4: Mount point list



Enable or disable a mount point

Choose **Action** > **Disable** on the right of a mount point to block access to the mount point from all clients.

Choose **Action** > **Enable** on the right of a mount point to enable the clients to access the mount point.

Delete a mount point

Choose **Action** > **Delete** on the right of a mount point to delete the mount point.

Note:
Once deleted, the mount point cannot be restored.

Modify a permission group

You must bind a permission group to each of your mount points. Each permission group has a source IP address whitelist used to restrict access to the mount point from ECS instances. You can modify the permission group which is bound to a mount point as needed.

Choose **Action** > **Modify Permission Group** on the right of a mount point to bind the mount point to a permission group.

Note:
The modified permission group takes effect after a delay of one minute at most.

15.6 Manage permission groups

Context

In NAS, each permission group has an IP address whitelist. You can add rules to the permission group to allow specified IP addresses or IP address segments to access the file system, and assign different levels of access permissions to different IP addresses or IP address segments.

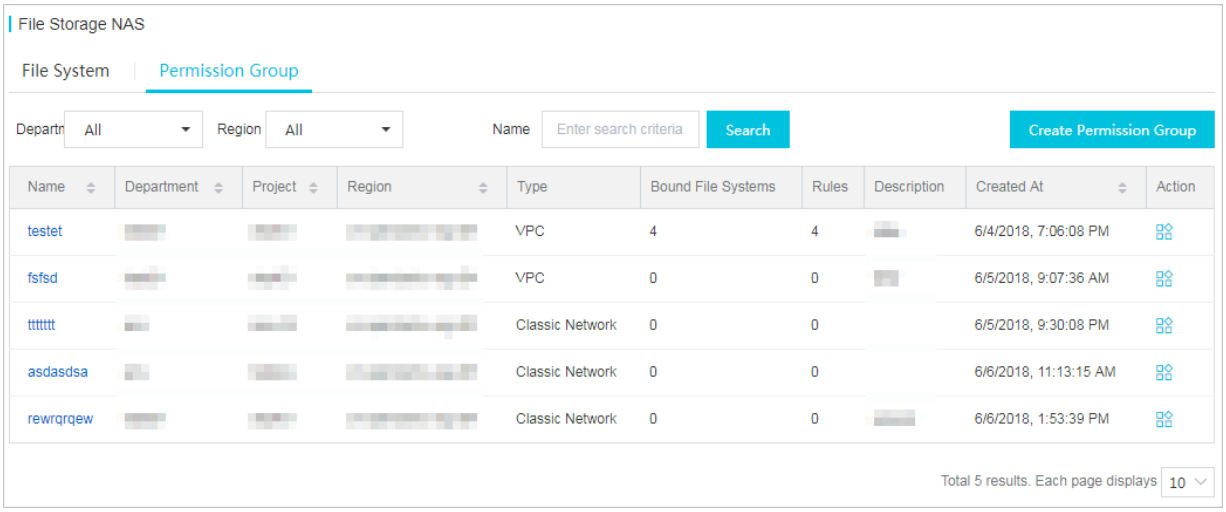
Warning:

To guarantee the security of your data, we recommend that you add permission group rules carefully and only authorize necessary IP addresses to access data.

View the permission group list

Log on to the NAS console. On the **File Storage NAS** page, click the **Permission Group** tab to view the permission group list, as shown in *Figure 15-5: Permission group list*.

Figure 15-5: Permission group list



Delete a permission group

Choose **Action** > **Delete** on the right of a permission group to delete the permission group.

Note:

Permission groups in use cannot be deleted.

Manage rules

Click the name of a permission group to go to the **Rules List** page of the permission group.

On the page, you can:

- *Create a permission group rule*.
- Choose **Action** > **Edit** on the right of a rule to modify the rule.

- Choose **Action** > **Delete** on the right of a rule to delete the rule.

15.7 Data migration

15.7.1 Migrate local files and files stored on Alibaba Cloud OSS instances to Alibaba Cloud NAS instances

The **`nasimport`** tool helps you synchronize files and data from your local data centers, your Alibaba Cloud OSS instances, and third-party cloud storage devices to your Alibaba Cloud NAS instance.

Context

nasimport has the following functions:

- Synchronizes files stored on local machines, Alibaba Cloud OSS instances and third-party cloud storage products, and HTTP-linked files to Alibaba Cloud NAS instances.
- Mounts NAS instances automatically.
- Synchronizes stored data (files can be modified after the specified time).
- Synchronizes incremental data automatically.
- Supports resumable data transfer.
- Lists, uploads, and downloads data in parallel.

To migrate a large volume of data (over 2 TB) to a NAS instance in a short time, you can contact Alibaba Cloud technical support for a multi-machine parallel synchronization solution in addition to using nasimport.

Runtime environment

You must run nasimport on an ECS virtual machine where you can mount the target NAS file system. For more information about whether an ECS instance supports NAS file system mounting and how to mount the NAS file system, see *Mount a file system*.

You must run nasimport in JDK 1.7 or later versions. The *Oracle version JDK* is recommended.

Note:

Before running the program, run **`ulimit -n`** to check the number of opened files allowed by the process. If the number is smaller than 10,240, modify the number.

Deployment and configuration

1. Create a working directory for synchronization on your local server and download the nasimport toolkit to this directory.

   Example: Create */root/ms* as the working directory and download the toolkit to this directory.

   ```
   export work_dir=/root/ms wget http://docs-aliyun.cn-hangzhou.oss.
   aliyun-inc.com/assets/attach/45306/cn_zh/1479113980204/nasimport_
   linux.tgz tar zxvf ./nasimport_linux.tgz -C "$work_dir"
   ```

2. Edit the configuration file config/sys.properties in the working directory $work_dir.

   ```
   vim $work_dir/config/sys.properties workingDir=/root/ms slaveUserN
   ame= slavePassword= privateKeyFile= slaveTaskThreadNum=60 slaveMaxTh
   roughput(KB/s)=100000000 slaveAbortWhenUncatchedException=false
   dispatcherThreadNum=5
   ```

   We recommend that you use the default configuration. If needed, you can edit the configuration field values. For more information, see *Table 15-4: Field description*.

   Table 15-4: Field description

Field	Description
workingDir	Working directory, which is the directory where the nasimport toolkit is extracted
slaveTaskThreadNum	Number of working threads that run synchronization simultaneously
slaveMaxThroughput (KB/s)	Upper limit of migration traffic
slaveAbortWhenUncatchedException	Whether to skip or abort an unknown error. By default, unknown errors are not aborted.
dispatcherThreadNum	Number of parallel threads in a dispatching job. Normally, you can keep the default value.

Migration

nasimport supports the following commands:

- Submit a job:

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 submit $jobConfigPath
```

- Cancel a job:

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 clean $jobName
```

- Check the status of a job:

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 stat detail
```

- Retry a job:

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 retry $jobName
```

Perform the following steps to run nasimport:

1. Run the following command to start nasimport:

```
cd $work_dir nohup java -Dskip_exist_file=false -jar $work_dir/
nasimport.jar -c $work_dir/config/sys.properties start > $work_dir/
nasimport.log 2>&1 &
```

Note:

The related log file is automatically generated in the directory where nasimport is started.

We recommend that you start nasimport in the working directory $work_dir. If the value of **`skip_exist_file`** is true when nasimport is started, a file that exists in the NAS file system with the same length as the source will be skipped in the upload.

2. Edit the sample job description file *`nas_job.cfg`*.

Table 15-5: Field description

Field name	Description
jobName	Customize the job name. It is a unique name that identifies a job. You can submit multiple jobs with different names.
jobType	You can set this field to import (to synchronize data) or audit (to verify the global consistency of the source data and target data in synchronization).
isIncremental=false	Whether to enable the automatic incremental mode. If it is set to true, incremental data is scanned at the interval

Field name	Description
	specified by incrementalModeInterval (in seconds) and synchronized to Alibaba Cloud NAS instances.
incrementalModeInterval=86400	Synchronization interval in incremental mode.
importSince	A time point expressed as a Unix timestamp (in seconds). If this field is set, data modified after this time point is synchronized. The default value is 0.
srcType	Synchronization source type. Currently, you can synchronize local files and files stored in Alibaba Cloud OSS instances or third-party cloud storage products.
srcAccessKey	AccessKey of the data source. You must set this field if srcType is set to Alibaba Cloud OSS or a third-party cloud storage product.
srcSecretKey	SecretKey of the data source. You must set this field if srcType is set to Alibaba Cloud OSS or a third-party cloud storage product.
srcDomain	Source endpoint.  **Note:** When you configure the migration service, set srcDomain to an internal network domain name with "internal" if the data source is an Alibaba Cloud OSS instance. In this way, you only need to pay for accessing the Alibaba Cloud OSS instance, saving the cost of the download traffic from the Alibaba Cloud OSS instance and enjoying a faster migration speed. You can retrieve the internal network domain name of the Alibaba Cloud OSS instance on the OSS console. If your NAS file system is in a VPC instance and the data source is an Alibaba Cloud OSS instance, set srcDomain to the VPC domain name provided by the OSS instance.
srcBucket	Name of the source bucket.
srcPrefix	Source prefix. The default value is null.

Field name	Description
	If srcType is set to local, enter the local directory to be synchronized. Note that the directory must be a full path ending with a slash (/). If srcType is set to an OSS instance or a third-party cloud storage device, enter the prefix of the object to be synchronized. To synchronize all files, you can leave the prefix blank.
destType	Synchronization target type (NAS by default).
destMountDir	Locally mounted directory of NAS.
destMountTarget	Domain name of a NAS mount point.
destNeedMount=true	Whether nasimport performs automatic mounting. The default value is true. You can set it to false and manually mount the NAS mount point to the destMountDir directory.
destPrefix	Prefix of the synchronization target file. The default value is null.
taskObjectCountLimit	Maximum number of files in each task. This field affects the parallel execution of tasks and is usually set to the total number of files or the number of download threads that you set. If you do not know the total number of files, you can keep the default value.
taskObjectSizeLimit	Maximum volume (in bytes) of the data downloaded in each task.
scanThreadCount	Number of threads that scan files in parallel. This field affects the file scan efficiency.
maxMultiThreadScanDepth	Maximum allowable depth of the directory in parallel scan. You can keep the default value.

Note:

- If you have configured the automatic incremental mode, the job runs periodically and permanently to scan the latest data.
- If srcType is set to a third-party cloud storage device, the List operation on files cannot implement checkpoints due to the API restrictions of the third-party cloud storage devices. Killing the process before the List operation is complete may lead to relisting of all files.

3. Submit the job.

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 submit $work_dir/nas_job.cfg
```

Note:

- If the job that you submit has the same name as a job in progress, the job submission fails.
- To pause a synchronization job, stop the nasimport process. You can restart the nasimport process to resume synchronization from where it was paused.
- To resynchronize all files, stop the nasimport process and call the following command to clear the current job. For example, if the job named nas_job (you can set the job name in the nas_job.cfg file) is running, run the following command:

```
ps axu | grep "nasimport.jar.* start" | grep -v grep | awk {
print "kill -9 "$2} | bash java -jar $work_dir/nasimport.jar -c $
work_dir/conf/sys.properties clean nas_job
```

4. Check the job execution status.

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.
properties stat detail --------------job stats begin---------------
 ----------------job stat begin------------------ JobName:nas_job
 JobState:Running PendingTasks:0 RunningTasks:1 SucceedTasks:0
FailedTasks:0 ScanFinished:true RunningTasks Progress: FD813E8B93
F55E67A843DBCFA3FAF5B6_1449307162636:26378979/26378979 1/1
 ----------------job stat end------------------ --------------job
stats end---------------
```

This shows the overall progress of the current job and the progress of the current task. In the preceding information, `26378979/26378979` indicates the total volume of data to be uploaded (26,378,979 bytes) and the volume of data already uploaded (26,378,979 bytes), respectively. `1/1` indicates the total number of files to be uploaded (1) and the number of files already uploaded (1), respectively.

The migration tool splits one job into multiple tasks for parallel execution. The job is considered complete after all tasks are complete. After the job is complete, JobState displays `Succeed` or `Failed`, indicating that the job is successful or has failed. If the job fails, run the following command to check the failure cause of each task.

In the following command, replace $jobName with the name of the actual job (you can set jobName in the nas_job.cfg file).

```
cat $work_dir/master/jobs/$jobName/failed_tasks/*/audit.log
```

As we have already attempted to retry failed jobs in nasimport, the failure may be due to the temporary unavailability of the source or target data. Use the following command to retry the failed task:

```
java -jar $work_dir/nasimport.jar -c $work_dir/config/sys.properties
 retry $jobNam
```

Common job failure causes

- The job configuration is incorrect, for example, the AccessKey or ID is incorrect or permissions are insufficient. In this case, generally all tasks fail. To confirm the causes, check the *$work_dir/nasimport.log* file.
- The method for encoding the source file name is inconsistent with the systems default method, for example, GBK in Windows and UTF-8 in Linux by default. This fault occurs if the NFS data sources are used.
- A file in the source directory is modified during the upload process. If this fault occurs, the `SIZE_NOT_MATCH` error is displayed in audit.log. In this case, the old file is successfully uploaded, but the changes are not synchronized to Alibaba Cloud NAS instances.
- The source file is deleted during the upload process, causing file download failure.
- An error occurs in the data source, causing source data download failure.
- The Clean operation is performed before the nasimport process is killed, which may cause a program execution.
- The nasimport program exits unexpectedly and the job status is Abort. If this fault occurs, contact Alibaba Cloud technical support.

15.7.2 Tool to migrate data in Windows

The NAS data migration tool nasimport for Windows is available after you download and decompress it. nasimport is used to synchronize files from object storage servers (such as OSS instances) and local disks to Alibaba Cloud NAS instances.

Context

nasimport has the following features:

- Synchronizes local files, files stored on Alibaba Cloud OSS instances and third-party cloud storage devices, and HTTP-linked files to Alibaba Cloud NAS instances.
- Synchronizes stored data (files can be modified after the specified time).
- Synchronizes incremental data automatically.
- Supports resumable data transfer.
- Uploads and downloads data in parallel.

Operation requirements

You must run nasimport on an ECS virtual machine where you can mount the target NAS file system. For more information about whether an ECS instance supports NAS file system mounting and how to mount the NAS file system, see *Mount a file system*.

Supported operating systems

- Windows Server 2008 standard edition SP2 32-bit
- Windows Server 2008 R2 data-center edition 64-bit
- Windows Server 2012 R2 data-center edition 64-bit
- Windows Server 2016 data-center edition 64-bit

Deployment and configuration

1. Download the *nasimport toolkit*.
2. Create a synchronization working directory (for example, C:\NasImport) on the local server and decompress the nasimport toolkit to this directory.
3. Edit the configuration file *`config/sys.properties`* in the working directory.

   We recommend that you use the default configuration. If needed, you can edit the configuration field values. *Description of fields* describes the fields.

   Table 15-6: Field description

Field	Description
workingDir	Working directory, which is the directory where the nasimport toolkit is extracted
slaveTaskThreadNum	Number of working threads that run synchronization simultaneously
slaveMaxThroughput (KB/s)	Upper limit of migration traffic

Field	Description
slaveAbortWhenUncatchedException	Whether to skip or abort an unknown error. By default, unknown errors are not aborted.
dispatcherThreadNum	Number of parallel threads in a dispatching job. Normally, you can keep the default value.

Migration

Commands supported by nasimport

- Submit a job: `nasimport -c config/sys.properties submit <your-job-configuration>`
- Cancel a job: `nasimport -c config/sys.properties clean <job-name>`
- View a job: `nasimport -c config/sys.properties stat detail`
- Retry a job: `nasimport -c config/sys.properties retry <job-name>`
- Start nasimport: `nasimport -c config/sys.properties start`

1. Start nasimport.

   Enter the working directory and open a CLI. Run the following command in the CLI:

   ```
   nasimport –c config/sys.properties start
   ```

   Figure 15-6: Start nasimport

Note:

- Keep nasimport running. You can also set nasimport as a background service in Windows.
- When starting nasimport, you can redirect the log to a file for later viewing.

```
nasimport -c config\sys.properties start > nasimport.log 2>&1
```

2. Define a job.

   Use the config\local_job.cfg template to define a job.

Table 15-7: Field description

Field name	Description
jobName	Customize the job name. It is a unique name that identifies a job. You can submit multiple jobs with different names.
jobType	You can set this field to import (to synchronize data) or audit (to verify the global consistency of the source data and target data in synchronization).
isIncremental=false	Whether to enable the automatic incremental mode. If it is set to true, incremental data is scanned at the interval specified by incrementalModeInterval (in seconds) and synchronized to Alibaba Cloud NAS instances.
incrementalModeInterval=86400	Synchronization interval in incremental mode.
importSince	A time point expressed as a Unix timestamp (in seconds). If this field is set, data modified after this time point is synchronized. The default value is 0.
srcType	Synchronization source type. Currently, you can synchronize local files and files stored in Alibaba Cloud OSS instances or third-party cloud storage products.
srcAccessKey	AccessKey of the data source. You must set this field if srcType is set to Alibaba Cloud OSS or a third-party cloud storage product.
srcSecretKey	SecretKey of the data source. You must set this field if srcType is set to Alibaba Cloud OSS or a third-party cloud storage product.
srcDomain	Source endpoint.  **Note:** When you configure the migration service, set srcDomain to an internal network domain name with "internal" if

Field name	Description
	the data source is an Alibaba Cloud OSS instance. In this way, you only need to pay for accessing the Alibaba Cloud OSS instance, saving the cost of the download traffic from the Alibaba Cloud OSS instance and enjoying a faster migration speed. You can retrieve the internal network domain name of the Alibaba Cloud OSS instance on the OSS console. If your NAS file system is in a VPC instance and the data source is an Alibaba Cloud OSS instance, set srcDomain to the VPC domain name provided by the OSS instance.
srcBucket	Name of the source bucket.
srcPrefix	Source prefix. The default value is null. If srcType is set to local, enter the local directory to be synchronized. Note that the directory must be a full path ending with a slash (/). If srcType is set to an OSS instance or a third-party cloud storage device, enter the prefix of the object to be synchronized. To synchronize all files, you can leave the prefix blank.
destType	Synchronization target type (NAS by default).
destMountDir	Locally mounted directory of NAS.
destMountTarget	Domain name of a NAS mount point.
destNeedMount=true	Whether nasimport performs automatic mounting. The default value is true. You can set it to false and manually mount the NAS mount point to the destMountDir directory.
destPrefix	Prefix of the synchronization target file. The default value is null.
taskObjectCountLimit	Maximum number of files in each task. This field affects the parallel execution of tasks and is usually set to the total number of files or the number of download threads that you set. If you do not know the total number of files, you can keep the default value.
taskObjectSizeLimit	Maximum volume (in bytes) of the data downloaded in each task.

Field name	Description
scanThreadCount	Number of threads that scan files in parallel. This field affects the file scan efficiency.
maxMultiThreadScanDepth	Maximum allowable depth of the directory in parallel scan. You can keep the default value.

Note:

- If you have configured the automatic incremental mode, the job runs periodically and permanently to scan the latest data.
- If srcType is set to a third-party cloud storage device, the List operation on files cannot implement checkpoints due to the API restrictions of the third-party cloud storage devices. Killing the process before the List operation is complete may lead to relisting of all files.

3. Submit the job.

   The following example shows how to copy the local *C:\Program Files\Internet Explorer* directory to a NAS instance.

   a. Edit a job: Copy *config\local_job.cfg* to the working directory and edit the following items:

srcType	local
srcPrefix	C:\\Program Files\\Internet Explorer
destMountDir	h:
destNeedMount	true
destMountTarget	xxxx-yyy.cn-beijing.nas.aliyuncs.com

   Note:
   You must specify a non-existent drive letter as destMountDir. Otherwise, destMountDir may conflict with an existing drive. destMountTarget is the NAS mount point.

   b. Submit a job: Restart a CLI in the working directory and run `nasimport -c config\sys.properties submit local_job.cfg`.

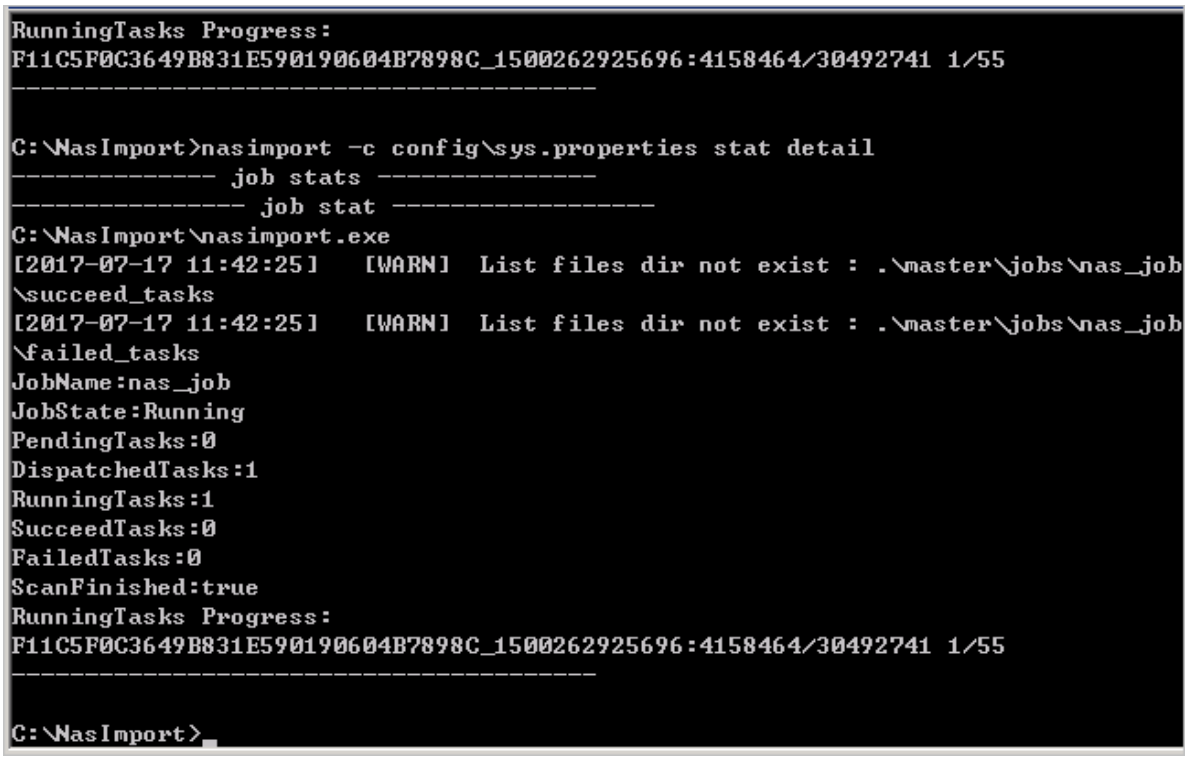
   Note:

   - If the job that you submit has the same name as a job in progress, the job submission fails.

- To pause a synchronization job, stop the nasimport process. You can restart the nasimport process to resume synchronization from where it was paused.

4. Check the job status. Run the following command in the CLI:

```
nasimport -c config\sys.properties stat detail
```



This shows the overall progress of the current job and the progress of the current task. In this example, `4158464/30492741` indicates the volume of data already uploaded (4,158,464 bytes) and the total volume of data to be uploaded (30,492,741 bytes), respectively. `1/1` indicates the total number of files to be uploaded (1) and the number of files already uploaded (1), respectively.

The migration tool splits one job into multiple tasks for parallel execution. The job is considered complete after all tasks are complete. After the job is complete, JobState displays `Succeed`

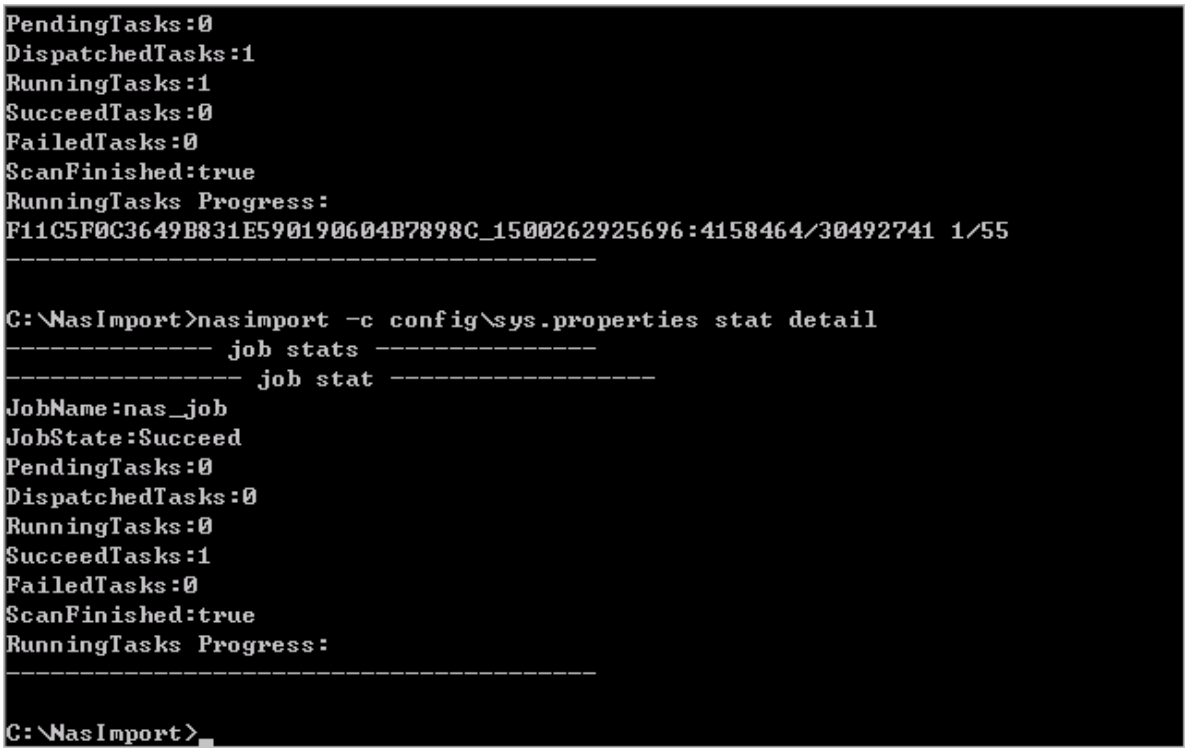
or `Failed`, indicating that the job is successful or has failed. If the job fails, run the following command to check the failure cause of each task:

```
master/jobs/$jobName/failed_tasks/*/audit.log
```

As we have already attempted to retry failed jobs in nasimport, the failure may be due to the temporary unavailability of the source or target data. Use the following command to retry the failed task:

```
nasimport -c config/sys.properties retry <job-name>
```

Run stat detail again after a while.



`SucceededTasks` is **1**, indicates that the job is complete. Open the file browser. The file is displayed in the *H:* drive.

Common failure causes

- The job configuration is incorrect, for example, the AccessKey or ID is incorrect or permissions are insufficient. In this case, generally all tasks fail. To confirm the causes, check the nasimport.log file in the working directory. (You must redirect the log to this file when starting nasimport, or directly view the log on the CLI where you run nasimport start.)

```
C:\NasImport\nasimport.exe
[2017-07-17 12:22:40]   [INFO]  JobDispatcher:Init
[2017-07-17 12:22:40]   [INFO]  job controller daemon start, working dir:.\
[2017-07-17 12:22:40]   [INFO]  watching job queue:.\master\jobqueue\
[2017-07-17 12:22:40]   [INFO]  JobDispatcher:Run
[2017-07-17 12:22:40]   [INFO]  try lock .\master\jobs\nas_job\.lock succeed
[2017-07-17 12:22:40]   [INFO]  start job:nas_job
[2017-07-17 12:22:40]   [INFO]  list checkpoint: .\master\jobs\nas_job\checkpoints\0, cpt
[2017-07-17 12:22:40]   [INFO]  scan task load checkpoint: [totalSize=0, totalCount=0, pre
[2017-07-17 12:22:40]   [INFO]  single thread scan start: nas_job
com.aliyun.oss.OSSException: The OSS Access Key Id you provided does not exist in our reco
        at com.aliyun.oss.common.utils.ExceptionFactory.createOSSException(ExceptionFactor
        at com.aliyun.oss.internal.OSSErrorResponseHandler.handle(OSSErrorResponseHandler.
        at com.aliyun.oss.common.comm.ServiceClient.handleResponse(ServiceClient.java:248)
        at com.aliyun.oss.common.comm.ServiceClient.sendRequestImpl(ServiceClient.java:130
        at com.aliyun.oss.common.comm.ServiceClient.sendRequest(ServiceClient.java:68)
        at com.aliyun.oss.internal.OSSOperation.send(OSSOperation.java:94)
        at com.aliyun.oss.internal.OSSOperation.doOperation(OSSOperation.java:149)
        at com.aliyun.oss.internal.OSSOperation.doOperation(OSSOperation.java:113)
        at com.aliyun.oss.internal.OSSBucketOperation.listObjects(OSSBucketOperation.java:
        at com.aliyun.oss.OSSClient.listObjects(OSSClient.java:526)
        at com.aliyun.ossimport2.master.scanner.OssLister.list(OssScanner.java:65)
        at com.aliyun.ossimport2.master.scanner.SingleThreadTask.run(SingleThreadTask.java
```

- The method for encoding the source file name is inconsistent with the default method of the system, for example, GBK in Windows and UTF-8 in Linux by default. This fault easily occurs if the NFS data sources are used.
- A file in the source directory is modified during the upload process. If this fault occurs, the `SIZE_NOT_MATCH` error is displayed in audit.log. In this case, the old file is successfully uploaded, but the changes are not synchronized to Alibaba Cloud NAS instances.
- The source file is deleted during the upload process, causing file download failure.
- An error occurs in the data source, causing source data download failure.
- The Clean operation is performed before the nasimport process is killed, which may cause a program execution.
- The nasimport program exits unexpectedly and the job status is Abort. If this fault occurs, contact Alibaba Cloud technical support.

16 ApsaraDB for Relational Database Service (RDS)

16.1 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.

16.2 Limits

16.2.1 Restrictions on MySQL

To guarantee the stability and security of ApsaraDB for MySQL, certain restrictions apply to the database and management properties, as shown in *Table 16-1: Restrictions on MySQL*.

Table 16-1: Restrictions on MySQL

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

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

Operation	Restriction
	• The master account cannot be rolled back to a common account.

16.2.2 Restrictions on SQL Server

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

Table 16-2: Restrictions on SQL Server

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

Feature	Description
Asynchronous messaging	Limited
Replication	Limited
Policy management	Limited

16.2.3 Restrictions on PostgreSQL

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

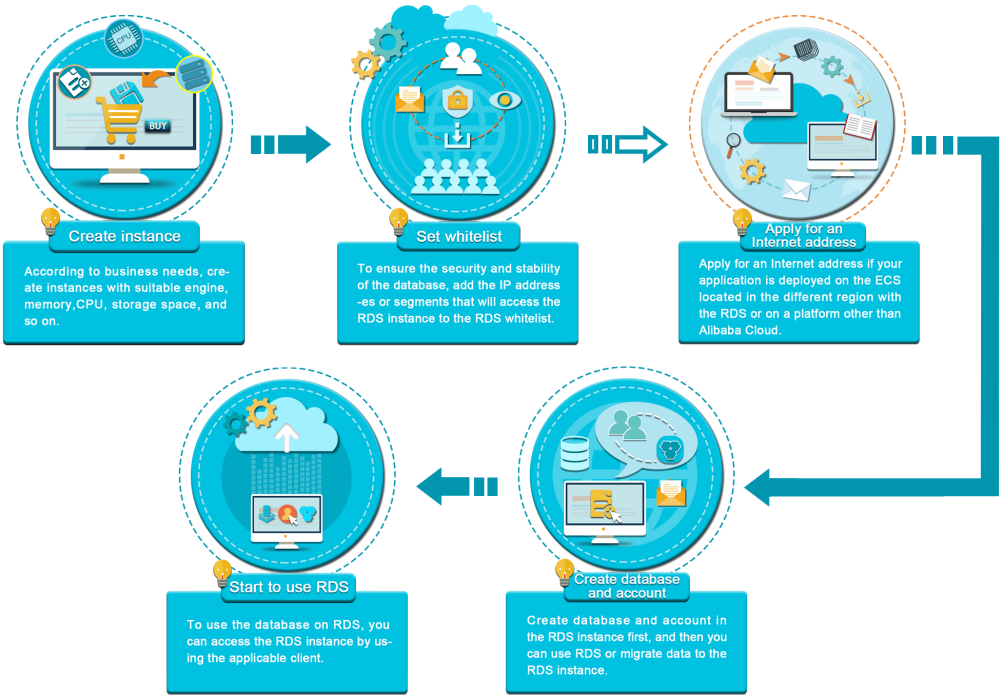
Table 16-3: Restrictions on PostgreSQL

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

16.3 Procedure

After you create an instance, you need to complete the following operations before you can start to use the instance.

Figure 16-1: Quick start flowchart



16.4 Log on to the RDS console

Take the Chrome browser as an example to describe how to log on to the RDS console through the Apsara Stack console as cloud product users.

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

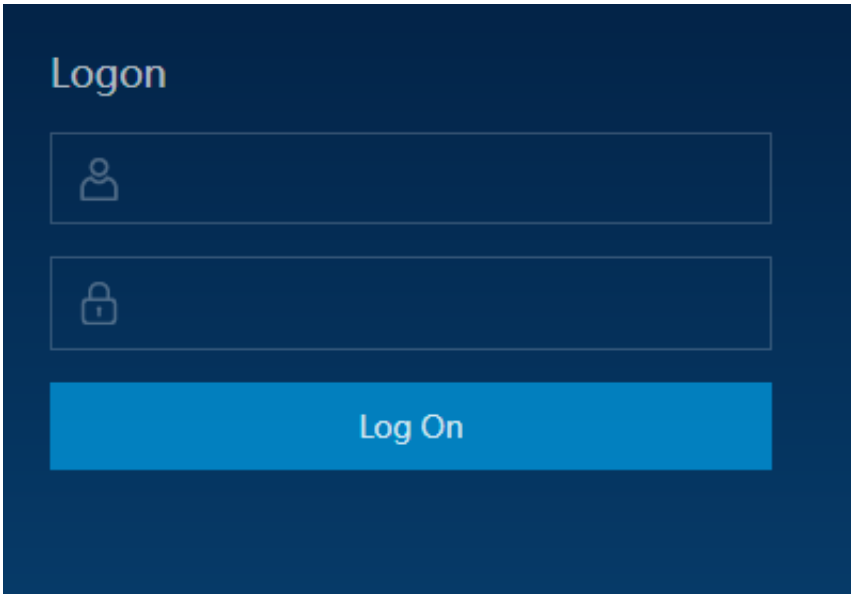
  The access address of the Apsara Stack console is http://`x.x.x.x`/manage, where `x.x.x.x` represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://`IP address or domain name address of the Apsara Stack console`/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 16-2: Log on to the Apsara Stack console*.

Figure 16-2: Log on to the Apsara Stack console



3. Enter the correct username and password.
   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **Log On** to go to the **Home** page.
5. In the menu bar, choose **Console** > **Database** > **Relational Database Service** .

16.5 Create an instance

This topic describes how to create an instance on the RDS console.

Prerequisites

Apply for an account of Alibaba Cloud console.

Procedure

1. *Log on to the RDS console*.

2. Click **Create Instance** in the upper-right corner of the page to go to the **Create Instance** page.
3. Select instance information such as**Configurations**, **Network Type**, **Specification Configuration**, **Connection Mode**, and **Quantity**, as shown in *Table 16-4: New instance configurations*.

Table 16-4: New instance configurations

Configurations	Description
Department	Select a department for the instance.
Project	Select a project for the instance.
Region	Select a region for the instance.
Zone	Common instances of ApsaraDB adopt the hot standby architectu re. A single-zone instance indicates that the master and slave nodes of the instance are all in the same zone.
Instance type	Select **Internal** or **External** to generate an internal IP address or an external IP address.
Network Type	Classic network: Cloud services in a classic network are not isolated. Unauthorized access to a cloud service is blocked only by the security group or whitelist policy of the service. Virtual Private Cloud (VPC): A VPC helps you build an isolated network environment in Alibaba Cloud. You can customize route tables, IP address ranges, and gateways in a VPC. We recommend that you use a VPC for higher security. You must create a VPC in advance. Alternatively, you can change the network type after creating an instance.
Access Mode	Select **Standard Mode** or **Safe Mode**: • **Standard Mode**: ApsaraDB for RDS uses Server Load Balancer (SLB) to eliminate the impact of database engine HA failovers on the application layer. This shortens the response time, but will slightly increase the probability of transient disconnections. • **Safe Mode**: This mode prevents 90% of transient disconnections. However, it increases the response time by 20% or more and incurs performance loss.
Instance Name	The instance name can be 2 to 256 characters in length and can contain letters, numbers, and underscores (_). It must start with a letter.

Configurations	Description
Database Type	Different database types are supported in different regions. For the specific database type, see the actual options on the page.
Database Version	Set the database version.
CPU/Memory	The maximum number of connections and the maximum IOPS are determined by the memory size.
Storage Space	The storage space contains the space for data, system files, binlog files, and transaction files.
Quantity	Number of RDS instances that can be created. You can create up to 20 RDS instances.

4. Click **Create** to create the instance.

16.6 Initial configuration

16.6.1 RDS for MySQL

16.6.1.1 Configure a whitelist

To guarantee database security and reliability, you need to modify its whitelist before you enable an instance. You need to add the IP addresses or IP address segments used for database access to the whitelist of the RDS instance. This topic describes how to configure the whitelist of an RDS instance.

Context

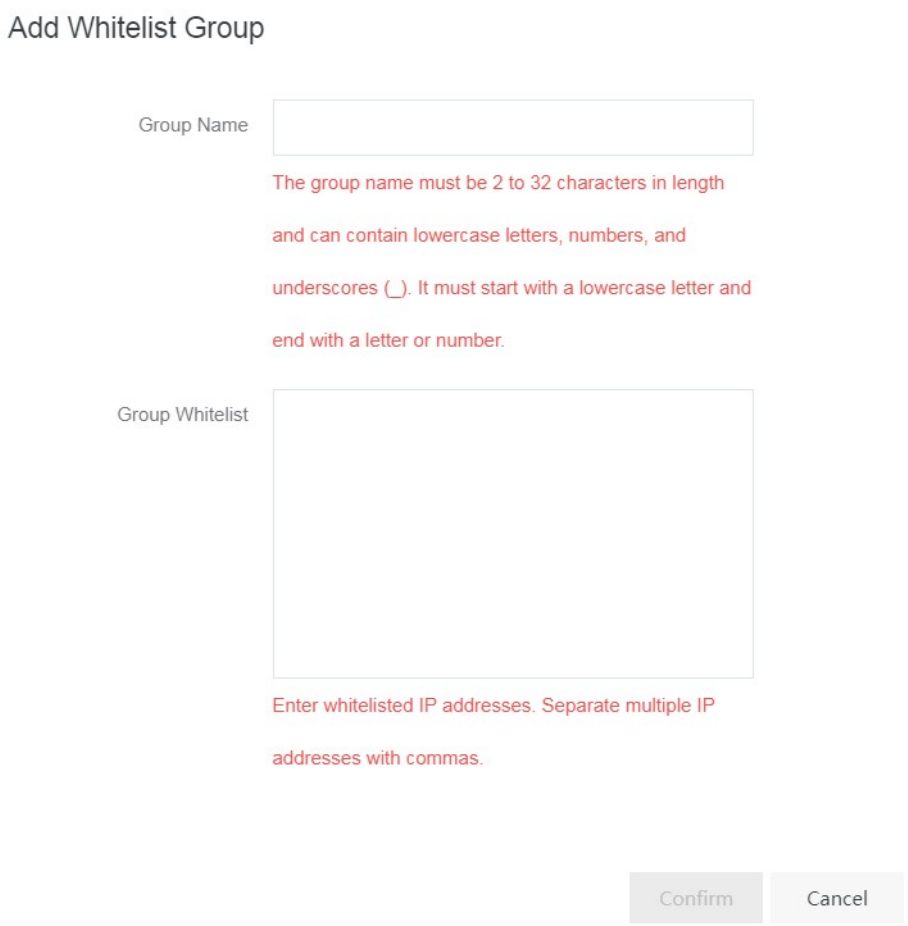
Note:

- The system creates a default whitelist group for each instance. This default whitelist group can only be modified or cleared, but cannot be deleted.
- For each newly created RDS instance, the local loopback IP address 0.0.0.0/0 is added to the **default** whitelist group by default. This means that any IP address is allowed to access the RDS instance. This configuration greatly reduces the security of the database. Delete 0.0.0.0/0 first.
- If the whitelist is configured to 127.0.0.1, all IP addresses or IP segments are prohibited to access the RDS instance. Therefore, you must delete 127.0.0.1 from the whitelist before you add other IP addresses or IP segments.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation bar, select **Security Control** > **Whitelist Settings**.
4. Click the ' ' icon of the **default** whitelist group to delete the default whitelist 0.0.0.0/0.
5. Add the IP addresses or IP segments allowed to access the RDS instance to the **default** whitelist group.
6. Click **Confirm**.
7. Click **Add Whitelist Group**.
8. In the **Add Whitelist Group** dialog box, enter the group name and the IP addresses or IP segments allowed to access the RDS instance, and click **Confirm**.

Figure 16-3: Add a whitelist group



Parameter description:

Table 16-5: Add a whitelist group

Parameter name	Description
`Group Name`	The name can be 2 to 32 characters in length and can contain lowercase letters, numbers, and underscores (_). It must start with a lowercase letter and end with a letter or number. You cannot modify the name of a created whitelist group.
`Group Whitelist`	Enter the IP addresses or IP address segments allowed to access the RDS instance. • If you enter an IP address segment, for example, 10.10.10.0/24, any IP address in the format of 10.10.10.X can access the RDS instance. • If you enter multiple IP addresses, separate them with commas (no space before or after each comma), for example, 192.168.0.1,172 .16.213.9.

What's next

Correct use of the whitelist can improve access security for your RDS instance. We recommend that you maintain the whitelist periodically. To modify the whitelist group, click . To clear the default whitelist group or delete a custom whitelist group, click .

16.6.1.2 Create a database and an account

To use ApsaraDB, you must create a database and an account in your instance. This topic describes how to create a database and an account on the RDS console.

Context

Note:

- Databases under the same instance share all resources of this instance. You can create up to 500 databases and 500 accounts under each instance in MySQL 5.6.
- To migrate the local database to ApsaraDB for RDS, you must create the same database and account in your RDS instance as those of the local database.
- When assigning database account permissions, follow the minimum permission principle and use service roles to create accounts and assign proper read-only and read/write permissions . When necessary, you may split database accounts and databases into smaller units, so that

each account can only access data for its own services. If the account does not need to write data to a database, assign read-only permissions.

- For database security, set strong passwords for the accounts and change the passwords periodically.

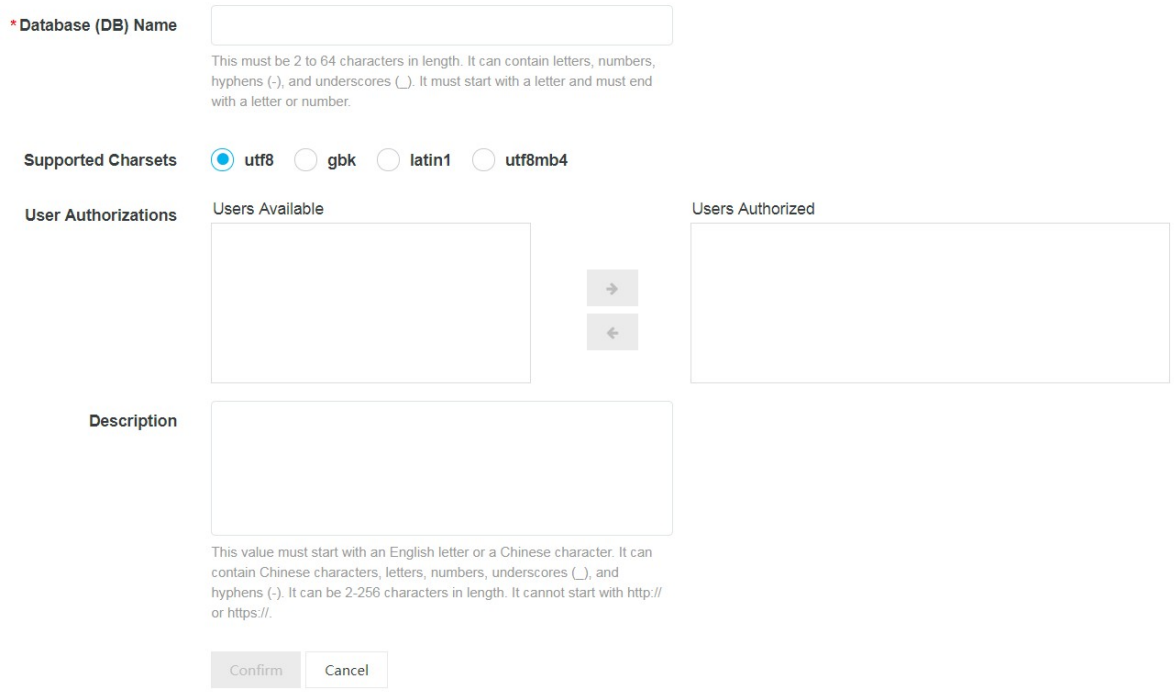
Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Database Management** > **Account List** to go to the **Account List** page.
4. Click **Add Account**.
5. Enter information about the account to be created.

   Parameter description:

   - **`Database Account`**: The account name can be 2 to 16 characters in length and can contain lowercase letters, numbers, and underscores (_). It must start with a letter and end with a letter or number, for example, *user4example*.
   - **`Enter a new password.`**: The password of the database account. The password can be 6 to 32 characters in length and can contain letters, numbers, and underscores (_). For example, password4example.
   - **`Reenter the password.`**: Re-enter the password to verify that it is correct, for example, *password4example.*
   - **`Remarks`**: You can enter related information of the account to assist in future account management. You can enter a maximum of 256 English characters.
6. Click **Confirm**.
7. In the left-side navigation pane, choose **Database Management** > **Database List** to go to the **Database List** page.
8. Click **Add Database**.
9. Enter information about the database to be created, as shown in *Figure 16-4: Create a database*.

Figure 16-4: Create a database



Parameter description:

- **`Database (DB) Name`**: The database name can be 2 to 64 characters in length and can contain lowercase letters, numbers, underscores (_), and hyphens (-). It must start with a letter and end with a letter or number.
- **`Supported Charsets`**: Set character sets for the database. The options are: utf8, gbk, latin1, and utf8mb4.
- **`User Authorizations`**: Select an account that is authorized to use the database. This parameter can be empty if no account is created.
- **`Description`**: You can enter related information about the database to assist in future database management. You can enter a maximum of 256 English characters.

10. Click **Connfirm**.

16.6.1.3 Create a master account

ApsaraDB for RDS supports classic mode and master mode. For MySQL 5.6 instances, you can create a master account to upgrade the account management mode from classic to master. Compared to the classic mode, the master mode enables more permissions to meet personalized and sophisticated permission management requirements. In master mode, you can use SQL to

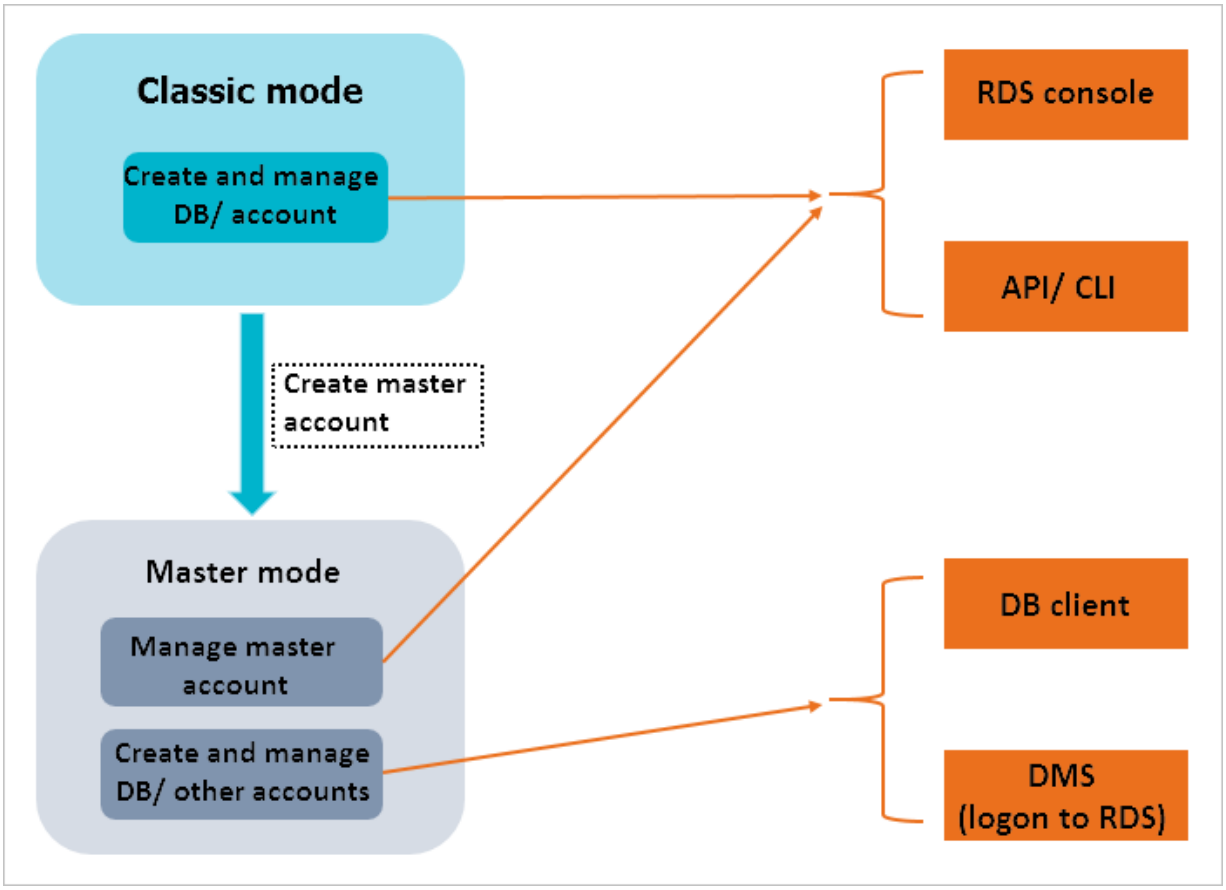
directly manage databases and accounts. Therefore, we recommend that you use the master mode.

Context

After a master account is created for a primary instance, the master account is synchronized to read-only instances. In master mode, you are not allowed to manage databases and ordinary accounts on the RDS console or through APIs. You must use SQL commands or Alibaba Cloud DMS to perform related operations. However, you can reset the permissions and password of the master account on the RDS console or through APIs. Other accounts in the instance are not affected.

*Figure 16-5: Comparison between account management modes*shows how to upgrade the account management mode of MySQL 5.6 from classic to master. It also shows the differences in database or account creation and management between the two modes.

Figure 16-5: Comparison between account management modes



Note:

- MySQL 5.6 instance accounts can only upgrade from the classic mode to the master mode, but cannot downgrade the account from master to classic.
- The following changes occur after an instance switches to the master account mode:
  - After a master account is created, you cannot create or manage databases on the console. The **Create Account** button is not displayed on the **Account List** page. However, this change only affects a single instance and does not affect the console of other instances.
  - In MySQL 5.6, you cannot directly access the mysql.user and mysql.db tables. However, you can view the existing account and permissions through mysql.user_view and mysql.db_view.
  - You cannot use the master account to change the passwords of ordinary accounts. To change the password of an ordinary account, you must delete the account and create a new one.
- When a master account is created, the instance restarts once, which causes a transient disconnection of less than 30 seconds. To avoid the service impacts from transient disconnect ions, create an account at a proper time and make sure that your application can be automatically reconnected.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Database Management** > **Account List** to go to the **Account List** page.
4. Click **Add Account**.
5. Enter account information. Set **`User Type`** to **`Administrator`**.

   Parameter description:

   Table 16-6: Master account creation parameters

Parameter name	Description
`Database Account`	The account name can be 2 to 16 characters in length and can contain lowercase letters, numbers, and underscores (_). It must start with a letter and end with a letter or number, for example, user4example.
`User Type`	Select **Administrator**, which indicates the master account.

Parameter name	Description
Enter a new password.	The password can be 6 to 32 characters in length and can contain letters, numbers, and underscores (_), for example, password4example.
Reenter the password.	Re-enter the password to verify that it is correct, for example, *password4example*.
Remarks	You can enter related information about the account to assist in future account management. You can enter a maximum of 256 English characters.

6. Click **Confirm**.

16.6.2 RDS for SQL Server

16.6.2.1 Configure a whitelist

To guarantee database security and reliability, you need to modify its whitelist before you enable an instance. You need to add the IP addresses or IP address segments used for database access to the whitelist of the RDS instance. This topic describes how to configure the whitelist of an RDS instance.

Context

Note:

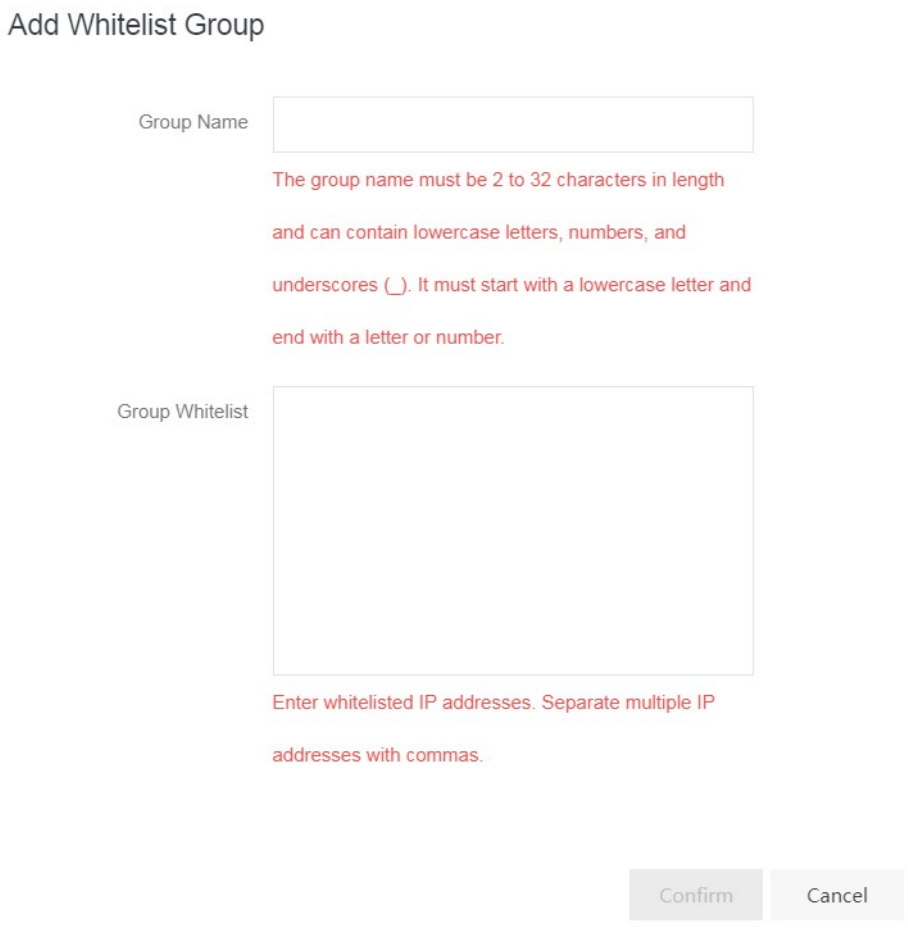
- The system creates a default whitelist group for each instance. This default whitelist group can only be modified or cleared, but cannot be deleted.
- For each newly created RDS instance, the local loopback IP address 0.0.0.0/0 is added to the **default** whitelist group by default. This means that any IP address is allowed to access the RDS instance. This configuration greatly reduces the security of the database. Delete 0.0.0.0/0 first.
- If the whitelist is configured to 127.0.0.1, all IP addresses or IP segments are prohibited to access the RDS instance. Therefore, you must delete 127.0.0.1 from the whitelist before you add other IP addresses or IP segments.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation bar, select **Security Control** > **Whitelist Settings**.

4. Click the icon of the **default** whitelist group to delete the default whitelist 0.0.0.0/0.
5. Add the IP addresses or IP segments allowed to access the RDS instance to the **default** whitelist group.
6. Click **Confirm**.
7. Click **Add Whitelist Group**.
8. In the **Add Whitelist Group** dialog box, enter the group name and the IP addresses or IP segments allowed to access the RDS instance, and click **Confirm**.

Figure 16-6: Add a whitelist group



Parameter description:

Table 16-7: Add a whitelist group

Parameter name	Description
Group Name	The name can be 2 to 32 characters in length and can contain lowercase letters, numbers, and underscores (_). It must start with a lowercase letter and end with a letter or number. You cannot modify the name of a created whitelist group.
Group Whitelist	Enter the IP addresses or IP address segments allowed to access the RDS instance. • If you enter an IP address segment, for example, 10.10.10.0/24, any IP address in the format of 10.10.10.X can access the RDS instance. • If you enter multiple IP addresses, separate them with commas (no space before or after each comma), for example, 192.168.0.1,172 .16.213.9.

What's next

Correct use of the whitelist can improve access security for your RDS instance. We recommend that you maintain the whitelist periodically. To modify the whitelist group, click . To clear the default whitelist group or delete a custom whitelist group, click .

16.6.2.2 Create a database and an account

Before using ApsaraDB for RDS, you must create a database and an account in your RDS instance.

Context

Note:

- To migrate the local database to ApsaraDB for RDS, you must create the same database and account in your RDS instance as those of the local database.
- Databases under the same instance share all resources of this instance.
- When you assign account permissions for each database, follow the minimum permission principle and service roles to create accounts and rationally assign read-only and read/write permissions. When necessary, you may split database accounts and databases into smaller

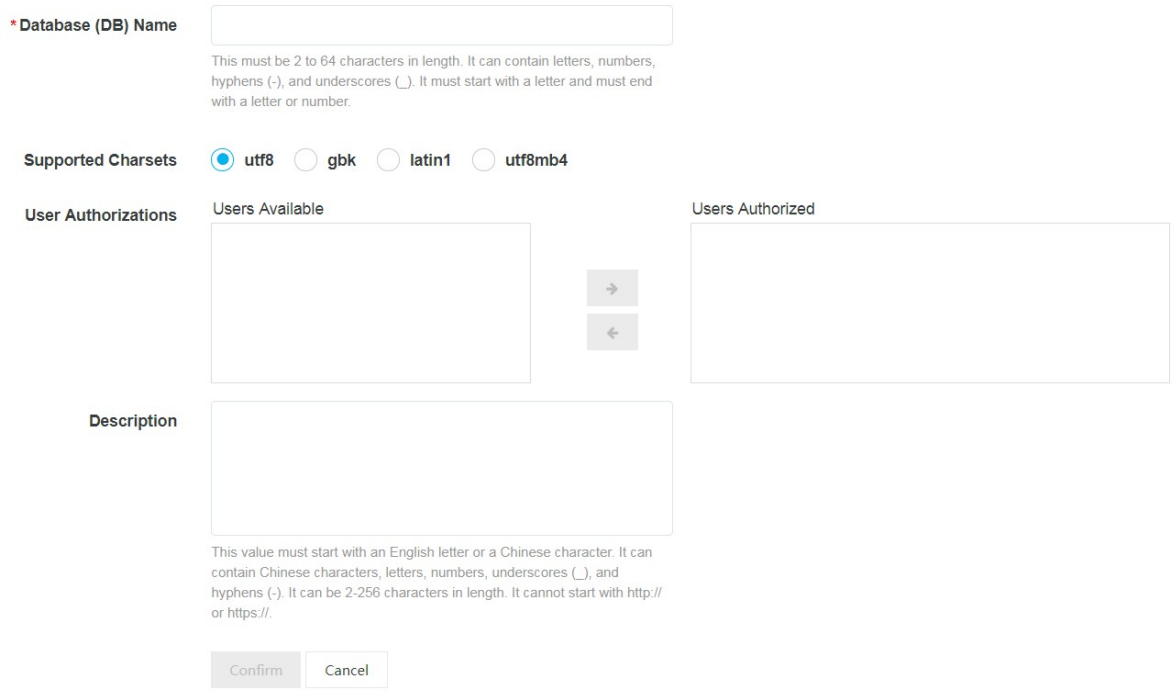
units, so that each account can only access data for its own services. If the account does not need to write data to a database, assign read-only permissions.

- For database security, set strong passwords for the accounts and change the passwords periodically.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the target instance to go to the **Basic Info** page.
3. In the left-side navigation pane, select **Database Management** > **Account List** to go to the **Account List** page.
4. Click **Add Account**.
5. Enter information about the account to be created.
6. Click **Confirm**.
7. In the left-side navigation pane, choose **Database Management** > **Database List** to go to the **Database List** page.
8. Click **Add Database**.
9. Enter information about the database to be created.

Figure 16-7: Create a database



Parameter description:

Table 16-8: Parameters for creating a database

Parameter name	Description
Database (DB) Name	The name can be 2 to 64 characters in length and can contain lowercase letters, numbers, underscores (_), or hyphens (_). It must start with a letter and end with a letter or number.
Supported Charsets	Set a character set for the database.
User Authorizations	Select an account authorized by the database. This parameter can be empty if no account is created. For more information about account authorization, see *Modify account permissions*.
Description	You can enter related information of the database to assist in future database management. You can enter a maximum of 256 English characters.

10. Click **Confirm**.

16.6.3 RDS for PostgreSQL/PPAS

16.6.3.1 Configure a whitelist

To guarantee database security and reliability, you need to modify its whitelist before you enable an instance. You need to add the IP addresses or IP address segments used for database access to the whitelist of the RDS instance. This topic describes how to configure the whitelist of an RDS instance.

Context

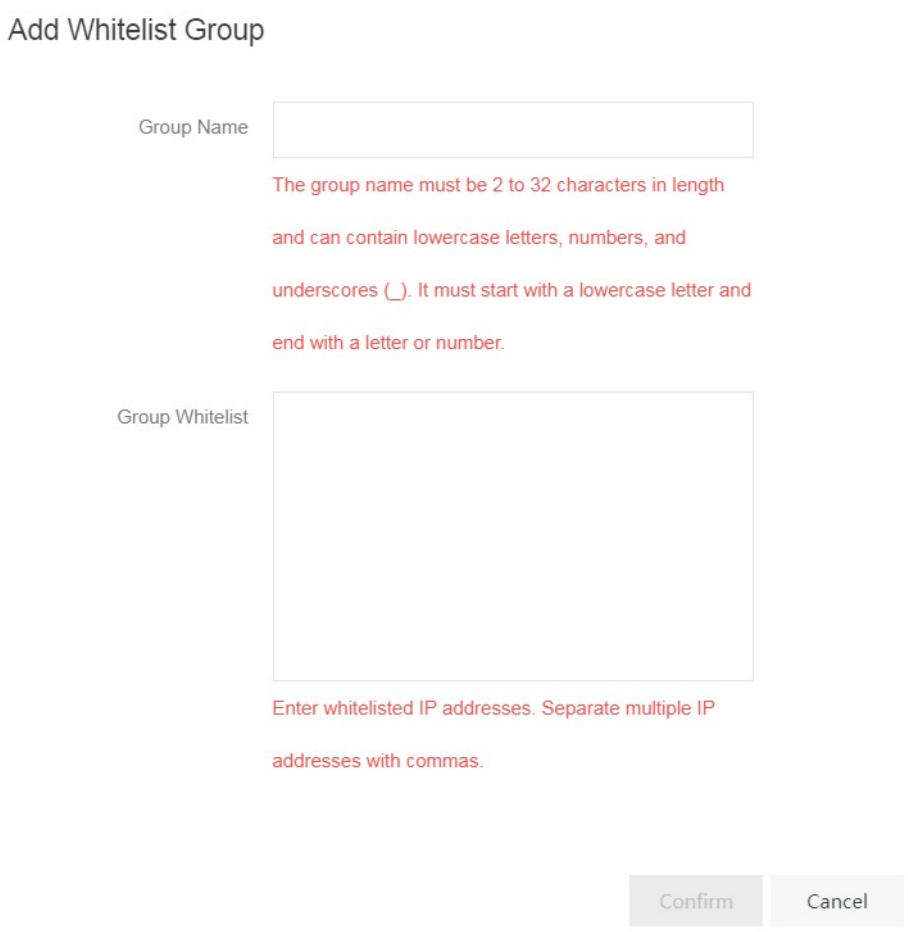
Note:

- The system creates a default whitelist group for each instance. This default whitelist group can only be modified or cleared, but cannot be deleted.
- For each newly created RDS instance, the local loopback IP address 0.0.0.0/0 is added to the **default** whitelist group by default. This means that any IP address is allowed to access the RDS instance. This configuration greatly reduces the security of the database. Delete 0.0.0.0/0 first.
- If the whitelist is configured to 127.0.0.1, all IP addresses or IP segments are prohibited to access the RDS instance. Therefore, you must delete 127.0.0.1 from the whitelist before you add other IP addresses or IP segments.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation bar, select **Security Control** > **Whitelist Settings**.
4. Click the icon of the **default** whitelist group to delete the default whitelist 0.0.0.0/0.
5. Add the IP addresses or IP segments allowed to access the RDS instance to the **default** whitelist group.
6. Click **Confirm**.
7. Click **Add Whitelist Group**.
8. In the **Add Whitelist Group** dialog box, enter the group name and the IP addresses or IP segments allowed to access the RDS instance, and click **Confirm**.

Figure 16-8: Add a whitelist group



Parameter description:

Table 16-9: Add a whitelist group

Parameter name	Description
`Group Name`	The name can be 2 to 32 characters in length and can contain lowercase letters, numbers, and underscores (_). It must start with a lowercase letter and end with a letter or number. You cannot modify the name of a created whitelist group.
`Group Whitelist`	Enter the IP addresses or IP address segments allowed to access the RDS instance. • If you enter an IP address segment, for example, 10.10.10.0/24, any IP address in the format of 10.10.10.X can access the RDS instance. • If you enter multiple IP addresses, separate them with commas (no space before or after each comma), for example, 192.168.0.1,172 .16.213.9.

What's next

Correct use of the whitelist can improve access security for your RDS instance. We recommend that you maintain the whitelist periodically. To modify the whitelist group, click . To clear the default whitelist group or delete a custom whitelist group, click .

16.6.3.2 Create a database and an account

Before using ApsaraDB for RDS, you must create a database and an account in your RDS instance. Before migrating your local database, you must create the same database and account in your RDS instance as those of the local database.

Context

Note:

- To migrate the local database to ApsaraDB for RDS, you must create the same database and account in your RDS instance as those of the local database.
- Databases under the same instance share all resources of this instance.
- When you assign account permissions for each database, follow the minimum permission principle and service roles to create accounts and rationally assign read-only and read/write permissions. When necessary, you may split database accounts and databases into smaller

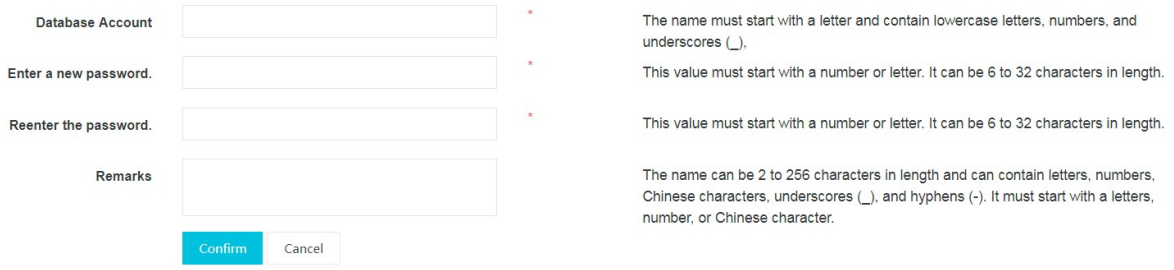
units, so that each account can only access data for its own services. If the account does not need to write data to a database, assign read-only permissions.

- For database security, set strong passwords for the accounts and change the passwords periodically.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Account List** to go to the **Account List** page.
4. Click **Add Account**.
5. Enter information about the account to be created.

Figure 16-9: Create an account



Parameter description:

Table 16-10: Account creation parameters

Parameter name	Description
Database Account	The account can be 2 to 16 characters in length and can contain lowercase letters, numbers, and underscores (_). It must start with a letter and end with a letter or number, for example, *user4example*.
Enter a new password.	The password can be 6 to 32 characters in length and can contain letters, numbers, and underscores (_), for example, *password4example*.
Reenter the password.	Re-enter the password to verify that it is correct, for example, *password4example*.
Remarks	You can enter related information of the account to assist in future account management. You can enter a maximum of 256 English characters.

6. Connect to your RDS instance from the client. For more information, see *Connect to a PostgreSQL or PPAS instance from a client*.
7. Run the following command to create a database.

   ```
   CREATE DATABASE"databasename"
   ```

   Where **databasename** is the name of the database to be created. For example, `CREATE DATABASE "mydatabase"`.

16.7 Instance connection

16.7.1 Connect to a MySQL instance from a client

This topic describes how to connect to an RDS instance from the MySQL-Front client.

Prerequisites

Install the MySQL-Front client.

Context

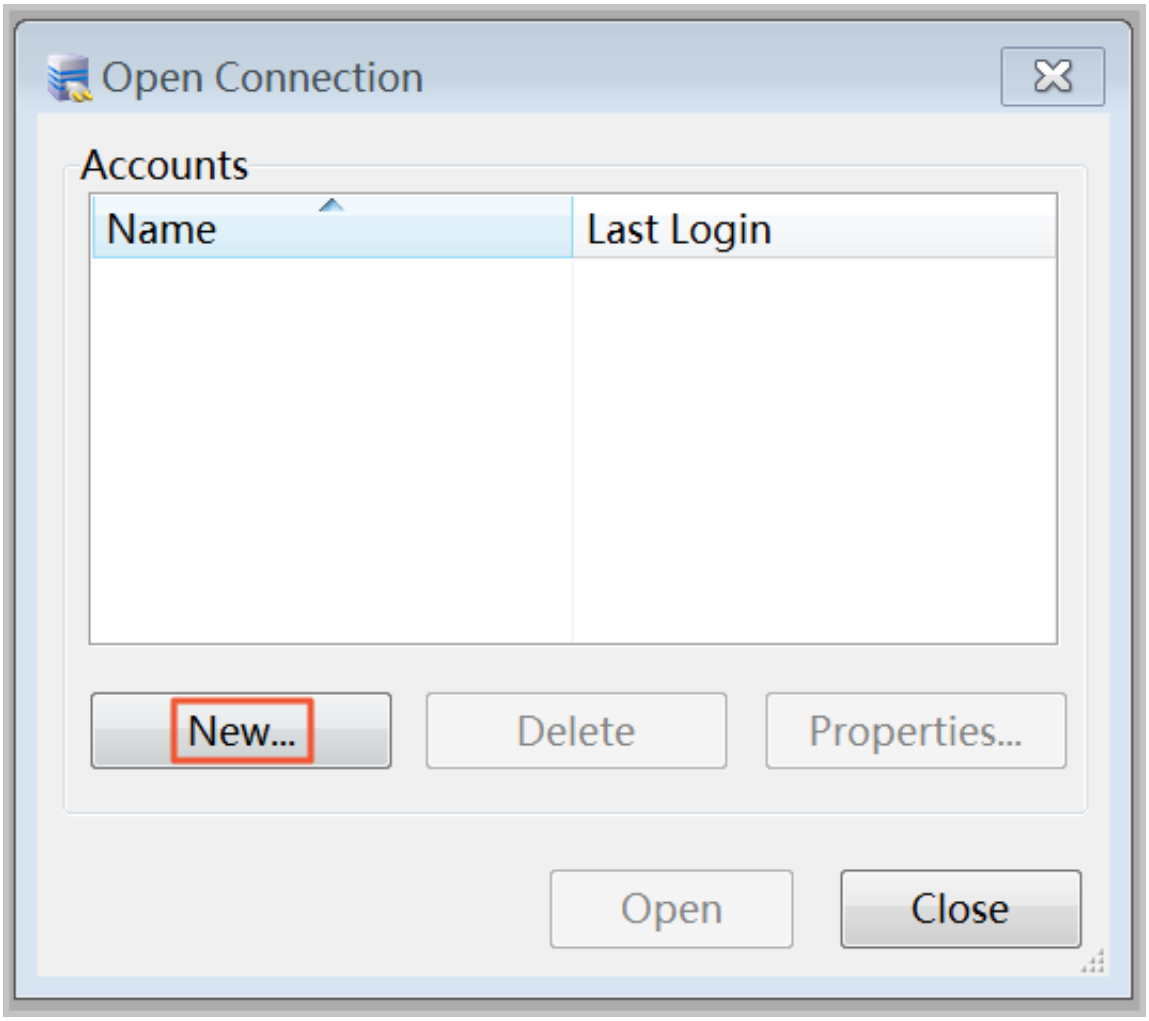
ApsaraDB for RDS for MySQL is fully compatible with MySQL, so you can connect to RDS in the same way you connect to an on-premise MySQL server. This topic describes how to connect to an RDS instance through the MySQL-Front client. You can refer to this topic as an example when connecting through other clients. When you connect to an RDS instance from a client, choose to use the internal address or the public address as follows:

- If your client is deployed on an ECS instance that is in the same region and has the same network type as your RDS instance, use the internal address.
- In other cases, use the public address.

Procedure

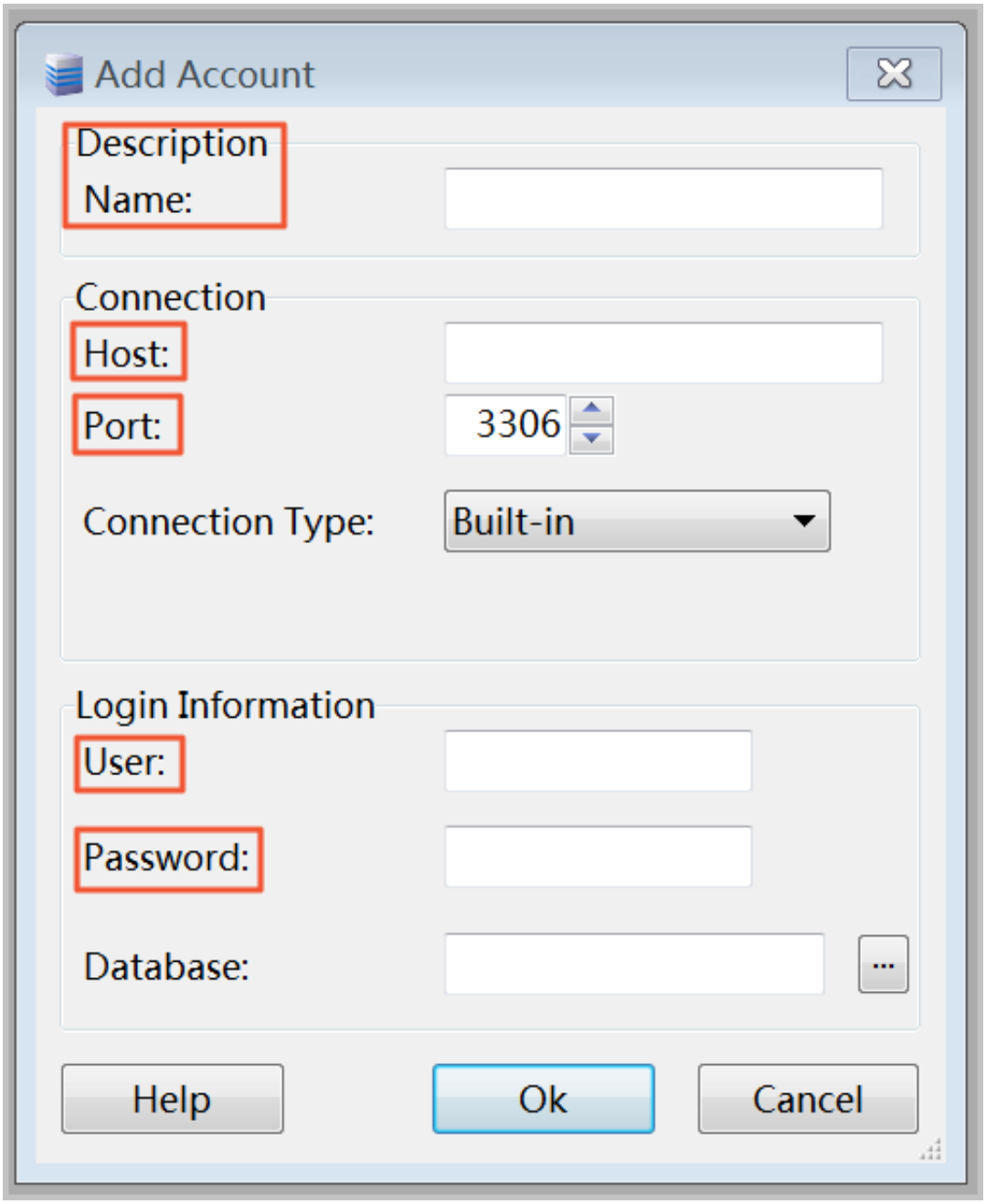
1. Add the IP address used to access the RDS instance to the RDS whitelist. For more information about how to set the whitelist, see *Configure a whitelist*.
2. Start the MySQL-Front client.
3. In the **Open Connection** window, click **New**.

Figure 16-10: New connection



4. Enter the RDS connection information.

Figure 16-11: Enter connection information



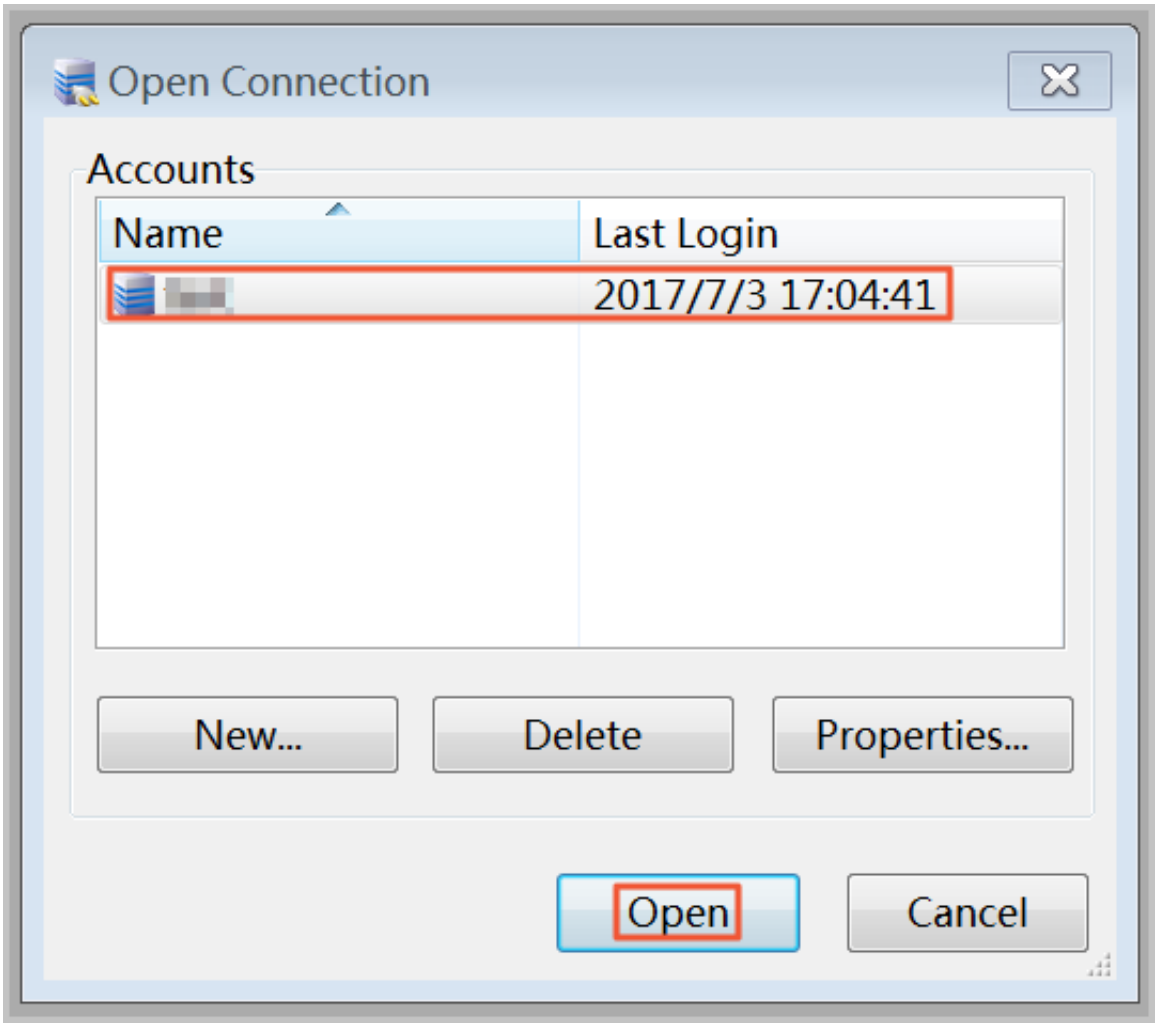
Parameter description:

Parameter name	Description
Name	Enter the connection task name. It is the same as the **Host** field by default.
Host	Enter the internal address of the RDS instance when you use an internal connection. Enter the public address of the RDS instance

Parameter name	Description
	when you use a public connection. You can view the address and port information as follows: 1. *Log on to the RDS console*. 2. Click the ID of the instance to go to the **Basic Information** page. 3. In the **Classic Network Connection Information** area, query the network connection address and port number of the instance.
Port	Enter the internal port number of the RDS instance when you use an internal connection. Enter the public port number of the RDS instance when you use a public connection.
User	Enter the account used to access the RDS instance.
Password	Enter the password for the account to access the RDS instance.

5. Click **OK**.
6. In the **Open Connection** window, select the created connection and click **Open**, as shown in *Figure 16-12: Instance connection*. If the connection information is correct, you can connect to the RDS instance successfully.

Figure 16-12: Instance connection



16.7.2 Connect to an SQL Server instance from a client

This topic describes how to connect to an RDS instance from a Microsoft SQL Server Management Studio (SSMS) client.

Prerequisites

Install the SSMS client.

Context

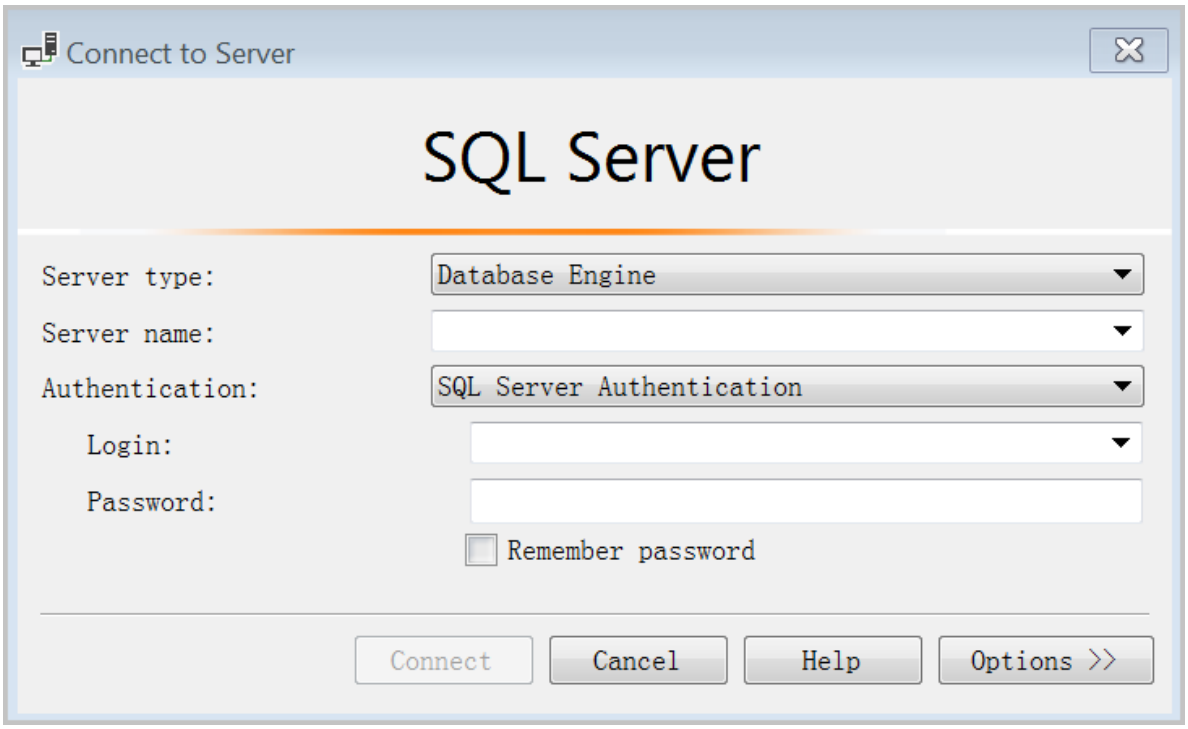
This topic describes how to connect to an SQL Server RDS instance through the SSMS client. You can refer to this topic as an example when connecting through other clients. When connecting to an RDS instance from a client, choose to use the internal address or the public address as follows:

- If your client is deployed on an ECS instance that is in the same region and has the same network type as your RDS instance, use the internal address.
- In other cases, use the public address.

Procedure

1. Add the IP address used to access the RDS instance to the RDS whitelist. For more information about how to set the whitelist, see *Configure a whitelist*.
2. Start the SSMS client.
3. Select **Connect** > **Database Engine**.
4. In the dialog box that appears, enter logon information.

Figure 16-13: Configure connection information



Parameter description:

Parameter name	Description
Server Name	Enter the internal or public address and port number of the RDS instance, and separate the address from the port number with a comma, for example, `rm-bptest.sqlserver.rds.aliyuncs.com,3433`. You can view the address and port information as follows: **1.** *Log on to the RDS console*.

Parameter name	Description
	2. Click the ID of the instance to go to the **Basic Information** page. 3. In the **Classic Network Connection Information** area, query the network connection address and port number of the instance.
`Identity Verification`	Select SQL Server identity verification.
`User name`	Enter the account used to access the RDS instance.
`Password`	Enter the password for the account used to access the RDS instance.

5. Click **Connect**. If the connection information is correct, you can connect to the RDS instance successfully.

16.7.3 Connect to a PostgreSQL or PPAS instance from a client

This topic describes how to connect to an RDS instance from a pgAdmind 4 client.

Prerequisites

Make sure that you have installed the pgAdmind 4 client.

Context

This topic describes how to connect to an RDS instance through the pgAdmind 4 client. You can refer to this topic as an example when connecting through other clients. When you connect to an RDS instance from a client, choose to use the internal address or the public address as follows:

- If your client is deployed on an ECS instance that is in the same region and has the same network type as your RDS instance, use the internal address.
- In other cases, use the public address.

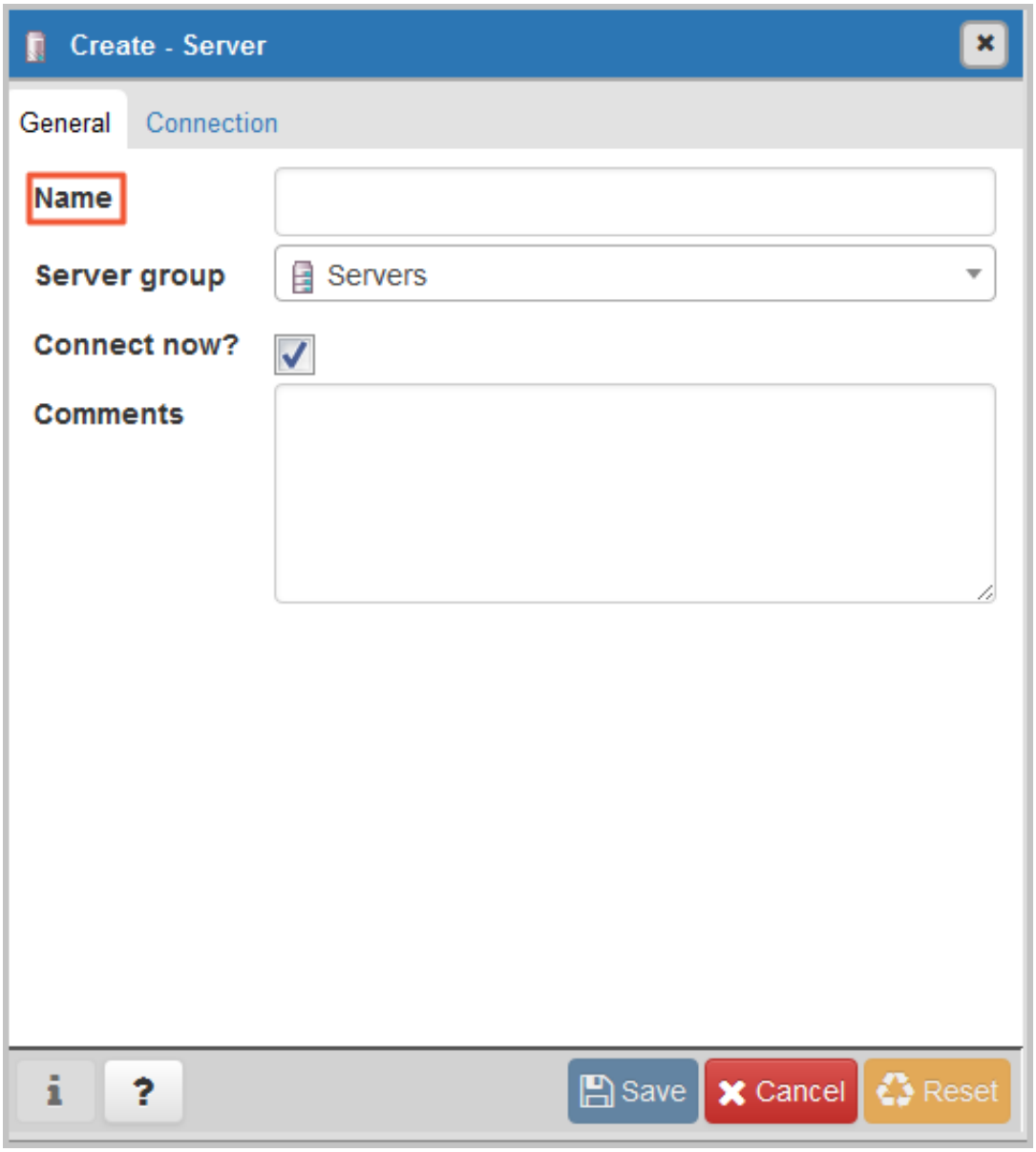
Procedure

1. Add the IP address used to access the RDS instance to the RDS whitelist. For more information about how to set the whitelist, see *Configure a whitelist*.
2. Start the pgAdmind 4 client.
3. Right-click **Servers** and select **Create** > **Server**.

Figure 16-14: Create a service

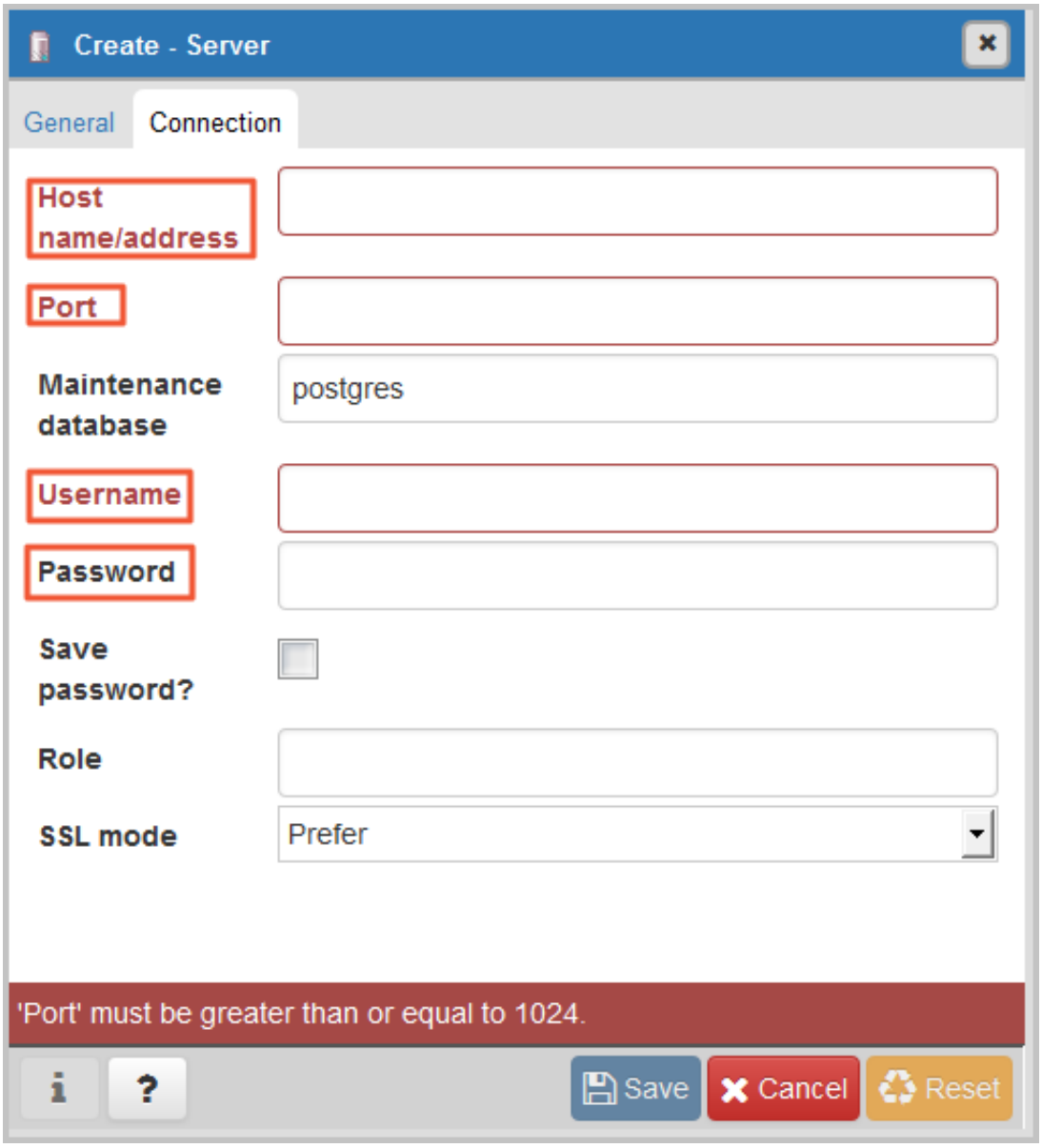
4. On the **Create - Server** dialog box, click the **General** tab and enter the server name.

Figure 16-15: Enter a server name



5. Click the **Connection** tab and enter instance connection information.

Figure 16-16: Configure instance connection information



Parameter description:

Parameter name	Description
`Host name/address`	Enter the internal address of the RDS instance when you use an internal connection. Enter the public address of the RDS instance when you use a public connection. You can view the address and port information as follows: 1. *Log on to the RDS console*. 2. Click the ID of the target instance to go to the **Basic Information** page.

Parameter name	Description
	3. In the **Classic Network Connection Information** area, query the network connection address and port number of the instance.
Port	Enter the internal port number of the RDS instance when you use an internal connection. Enter the public port number of the RDS instance when you use a public connection.
Username	Enter the account used to access the RDS instance.
Password	Enter the password of the account used to access the RDS instance.

6. Click **Save**.
7. If the connection information is correct, select **Servers** > **Server Name** > **Database** > **postgres**.

16.8 Read-only instances

16.8.1 Read-only instances

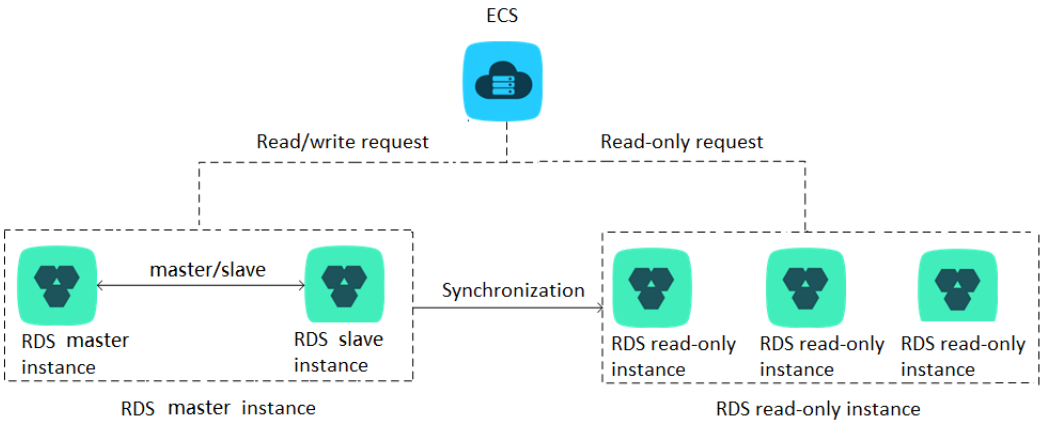
Introduction

Currently, only ApsaraDB for MySQL 5.6 supports read-only instances.

A single instance may be unable to address the read pressure in scenarios where there are massive read requests. In this case, main services may be affected. To achieve auto scaling of read capability and relieve database pressure, you can create one or more read-only instances in a region. In this case, massive data can be read from the database and application throughput can be increased.

A read-only instance uses a single physical node without backup nodes. It uses the native replication capability of MySQL to synchronize changes in the primary instance to all relevant read -only instances. Read-only instances must be in the same region as the primary instance, but not necessarily in the same zone as the primary instance. The following figure shows the topology of a read-only instance.

Figure 16-17: Topology of a read-only instance



Features

Read-only instances have the following features:

- The instance type can be different from that of the primary instance and can be changed at any time, which assists in elastic upgrade or downgrade.
- No account or database maintenance is required for a read-only instance. Both the account and database of the read-only instance are synchronized from the primary instance.
- Read-only instances support independent whitelist configuration.
- System performance monitoring is provided.

  ApsaraDB for RDS provides nearly 20 system performance monitoring views, such as those for disk capacity, IOPS, connections, CPU usage, and network traffic. You can easily view the load of the instances.
- Optimization recommendations: ApsaraDB for RDS provides a variety of optimization recommendations, such as storage engine check, primary key check, large table check, and the check for excessive indexes and missing indexes. You can optimize your databases based on the optimization recommendations and their specific applications.

Restrictions

Read-only instances have the following functional limitations:

- A maximum of five read-only instances can be created for each primary instance.
- Backup configuration: Backup configuration and temporary backup are not supported.
- Instance recovery:

  - Temporary instances cannot be created through backup files or any point in time. Instances cannot be overwritten through backup sets.
  - The primary instance cannot use backup sets to overwrite a created read-only instance to recover its data.
- Data migration: Data cannot be migrated to read-only instances.
- Database management: Databases cannot be created or deleted.
- Account management: Accounts cannot be created, deleted, or granted permissions. Account passwords cannot be changed.

16.8.2 Create a read-only instance

Context

Note:

- A maximum of five read-only instances can be created for each primary instance.
- This operation is only applicable to instances of MySQL 5.6.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Performance Optimization** > **Read-Only Instance**.
4. Click **Create Read-only Instance**.
5. Select instance configurations and click **Create**.

Note:

- If a VPC is used, we recommend that you choose the same VPC as that of the primary instance.
- To ensure sufficient I/O for data synchronization, we recommend that the memory configuration of the read-only instance is at least the same as the primary instance.

16.8.3 Read-only instance management

16.8.3.1 Access the read-only instance management page through a read-only instance

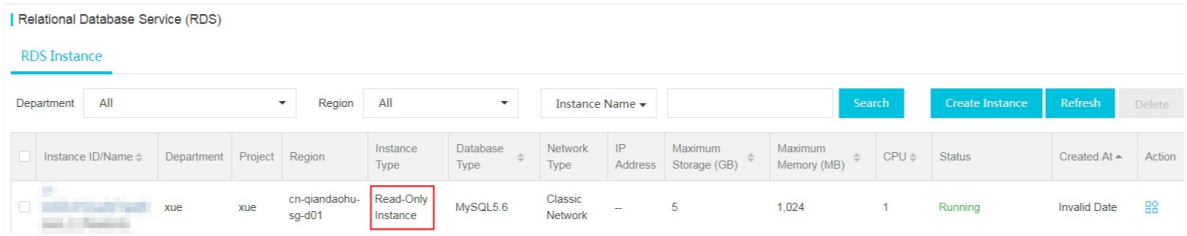
You can go to the read-only instance management page from the instance list page. You can also access the management page from the read-only instance list page of the primary instance. You can manage read-only instances in a way similar to ordinary instance management methods. The page shows the management operations that can be performed. This topic describes how to go to the read-only instance management page from the instance list page.

Procedure

1. *Log on to the RDS console*.
2. On the **RDS Instances** list page, click the ID of the read-only instance to go to the **Basic Information** page. This page allows you to manage the read-only instance.

   In the instance list, read-only instances are displayed with **Instance Type** as **`Read-Only Instances`**, as shown in *Figure 16-18: View read-only instances*.

Figure 16-18: View read-only instances



16.8.3.2 Access the read-only instance management page through the primary instance

You can go to the read-only instance management page from the instance list page. You can also access the management page from the read-only instance list page of the primary instance. You can manage read-only instances in a way similar to ordinary instance management methods. The page shows the management operations that can be performed. This topic describes how to go to the read-only instance management page from the read-only instance list page of the primary instance.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.

3. In the left-side navigation pane, choose **Performance Optimization** > **Read-Only Instance**.
4. Click the ID of the read-only instance to go to the **Basic Information** page. This page allows you to manage the read-only instance.

16.9 Read/write splitting

16.9.1 Read/Write splitting

This topic describes principles, benefits, and restrictions of read/write splitting.

Function overview

Currently, read/write splitting is only available in MySQL 5.6 because only read-only instances support read/write splitting. When read/write splitting is enabled, an instance provides three connection addresses:

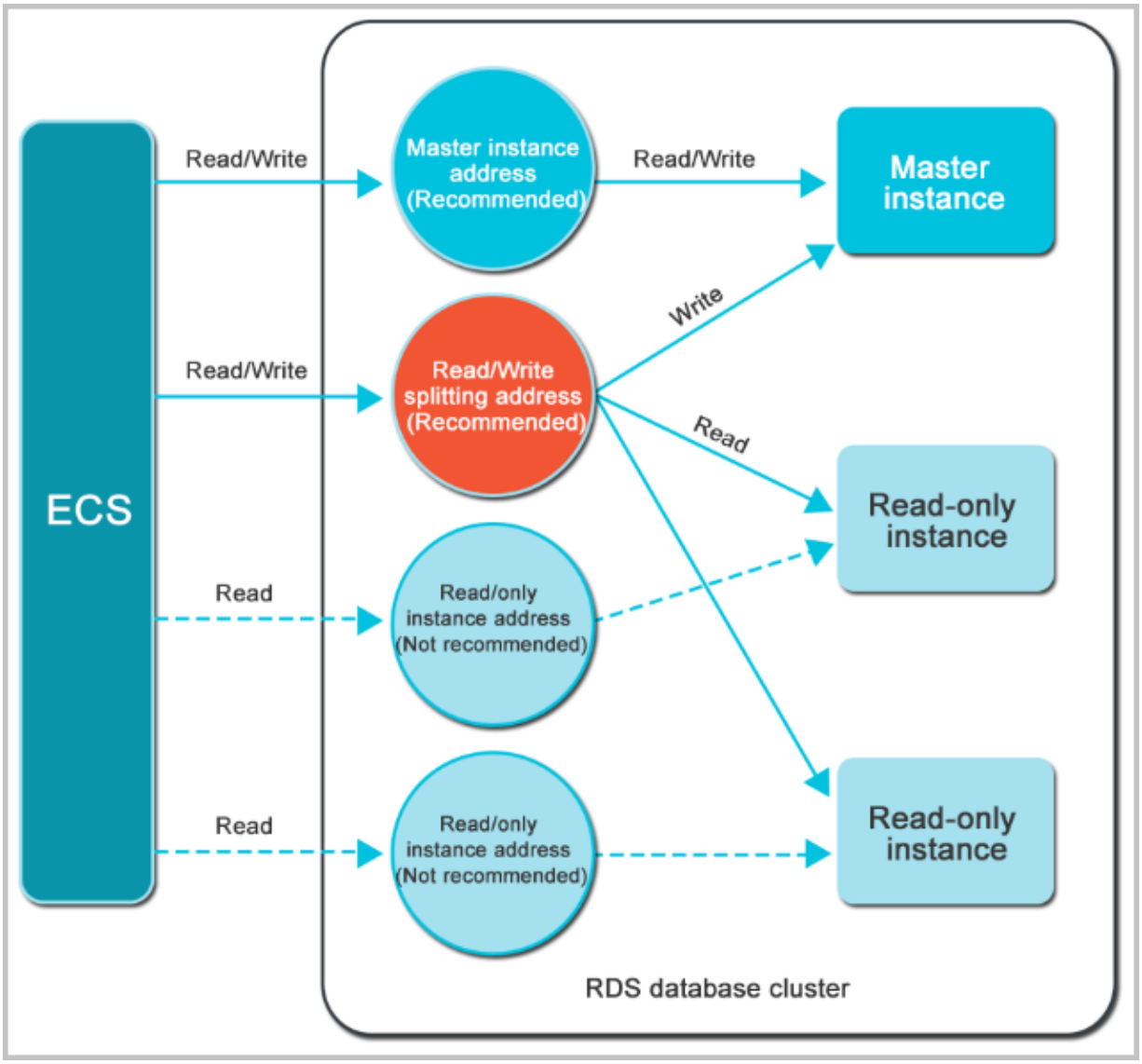
- Connection address of the primary instance
- Connection address of the read-only instance
- Connection address of read/write splitting

The primary instance and read-only instance have independent connection addresses. Currently , instance connection addresses are automatically configured in the application to split read and write operations.

The read/write splitting function provides an extra read/write splitting address. This address links the primary instance with all its read-only instances to enable an automatic link for read/write splitting. The application can use this method to read and write data by connecting to the same read/write splitting address. Write requests are automatically routed to the primary instance, and read requests are routed to each read-only instance based on their weight. You can scale up the processing capability of the system by adding more read-only instances. No application change is required.

*Figure 16-19: Principle of read/write splitting*shows how the application uses different types of connection addresses to access the database.

Figure 16-19: Principle of read/write splitting



Benefits

- Easy maintenance with a single read/write splitting address

  In the existing read-only instance model, the primary instance and read-only instances require independent connection addresses. You need to configure and manage each of these addresses in the application, so that write requests can go directly to the primary instance and read requests to read-only instances.

  This function provides an extra read/write splitting address. You can connect to this address to perform read/write operations on the primary and read-only instances. In addition, the forwarding logic of read/write statements is visible to you, which reduces maintenance costs.
- Improved performance with native RDS highly secure links

If you build a proxy layer to implement read/write splitting in the cloud, data has to go through multiple components for statement parsing and forwarding before it reaches the database, which affects response latency significantly. RDS read/write splitting is built into the existing highly secure link. No extra component is used. This reduces latency and increases the processing speed.

- Wide usage scenarios with customizable weights and thresholds

  RDS read/write splitting allows you to set read request weights for the primary and read-only instances. You can also set latency thresholds for read-only instances.

- Enhanced database availability with instance health check

  RDS read/write splitting performs health checks automatically for all instances in the distribution system. If any instance fails or its latency exceeds the threshold, ApsaraDB for RDS automatically removes the instance from the distribution system. The removed instance is marked as unavailable and no longer allocated with read requests. Read and write requests are allocated among the remaining healthy instances based on the predefined weights. This method ensures that the application can still run properly when a single-node read-only instance fails. After the instance recovers, ApsaraDB for RDS automatically reclaims the instance into the request distribution system.

  Note:
  To prevent a single point of failure, we recommend that you create at least two read-only instances for each primary instance.

Restrictions

- Currently, the following commands or functions cannot be forwarded to a read-only instance:
  - The `stmt prepare sql` command is automatically executed on the primary instance.
  - The `stmt prepare` command cannot be forwarded to a read-only instance before the `stmt close` command is run.
  - The environment variable configurations of `set global`, `set user`, and `set once` are automatically executed on the primary instance.
- The following commands or functions are not currently supported:
  - SSL encryption
  - Compression protocols
  - com_dump_table and com_change_user protocols

- `kill connection [query]`
- `change user`

- Execution results of the following commands are random:

  The `show processlist`, `show master status`, and `com_process_info` commands return results based on the instance connected during execution.
- All transactions are routed to the primary database.
- Consistent non-transaction reads are not guaranteed in read/write splitting. If you require consistent reads, add hints to route query requests to the primary database or encapsulate query requests in transactions.

16.9.2 Enable read/write splitting

In scenarios where there are few write requests but a great number of read requests, you can enable read/write splitting to distribute read pressure of the primary instance. This topic describes how to enable the read/write splitting function.

Prerequisites

- Read/write splitting can only be enabled on a high-availability MySQL 5.6 primary instance.
- A read-only instance has been created under the primary instance. If there is no read-only instance, contact the administrator.
- The primary instance has been switched to the safe mode.

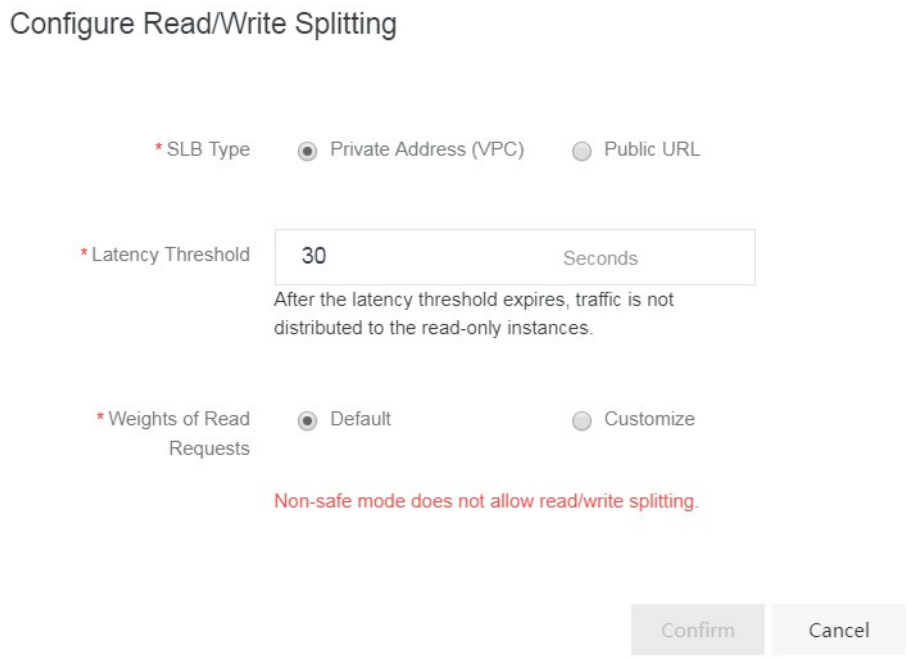
> **Note:**
> When you enable the read/write splitting function for the first time, the system automatically upgrades the backend control system of the primary instance and all associated read-only instances to the latest version to guarantee normal service operations. The primary instance and read-only instances automatically restart once when the function is enabled. During the restart process, the primary instance incurs a transient disconnection of 30 seconds or less, and the read-only instances are inaccessible. To avoid service impacts from transient disconnections, we recommend that you enable read/write splitting during off-peak hours and make sure that your application can be automatically reconnected.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Performance Optimization** > **Read/Write Splitting**.

4. Click **Enable** to open the **Configure Read/Write Splitting** dialog box.
5. Enter configuration information, as shown in *Figure 16-20: Configure read/write splitting*.

Figure 16-20: Configure read/write splitting



Parameter description:

Table 16-11: Read/Write splitting parameters

Name	Description
SLB Type	Read/write splitting address, which can be an internal address or a public address. If an internal address is selected, the internal network type of read/write splitting automatically is consistent with that of the primary instance. For example, if the internal network type of the primary instance is VPC, the internal network type of read/write splitting is also VPC.
Latency Threshold	Maximum allowed latency when read-only instances synchronize data from the primary instance. The value ranges from 0 to 7200 seconds. If the latency of a read-only instance exceeds this threshold, read requests are not forwarded to this instance regardless of its weight. Read-only instances probably incur latency based on SQL execution. We recommend that you set the parameter to a value not less than 30 seconds.
Weights of Read Requests	Read request weight of each instance. An instance with a higher weight can process more read requests. For example, a read/write splitting address is associated with one primary instance and three read-only instances. Their read weights are 0, 100, 200, and 200 respectively. The

Name	Description
	primary instance does not process read requests, and write requests are automatically forwarded to the primary instance. The three read-only instances process read requests by a ratio of 1:2:2. You can use either of the following weight configuration methods: • **`Default`**: The system automatically distributes weights to instances based on the instance type. Any new read-only instances under the primary instance are automatically added to the read/write splitting link based on the weights distributed by the system. • **`Customize`**: You can customize the read request weight of each instance in the range of 0 to 10000. If you select this method, the weights of new read-only instances under the primary instance are 0 by default. You need to set the weights manually.

6. Click **Confirm**.

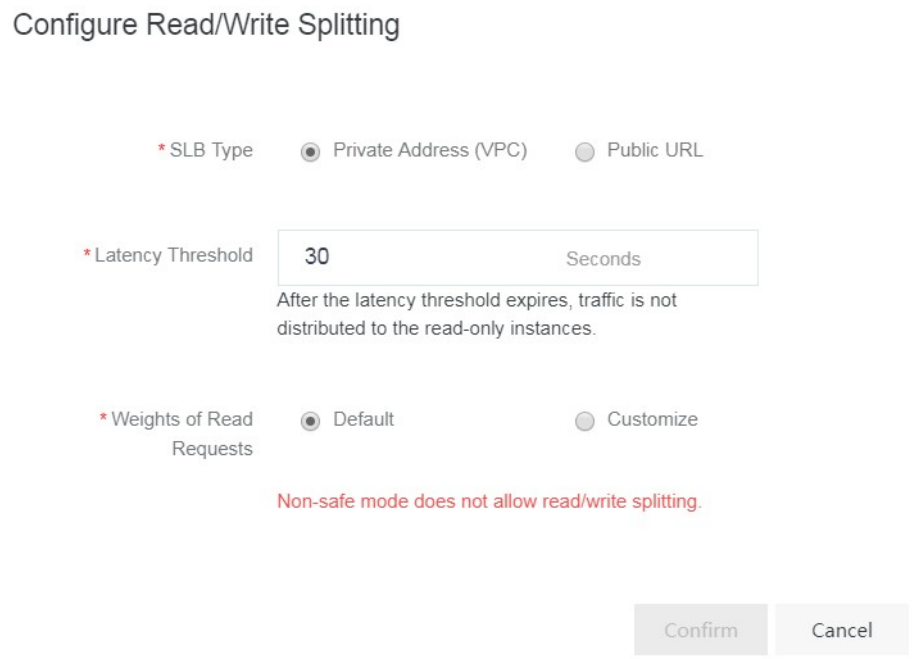
16.9.3 Modify the latency threshold and read weight distribution

After read/write splitting is enabled, you can configure it as needed. This topic describes how to modify the latency threshold and read weight distribution of read/write splitting.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Performance Optimization** > **Read/Write Splitting**.
4. Click **Enable Read/Write Splitting** to open the **Configure Read/Write Splitting** dialog box.
5. Modify configurations as follows.

Figure 16-21: Configure read/write splitting



Parameter Description:

Table 16-12: Read/Write splitting parameters

Parameter Name	Description
Latency Threshold	Maximum allowed latency when read-only instances synchronize data from the primary instance. The value ranges from 0 to 7200 seconds. If the latency of a read-only instance exceeds this threshold, read requests are not forwarded to this instance regardless of its weight. Read-only instances incur latencies based on SQL execution. We recommend that you set the parameter to a value of not less than 30 seconds.
Weights of Read Requests	Read request weight of each instance. An instance with a higher weight can process more read requests. For example, a read/write splitting address is associated with one primary instance and three read-only instances. Their read weights are 0, 100, 200, and 200 respectively. The primary instance does not process read requests, and write requests are automatically forwarded to the primary instance. The three read-only instances process read requests by a ratio of 1:2:2. You can use either of the following weight configuration methods: • **Default**: The system automatically distributes weights to instances based on the instance type. Any new read-only instances under the

Parameter Name	Description
	primary instance are automatically added to the read/write splitting link based on the weights distributed by the system. • **`Customize`**: You can customize the read request weight of each instance in the range of 0 to 10000. If you select this method, the weights of all new read-only instances under the primary instance are set to 0 by default. You need to set the weights manually.

6. Click **Confirm**.

16.9.4 Disable read/write splitting

This topic describes how to disable the read/write splitting function.

Context

You can disable read/write splitting if this function is no longer required. The read/write splitting function can only be used when at least one read-only instance is available. Therefore, you must disable the read/write splitting function before you delete the last read-only instance. Otherwise, the deletion fails.

Note:
After read/write splitting is disabled, your application can no longer connect to the read/write splitting address. Make sure that your database connection configuration does not include this connection address.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Performance Optimization** > **Read/Write Splitting**.
4. Click **Disable Read/Write Splitting**.
5. Click **Confirm**.

16.9.5 Monitor read/write splitting performance

You can view the read/write splitting performance on the monitoring page of the RDS console.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.

3. In the left-side navigation pane, choose **System Resource Monitoring** > **Database Performance**.
4. Select **QPS/TPS** to view transaction per second (TPS) and query per second (QPS). You can view the number of reads and writes of all databases, which includes the primary database and read-only databases involved in read/write splitting.

16.9.6 Rules of system weight distribution

This topic describes the rules for the system to distribute read weights.

Table of weight values

The system automatically configures fixed read weight values for instances, as shown in the following table.

Code	Type	Memory	CPU	Weight
rds.mysql.t1.small	General	1GB	1	100
rds.mysql.s1.small	General	2GB	1	100
rds.mysql.s2.large	General	4GB	2	200
rds.mysql.s2.xlarge	General	8GB	2	200
rds.mysql.s3.large	General	8GB	4	400
rds.mysql.m1.medium	General	16GB	4	400
rds.mysql.c1.large	General	16GB	8	800
rds.mysql.c1.xlarge	General	32GB	8	800
rds.mysql.c2.xlarge	General	64GB	16	1600
rds.mysql.c2.xlp2	General	96GB	16	1600
mysql.x8.medium.2	Exclusive package	16GB	2	200
mysql.x8.large.2	Exclusive package	32GB	4	400

Code	Type	Memory	CPU	Weight
mysql.x8.xlarge.2	Exclusive package	64GB	8	800
mysql.x8.2xlarge.2	Exclusive package	128GB	16	1600
rds.mysql.st.d13	Exclusive host	220GB	30	3,000

Specify whether an SQL statement is sent to the master instance or a read-only instance with hints

In addition to the weight distribution of read/write splitting, a hint provides complementary SQL syntax to specify whether an SQL statement is executed in the master instance or a read-only instance.

RDS read/write splitting supports the following hint formats:

- `/*FORCE_MASTER*/`: specifies that subsequent SQL statements are executed in the master instance.
- `/*FORCE_SLAVE*/`: specifies that subsequent SQL statements are executed in the read-only instance.

For example, after a hint is prefixed to the following statement, the statement is always routed to and executed in the primary instance regardless of the weight value.

```
/*FORCE_MASTER*/ SELECT * FROM table_name;
```

16.10 Instance management

16.10.1 Query details

You can view the details of an instance, such as the basic information, internal network connection information, running status, and its configuration. This topic describes how to query the details of an instance.

Procedure

1. *Log on to the RDS console*.
2. You can use either of the following ways to go to the instance details page:
   - Click the ID of the instance to go to the **Basic Information** page.
   - In the **Action** column of the instance, choose > **View Details**.

16.10.2 Restart an instance

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. Click **Restart Instance** in the upper-right corner of the instance management page.
4. In the confirmation dialog box that appears, click **Confirm**.

16.10.3 Modify configurations

You can modify the configurations of your instance, such as memory and storage space, if the configurations are either too high or it cannot meet application requirements.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. Click **Modify Configuration** in the upper-right corner of the page.
4. On the configuration page, select required configurations.
5. Click **Confirm**.

16.10.4 Release an instance

Procedure

1. *Log on to the RDS console*.
2. In the **Action** column of the instance, choose  > **Delete**.
3. Click **Confirm**.

16.10.5 Set parameters

ApsaraDB for RDS allows you to define some instance parameters. For more information about the parameters that can be modified, see **Parameter Settings** on the RDS console.

Context

 **Note:**

- Set parameters on the **Parameter Settings** page in accordance with the specified `Parameter Range`.

- Modification of some parameters will require you to restart the instance. Go to the **Parameters** page and check the **`Requires Restart`** option to see if a restart is required. Before you restart the instance, make sure that the instance restart does not affect other services.

Because ApsaraDB for RDS is fully compatible with MySQL, their parameter setting methods are similar. Refer to this example to modify parameters on the RDS console.You can also run commands in API mode to modify the parameters.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Performance Optimization** > **Parameter Settings**.
4. In the **Action** column of the parameter, select  > **Edit**.
5. In the window that appears, enter a new value and click **Confirm**.

16.10.6 Change ownership

You can change the ownership (department and project) of an instance based on your service requirements. This topic describes how to change the ownership of an instance.

Procedure

1. *Log on to the RDS console*.
2. In the **Action** column of the instance, choose  > **Change Ownership**.
3. On the **Change Ownership** page, select a new department and project for the instance.
4. Click **Confirm**.

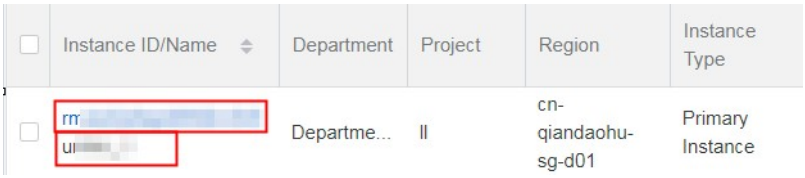
16.10.7 Modify an instance name

You can modify instance names to assist in management. This topic describes how to modify the name of an instance.

Context

In the instance list, the **`Instance ID/Name`** column shows instance IDs in the upper part and instance names in the lower part, as shown in *Figure 16-22: Instance ID/Instance name*. You can modify instance names but cannot modify instance IDs.

Figure 16-22: Instance ID/Instance name



Procedure

1. *Log on to the RDS console*.
2. In the **Action** column of the instance, choose > **Edit Instance Name**.
3. In the **`Instance Name`** text box, enter a new name for the instance.
4. Click **Confirm**.

16.10.8 Typical parameter configuration

16.10.8.1 Modifiable MySQL instance parameters

*Table 16-13: Modifiable MySQL instance parameters*lists the modifiable MySQL instance parameters. For more information, see the MySQL official documentation at *https://dev.mysql.com /doc/*.

Table 16-13: Modifiable MySQL instance parameters

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
auto_increment_increment	1	1	No	[1-65535]	auto_increment_increment and auto_increment_offset are intended for use with master-to-master replication, and can be used to control the operation of AUTO_INCREMENT columns. Both variables have global and session values, and each can assume an integer value between 1 and 65,535 inclusive. Setting the value of either of these

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
					two variables to 0 causes its value to be set to 1 instead. Attempting to set the value of either of these two variables to an integer greater than 65, 535 or less than 0 causes its value to be set to 65, 535 instead. Attempting to set the value of auto_increment_increment or auto_increment_offset to a noninteger value produces an error, and the actual value of the variable remains unchanged.
auto_increment_offset	1	1	No	[1-65535]	Determines the starting point for the AUTO_INCREMENT column value.
back_log	3000	3000	Yes	[0-65535]	The number of outstanding connection requests that MySQL can have.
binlog_cache_size	209715228 KB		No	[4096-16777216]	The size of the cache to hold changes to the binary log during a transaction.
binlog_checksum	CRC32	CRC32	Yes	[CRC32|NONE]	The master to write a checksum for each event in the binary log.
binlog_row_image	Full	Full	No	[full|minimal]	Binlog save every column or actually required column in binlog images.
binlog_stmt_cache_size	32768	32768	No	[4096-16777216]	The size of the statement cache for updates to non-transactional engines for the binary log.

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
character_set_server	utf8	utf8	Yes	[utf8|latin1|gbk|utf8mb4]	The server's default character set.
concurrent_insert	1	1	No	[0|1|2]	NEVER-Disables concurrent inserts; AUTO-(Default) Enables concurrent insert for MyISAM tables that do not have holes; ALWAYS-Enables concurrent inserts for all MyISAM tables, even those that have holes. For a table with a hole, new rows are inserted at the end of the table if it is in use by another thread. Otherwise, MySQL acquires a normal write lock and inserts the row into the hole.
connect_timeout	10	10	No	[1-3600]	The number of seconds that the mysqld server waits for a connect packet before responding with Bad handshake. The default value is 10 seconds as of MySQL 5.1.23 and 5 seconds before that. Increasing the connect_timeout value might help if clients frequently encounter errors of the form Lost connection to MySQL server at 'XXX', system error: errno.
default_storage_engine	InnoDB	TokuDB	Yes	[InnoDB|TokuDB|innodb|tokudb]	The default storage engine for new tables.
default_time_zone	SYSTEM	SYSTEM	Yes	[SYSTEM|-12:00|-11:00	The default time zone for the database.

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
				|-10:00|-9:00|-8:00|-7:00|-6:00|-5:00|-4:00|-3:00|-2:00|-1:00|+0:00|+1:00|+2:00|+3:00|+4:00|+5:00|+5:30|+6:00|+6:30|+7:00|+8:00|+9:00|+10:00|+11:00|+12:00|+13:00]	
default_week_format	0	0	No	[0-7]	The default mode value to use for the WEEK() function.
delayed_insert_limit	100	100	No	[1-4294967295]	After inserting delayed_insert_limit delayed rows, the INSERT DELAYED handler thread checks whether there are any SELECT statements pending. If so, it permits them to execute before continuing to insert delayed rows.
delayed_insert_timeout	300	300	No	[1-3600]	How many seconds an INSERT DELAYED handler thread should wait for INSERT statements before terminating.
delayed_queue_size	1000	1000	No	[1-4294967295]	This is a per-table limit on the number of rows to queue when handling INSERT DELAYED statements. If the queue becomes full, any client that issues an INSERT DELAYED statement waits

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
					until there is room in the queue again.
delay_key_write	ON	ON	No	[ON|OFF|ALL]	This option applies only to MyISAM tables. It can have one of the following values to affect handling of the DELAY_KEY_WRITE table option that can be used in CREATE TABLE statements.
div_precision_increment	4	4	No	[0-30]	This variable indicates the number of digits by which to increase the scale of the result of division operations performed with the / operator. The default value is 4. The minimum and maximum values are 0 and 30, respectively. The following example illustrates the effect of increasing the default value.
eq_range_index_dive_limit	10	10	No	[1-200]	This variable indicates the number of equality ranges in an equality comparison condition when the optimizer should switch from using index dives to index statistics in estimating the number of qualifying rows.
explicit_defaults_for_timestamp	False	False	Yes	True/false	As indicated by the warning, to turn off the nonstandard behaviors, enable the explicit_defaults_for_timestamp system variable at server startup.

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
ft_min_word_len	4	4	Yes	[1-3600]	The minimum length of the word to be included in a FULLTEXT index.
ft_query_expansion_limit	20	20	Yes	[0-1000]	The number of top matches to use for full-text searches performed using WITH QUERY EXPANSION.
group_concat_max_len	1024	1024	No	[4-18446744073709547520]	The maximum permitted result length in bytes for the GROUP_CONCAT() function. The default is 1024 , Unit:Byte.
innodb_adaptive_hash_index	ON	ON	No	[ON|OFF]	It may be desirable, depending on your workload , to dynamically enable or disable adaptive hash indexing to improve query performance. The size in bytes of a memory pool InnoDB uses to store data dictionary information and other internal data structures. The more tables you have in your application, the more memory you allocate here. If InnoDB runs out of memory in this pool, it starts to allocate memory from the operating system and writes warning messages to the MySQL error log. The default value is 8MB.
innodb_additional_mem_pool_size	2097152	2097152	Yes	[2097152-104857600]	The locking mode to use for generating auto-increment values. The permissible values are 0, 1, or 2.

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
innodb_autoinc_lock_mode	1	1	Yes	[0|1|2]	The number of threads that can enter InnoDB concurrently is determined by the innodb_thread_concurrency variable.
innodb_concurrency_tickets	5000	5000	No	[1-4294967295]	The number of threads that can enter InnoDB concurrently is determined by the innodb_thread_concurrency variable.
innodb_ft_max_token_size	84	84	Yes	[10-84]	Maximum length of words that are stored in an InnoDB FULLTEXT index.
innodb_ft_min_token_size	3	3	Yes	[0-16]	Minimum length of words that are stored in an InnoDB FULLTEXT index.
innodb_large_prefix	OFF	OFF	No	[ON|OFF]	Enable this option to allow index key prefixes longer than 767 bytes (up to 3072 bytes), for InnoDB tables that use the DYNAMIC and COMPRESSED row formats.
innodb_lock_wait_timeout	50	50	No	[1-1073741824]	The timeout in seconds an InnoDB transaction may wait for a row lock before giving up. The default value is 50 seconds. Unit: Second.
innodb_max_dirty_pages_pct	75	75	No	[50-90]	This is an integer in the range from 0 to 100. The default value is 90 for the built-in InnoDB, 75 for InnoDB Plugin. The main thread in InnoDB tries to write pages from the buffer pool so that the percentage of dirty (not yet written)

Parameter	Defaul value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
					pages will not exceed this value.
innodb_old_blocks_pct	37	37	No	[5-95]	(InnoDB Plugin only) Specifies the approximate percentage of the InnoDB buffer pool used for the old block sublist. The range of values is 5 to 95. The default value is 37 (that is, 3/8 of the pool).
innodb_old_blocks_time	1000	1000	No	[0-1024]	(InnoDB Plugin only) Specifies how long in milliseconds (ms) a block inserted into the old sublist must stay there after its first access before it can be moved to the new sublist. The default value is 0: A block inserted into the old sublist moves immediately to the new sublist the first time it is accessed, no matter how soon after insertion the access occurs. If the value is greater than 0, blocks remain in the old sublist until an access occurs at least that many ms after the first access. Unit: ms.
innodb_online_alter_log_max_size	134217728	134217728	No	[134217728-2147483647]	Specifies an upper limit on the size of the temporary log files used during online DDL operations for InnoDB tables.
innodb_open_files	3000	3000	Yes	[1-8192]	This variable is relevant only if you use multiple InnoDB tablespaces. It specifies

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
					the maximum number of . ibd files that MySQL can keep open at one time. The minimum value is 10. The default value is 300.
innodb_print_all_deadlocks	OFF	OFF	No	[OFF|ON]	When this option is enabled , information about all deadlocks in InnoDB user transactions is recorded in the mysqld error log.
innodb_purge_batch_size	300	300	Yes	[1-5000]	The granularity of changes , expressed in units of redo log records, that trigger a purge operation, flushing the changed buffer pool blocks to disk.
innodb_purge_threads	1	1	Yes	[1-32]	The number of background threads devoted to the InnoDB purge operation.
innodb_read_ahead_threshold	56	56	No	[0-64]	(InnoDB Plugin only) Controls the sensitivity of linear read-ahead that InnoDB uses to prefetch pages into the buffer pool . If InnoDB reads at least innodb_read_ahead_threshold pages sequentially from an extent (64 pages), it initiates an asynchronous read for the entire following extent.
innodb_read_io_threads	4	4	Yes	[1-64]	(InnoDB Plugin only) The number of I/O threads for read operations in InnoDB. The default value is 4.

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
innodb_rollback_on_timeout	OFF	OFF	Yes	[OFF|ON]	InnoDB rolls back only the last statement on a transaction timeout by default. If --innodb_rollback_on_timeout is specified, a transaction timeout causes InnoDB to abort and roll back the entire transaction (the same behavior as in MySQL 4.1). This variable was added in MySQL 5.1.15
innodb_stats_method	nulls_equal	nulls_equal	No	[nulls_equal|nulls_unequal| nulls_ignored]	How the server treats NULL values when collecting statistics about the distribution of index values for InnoDB tables. This variable has three possible values , nulls_equal, nulls_unequal, and nulls_ignored. For nulls_equal, all NULL index values are considered equal and form a single value group that has a size equal to the number of NULL values. For nulls_unequal, NULL values are considered unequal, and each NULL forms a distinct value group of size 1. For nulls_ignored, NULL values are ignored.
innodb_stats_on_metadata	OFF	OFF	No	[ON|OFF]	When this variable is enabled (which is the default , as before the variable was created), InnoDB updates statistics during metadata statements such as SHOW TABLE STATUS or SHOW INDEX, or when

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
					accessing the INFORMATION_SCHEMA tables TABLES or STATISTICS. (These updates are similar to what happens for ANALYZE TABLE.) When disabled, InnoDB does not update statistics during these operations. Disabling this variable can improve access speed for schemas that have a large number of tables or indexes. It can also improve the stability of execution plans for queries that involve InnoDB tables.
innodb_stats_sample_pages	8	8	No	[1-4294967296]	(InnoDB Plugin only) The number of index pages to sample for index distribution statistics such as are calculated by ANALYZE TABLE. The default value is 8.
innodb_strict_mode	OFF	OFF	No	[ON|OFF]	(InnoDB Plugin only) Whether InnoDB returns errors rather than warnings for certain conditions. This is analogous to strict SQL mode. The default value is OFF. See InnoDB Strict Mode for a list of the conditions that are affected.
innodb_table_locks	ON	ON	No	[ON|OFF]	If autocommit = 0, InnoDB honors LOCK TABLES; MySQL does not return from LOCK TABLES ... WRITE until all other threads have released all their locks to

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
					the table. The default value of innodb_table_locks is 1 , which means that LOCK TABLES causes InnoDB to lock a table internally if autocommit = 0.
innodb_thread_concurrency	0	0	No	[0-128]	InnoDB tries to keep the number of operating system threads concurrently inside InnoDB less than or equal to the limit given by this variable. Once the number of threads reaches this limit, additional threads are placed into a wait state within a FIFO queue for execution.
innodb_thread_sleep_delay	10000	10000	No	[1-3600000]	How long InnoDB threads sleep before joining the InnoDB queue, in microseconds. The default value is 10,000. A value of 0 disables sleep.Unit:ms
innodb_write_io_threads	4	4	Yes	[1-64]	(InnoDB Plugin only) The number of I/O threads for write operations in InnoDB. The default value is 4.
interactive_timeout	7200	7200	No	[10-86400]	The number of seconds the server waits for activity on an interactive connection before closing it. An interactive client is defined as a client that uses the CLIENT_INTERACTIVE option to mysql_real _connect(). Unit:second.

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
key_cache_age_threshold	300	300	No	[100-4294967295]	This value controls the demotion of buffers from the hot sublist of a key cache to the warm sublist. Lower values cause demotion to happen more quickly. The minimum value is 100. The default value is 300.Unit: Second.
key_cache_block_size	1024	1024	No	[512-16384]	The size in bytes of blocks in the key cache. The default value is 1024.Unit: Byte.
key_cache_division_limit	100	100	No	[1-100]	The division point between the hot and warm sublists of the key cache buffer list. The value is the percentage of the buffer list to use for the warm sublist. Permissible values range from 1 to 100. The default value is 100.
log_queries_not_using_indexes	OFF	OFF	No	[ON|OFF]	If a query takes longer than this many seconds, the server increments the Slow_queries status variable. If the slow query log is enabled, the query is logged to the slow query log file; Unit: Second.
long_query_time	1	1	No	[0.03-10]	If a query takes longer than this many seconds, the server increments the Slow_queries status variable. If the slow query log is enabled, the query is logged to the slow query log file; Unit:Second.

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
loose_max_statement_time	0	0	No	[0-4294967295]	statement be interrupted if the executing time exceeds this value.
loose_rds_indexstat	OFF	OFF	No	[ON|OFF]	If ON, start to collect index information.
loose_rds_max_tmp_disk_space	10737418240	10737418240	No	[10737418240-1073741824 0]	RDS maximum temp disk space.
loose_rds_tablestat	OFF	OFF	No	[ON|OFF]	RDS table statistics.
loose_rds_threads_running_high_watermark	50000	50000	No	[0-50000]	Max concurrency allowed for SELECT.
loose_tokudb_buffer_pool_ratio	0	0	Yes	[0-100]	TokuDB buffer pool size ratio.
low_priority_updates	0	0	No	[0|1]	If set to 1, all INSERT, UPDATE, DELETE, and LOCK TABLE WRITE statements wait until there is no pending SELECT or LOCK TABLE READ on the affected table. This affects only storage engines that use only table-level locking (such as MyISAM, MEMORY , and MERGE). This variable previously was named sql_low_priority_updates.
max_allowed_packet	1073741824	1024M	No	[16384-1073741824]	The maximum size of one packet or any generated/ intermediate string. Unit: Byte.

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
max_connect_errors	100	100	No	[1-4294967295]	If more than this many successive connection requests from a host are interrupted without a successful connection, the server blocks that host from further connections. You can unblock blocked hosts by flushing the host cache.
max_length_for_sort_data	1024	1024	No	[0-838860]	The cutoff on the size of index values that determines which filesort algorithm to use.
max_prepared_stmt_count	16382	16382	No	[0-1048576]	This variable limits the total number of prepared statements in the server.
max_write_lock_count	102400	102400	No	[1-102400]	After this many write locks , permit some pending read lock requests to be processed in between.
myisam_sort_buffer_size	262144	262144	No	[262144-16777216]	The size of the buffer that is allocated when sorting MyISAM indexes during a REPAIR TABLE or when creating indexes with CREATE INDEX or ALTER TABLE.
net_read_timeout	30	30	No	[1-31536000]	The number of seconds to wait for more data from a connection before aborting the read.
net_retry_count	10	10	No	[1-4294967295]	If a read or write on a communication port is interrupted, retry this many times before giving up.

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
net_write_timeout	60	60	No	[1-31536000]	The number of seconds to wait for a block to be written to a connection before aborting the write.
open_files_limit	65535	65535	Yes	[4000-65535]	The number of files that the operating system permits mysqld to open. The value of this variable at runtime is the real value permitted by the system and might be different from the value you specify at server startup. The value is 0 on systems where MySQL cannot change the number of open files.
performance_schema	OFF	OFF	Yes	[ON|OFF]	Enable performance_schema or not.
query_alloc_block_size	8192	8192	No	[1024-16384]	The allocation size of memory blocks that are allocated for objects created during statement parsing and execution. Unit: Byte.
query_cache_limit	1048576	1048576	No	[1-1048576]	Do not cache results that are larger than this number of bytes. The default value is 1MB.
query_cache_size	31457280	31457280	No	[0-104857600]	The amount of memory allocated for caching query results. The default value is 0, which disables the query cache. The permissible values are multiples of 1024 ; other values are rounded down to the nearest multiple . Unit: Byte.

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
query_cache_type	0	0	Yes	[0|1|2]	Set the query cache type. Setting the GLOBAL value sets the type for all clients that connect thereafter. Individual clients can set the SESSION value to affect their own use of the query cache. Possible values are shown in the following table. • 0: Do not cache results in or retrieve results from the query cache. Note that this does not deallocate the query cache buffer. To do that, you should set query_cache_size to 0. • 1: Cache all cacheable query results except for those that begin with SELECT SQL_NO_CACHE. • 2: Cache results.
query_cache_wlock_invalidate	OFF	OFF	No	[ON|OFF]	Normally, when one client acquires a WRITE lock on a MyISAM table, other clients are not blocked from issuing statements that read from the table if the query results are present in the query cache. Setting this variable to 1 causes acquisition of a WRITE lock for a table to invalidate any queries in the query cache that refer to the table. This forces other clients that attempt

Parameter	Defaul value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
					to access the table to wait while the lock is in effect.
query_prea lloc_size	8192	8192	No	[8192-1048576]	The size of the persistent buffer used for statement parsing and execution. This buffer is not freed between statements. If you are running complex queries, a larger query_prealloc_size value might be helpful in improving performance , because it can reduce the need for the server to perform memory allocation during query execution operations. Unit: Byte.
rds_reset_ all_filter	0	0	No	[0|1]	rds_reset_all_filter=1 , means reset the rule of filter.
slow_launc h_time	2	2	No	[1-1024]	If creating a thread takes longer than this many seconds, the server increments the Slow_launc h_threads status variable.
sql_mode	\s	\s	No	(Support space and | REAL_AS_FL OAT| PIPES_AS_C ONCAT| ANSI_QUOTE S| IGNORE_SPA CE| ONLY_FULL_ GROUP_BY| NO_UNSIGNE D_SUBTRACT ION|	Modes define what SQL syntax MySQL should support and what kind of data validation checks it should perform.

Parameter	Defaul value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
				NO_DIR_IN_ CREATE| POSTGRESQL |ORACLE| MSSQL|DB2 |MAXDB| NO_KEY_OPT IONS| NO_TABLE_O PTIONS| NO_FIELD_O PTIONS| MYSQL323 |MYSQL40 |ANSI| NO_AUTO_VA LUE_ON_ZER O| NO_BACKSLA SH_ESCAPES | STRICT_TRA NS_TABLES | STRICT_ALL _TABLES| NO_ZERO_IN _DATE| NO_ZERO_DA TE| ALLOW_INVA LID_DATES| ERROR_FOR_ DIVISION_B Y_ZERO| TRADITIONA L| HIGH_NOT_P RECEDENCE |	

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
				NO_ENGINE_SUBSTITUTION|PAD_CHAR_TO_FULL_LENGTH)(,REAL_AS_FLOAT|,PIPES_AS_CONCAT|,ANSI_QUOTES|,IGNORE_SPACE|,ONLY_FULL_GROUP_BY|,NO_UNSIGNED_SUBTRACTION|,NO_DIR_IN_CREATE|,POSTGRESQL|,ORACLE|,MSSQL|,DB2|,MAXDB|,NO_KEY_OPTIONS|,NO_TABLE_OPTIONS|,NO_FIELD_OPTIONS|,MYSQL323|,MYSQL40|,ANSI|,NO_AUTO_VALUE_ON_ZERO|,NO_BACKSLASH_ESCAPES	

Parameter	Defaul value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
				|, STRICT_TRANS_TABLES|, STRICT_ALL_TABLES|, NO_ZERO_IN_DATE|, NO_ZERO_DATE|, ALLOW_INVALID_DATES|, ERROR_FOR_DIVISION_BY_ZERO|, TRADITIONAL|, HIGH_NOT_PRECEDENCE|, NO_ENGINE_SUBSTITUTION|, PAD_CHAR_TO_FULL_LENGTH)*	
table_definition_cache	512	512	No	[400-80480]	The number of table definitions (from .frm files) that can be stored in the definition cache. If you use a large number of tables, you can create a large table definition cache to speed up opening of tables. The table definition cache takes less space and does not use file descriptors, unlike the normal table cache. The

Parameter	Default value	Running parameter value	Require restart or not	Modifiable parameter value	Parameter description
					minimum and default values are both 400.
table_open_cache	2000	2000	No	[1-524288]	The stack size of each thread.
thread_stack	262144	262144	Yes	[131072-18446744073709551615]	The stack size of each thread.
tmp_table_size	2097152	2097152	No	[262144-67108864]	The maximum size of internal in-memory temporary tables.
transaction_isolation	READ-COMMITTED	READ-COMMITTED	Yes	[READ-COMMITTED | REPEATABLE-READ]	The default transaction isolation level.
wait_timeout	86400	86400	No	[60-259200]	The number of seconds the server waits for activity on a noninteractive connection before closing it.

16.10.8.2 Best practice for MySQL instance parameter optimization

16.10.8.2.1 Preface

You can optimize MySQL parameters for ApsaraDB RDS instances. This topic describes the best practices for modifiable and non-modifiable MySQL parameters. It also describes how to optimize modifiable parameters to improve instance performance.

16.10.8.2.2 Non-modifiable MySQL instance parameters

The maximum number of connections allowed and memory size vary with the RDS instance type. Therefore, parameters related to the instance type, such as connections and memory, are restricted and cannot be modified. If you encounter connection or memory bottlenecks, they can be resolved as follows:

- Memory bottleneck: An out of memory (OOM) error will appear in your instance, and will switch the primary and secondary instances.
- Connection bottleneck: If applications cannot establish new connections to the database, you need to optimize the applications or slow SQL statements, or upgrade the instance type.

To guarantee data security of the primary and secondary instances, disable the modification of parameters related to data security, such as innodb_flush_log_at_trx_commit, sync_binlog, gtid_mode, semi_sync, and binlog_format.

16.10.8.2.3 Modifiable MySQL instance parameters

Other than the non-modifiable parameters described in *Non-modifiable MySQL instance parameters*, most of the parameters of your ApsaraDB for RDS instance have been optimized by the DBA and source code teams. This helps you run your database without the need to adjust any parameters. For more information about the modifiable MySQL instance parameters, see *Modifiable MySQL instance parameters*. These parameters are applicable in most scenarios. You need to adjust some parameters in some special cases. For example:

- If you use the TokuDB storage engine, you need to use the tokudb_buffer_pool_ratio parameter to adjust the percentage of the memory available for the engine.
- If your applications require a relatively long lock time-out time, you need to adjust the innodb_lock_wait_timeout parameter.

16.10.8.2.4 How to configure parameters

This topic describes how to configure important parameters on the RDS console. If these parameters are incorrectly configured, your instances may encounter performance problems or applications may report errors.

open_files_limit

Function: Controls the number of file handles that can be simultaneously enabled by each MySQL instance.

Cause: More file handles (allocated to each instance) are consumed when more database tables are opened. ApsaraDB for RDS sets open_files_limit to 8192 when initializing instances. When the number of opened tables exceeds this value, errors are returned for all database requests.

Note:

File descriptors are consumed when MyISAM engine tables are accessed. The InnoDB storage engine manages opened tables based on the table_open_cache parameter.

Symptom: If open_files_limit is set to an excessively small value, applications may report the following error:

```
[ERROR] /mysqld: Can't open file: './mysql/user.frm' (errno: 24 -Too
many open files);
```

Suggestion: Increase the value of open_files_limit. Currently, ApsaraDB for RDS supports a value of up to 65535 for this parameter. We also recommend that you replace the MyISAM storage engine with the InnoDB storage engine.

back_log

Function: MySQL creates a thread for every connection request that it processes. If front-end applications initiate too many transient connection requests to the database when a new thread is created, MySQL prevents new connection requests from entering the request queue based on the back_log parameter. MySQL denies new connection requests when the number of connection requests in the waiting state exceeds the value of back_log. If you want MySQL to process a large number of transient connection requests, increase the value of back_log.

Symptom: If back_log is set to an excessively small value, applications may report the following error:

```
SQLSTATE[HY000] [2002] Connection timed out;
```

Suggestion: Increase the value of back_log. The initial value of back_log used to be 50, but has now been increased to 3000.

Note:
You must restart your instances after you change the parameter value.

innodb_autoinc_lock_mode

Function: In MySQL versions MySQL 5.1.22 and later, the innodb_autoinc_lock_mode parameter is introduced to solve the auto-increment lock problem. This parameter controls the auto-increment lock behavior. This parameter can be set to 0, 1, and 2. The default value is 1 in ApsaraDB for RDS. This indicates that InnoDB uses the lightweight mutex lock to obtain auto-increment locks in place of table-level locks. However, when auto-increment table locks are used for loading data through the INSERT … SELECT statement or REPLACE … SELECT statement, applications may encounter deadlocks during the concurrent data import process.

Symptom: Applications may encounter deadlocks during the concurrent data import process when auto-increment table locks are used to load data through the INSERT … SELECT statement or REPLACE … SELECT statement. The following error is reported:

```
RECORD LOCKS space id xx page no xx n bits xx index PRIMARY of table
 xx.xx trx id xxx lock_mode X insert intention waiting. TABLE LOCK
table xxx.xxx trx id xxxx lock mode AUTO-INC waiting;
```

Suggestion: We recommend that you change the value of innodb_autoinc_lock_mode to 2 to enable the use of the lightweight mutex lock (only in row mode) for all types of insert operations. This avoids auto_inc deadlocks and greatly improves the performance of the INSERT … SELECT statement.

Note:
If you set the parameter value to 2, you must also set the format of binlog to row.

query_cache_size

Function: Controls the memory size of the MySQL query cache. If the query cache is enabled, MySQL locks the query cache when it executes a query and then determines whether the query cache contains the queried data. If so, MySQL returns results directly; if not, MySQL proceeds to engine query and other operations. The INSERT, UPDATE, and DELETE statements can invalidate the query cache and any changes made to schemas and indexes. It costs a lot to maintain the invalid query cache, which brings great pressure to MySQL. The query cache helps improve instance performance when the database is not frequently updated. However, when data is frequently written to several tables of the database, the query cache lock results in frequent lock conflicts. The write and read operations of a specific table has to wait for the query cache lock to unlock, which reduces the query efficiency of the SELECT statement.

Symptom: The database goes through several different statuses, which are: checking query cache, waiting for query cache lock, and storing results in query cache.

Suggestion: ApsaraDB for RDS disables the query cache by default. If your instances enable the query cache, you can disable it when the preceding problem occurs. However, you can enable the query cache to solve database performance problems in some cases.

net_write_timeout

Function: Controls the time-out time when a block is sent to a client.

Symptom: If the parameter is set to an excessively small value, the client may report the following error:

```
the last packet successfully received from the server was milliseconds
 ago , the last packet sent successfully to the server was milliseconds
 ago.
```

Suggestion: The default value is 60 seconds in ApsaraDB for RDS. A small value of net_write_timeout may result in frequent disconnections when the network condition is poor or it takes a long time for the client to process each block. In this case, we recommend that you increase the value of this parameter.

tmp_table_size

Function: Determines the maximum value of the internal temporary memory table, which is assigned to each thread. (The minimum values of tmp_table_size and max_heap_table_size decide the actual value) If the temporary memory table exceeds the parameter value, MySQL automatically converts it to a disk-based MyISAM table. Avoid the use of temporary tables when you optimize query statements. If you need to use a temporary table, make sure that the temporary table is created in the memory.

Symptom: If a complex SQL statement contains GROUP BY and DISTINCT clauses, and cannot be optimized through indexes, temporary tables are used. In this case, SQL execution takes a longer time.

Suggestion: If the application involves many GROUP BY and DISTINCT clauses and the database has enough memory, you can increase the values of tmp_table_size and max_heap_table_size to improve query performance.

16.10.8.2.5 New MySQL parameters

oose_rds_max_tmp_disk_space

Function: Controls the temporary file size available for MySQL. The default value is 10 GB in ApsaraDB for RDS.

Symptom: If the temporary file size exceeds the limit indicated by this parameter, applications may report the following error:

```
The table '/home/mysql/dataxxx/tmp/#sql_2db3_1' is full.
```

Suggestion: Check whether the SQL statements that cause additional temporary files can be optimized by indexing or other means. If your instance has enough space, you can increase the value of this parameter to guarantee normal execution of SQL statements.

Note:

You must restart your instances after you change the parameter value.

loose_tokudb_buffer_pool_ratio

Function: Controls the buffer size available for the TokuDB storage engine. For example, if you set innodb_buffer_pool_size to 1000 MB and tokudb_buffer_pool_ratio to 50 (which indicates 50% of the buffer size), then the TokuDB storage engine can use up to 500 MB of the buffer space.

Suggestion: The default value is 0 in ApsaraDB for RDS. If you use the TokuDB storage engine in your RDS instance, we recommend that you increase the value of the parameter to improve the access performance of the TokuDB engine table.

Note:

You must restart your instances after you change the parameter value.

loose_max_statement_time

Function: Controls the maximum query time in MySQL.

Symptom: By default, the query time is not limited. If the query time exceeds the limit indicated by this parameter, the query fails as follows:

```
ERROR 3006 (HY000): Query execution was interrupted, max_statem
ent_time exceeded
```

Suggestion: Modify this parameter if you want to control the SQL execution time (in milliseconds) of your database.

loose_rds_threads_running_high_watermark

Function: Controls the number of concurrent MySQL queries. For example, if you set rds_threads_running_high_watermark to 100, then 100 MySQL queries can be

initiated concurrently. Additional queries are denied. This parameter is used with rds_threads_running_ctl_mode (default value: select).

Suggestion: This parameter is often used to handle peak-hour or highly concurrent requests, which provides effective database protection.

16.11 Account management

16.11.1 Create an account

This topic describes the features that are available for accounts in classic and master modes, as well as how to create accounts in different modes.

You need to create an account in the ApsaraDB instance before you can use the database . ApsaraDB for RDS supports classic mode and master mode. Classic mode is an earlier management mode. You cannot use SQL to manage databases and accounts in classic mode . Master mode is a newer management mode. You can use SQL to manage databases and accounts in master mode. You also have more permissions in this mode. We recommend that you use master mode if you have personalized and sophisticated permission management requirements.

Account modes

In classic mode, all accounts are created through the RDS console or API, instead of through SQL . All accounts are created equally. The RDS console is used to create and manage all accounts and databases.

In master mode, the first account you create is the initial account. You must use the RDS console or API to create and manage it. Log on to a database with your initial account. You can then create and manage other ordinary accounts through SQL commands or Alibaba Cloud's DMS. However, you cannot use your initial account to change the passwords of other ordinary accounts . To change the password of an ordinary account, you must delete the ordinary account and create a new one. In the following example, the initial account is used as root to log on to the database. Then, an ordinary account "jeffrey" is created.

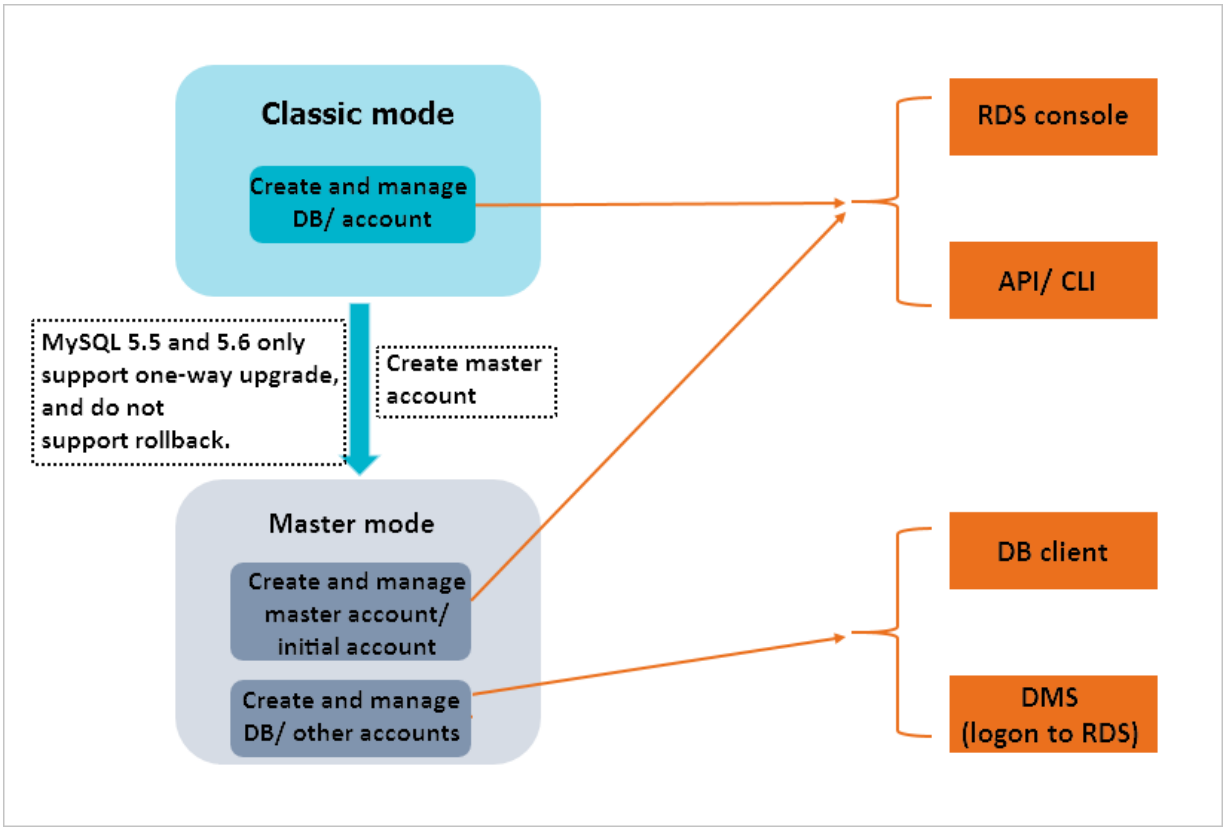
```
mysql -hxxxxxxxxx.mysql.rds.aliyuncs.com -uroot -pxxxxxx -e "
    CREATE USER 'jeffrey'@'%' IDENTIFIED BY 'password';
    CREATE DATABASE DB001;
```

"

In master mode, the database management page is not available on the RDS console. APIs such as CreateDatabase cannot be used to manage databases. You must use SQL commands or DMS to create and manage databases.

*Figure 16-23: Difference between ordinary and master accounts*shows how to create and manage databases and accounts in classic and master modes.

Figure 16-23: Difference between ordinary and master accounts



How to create an account

Note:

- When you assign database account permissions, follow the minimum permission principle and use service roles to create accounts and assign proper read-only and read/write permissions. When necessary, you may split database accounts and databases into smaller units so that each database account can only access data for its own services. If the account does not need to write data to a database, assign read-only permissions.
- Set strong passwords for the accounts and change the passwords periodically.

Procedure

- For more information about how to create an ordinary account for MySQL, see *Create a database and an account*.
- For more information about how to create a master account for MySQL, see *Create a master account*.
- For more information about how to create an account for SQL Server, see *Create a database and an account*.
- For more information about how to create an account for PostgreSQL and PPAS, see *Create a database and an account*.

16.11.2 Reset your password

When using ApsaraDB for RDS, you can reset the password on the RDS console if you forget the password of your database account.

Context

Note:
For data security considerations, we recommend that you change your password periodically.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Database Management** > **Account List** to go to the **Account List** page.
4. In the **Action** column of the instance, select > **Reset Password**.
5. On the **Reset Password** page, enter a new password and click **Confirm**.

Note:
The password can be 6 to 32 characters in length and can contain letters, numbers, and underscores (_). We recommend that you do not use a previous password.

16.11.3 Modify account permissions

When using ApsaraDB for RDS, you can modify account permissions of your instance at any time.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Database Management** > **Account List** to go to the **Account List** page.
4. In the **Action** column of the instance, select > **Modify Permissions**.
5. On the displayed page, modify account permissions and click **Confirm**.

16.11.4 Delete an account

If your RDS instance uses ordinary accounts, you can use the console to delete unnecessary accounts. This topic describes how to delete an ordinary account.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Database Management** > **Account List** to go to the **Account List** page.
4. In the **Action** column of the account, choose > **Delete**.
5. In the dialog box that appears, click **Confirm**.

16.11.5 Modify descriptions

You can add descriptions to assist in account management when you create an account. You can also modify descriptions after the account is created. This topic describes how to modify the account description.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Database Management** > **Account List** to go to the **Account List** page.
4. In the **Action** column of the account, choose > **Edit Description**.
5. Modify the information in **`Description`**.
6. Click **Confirm**.

16.12 Database management

16.12.1 Create a database

You can use the RDS console to create a database. Each database name within an instance must be unique, but duplicate names are allowed across instances. Before you migrate your local database, you must create a database that is the same as the local database in your RDS instance.

Context

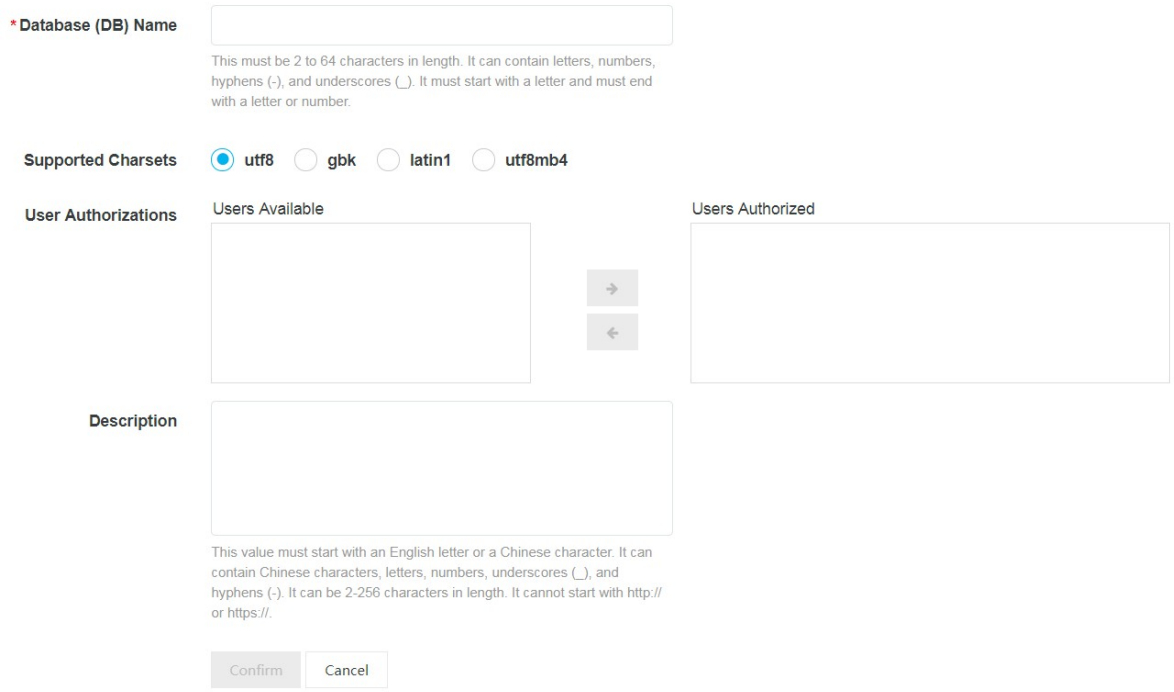
Note:

- This task only applies to MySQL and SQL Server instances that use ordinary accounts.
- To migrate the local database to ApsaraDB for RDS, create a database that is the same as the local database in your RDS instance.
- When you assign database account permissions, follow the minimum permission principle and use service roles to create accounts and assign proper read-only and read/write permissions . When necessary, you may split database accounts and databases into smaller units, so that each account can only access data for its own services. If the account does not need to write data to a database, assign read-only permissions.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Database Management** > **Database List** to go to the **Database List** page.
4. Click **Add Database**.
5. Enter information about the database to be created, as shown in *Figure 16-24: Create a database*.

Figure 16-24: Create a database



Parameter description:

Table 16-14: Database creation parameters

Parameter name	Description
Database (DB) name	The name can be 2 to 64 characters in length and can contain lowercase letters, numbers, and underscores (_). It must start with a letter and end with a letter or number.
Supported Charsets	Set a character set for the database.
User Authorizations	Select an account authorized by the database. This parameter can be empty if no account is created. For more information about account authorization, see *Modify account permissions*.
Description	You can enter related information of the database to assist in future database management. You can enter a maximum of 256 English characters.

6. Click **Confirm**.

16.12.2 Delete a database

This topic describes how to use the RDS console to delete a database from an instance.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Database Management** > **Database List** to go to the **Database List** page.
4. In the **Action** column of the database, select > **Delete**.
5. In the confirmation box that appears, click **Confirm**.

16.13 Set an access mode

ApsaraDB for RDS supports two access modes: **`Standard Mode`** and **`Safe Mode`**. This topic describes the differences between the two access modes and their configuration methods.

Prerequisites

Set the network type to **Classic Network**.

Context

The **`Standard Mode`** and the **`Safe Mode`** have the following differences:

- Standard mode: ApsaraDB for RDS uses Server Load Balancer to eliminate the impact from HA switching of the database engine on the application layer. This shortens the response time, but will slightly increase the probability of transient disconnections and disable SQL interception . This mode supports only one connection address. If the instance has both an internal address and a public address, release one of them before switching to the standard mode.
- Safe mode: This mode prevents 90% of transient disconnections and supports SQL interception (SQL injection attacks are prevented based on SQL semantic analysis), but increases the response time by 20% or more. This mode supports connections to an internal address and a public address at the same time.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. Click **Change Connection Mode** in the upper-right corner of the page.

Note:

When the access mode change is in progress, the **Status** of the instance changes to **Creating**. When the **Status** changes to **Running**, the access mode is successfully changed.

16.14 Security management

16.14.1 Configure a whitelist

To guarantee database security and reliability, you need to modify its whitelist before you enable an instance. You need to add the IP addresses or IP address segments used for database access to the whitelist of the RDS instance. This topic describes how to configure the whitelist of an RDS instance.

Context

Note:

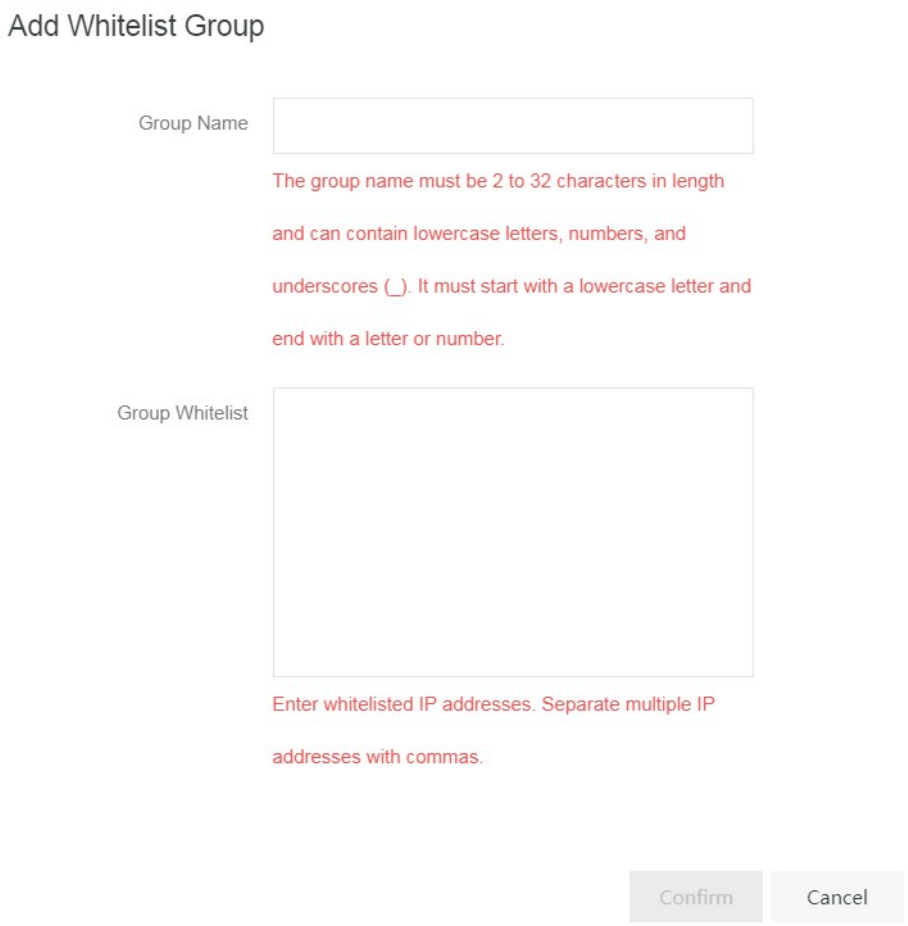
- The system creates a default whitelist group for each instance. This default whitelist group can only be modified or cleared, but cannot be deleted.
- For each newly created RDS instance, the local loopback IP address 0.0.0.0/0 is added to the **default** whitelist group by default. This means that any IP address is allowed to access the RDS instance. This configuration greatly reduces the security of the database. Delete 0.0.0.0/0 first.
- If the whitelist is configured to 127.0.0.1, all IP addresses or IP segments are prohibited to access the RDS instance. Therefore, you must delete 127.0.0.1 from the whitelist before you add other IP addresses or IP segments.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation bar, select **Security Control** > **Whitelist Settings**.
4. Click the icon of the **default** whitelist group to delete the default whitelist 0.0.0.0/0.
5. Add the IP addresses or IP segments allowed to access the RDS instance to the **default** whitelist group.
6. Click **Confirm**.
7. Click **Add Whitelist Group**.

8. In the **Add Whitelist Group** dialog box, enter the group name and the IP addresses or IP segments allowed to access the RDS instance, and click **Confirm**.

Figure 16-25: Add a whitelist group



Parameter description:

Table 16-15: Add a whitelist group

Parameter name	Description
Group Name	The name can be 2 to 32 characters in length and can contain lowercase letters, numbers, and underscores (_). It must start with a lowercase letter and end with a letter or number. You cannot modify the name of a created whitelist group.
Group Whitelist	Enter the IP addresses or IP address segments allowed to access the RDS instance.

Parameter name	Description
	• If you enter an IP address segment, for example, 10.10.10.0/24, any IP address in the format of 10.10.10.X can access the RDS instance. • If you enter multiple IP addresses, separate them with commas (no space before or after each comma), for example, 192.168.0.1,172.16.213.9.

What's next

Correct use of the whitelist can improve access security for your RDS instance. We recommend that you maintain the whitelist periodically. To modify the whitelist group, click  . To clear the default whitelist group or delete a custom whitelist group, click  .

16.14.2 Audit logs

You can query the SQL logs, operation logs, and error logs of an instance on the RDS console to locate and analyze faults. This topic describes how to query audit logs.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Security Control** > **Audit Log** to go to the **Audit Log** page.
4. Click the **SQL Log**, **Error Log**, or **Operation Log** tab.
5. Select a time range and click **Apply**.

What's next

On the **Operation Log** page, you can click **Export** to export operation logs for offline analysis.

16.14.3 Configure SSL

Context

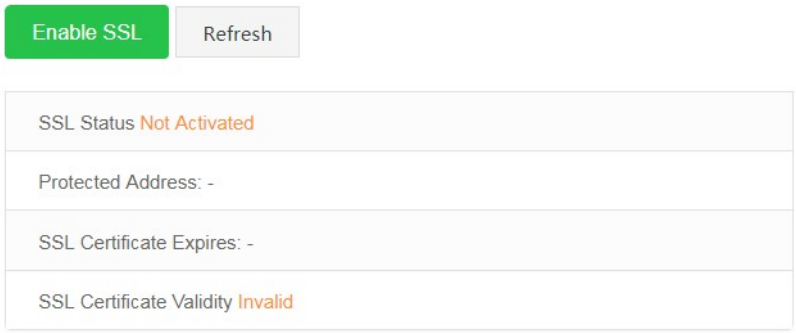
To increase link security, you can enable Secure Sockets Layer (SSL) encryption and install SSL certificates on the necessary application services. SSL is used on the transport layer to encrypt network connections.It increases the security and integrity of communication data, but it also increases the network connection response time.

Note:

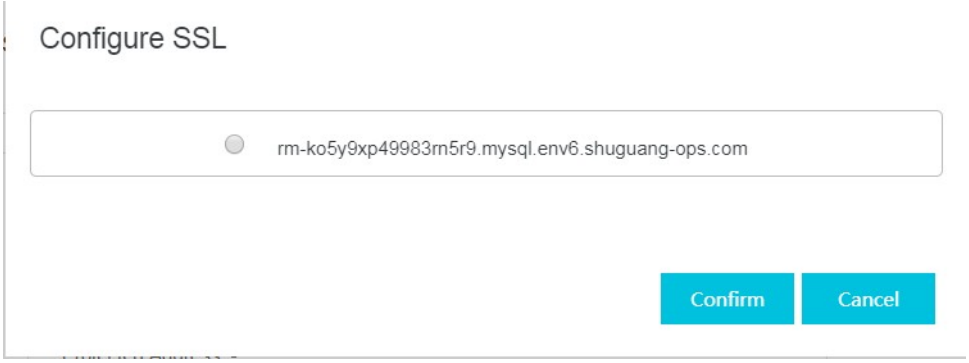
- Due to the inherent drawbacks of SSL encryption, it significantly increases your CPU usage . We recommend that you only enable SSL encryption for public connections that require encryption. Internal network connections are relatively secure, and generally do not require encryption.
- Exercise caution when enabling SSL encryption because it cannot be disabled once it is enabled.

Procedure

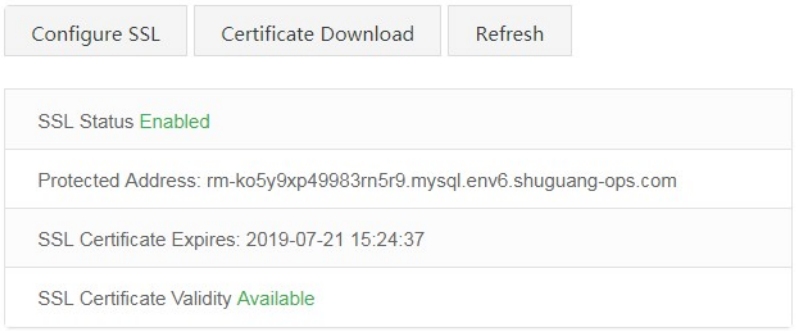
1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Security Control** > **Configure SSL Encryption** to go to the **Configure SSL Encryption** page. SSL details about the instance are displayed.
4. Click **Enable SSL**, as shown in the following figure.



5. In the **Configure SSL** dialog box that appears, select an instance ID.



6. Click **Confirm** to enable SSL encryption, as shown in the following figure.



16.14.4 Download SSL CA certificates

Context

To increase link security, you can enable SSL encryption and install SSL CA certificates on the necessary application services. SSL is used on the transport layer to encrypt network connection s.It increases the security and integrity of communication data, but it also increases the network connection response time.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Security Control** > **Configure SSL Encryption** to go to the **Configure SSL Encryption** page. The page displays the SSL details of the instance.
4. Click **Certificate Download**.
5. In the displayed dialog box, click **Confirm** to download SSL CA certificates.

   The downloaded package includes three files:

   - P7b file: used to import CA certificates to the Windows system.
   - PEM file: used to import CA certificates to other operating systems or applications.
   - JKS file: stores truststore certificates in Java. The password is apsaradb. It is used to import the CA certificate chain to Java programs.

   Note:

   When the JKS file is used in Java, you need to modify the default JDK security configuration in JDK7 and JDK8. Open the `jre/lib/security/java.security` file on the computer where the database that needs SSL access resides, and modify two configurations as follows:

   ```
   jdk.tls.disabledAlgorithms=SSLv3, RC4, DH keySize < 224
   ```

```
jdk.certpath.disabledAlgorithms=MD2, RSA keySize < 1024
```

If no JDK security configuration is modified, the following error is reported. Other similar errors are also caused by Java security configurations.

```
javax.net.ssl.SSLHandshakeException: DHPublicKey does not comply to
 algorithm constraints
```

16.15 Performance management

16.15.1 Slow SQL statistics

You can use the RDS console to query slow SQL statistics to locate and analyze faults.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation bar, select **Performance Optimization** > **Slow SQL Statistics** .
4. Select a time range and click **Search**.

Note:

The system does not list slow logs from the past two hours. These logs are contained in the slow_log table of the MySQL database.

16.15.2 Missing index

Based on the SQL statement execution status and performance of your RDS instance, the system prompts you about database tables with missing indexes, and provides you with a statement to add indexes. This topic describes how to query missing indexes.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Performance Optimization** > **Missing Indexes**. This page allows you to query all missing indexes.

16.16 Backup and recovery

16.16.1 RDS data backup

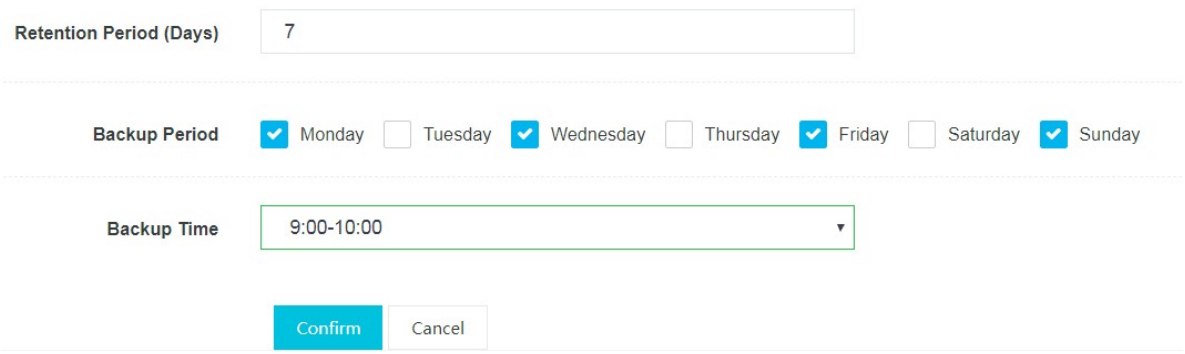
16.16.1.1 Automatic backup

Automatic backup supports full physical backups. After you configure a backup policy, ApsaraDB for RDS automatically backs up databases in accordance with the policy. This topic describes how to configure a policy for automatic backup.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Backup and Recovery** > **Backup List**.
4. Click the **Backup Settings** tab and click **Set**.
5. On the **Backup Settings** page, set backup specifications, as shown in *Figure 16-26: Configure a backup policy*.

   Figure 16-26: Configure a backup policy

   Parameter description:

   - **`Retention Period (Days)`**: Set the number of days for which backup files are retained. The default value is 7 days. The value can be 1 to 30 days.
   - **`Backup Period`**: Set it to one or multiple days in a week.
   - **`Backup Time`**: Set it to any time range, in hours.
6. Click **Confirm**.

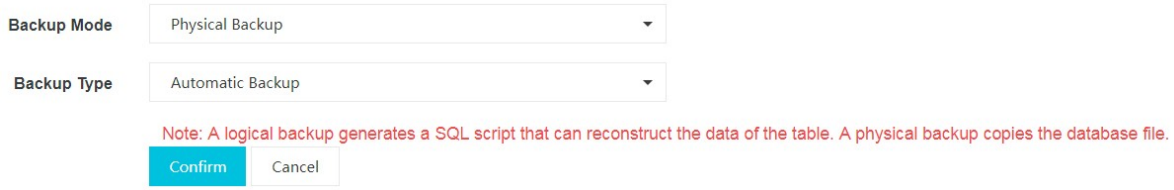
16.16.1.2 Manual backup

Manual backup supports full physical backups and full logical backups. This topic describes how to manually back up RDS data.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. Click **Back up Instance** on the right side of the page.
4. On the **Back up Instance** page, select a backup mode and a backup type, as shown in *Figure 16-27: Configure manual backup*.

   Figure 16-27: Configure manual backup

   Parameter description:

   - **Backup Mode**: Select physical backup or logical backup.
   - **Backup Type**: Select full backup or automatic backup.
5. Click **Confirm**.

16.16.2 RDS data recovery

16.16.2.1 Recover data directly to the primary instance

During direct data recovery, the backup data overwrites the data of the primary instance, and the data generated after creation of the backup data is lost. We recommend that you create a temporary instance for data recovery and migration to guarantee higher security. This topic describes how to use backup data to overwrite the data of the primary instance.

Context

Note:

If a read-only instance exists, the backup data cannot directly overwrite the original data of the primary instance. In this case, create a temporary instance for data recovery. For more information, see *Recover data to the primary instance through a temporary instance*.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Backup and Recovery** > **Backup List**.
4. Click the **Backups** tab.
5. Select a time range and click **Search**.
6. Find the backup set to be recovered. In the **Action** column, choose > **Restore and Replace**.

16.16.3 Binary log (binlog)

This topic describes how to query and download the binlogs of an RDS instance.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Backup and Recovery** > **Binlog List**.
4. Select a time range and click **Search** to query the binlogs generated within the selected time range.
5. To download a binlog, choose > **Download** in the **Action** column.

16.16.4 Create a clone instance

Prerequisites

When you create a clone instance, the primary instance must meet the following conditions:

- It must be in the running state and unlocked.
- No ongoing migration tasks.
- Data backup and log backup are enabled.
- If you want to create a clone instance by restoring a backup set, the primary instance must have at least one completed backup set.

Context

You can create a clone instance that has the same data and configurations as the primary instance. This function allows you to restore data of the primary instance or create multiple instances that are the same as the primary instance.

You can specify a backup set or any point in time within the backup retention period to create a clone instance. A clone instance only copies the data of the primary instance, but not the content of read-only or disaster recovery instances under the primary instance. The copied data includes database data, account information, and instance configurations (such as whitelist configurations, backup configurations, parameter configurations, and alarm threshold configurations).

The database type of a clone instance must be the same as that of the primary instance. Other settings (such as the instance series, zone, network type, instance type, and storage space) can be different. If you want to create a clone instance to restore data of the primary instance, we recommend that you select the same instance type and storage space as the primary instance. Otherwise, data restoration takes a longer time to complete.

The clone and primary instances have the same account mode. You can change the account password of the clone instance. For example, if you create a clone instance for a primary instance that uses a primary account, the clone instance also uses a primary account.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **Backup and Recovery** > **Backup List**.
4. On the **Backups** page, click **Create Duplicate Instance**.
5. Select the **`Specifications`**, **`Restore Mode`**, and **`Restore Ponit Time`** for the clone instance, and click **Create**.

16.17 Check the job execution status

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, select **Task List** to go to the **Instance Task List** page.
4. Set **Start Date**, **End Date**, or **Status** of the job. Click **Search** to check the job execution status of an instance.

16.18 Monitor system resources

The RDS console provides a variety of performance metrics for you to monitor the status of your instance.

Procedure

1. *Log on to the RDS console*.
2. Click the ID of the instance to go to the **Basic Information** page.
3. In the left-side navigation pane, choose **System Resource Monitoring**.
4. Select the monitored data you wish to view, such as system resources, performance, and InnoDB and MyISAM engines. The data is displayed as shown in *Table 16-16: Metrics*.

Table 16-16: Metrics

Page	Metric	Description	Monitoring frequency	Monitoring cycle
System Resources	Storage Space	Disk space usage of the instance, such as overall usage of the disk space, data space, log space, temporary file space, and system file space. Unit: MB	60s 300s	30 days
	IOPS	I/O requests of the instance per second. Unit: Times/ second	60s 300s	30 days
	CPU Usage	CPU usage of the instance (excluding the CPU resources used by the operating system).	60s 300s	30 days

Page	Metric	Description	Monitoring frequency	Monitoring cycle
	Network traffic	Inbound and outbound traffic of the instance per second. Unit: KB	60s 300s	30 days
Database Performance	QPS/TPS	Number of SQL statements executed and transactions processed per second.	60s 300s	30 days
	Temporary Tables	Number of temporary tables automatically created on the hard disk when the database runs SQL statements.	60s 300s	30 days
	COMDML	Number of times the database runs SQL statements per second. The statements include INSERT, DELETE, INSERT_SELECT, REPLACE, REPLACE_SELECT, SELECT, and UPDATE.	60s 300s	30 days
	ROWDML	Number of operations performed on InnoDB per second, such as the number of physical writes	60s 300s	30 days

Page	Metric	Description	Monitoring frequency	Monitoring cycle
		to the log file, and the number of InnoDB table rows that are read, updated, deleted, and inserted.		
InnoDB Engine	InnoDB Buffer Pool	Read hit rate, usage, and dirty data block percentage of the InnoDB buffer pool.	60s 300s	30 days
	InnoDB Read/ Write Volume	Volume of InnoDB data that is read and written per second. Unit: KB	60s 300s	30 days
	InnoDB Reads and Writes	Number of InnoDB reads and writes per second.	60s 300s	30 days
	InnoDB Log	Number of InnoDB physical writes to the log file, log write requests, and FSYNC writes to the log file.	60s 300s	30 days
MyISAM Engine	MyISAM Key Buffer	Read hit rate, write hit rate, and usage of the MyISAM key buffer per second.	60s 300s	30 days
	MyISAM Reads and Writes	Number of MyISAM reads	60s 300s	30 days

Page	Metric	Description	Monitoring frequency	Monitoring cycle
		and writes from/ to the buffer pool and hard disk per second.		

16.19 Local database migration to RDS

16.19.1 Compress data

Context

ApsaraDB for MySQL 5.6 allows you to compress data with the TokuDB storage engine. A large number of tests showed that after data tables are switched from the InnoDB storage engine to the TokuDB storage engine, the amount of data can be reduced by 80% to 90%. For example, 2 TB of data can be compressed to 400 GB or lower. The TokuDB storage engine also supports transaction and online DDL operations, which are compatible with applications running on a MyISAM or an InnoDB storage engine.

Restrictions on TokuDB:

- The TokuDB storage engine does not support foreign keys.
- The TokuDB storage engine is not applicable to scenarios where frequent reading of large amounts of data is required.

Procedure

1. Run the following command to check the MySQL version.

```
SELECT version();
```

2. Set the **loose_tokudb_buffer_pool_ratio** to indicate the proportion that TokuDB occupies in the shared cache of TokuDB and InnoDB.

```
select sum(data_length) into @all_size from information_schema.
tables where engine='innodb';
select sum(data_length) into @change_size from information_schema
.tables where engine='innodb' and concat(table_schema, '.',
table_name) in ('XX.XXXX', 'XX.XXXX', 'XX.XXXX');
select round(@change_size/@all_size*100);
```

Note:

In the preceding command, **XX.XXXX** refers to the database or table that is to be transferred to the TokuDB storage engine.

3. Restart the instance. For more information, see *Restart an instance*.
4. Run the following command to modify the storage engine. You can also log on to DMS to modify the data table storage engine. For more information, see *Cite LeftDMS Product DocumentationCite Right*.

```
ALTER TABLE XX.XXXX ENGINE=TokuDB
```

Note:
In the preceding command, **XX.XXXX** refers to the database or table that is to be transferred to the TokuDB storage engine.

16.19.2 MySQL data migration

16.19.2.1 Use mysqldump to migrate MySQL data

The mysqldump tool is easy to use but has a long service downtime. Use mysqldump if the data volume is small or if a long service downtime is allowed.

Prerequisites

An ECS instance must be activated.

Context

ApsaraDB for RDS is fully compatible with MySQL. The procedure for migrating the original database to an RDS instance is similar to the procedure for migrating data from one MySQL server to another.

Before data migration, create a migration account in the local database, and grant the read/write permissions of the database to the migration account.

Procedure

1. Run the following command to create a migration account in the local database.

   ```
   CREATE USER 'username'@'host' IDENTIFIED BY 'password';
   ```

   Parameter description:

   - username: specifies the account to be created.

- host: specifies the host from which you log on to the database using the account. As a local user, you can use **localhost** to log on to the database. To log on from any host, you can use the wildcard %.
- password: specifies the logon password for this account.

In the following example, an account named **William** with the password **Changme123** is allowed to log on to the local database from any host.

```
CREATE USER 'William'@'%' IDENTIFIED BY 'Changme123';
```

2. Run the following command to grant permissions to the migration account in the local database.

```
GRANT SELECT ON databasename.tablename TO 'username'@'host' WITH
GRANT OPTION; GRANT REPLICATION SLAVE ON databasename.tablename TO '
username'@'host' WITH GRANT OPTION;
```

```
GRANT REPLICATION SLAVE ON databasename.tablename TO 'username'@'host
' WITH GRANT OPTION;
```

Parameter description:

- privileges: specifies the operation permissions of the account, such as SELECT, INSERT, and UPDATE. To grant all permissions to the account, use **ALL**.
- databasename: specifies the name of the database. To grant all database permissions to the account, use the wildcard *.
- tablename: specifies the name of the table. To grant all table permissions to the account, use the wildcard *.
- username: specifies the name of the account to be granted permissions.
- host: specifies the host authorized for the account to log on to the database. As a local user, you can use **localhost** to log on to the database. To log on from any host, you can use the wildcard %.
- WITH GRANT OPTION: specifies an optional parameter that enables the account to use the GRANT command.

In the following example, an account named **William** is granted with all database and table permissions, and allowed to log on to the local database from any host.

```
GRANT ALL ON *.* TO 'William'@'%';
```

3. Use the data export tool of mysqldump to export data in the database as data files.

> **Note:**
> Do not update data during data export. This step exports data only. It does not export stored procedures, triggers, and functions.

```
mysqldump -h localIp -u userName -p --opt --default-character-set=utf8
 --hex-blob dbName --skip-triggers > /tmp/dbName.sql
```

Parameter description:

- localIp: specifies the IP address of the local database server.
- userName: specifies the migration account of the local database.
- dbName: specifies the name of the database to be migrated.
- /tmp/dbName.sql: specifies the name of the backup file.

4. Use mysqldump to export stored procedures, triggers, and functions.

> **Note:**
> Skip this step if no stored procedures, triggers, and functions are used in the database.
> When exporting stored procedures, triggers, and functions, you must remove "definer" to be compatible with RDS.

```
mysqldump -h localIp -u userName -p --opt --default-character-set=utf8
 --hex-blob dbName -R | sed -e 's/DEFINER[ ]*=[ ]*[^*]*\*/\*/' > /tmp/
triggerProcedure.sql
```

Parameter description:

- localIp: specifies the IP address of the local database server.
- userName: specifies the migration account of the local database.
- dbName: specifies the name of the database to be migrated.
- /tmp/triggerProcedure.sql: specifies the name of the backup file.

5. Upload the data files and stored procedure files to ECS.

   The example in this topic illustrates how to upload files to the following path.

   `/tmp/dbName.sql`

   `/tmp/triggerProcedure.sql`

6. Log on to ECS and import the data files and stored procedure files to the target RDS instance.

```
mysql -h intranet4example.mysql.rds.aliyuncs.com –u userName -p dbName
 &lt; /tmp/dbName.sql
```

```
mysql -h intranet4example.mysql.rds.aliyuncs.com -u userName -p dbName
 &lt; /tmp/triggerProcedure.sql
```

Parameter description:

- intranet4example.mysql.rds.aliyuncs.com: specifies the RDS instance connection address. An internal address is used as an example.
- userName: specifies the migration account of the RDS database.
- dbName: specifies the name of the database to be imported.
- /tmp/dbName.sql: specifies the name of the data file to be imported.
- /tmp/triggerProcedure.sql: specifies the name of the stored procedure file to be imported.

16.20 Typical application

16.20.1 Store multi-structure data

Context

Object Storage Service (OSS) is an Alibaba Cloud storage service that features massive capacity , robust security, low cost, and high reliability. ApsaraDB for RDS can work with OSS to form multiple types of data storage solutions.
For example, ApsaraDB for RDS and OSS can be used in a forum. Resources such as the images of registered users and those posted on the forum can be stored on OSS, which reduces the storage pressure on ApsaraDB for RDS.

An example of combined use of ApsaraDB for RDS and OSS is as follows.

Procedure

1. Run the following command to initialize OssAPI:

```
from oss.oss_api import * endpoint=" oss-cn-hangzhou.aliyuncs.com"
accessKeyId, accessKeySecret="your id","your secret" oss = OssAPI(
endpoint, accessKeyId, accessKeySecret)
```

2. Run the following command to create a bucket:

```
#Set the bucket ACL to Private: res = oss.create_bucket(bucket,"
private") print "%s\n%s" % (res.status, res.read())
```

3. Run the following command to upload an object:

```
res = oss.put_object_from_file(bucket, object, "test.txt") print "%s
\n%s" % (res.status, res.getheaders())
```

4. Run the following command to obtain the corresponding object:

```
res = oss.get_object_to_file(bucket, object, "/filepath/test.txt")
print "%s\n%s" % (res.status, res.getheaders())
```

In the ECS application code, ApsaraDB for RDS stores the ID of each user, and OSS stores the avatar resource of the user. The Python code is as follows:

```
#!/usr/bin/env python from oss.oss_api import * endpoint" oss-cn-
hangzhou.aliyuncs.com" accessKeyId, accessKeySecret="your id", "your
 secret" oss = OssAPI(endpoint, accessKeyId, accessKeySecret)
user_id = mysql_client.fetch_one(sql)#Search for user_id in ApsaraDB
 for RDS.
#Obtain the user's avatar and download it to the corresponding path.
oss.get_object_to_file(bucket, object, your_path/user_id+’.png’)
#Process the avatar uploaded by the user.
oss.put_object_from_file(bucket, object, your_path/user_id+’.png’)
```

17 ApsaraDB for Redis

17.1 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.

17.2 Quick start

17.2.1 Log on to the ApsaraDB for Redis console

Take the Chrome browser as an example to describe how to log on to the ApsaraDB for Redis console through the Apsara Stack console as cloud product users.

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

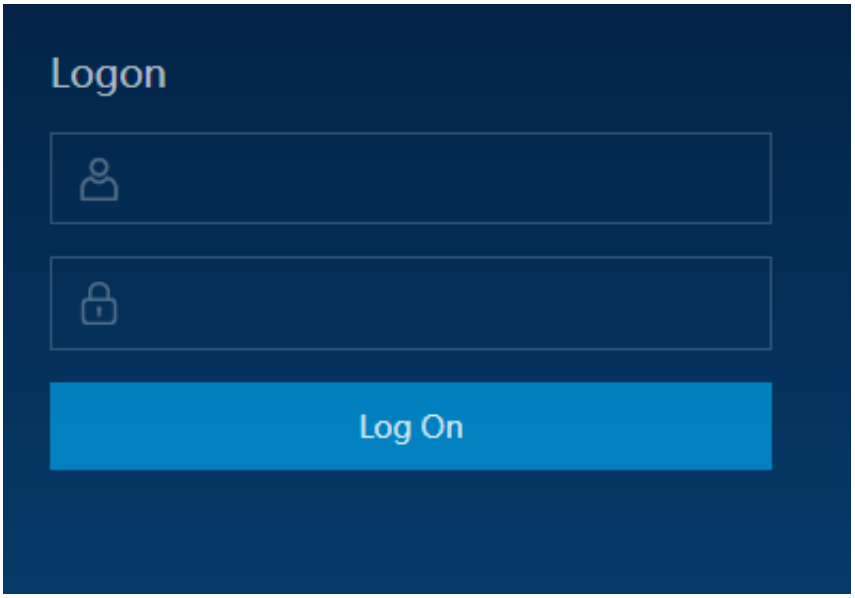
  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 17-1: Log on to the Apsara Stack console*.

Figure 17-1: Log on to the Apsara Stack console



3. Enter the correct username and password.
   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **Log On** to go to the **Home** page.
5. In the menu bar, choose **Console** > **Database** > **ApsaraDB for Redis**.

17.2.2 Create an instance

ApsaraDB supports both the classic network and VPC. You can create ApsaraDB for Redis instances on different networks.

Prerequisites

To create an ApsaraDB for Redis instance on VPC, you must create a VPC instance and create an ApsaraDB for Redis instance in the same region as the VPC instance.

Context

Note:
The network type is specified when the instance is created, which cannot be modified.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. Click **Create Instance** in the upper right corner of the instance list.
3. On the **Create Redis Instance** page, select a network type and complete other settings.

   If you select VPC, create a VPC instance first. For more information, see **Create a VPC instance and VSwitch** in *Cite LeftVPC User GuideCite Right*.

   Table 17-1: ApsaraDB for Redis configuration parameters

Parameter	Description
Region	Select a region for the ApsaraDB for Redis instance.
Zone	Select a zone for the ApsaraDB for Redis instance.
Department	Select a department for the ApsaraDB for Redis instance.
Project	Select a project for the ApsaraDB for Redis instance. **Note:** After a project is selected, only the members of the project can access the ApsaraDB for Redis instance. For more information, see **View project members** in the *Cite LeftApsara Stack Console User GuideCite Right*.
Architecture	Select an architecture type for the ApsaraDB for Redis instance. ApsaraDB for Redis provides the Clustered and Standard architectures. The Clustered architecture is applicable to businesses that require large ApsaraDB for Redis capacity or high performance. As ApsaraDB for Redis runs under a single thread, the Standard architecture is recommended for businesses that require a performance lower than 100,000 QPS. For higher performance, select the Clustered version. For more information, see **`Product Architecture`** in *Cite LeftApsara Stack Product IntroductionCite Right*.
Node Type	Select a node type for the ApsaraDB for Redis instance. ApsaraDB for Redis supports the Primary/Secondary dual-node architecture.
Service Plan	Select the Standard or Premium service plan.

Parameter	Description
	A Premium service plan provides instances with premium configuration.
Instance Specifications	Select the instance specification. The maximum connections and maximum internal bandwidth vary with instance specifications.
Network Type	On Alibaba Cloud platform, a classic network and a VPC instance have the following differences: • Classic network: The cloud services in a classic network are not isolated, and unauthorized access to a cloud service is blocked only by the security group or whitelist policy of the service. • VPC: VPC helps you build an isolated network environment in Alibaba Cloud. You can customize the route table, IP address range, and gateway in VPC. In addition, you can combine your data center and cloud resources in Alibaba Cloud VPC into a virtual data center through a leased line or VPN to migrate applications to the cloud smoothly.
Set Password	Set a password for accessing the instance. The text can be 8 to 30 characters in length and must contain a mixture of uppercase letters, lowercase letters, and numbers. Special characters are not supported.
Instance Name	Enter a name of the instance. The value must be 2 to 128 characters in length, start with a letter or a Chinese character, and can contain numbers, letters, Chinese characters, underscores (_), and hyphens (-).

4. Click **Create**.

   After the instance is created, wait until the instance status becomes **Normal**.

17.2.3 Set an IP address whitelist

Before using an ApsaraDB for Redis instance, you must add IP addresses or IP address segments used for database access to the whitelist of the target instance. This ensures database security and stability.

Context

Correct use of the whitelist improves access protection for ApsaraDB for Redis instances. We recommend that you regularly maintain the whitelist.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details**.
3. On the **Instance Information** page, click **Modify Whitelist**.
4. In the **Modify Whitelist** window, set the IP address whitelist, and click **Confirm**.

> **Note:**
> To allow all IP addresses to access the database, set the IP address whitelist to 0.0.0.0/0. To disable database access from all IP addresses, set the IP address whitelist to 127.0.0.1. We recommend that you delete the default IP address 127.0.0.1. Otherwise, the new IP addresses will be invalid.

17.2.4 Connect to an instance

17.2.4.1 Connect to ApsaraDB for Redis from a Redis client

Context

As services provided by ApsaraDB for Redis are completely compatible with those of a native database, the databases are connected in similar ways. Any clients compatible with the Redis protocol can access ApsaraDB for Redis. You can select any Redis clients based on their application features.

> **Note:**
> ApsaraDB for Redisonly supports access requests from the Apsara Stack internal network rather than those from the Internet. That means only Redis clients installed on ECS instances of the same node can be connected to ApsaraDB for Redis for data operations.

To use Redis clients, see *http://redis.io/clients*.

17.2.4.1.1 Jedis client

Download Jedis

Click *Reference URL*.

Example of Jedis single-connection

```
import redis.clients.jedis.Jedis; public class jedistest {
public static void main(String[] args) { try { String host = "xx
```

```
.kvstore.aliyuncs.com";//Access URL displayed on the console int
 port = 6379; Jedis jedis = new Jedis(host, port); //Authentica
tion information jedis.auth("password");//password String key
 = "redis"; String value = "aliyun-redis"; //Select a database
 (The default value is 0.) jedis.select(1); //Set a key jedis.
set(key, value); System.out.println("Set Key " + key + " Value
: " + value); //Get the set key String getvalue = jedis.get(
key); System.out.println("Get Key " + key + " ReturnValue: " +
getvalue); jedis.quit(); jedis.close(); } catch (Exception e) {
e.printStackTrace(); } } }
```

Example of JedisPool

Configuration file

You can configure the pom configuration file based on the client version that you select.

The configuration is as follows:

```
<dependency> <groupId>redis.clients</groupId> <artifactId>jedis
</artifactId> <version>2.7.2</version> <type>jar</type> <scope>
compile</scope> </dependency>
```

Reference to be added

```
import org.apache.commons.pool2.PooledObject; import org.apache
.commons.pool2.PooledObjectFactory; import org.apache.commons
.pool2.impl.DefaultPooledObject; import org.apache.commons.
pool2.impl.GenericObjectPoolConfig; import redis.clients.jedis
.HostAndPort; import redis.clients.jedis.Jedis; import redis.
clients.jedis.JedisPool; import redis.clients.jedis.JedisPoolC
onfig;
```

Example of Jedis-2.7.2

```
JedisPoolConfig config = new JedisPoolConfig(); //Maximum number
 of idle connections, which is evaluated by the application.
Do not set it to a value greater than the maximum number of
connections of an ApsaraDB for Redis instance config.setMaxIdle
(200); //Maximum number of connections, which is evaluated by
 the application. Do not set it to a value greater than the
maximum number of connections of an ApsaraDB for Redis instance
 config.setMaxTotal(300); config.setTestOnBorrow(false); config
.setTestOnReturn(false); String host = "*.aliyuncs.com"; String
 password = "Password"; JedisPool pool = new JedisPool(config,
host, 6379, 3000, password); Jedis jedis = null; try { jedis =
pool.getResource(); /// ... do stuff here ... for example jedis
.set("foo", "bar"); String foobar = jedis.get("foo"); jedis.zadd
("sose", 0, "car"); jedis.zadd("sose", 0, "bike"); Set<String
> sose = jedis.zrange("sose", 0, -1); } finally { if (jedis !=
```

```
null) { jedis.close(); } } /// ... when closing your application
: pool.destroy();
```

Examples of Jedis-2.6 and Jedis-2.5

```
JedisPoolConfig config = new JedisPoolConfig(); //Maximum number
 of idle connections, which is evaluated by the application.
Do not set it to a value greater than the maximum number of
connections of an ApsaraDB for Redis instance config.setMaxIdle
(200); //Maximum number of connections, which is evaluated by
 the application. Do not set it to a value greater than the
maximum number of connections of an ApsaraDB for Redis instance
 config.setMaxTotal(300); config.setTestOnBorrow(false); config
.setTestOnReturn(false); String host = "*.aliyuncs.com"; String
 password = "Password"; JedisPool pool = new JedisPool(config,
host, 6379, 3000, password); Jedis jedis = null; boolean broken
 = false; try { jedis = pool.getResource(); /// ... do stuff
here ... for example jedis.set("foo", "bar"); String foobar =
jedis.get("foo"); jedis.zadd("sose", 0, "car"); jedis.zadd("
sose", 0, "bike"); Set<String> sose = jedis.zrange("sose", 0, -
1); } catch(Exception e) { broken = true; } finally { if (broken
) { pool.returnBrokenResource(jedis); } else if (jedis != null)
 { pool.returnResource(jedis); } }
```

17.2.4.1.2 phpredis client

Download phpredis

Click *Reference URL*.

Sample connection code

```
<?php /* Replace the following parameter values with the host
and port number of the connected instance */ $host = "localhost
"; $port = 6379; /* Replace the following parameter values with
 the instance ID and password */ $user = "test_username"; $pwd
 = "test_password"; $redis = new Redis(); if ($redis->connect
($host, $port) == false) { die($redis->getLastError()); } if ($
redis->auth($pwd) == false) { die($redis->getLastError()); } /*
Database operations can be performed after authentication. For
more information about the documentation, visit https://github.
com/phpredis/phpredis */ if ($redis->set("foo", "bar") == false
) { die($redis->getLastError()); } $value = $redis->get("foo");
echo $value; ?>
```

17.2.4.1.3 redis-py client

Download redis-py

Click *Reference URL*.

Sample connection code

```
#!/usr/bin/env python #-*- coding: utf-8 -*- import redis #
Replace the following parameter values with the host and port
number of the connected instance host = localhost port = 6379 #
Replace the following parameter value with the instance password
 pwd = test_password r = redis.StrictRedis(host=host, port=port
, password=pwd) #Database operations can be performed after
authentication. For more information about the documentation,
visit https://github.com/andymccurdy/redis-py r.set(foo, bar);
print r.get(foo)
```

17.2.4.1.4 C/C++ client

The following example describes how to use ApsaraDB for Redis on a C/C++ client.

Download, compile, and install the C client

```
git clone https://github.com/redis/hiredis.git cd hiredis make
sudo make install
```

Compile the test code

```
 #include <stdio.h> #include <stdlib.h> #include <string.h> #
include <hiredis.h> int main(int argc, char **argv) { unsigned
 int j; redisContext *c; redisReply *reply; if (argc < 4) {
printf("Usage: example xxx.kvstore.aliyuncs.com 6379 instance_i
d password\n"); exit(0); } const char *hostname = argv[1];
const int port = atoi(argv[2]); const char *instance_id = argv
[3]; const char *password = argv[4]; struct timeval timeout
 = { 1, 500000 }; // 1.5 seconds c = redisConnectWithTimeout(
hostname, port, timeout); if (c == NULL || c->err) { if (c) {
printf("Connection error: %s\n", c->errstr); redisFree(c); }
else { printf("Connection error: cant allocate redis context\n
"); } exit(1); } /* AUTH */ reply = redisCommand(c, "AUTH %s",
password); printf("AUTH: %s\n", reply->str); freeReplyObject(
reply); /* PING server */ reply = redisCommand(c,"PING"); printf
("PING: %s\n", reply->str); freeReplyObject(reply); /* Set a
key */ reply = redisCommand(c,"SET %s %s", "foo", "hello world
"); printf("SET: %s\n", reply->str); freeReplyObject(reply); /*
Set a key using binary safe API */ reply = redisCommand(c,"SET
 %b %b", "bar", (size_t) 3, "hello", (size_t) 5); printf("SET (
binary API): %s\n", reply->str); freeReplyObject(reply); /* Try
 a GET and two INCR */ reply = redisCommand(c,"GET foo"); printf
("GET foo: %s\n", reply->str); freeReplyObject(reply); reply =
redisCommand(c,"INCR counter"); printf("INCR counter: %lld\n",
reply->integer); freeReplyObject(reply); /* again ... */ reply
 = redisCommand(c,"INCR counter"); printf("INCR counter: %lld\n
", reply->integer); freeReplyObject(reply); /* Create a list of
 numbers, from 0 to 9 */ reply = redisCommand(c,"DEL mylist");
freeReplyObject(reply); for (j = 0; j < 10; j++) { char buf[64
]; snprintf(buf,64,"%d",j); reply = redisCommand(c,"LPUSH mylist
 element-%s", buf); freeReplyObject(reply); } /* Lets check
what we have inside the list */ reply = redisCommand(c,"LRANGE
 mylist 0 -1"); if (reply->type == REDIS_REPLY_ARRAY) { for (j
 = 0; j < reply->elements; j++) { printf("%u) %s\n", j, reply->
```

```
element[j]->str); } } freeReplyObject(reply); /* Disconnects and
 frees the context */ redisFree(c); return 0; }
```

Compile the code

```
 gcc -o example -g example.c -I /usr/local/include/hiredis –
lhiredis
```

Test and run the code

```
 example xxx.kvstore.aliyuncs.com 6379 instance_id password
```

17.2.4.1.5 .net client

The following example describes how to use ApsaraDB for Redis on a .net client.

1. Download and use the .net client.

   ```
   git clone https://github.com/ServiceStack/ServiceStack.Redis
   ```

2. Create a *.net* project.
3. Add the reference file stored in the library file directory *ServiceStack.Redis/lib/tests* to the client.

Sample test code

```
using System; using System.Collections.Generic; using System.
Linq; using System.Text; using System.Threading.Tasks; using
ServiceStack.Redis; namespace ServiceStack.Redis.Tests { class
Program { public static void RedisClientTest() { string host = "
127.0.0.1";/*IP address of the access host*/ string password = "
password";/*Password*/ RedisClient redisClient = new RedisClien
t(host, 6379, password); string key = "test-aliyun"; string
value = "test-aliyun-value"; redisClient.Set(key, value); string
 listKey = "test-aliyun-list"; System.Console.WriteLine("set
key " + key + " value " + value); string getValue = System.Text.
Encoding.Default.GetString(redisClient.Get(key)); System.Console
.WriteLine("get key " + getValue); System.Console.Read(); }
public static void RedisPoolClientTest() { string[] testReadWr
iteHosts = new[] { "redis://password@127.0.0.1:6379"/*redis
://password@access IP address:port*/ }; RedisConfig.VerifyMast
erConnections = false;//You need to set the parameter PooledRedi
sClientManager redisPoolManager = new PooledRedisClientManager
(10/*Number of connection pools*/, 10/*Connection pool timeout
 time*/, testReadWriteHosts); for (int i = 0; i < 100; i++)
 { IRedisClient redisClient = redisPoolManager.GetClient();//
Obtain the connection RedisNativeClient redisNativeClient =
 (RedisNativeClient)redisClient; redisNativeClient.Client =
null;//ApsaraDB for Redis does not support client setname.
Therefore, set the client object to null try { string key =
 "test-aliyun1111"; string value = "test-aliyun-value1111";
redisClient.Set(key, value); string listKey = "test-aliyun-list
"; redisClient.AddItemToList(listKey, value); System.Console.
WriteLine("set key " + key + " value " + value); string getValue
```

```
 = redisClient.GetValue(key); System.Console.WriteLine("get key
 " + getValue); redisClient.Dispose();// } catch (Exception e)
 { System.Console.WriteLine(e.Message); } } System.Console.Read
(); } static void Main(string[] args) { //Single-connection mode
 RedisClientTest(); //Connection pool mode RedisPoolClientTest
(); } } }
```

For more information about how to use the APIs, visit *https://github.com/ServiceStack/ServiceStack.Redis*.

17.2.4.1.6 node-redis client

1. Install node-redis.

```
npm install hiredis redis
```

2. Connect to ApsaraDB for Redis.

```
 var redis = require("redis"), client = redis.createClient({
detect_buffers: true}); client.auth("password", redis.print)
```

3. Use ApsaraDB for Redis.

```
// Write data client.set("key", "OK"); // Obtain data. A
string is returned client.get("key", function (err, reply)
 { console.log(reply.toString()); // print `OK` }); // If a
 buffer is transmitted, a buffer is returned client.get(new
 Buffer("key"), function (err, reply) { console.log(reply.
toString()); // print `<Buffer 4f 4b>` }); client.quit();
```

17.2.4.2 Connect to ApsaraDB for Redis through redis-cli

Context

> **Note:**
> ApsaraDB for Redis only supports access requests from the Alibaba Cloud internal network rather than those from the Internet. This means only redis-cli installed on ECS instances of the same node can be connected to ApsaraDB for Redis for data operations.

Procedure

You can run the following command to connect to ApsaraDB for Redis through redis-cli:

```
redis-cli -h instance connection address -a password
```

17.2.4.3 Connect to an instance over the Internet

Context

Currently, ApsaraDB for Redis is accessible through the ECS internal network. To access ApsaraDB for Redis through the Internet, run netsh on the ECS Windows server to create a port mapping or install rinetd on the ECS Linux server for forwarding.

- ECS Windows

  a) Log on to the ECS Windows server and run the following command at the command prompt:

  ```
  netsh interface portproxy add v4tov4 listenaddress=IP address
   of the ECS instance listenport=6379 connectaddress=connection
   address of ApsaraDB for Redis connectport=6379
  ```

  Supplement:

  - `netsh interface portproxy delete v4tov4 listenaddress=IP address of the ECS instance listenport=6379` You can delete unnecessary mappings
  - `netsh interface portproxy show all`You can check the mappings on the current server.

  b) Perform a verification test after the configuration is complete.

  - Perform a connection test on the ECS Windows server. If the IP address of the server is **1.1.1.1**, run `telnet 1.1.1.1 6379`.
  - Alternatively, run the redis-cli command locally to connect to the ECS Windows server and perform data write and query.

  After performing the preceding steps, you can use a local PC or server to connect to port 6379 of the ECS Windows server through the Internet and access ApsaraDB for Redis.

   **Note:**
  portproxy is provided by Microsoft rather than open source software. See the netsh official documentation on portproxy or consult with Microsoft engineers if you have any problems in configuration and use. Alternatively, use another scheme. For example, use portmap to configure proxy mappings.

- ECS Linux

a) Install rinetd on the ECS Linux server.

```
 wget http://www.boutell.com/rinetd/http/rinetd.tar.gz&&tar -xvf
rinetd.tar.gz&&cd rinetd sed -i s/65536/65535/g rinetd.c (Modify
the port range) mkdir /usr/man&&make&&make install
```

Note:

The rinetd installation package obtained from the download URL may be unavailable. You can download the rinetd installation package from other sources.

b) Create the configuration file.

```
 vi /etc/rinetd.conf
```

c) Enter the following information: `0.0.0.0 6379 URL to the Internet of ApsaraDB for Redis 6379 logfile /var/log/rinetd.log`

d) Run `rinetd` to start rinetd.

Note:

- Run `echo rinetd >>/etc/rc.local` to make rinetd autorun.
- Run `pkill rinetd` to kill the rinetd process.

e) Perform a verification test.

Run redis-cli locally to connect to the ECS Linux server for logon authentication. For example, if the IP address of the server with rinetd installed is 1.1.1.1, you can run the following command:

```
redis-cli -h 1.1.1.1 -a password of ApsaraDB for Redis
```

After performing the preceding steps, you can use a local PC or server to connect to port 6379 of the ECS Linux server through the Internet and access ApsaraDB for Redis.

Note:

You can use the above scheme to test and use rinetd. As rinetd is open source software, you can read its official documentation or contact rinetd for support if you have any problems in use.

17.3 Manage instances

17.3.1 Edit the password of an instance

You can reset the password of an instance if you did not set a password when creating the instance, or you forget or need to edit the password.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click **Reset Password**.
4. In the **Reset Password** window, enter a new logon password, confirm it, and click **Submit**.

17.3.2 View details of an instance

After creating an instance, you can view details of the instance on Apsara Stack console.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details**. On the **Instance Information** page, view the instance details.

   The **Instance Information** page displays **Basic Information**, **Specifications**, and **Connection Information**. The following table lists the items in each part.

   Table 17-2: Instance Information

Region	Item
Basic Information	• Instance ID • Name • Status • Region • Department • Project • Created At: • Zone • Network Type

Region	Item
	• VPC (displayed only when the instance is of the VPC type)
Specifications	• Instance Specifications • Maximum Connections • Maximum Internal Network Bandwidth • Maintenance Period • Access Control Whitelist
Connection Information	• Connection Address (Host) • Port Number • SSL Status • SSL Expires:

17.3.3 Change an instance name

After creating an instance, you can change the instance name on Apsara Stack console.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and choose > **Edit**.
3. In the **Edit Instance Information** window, set the name for the instance, and click **Confirm**.

17.3.4 Modify configurations

ApsaraDB for RedisConfiguration can be modified.

Context

Note:
The instance may be interrupted intermittently for several seconds when the configuration is modified.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click > **Change**. Alternatively, click the instance ID. On the **Instance Information** page, click **Change Instance**.
3. In the **Change Instance** window, set **Instance Specifications** and click **Confirm**.

`Instance changed` is displayed after the configuration is successfully modified. You can use the instance only when the instance status becomes **Normal** after a while.

17.3.5 Set the O&M time

To ensure the stability of ApsaraDB for Redis instances, the backend system irregularly maintains instances and machines.

Context

To guarantee stability of the entire maintenance process, instances enter the **Being Maintained** state before the preset O&M time on the day of maintenance. When an instance is in this state, access to data in the database is not affected. However, change-related functions (for example, configuration change) are temporarily unavailable for this instance on the console, whereas query-related functions such as performance monitoring are still available.

Note:
After the preset O&M time is reached, instances may be interrupted intermittently. We recommend that you maintain instances during off-peak hours.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click **Modify O&M Time**.
4. In the **Modify O&M Time** window, select the O&M time, and click **Confirm**.

17.3.6 Clear data of an instance

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click > **Clear**.
3. In the **Clear Instance** window, click **Confirm**.

17.3.7 Release an instance

Procedure

1. *Log on to the ApsaraDB for Redis console*.

2. In the instance list, locate the target instance and choose > **Release**.
3. In the **Delete Instance** window, click **Confirm**.

17.3.8 Enable data transmission encryption

To ensure instance security, you can enable the SSL/TLS encrypted connection after creating an instance.

Prerequisites

> **Note:**
> Clustered instances do not support SSL/TLS encryption.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click **Enable SSL**. A message is displayed, indicating that the operation is successful. After a while, refresh the instance information page. **Disable SSL** and **Download SSL Certificate** are displayed on the page, indicating that the operation is successful.

17.4 Import data

Context

redis-cli is a native Redis command line tool. ApsaraDB for Redis allows you to use redis-cli to seamlessly import existing Redis data to ApsaraDB for Redis.
NOTE:

- Because ApsaraDB for Redis supports access only from the Apsara Stack internal network, data importing takes effect only on Apsara Stack ECS instances. If your ApsaraDB for Redis instance is not on the Apsara Stack ECS instance, you must copy the existing append-only file (AOF) to the ECS instance before importing data.
- redis-cli is a native Redis command line tool. If redis-cli does not work on your ECS instance, you need to first download and install ApsaraDB for Redis.

Perform the following steps if you have created an ApsaraDB for Redis instance on the Apsara Stack ECS instance:

Procedure

1. Enable the Append-only file (AOF) function for the existing ApsaraDB for Redis instance (skip this step if the AOF function has been enabled).

```
# redis-cli -h old_instance_ip -p old_instance_port config set
 appendonly yes
```

2. Use the AOF to import data to the new ApsaraDB for Redis instance (assume that the generated AOF is named append.aof).

```
# redis-cli -h aliyun_redis_instance_ip -p 6379 -a password --pipe <
 appendonly.aof
```

Note:

If the AOF function does not need to always be enabled for the existing ApsaraDB for Redis instance, run the following command to disable the function after the data is imported:

```
# redis-cli -h old_instance_ip -p old_instance_port config set
 appendonly no
```

17.5 Backup and Restore

As an increasing number of businesses use ApsaraDB for Redis as the ultimate persistent storage engine, users have higher data reliability requirements. The ApsaraDB for Redis backup and restore solution ensures comprehensive upgrade of the ApsaraDB for Redis data reliability.

17.5.1 Set an automatic backup policy

Context

As an increasing number of applications use ApsaraDB for Redis for persistent storage, conventional backup is required to quickly restore data in the case of misoperations. Alibaba Cloud implements RDB snapshot backup on slave nodes to protect the performance of your instance in the backup process. Alibaba Cloud also provides convenient console operations, allowing you to customize the backup settings.

Prerequisites

Note:

Clustered instances do not support the backup and restore operations.

Procedure

1. *Log on to the ApsaraDB for Redis console*.

2. In the instance list, locate the target instance and click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click the **Backup and Restore** tab.
4. Click **Backup Settings**.
5. Click **Settings**. In the **Backup Settings** window, set the automatic backup period and time.

   Note:
   Backup data is retained for seven days by default. This setting cannot be modified.
6. Click **Confirm** to complete automatic backup setting.

17.5.2 Manual data backup

In addition to the general backup settings, you can initiate a manual backup request on the console at any time.

Prerequisites

Note:
Clustered instances do not support the backup and restore operations.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click the **Backup and Restore** tab.
4. Click **Backup Data**.
5. Click **Create Backup** in the upper right corner.
6. In the **Back Up Instance** window, click **Confirm** to back up the instance immediately.

   Note:
   On the **Backup Data** page, you can select time ranges and query historical backup data. Backup data is retained for seven days by default. You can query historical backup data of the last seven days.

17.5.3 Archive backups

Context

Due to industry regulatory or corporate policy requirements, you may need to regularly back up and archive ApsaraDB for Redis data. ApsaraDB for Redis provides a backup archiving function that automatically saves automatic and manual backup files to OSS instances. Currently, Alibaba Cloud stores your backup files on OSS instances for seven days. After seven days, the backup files are automatically deleted.

To archive these backup files for a longer period, you can copy the link on the console and download the database backup files for local storage.

Prerequisites

Note:
Clustered instances do not support the backup and restore operations.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click the **Backup and Restore** tab.
4. On the **Backup Data** page, select the backup dataset to be downloaded, click the icon and select **Download**.
5. In the displayed dialog box, click **Confirm**. The file is downloaded to the default local directory.

17.5.4 Restore data

The data restore function minimizes the loss caused by database misoperations. Currently, ApsaraDB for Redis supports data restore by backup data.

Prerequisites

Note:

- As the data restore operation is highly risky, and perform data restore after verifying that the data is correct.
- Clustered instances do not support the backup and restore operations.

Before restoring data, make sure that you have backed up the data. After the data is backed up, ApsaraDB for Redis retains the backup data for seven days by default.

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click the **Backup and Restore** tab. The **Backup Data** page is displayed by default.
4. Select the time range for restore and click **Search**. The backup datasets within the time range are listed.

   Backup data is available only when you have backed up the data in the period.
5. Select the target backup file and choose > **Data Restore**.
6. In the data restore window, click **Confirm** to restore the data of the original instance.

17.6 Set parameters

ApsaraDB for RedisAllows you to set some instance parameters. For more information about the parameters that can be modified, see **Parameter settings** on the ApsaraDB for Redis instance.

Context

ApsaraDB for Redis is completely compatible with the native database service. The setting method of the cloud database parameters is similar to that of the local database parameters. You can modify parameters on the ApsaraDB for Redis console by referring to this example or using other methods such as redis-cli.

For more information about the database parameter descriptions, see the official documentations for the corresponding database version by clicking the following links.

- *redis.conf for Redis 3.0*
- *redis.conf for Redis 2.8*

Procedure

1. *Log on to the ApsaraDB for Redis console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click the **Parameters** tab.

4. Select the parameter to be modified and choose > **Edit**.

5. Modify the parameter value and click **Confirm**.

17.7 Commands supported by ApsaraDB for Redis

ApsaraDB for Redis is compatible with Redis 3.0. Supports Redis 3.0 GEO commands. For more information about the commands of ApsaraDB for Redis, see *http://redis.io/commands*.

Supported command operations

Key	String	Hash	List	Set	SortedSet
DEL	APPEND	HDEL	BLPOP	SADD	ZADD
DUMP	BITCOUNT	HEXISTS	BRPOP	SCARD	ZCARD
EXISTS	BITOP	HGET	BRPOPLPUSH	SDIFF	ZCOUNT
EXPIRE	BITPOS	HGETALL	LINDEX	SDIFFSTORE	ZINCRBY
EXPIREAT	DECR	HINCRBY	LINSERT	SINTER	ZRANGE
MOVE	DECRBY	HINCRBYFLOAT	LLEN	SINTERSTORE	ZRANGEBYSCORE
PERSIST	GET	HKEYS	LPOP	SISMEMBER	ZRANK
PEXPIRE	GETBIT	HLEN	LPUSH	SMEMBERS	ZREM
PEXPTREAT	GETRANGE	HMGET	LPUSHX	SMOVE	ZREMRANGEBYRANK
PTTL	GETSET	HMSET	LRANGE	SPOP	ZREMRANGEBYSCORE
RANDOMKEY	INCR	HSET	LREM	SRANDMEMBER	ZREVRANGE
RENAME	INCRBY	HSETNX	LSET	SREM	ZREVRANGEBYSCORE
RENAMENX	INCRBYFLOAT	HVALS	LTRIM	SUNION	ZREVRANK
RESTORE	MGET	HSCAN	RPOP	SUNIONSTORE	ZSCORE
SORT	MSET	-	RPOPLPUSH	SSCAN	ZUNIONSTORE
TTL	MSETNX	-	RPUSH	-	ZINTERSTORE
TYPE	PSETEX	-	RPUSHX	-	ZSCAN

Key	String	Hash	List	Set	SortedSet
SCAN	SET	-	-	-	ZRANGEBYLEX
OBJECT	SETBIT	-	-	-	ZLEXCOUNT
-	SETEX	-	-	-	ZREMRANGEBYLEX
-	SETNX	-	-	-	-
-	SETRANGE	-	-	-	-
-	STRLEN	-	-	-	-

And

HyperLogLog	Pub/Sub	Transaction	Connection	Server	Scripting	Geo
PFADD	PSUBSCRIBE	DISCARD	AUTH	FLUSHALL	EVAL	GEOADD
PFCOUNT	PUBLISH	EXEC	ECHO	FLUSHDB	EVALSHA	GEOHASH
PFMERGE	PUBSUB	MULTI	PING	DBSIZE	SCRIPT EXISTS	GEOPOS
-	PUNSUBSCRIBE	UNWATCH	QUIT	TIME	SCRIPT FLUSH	GEODIST
-	SUBSCRIBE	WATCH	SELECT	INFO	SCRIPT KILL	GEORADIUS
-	UNSUBSCRIBE	-	-	KEYS	SCRIPT LOAD	GEORADIUSBYMEMBER
-	-	-	-	CLIENT KILL	-	-
-	-	-	-	CLIENT LIST	-	-
-	-	-	-	CLIENT GETNAME	-	-
-	-	-	-	CLIENT SETNAME	-	-
-	-	-	-	CONFIG GET	-	-
-	-	-	-	MONITOR	-	-

HyperLogLog	Pub/Sub	Transaction	Connection	Server	Scripting	Geo
-	-	-	-	SLOWLOG	-	-

Commands temporarily unavailable

Keys	Server
MIGRATE	BGREWRITEAOF
-	BGSAVE
-	CONFIG REWRITE
-	CONFIG SET
-	CONFIG RESETSTAT
-	COMMAND
-	COMMAND COUNT
-	COMMAND GETKEYS
-	COMMAND INFO
-	DEBUG OBJECT
-	DEBUG SEGFAULT
-	LASTSAVE
-	ROLE
-	SAVE
-	SHUTDOWN
-	SLAVEOF
-	SYNC

Commands not supported by cluster instances

Transaction	Scripting	Connection	Keys	List
DISCARD	EVAL	SELECT	MOVE	BLPOP
EXEC	EVALSHA	-	SCAN	BRPOP
MULTI	SCRIPT EXISTS	-	-	BRPOPLPUSH
UNWATCH	SCRIPT FLUSH	-	-	-
WATCH	SCRIPT KILL	-	-	-

Transaction	Scripting	Connection	Keys	List
-	SCRIPT LOAD	-	-	-

Commands restricted for cluster instances

Keys	Strings	Lists	Sets	Sorted Sets	HyperLogLog
RENAME	MSETNX	RPOPLPUSH	SINTERSTORE	ZUNIONSTORE	PFMERGE
RENAMENX	-	-	SINTER	ZINTERSTORE	-
-	-	-	SUNIONSTORE	-	-
-	-	-	SUNION	-	-
-	-	-	SDIFFSTORE	-	-
-	-	-	SDIFF	-	-
-	-	-	SMOVE	-	-

Note:

Restricted commands only support scenarios where keys to be operated are evenly distributed in a single hash slot, and data in multiple hash slots is not merged. Therefore, use the hash tag to ensure that keys to be operated are evenly distributed in one hash slot. For example, if key1, aakey, and abkey3 are to be operated, store them in {key}1, aa{key}, and ab{key}3 mode. In this case, restricted commands can take effect when being called. For more information about how to use the hash tag, see the *Redis documentation*.

18 ApsaraDB for Memcache

18.1 What is ApsaraDB for Memcache

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

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

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

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

18.2 Quick start

18.2.1 Log on to the ApsaraDB for Memcache console

Take the Chrome browser as an example to describe how to log on to the ApsaraDB for Memcache console through the Apsara Stack console as cloud product users.

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

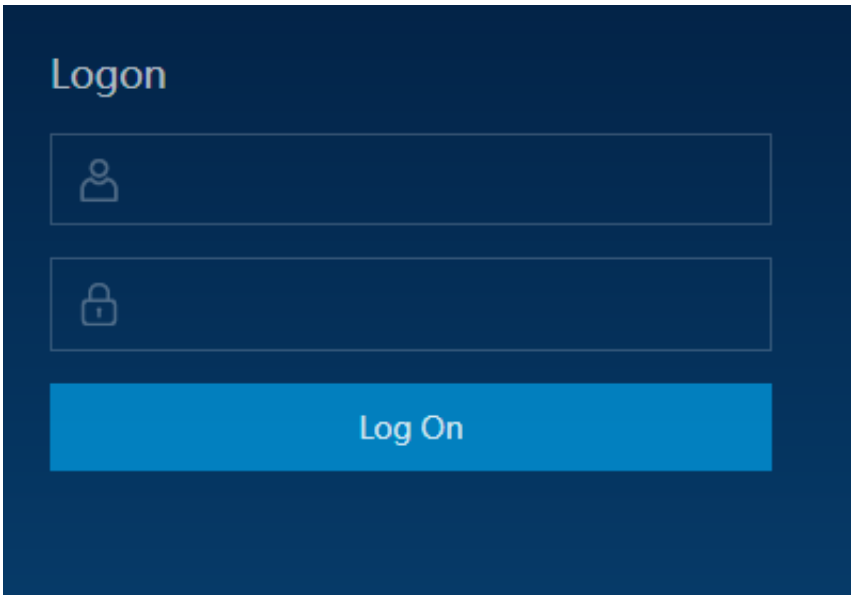
Procedure

1. Open your Chrome browser.

2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 18-1: Log on to the Apsara Stack console*.

   Figure 18-1: Log on to the Apsara Stack console



3. Enter the correct username and password.

   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click **Log On** to go to the **Home** page.
5. In the menu bar, choose **Console** > **Database** > **ApsaraDB for Memcache**.

18.2.2 Create an instance

ApsaraDB supports both the classic network and Virtual Private Cloud (VPC). You can create ApsaraDB for Memcache instances of different network types.

Prerequisites

- At least an ECS instance is required for activating ApsaraDB for Memcache.
- To create a VPC-type ApsaraDB for Memcache instance, you must first create a VPC instance. The ApsaraDB for Memcache instances and VPC must be in the same region.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. Click **Create Instance** in the upper right corner of the instance list.
3. On the **Create Memcache Instance** page, select a network type and complete other settings.

Table 18-1: Parameters of ApsaraDB for Memcache

Parameters	Description
Region	Select a region for the ApsaraDB for Memcache instance. ApsaraDB for Memcache is only accessible through the internal network. We recommend that you configure the ApsaraDB for Memcache instance and the ECS instances in the same zone of the same region.
Zone	Select a zone for the ApsaraDB for Memcache instance. ApsaraDB for Memcache is only accessible through the internal network. We recommend that you configure the ApsaraDB for Memcache instance and the ECS instances in the same zone of the same region.
Department	Select a department for the ApsaraDB for Memcache instance.
Project	Select a project for the ApsaraDB for Memcache instance.  **Note:** After a project is selected, only the members of the project can access the ApsaraDB for Redis instance. For more information, see **View project members** in the *Cite LeftApsara Stack Console User GuideCite Right*.
Instance Specifications	Select Instance Specification. The maximum number of connections and maximum internal network bandwidth vary with instance types.

Parameters	Description
Network Type	On Alibaba Cloud platform, a classic network and a VPC have the following differences: • Classic network: The cloud services in a classic network are not isolated. Unauthorized access to a cloud service is blocked only by the security group or whitelist policy of the service. • Virtual Private Cloud (VPC): VPC helps you build an isolated network environment in Alibaba Cloud. You can customize the route table, IP address range, and gateway in VPC. In addition , you can combine your data center and cloud resources in Alibaba Cloud VPC into a virtual data center through a leased line or VPN to migrate applications to the cloud smoothly. If you select VPC, create a VPC instance first. For more information, see **Create a VPC instance and VSwitch** in *Cite LeftVPC User GuideCite Right*.
Set Password	Set a password for accessing the instance. The password consists of 8–30 characters. It must contain uppercase letters, lowercase letters, and numbers. Special characters are not allowed. We recommend that you set a password for accessing the cloud instance when creating the instance. If no password is set upon creation, you must reset the password by choosing **Instances** > **Instance ID** > **Reset Password**.
Instance Name	Enter a name of the instance. The instance name consists of 2–128 characters. It can contain only letters, numbers, underlines (_), and hyphens (-). It must start with an uppercase or lowercase letter or a Chinese character.

4. Click **Create**.

18.2.3 Set an IP address whitelist

Context

Before using an ApsaraDB for Memcache instance, you must add IP addresses or IP address segments used for database access to the whitelist of the target instance. This guarantees database security and stability. Correct use of the whitelist improves access security for ApsaraDB for Memcache. We recommend that you maintain the whitelist on a regular basis. This document describes how to set a whitelist.

Note:

- Check that the ECS instance and ApsaraDB for Memcache instance added to the whitelist are in the same region.
- To enable applications to access multiple ApsaraDB for Memcache instances from the same ECS instance, you can bind one IP address to multiple ApsaraDB for Memcache instances.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click the **Security Settings** tab.
4. Click **Add Whitelist Group** or the modification icon after the default whitelist group.
5. In the **Modify Whitelist** window, enter IP addresses in **Group Whitelist** and click **Confirm**.

Note:
To allow all IP addresses to access the database, set the IP address whitelist to 0.0.0.0/0. To disable database access from all IP addresses, set the IP address whitelist to 127.0.0.1. We recommend that you delete the default IP address 0.0.0.0/0. Otherwise, the new IP addresses will be invalid.

18.2.4 Connect to an instance from a client

18.2.4.1 Client description

Any clients compatible with the Memcached protocol can access ApsaraDB for Memcache. Each Memcached client has its own features. You can select any Memcached client supporting SASL or Memcached Binary Protocol based on the application features.

The following Memcached clients can interact smoothly with ApsaraDB for Memcache, and therefore are recommended.

Note:

- You can remotely log on to ECS to access your ApsaraDB for Memcache instance. For more information, see *Connect to an instance over the Internet*.
- The following third-party open source clients are not provided by Alibaba Cloud and may contain bugs. The developer must guarantee the quality of the client. Alibaba Cloud is not held liable for any direct or indirect faults or losses arising from the client.

18.2.4.2 Java: Spymemcache

Context

Download a client

Client download URL

About the client

Client version

Java sample code

Procedure

1. Prepare the Java development environment. Log on to an existing Alibaba Cloud ECS instance and install the Java Development Kit (JDK) and commonly used integrated development environment (IDE) (such as Eclipse) on the instance.

   JDK *Download URL*

   Eclipse (*Download URL 1*, *Download URL 2*)

2. The first sample code is as follows. Copy the Java code to the Eclipse project.

   Note:
   Before you can compile the code successfully, you must download a JAR package from a third party to call the ApsaraDB for Memcache cache service. With this JAR package added, the code can be compiled.

   OcsSample1.java sample code (user name and password required)

```
import java.io.IOException; import java.util.concurrent.ExecutionE
xception; import net.spy.memcached.AddrUtil; import net.spy.
memcached.ConnectionFactoryBuilder; import net.spy.memcached
.ConnectionFactoryBuilder.Protocol; import net.spy.memcached.
MemcachedClient; import net.spy.memcached.auth.AuthDescriptor;
import net.spy.memcached.auth.PlainCallbackHandler; import net.spy.
memcached.internal.OperationFuture; public class OcsSample1 { public
 static void main(String[] args) { final String host = "xxxxxxxx.m
.yyyyyyyyyy.ocs.aliyuncs.com";//"Internal network address" on the
console final String port ="11211"; //Default port 11211, not need
 to change final String username = "xxxxxxxxx";//"Access account"
on the console final String password = "my_password";//"Password"
provided in the email MemcachedClient cache = null; try { AuthDescri
ptor ad = new AuthDescriptor(new String[]{"PLAIN"}, new PlainCallb
ackHandler(username, password)); cache = new MemcachedClient( new
ConnectionFactoryBuilder().setProtocol(Protocol.BINARY) .setAuthDes
criptor(ad) .build(), AddrUtil.getAddresses(host + ":" + port));
System.out.println("OCS Sample Code"); //Save a value with the "ocs
" key to ApsaraDB for Memcache to facilitate data verification and
 reading String key = "ocs"; String value = "Open Cache Service,
```

```
from www.Aliyun.com"; int expireTime = 1000; // Expiration time, in
 seconds. Timing starts from when data is written. After expireTime
elapses, the data expires and cannot be read OperationFuture<Boolean
> future = cache.set(key, expireTime, value); future.get(); //
The spymemcached set() method is asynchronous. The future.get()
operation starts after the cache.set() operation completes execution
. You can also have them to execute at the same time. //Save several
 values to ApsaraDB for Memcache and you can view the statistics
 on the ApsaraDB for Memcache console for(int i=0;i<100;i++){ key
="key-"+i; value="value-"+i; //Perform the Set operation and save
 the value to the cache expireTime = 1000; // Expiration time,
in seconds future = cache.set(key, expireTime, value); future.get
(); // Make sure that the previous (cache.set()) operation has
been completed } System.out.println("Set operation completed!"); //
Perform the Get operation and read the value with the "ocs" key from
 the cache System.out.println("Get operation"+cache.get(key)); }
catch (IOException e) { e.printStackTrace(); } catch (Interrupte
dException e) { e.printStackTrace(); } catch (ExecutionException
 e) { e.printStackTrace(); } if (cache != null) { cache.shutdown
(); } }//eof }
```

OcsSample2.java sample code (user name and password not required)

```
import java.io.IOException; import java.util.concurrent.ExecutionE
xception; import net.spy.memcached.AddrUtil; import net.spy.
memcached.BinaryConnectionFactory; import net.spy.memcached.
MemcachedClient; import net.spy.memcached.internal.OperationFuture
; public class OcsSample2 { public static void main(String[] args)
 { final String host = "xxxxxxxx.m.yyyyyyyyyy.ocs.aliyuncs.com"; //"
Internal network address" on the console final String port = "11211
"; //Default port 11211, not need to change MemcachedClient cache
 = null; try { cache = new MemcachedClient(new BinaryConnectionFact
ory(), AddrUtil.getAddresses(host + ":" + port)); System.out.println
("OCS Sample Code"); //Save a value with the "ocs" key to ApsaraDB
for Memcache to facilitate data verification and reading String key
 = "ocs"; String value = "Open Cache Service, from www.Aliyun.com
"; int expireTime = 1000; // Expiration time, in seconds. Timing
starts from when data is written. After expireTime elapses, the data
 expires and cannot be read OperationFuture<Boolean> future = cache
.set(key, expireTime, value); future.get(); //Save several values
 to ApsaraDB for Memcache and you can view the statistics on the
ApsaraDB for Memcache console for (int i = 0; i < 100; i++) { key
 = "key-" + i; value = "value-" + i; //Perform the Set operation
and save the value to the cache expireTime = 1000; // Expiration
time, in seconds future = cache.set(key, expireTime, value); future
.get(); } System.out.println("Set operation completed!"); //Perform
 the Get operation and read the value with the "ocs" key from the
 cache System.out.println("Get operation: " + cache.get(key)); }
catch (IOException e) { e.printStackTrace(); } catch (Interrupte
dException e) { e.printStackTrace(); } catch (ExecutionException
 e) { e.printStackTrace(); } if (cache != null) { cache.shutdown
(); } }//eof }
```

3. Modify the instance ID and internal network address in OcsSample1.java opened in Eclipse according to your instance information.

4. After the information is modified, you can run your program. Run the main function. The following result is displayed in the console window under Eclipse (ignore the red INFO debugging information that may be displayed):

```
OCS Sample Code Set operation completed! Get operation: Open Cache
Service, from www.Aliyun.com
```

18.2.4.3 PHP: memcached

Context

Download a client

Download a client

About the client

Client version

System requirements and environment configuration

Note:
If you already have a PHP Memcache environment, pay attention to the tips in the tutorial; otherwise, your production environment may be overwritten and services may become unavailable. We recommend that you back up your data before upgrading or compiling the environment.

ApsaraDB for Memcache for Windows

If the environment cannot be established using the standard PHP Memcached extensions, you can splice packets manually to access ApsaraDB for Memcache. For connection methods, see the following link. The sample code is simple. In comparison with PHP Memcached, it only supports mainstream interfaces, so you need to perform additional operations to use it with other specific interfaces. For installation and usage methods, click *Here*.

On a CentOS or Alibaba Cloud Linux 6 operating system

Note:
Memcached 2.2.0 extensions must use Libmemcached 1.0.x libraries. Libraries earlier than 1.0 cannot be compiled. The version of GCC used to compile Libmemcached must be 4.2 or later.

1. Check whether gcc-c++ and other components are installed (use `gcc -v` to check whether the GCC version is 4.2 or later). If the components are not installed, run `yum install gcc+ gcc-c++`.
2. Run `rpm -qa | grep php` to check whether the PHP environment is ready in the system. If not, run `yum install php-devel,php-common,php-cli` to install PHP with source code compiling.

   PHP 5.3 or later is recommended. PHP 5.2 contains the zend_parse_parameters_none function in the source code, which may have errors. If you need to use the function, read the PHP official documentation. If you compile the source code, follow the official PHP compiling and upgrading methods.
3. Check whether SASL-related environment packages are installed. If not, run `yum install cyrus-sasl-plain cyrus-sasl cyrus-sasl-devel cyrus-sasl-lib` to install SASL-related environments.
4. Check whether the Libmemcached source code package is installed. If not, run the following command to install it (Libmemcached 1.0.18 is recommended):

   ```
   wget https://launchpad.net/libmemcached/1.0/1.0.18/+download/
   libmemcached-1.0.18.tar.gz tar zxvf libmemcached-1.0.18.tar.gz cd
   libmemcached-1.0.18 ./configure --prefix=/usr/local/libmemcached --
   enable-sasl make make install cd ..
   ```
5. Run `yum install zlib-devel` to install the Memcached source code package (Memcached 2.2.0 is recommended).

   Note:

   - Before installing Memcached, check if there are any zlib-devel packages to be executed.
   - You must first check whether the Memcached client package (including the source code package) is installed. If the Memcached client package is installed, recompile it to add the -enable-memcached-sasl extension.

     ```
     wget http://pecl.php.net/get/memcached-2.2.0.tgz tar zxvf
      memcached-2.2.0.tgz cd memcached-2.2.0 phpize (If the system has
      two PHP environments, you must call the command by specifying
      the absolute path /usr/bin/phpize, which is the PHP environment
      path for using ApsaraDB for
                    Memcache) ./configure --with-libmemcached-dir=/
     usr/local/libmemcached --enable-memcached-sasl (Pay attention to
      this parameter) make make install
     ```
6. Add `extension=memcached.so memcached.use_sasl = 1` to the php.ini file. (Run the locate command to find this file. If the system has two PHP environments, you must find the

PHP environment path for using ApsaraDB for Memcache and modify the php.ini file in this path accordingly.)

7. Test whether the production environment is successfully deployed by using the test code provided at the end of the page. Replace the address, port, user name, and password in the test code with actual values.

On a CentOS or Alibaba Cloud Linux 5 (64-bit) operating system

1. Check whether gcc-c++ and other components are installed. If the components are not installed, run `yum install gcc+ gcc-c++`.
2. Run `rpm –qa | grep php` to check whether the PHP environment is ready in the system. If not, run `yum install php53 php53-devel` to install PHP with source code compiling. If the PHP environment has been prepared, skip this step. PHP 5.3 or later is recommended.

   PHP 5.2 contains the zend_parse_parameters_none function in the source code, which may have errors. If you need to use the function, read the PHP official documentation.
3. Run `yum install cyrus-sasl-plain cyrus-sasl cyrus-sasl-devel cyrus-sasl-lib` to install SASL-related environments.
4. Check whether Libmemcached (including the source code package) is installed. If not, run the following command to install Libmemcached (Libmemcached 1.0.2 is recommended):

```
 wget http://launchpad.net/libmemcached/1.0/1.0.2/+download/
libmemcached-1.0.2.tar.gz tar -zxvf libmemcached-1.0.2.tar.gz cd
libmemcached-1.0.2 ./configure --prefix=/usr/local/libmemcached --
enable-sasl make make install cd ..
```

5. Run `yum install zlib-devel` to install the Memcached source code package (Memcached 2.0 is recommended).

   Note:

   - Before installing Memcached, check if there are any zlib-devel packages to be executed.
   - You must first check whether the Memcached client package (including the source code package) is installed. If the Memcached client package is installed, recompile it to add the -enable-memcached-sasl extension.

```
wget http://pecl.php.net/get/memcached-2.0.0.tgz tar -zxvf
 memcached-2.0.0.tgz cd memcached-2.0.0 phpize (If the system has
 two PHP environments, you must call the command by specifying
 the absolute path /usr/bin/phpize, which is the PHP environment
 path for using ApsaraDB for
                Memcache. Run phpize in the Memcached source
 code directory) ./configure --with-libmemcached-dir=/usr/local/
```

```
libmemcached --enable-memcached-sasl (Pay attention to this
 parameter) make make install
```

6. Add `extension=memcached.so memcached.use_sasl = 1` to the php.ini file. (Run the locate command to find this file, which is in /etc/php.ini for yum installation. If the system has two PHP environments, you must find the PHP environment path for using ApsaraDB for Memcache and modify the php.ini file in this path accordingly.)
7. Run `php –m |grep ,memcached`. If the displayed result includes "memcache", ApsaraDB for Memcache is supported in the environment.
8. Test whether the production environment is successfully deployed by using the test code provided at the end of the page. Replace the address, port, user name, and password in the test code with actual values.

On an Ubuntu Debian operating system

1. Change the Ubuntu source.

   Solution 1: Run `vim /etc/apt/source.list` and add the following content at the beginning of the file:

   ```
   deb http://mirrors.aliyun.com/ubuntu/ precise main restricted
   universe multiverse deb http://mirrors.aliyun.com/ubuntu/ precise
   -security main restricted universe multiverse deb http://mirrors
   .aliyun.com/ubuntu/ precise-updates main restricted universe
   multiverse deb http://mirrors.aliyun.com/ubuntu/ precise-proposed
    main restricted universe multiverse deb http://mirrors.aliyun.
   com/ubuntu/ precise-backports main restricted universe multiverse
    deb-src http://mirrors.aliyun.com/ubuntu/ precise main restricted
    universe multiverse deb-src http://mirrors.aliyun.com/ubuntu/
   precise-security main restricted universe multiverse deb-src http://
   mirrors.aliyun.com/ubuntu/ precise-updates main restricted universe
    multiverse deb-src http://mirrors.aliyun.com/ubuntu/ precise-
   proposed main restricted universe multiverse deb-src http://mirrors
   .aliyun.com/ubuntu/ precise-backports main restricted universe
   multiverse apt-get update //Update the list
   ```

   Solution 2: Download the update_source package at `wget http://oss.aliyuncs.com/aliyunecs/update_source.zip`, decompress the package, run `chmod 777 file name` to grant the file execution permission, and run the script to change the source automatically.

2. Run `ape-get` to configure GCC and G++.

   You must first run dpkg –s installation package name (for example, dpkg –s gcc) to check whether gcc-c++ and other components are installed. If the components are not installed, run `apt-get build-dep gcc apt-get install build-essential`.

3. Install php5 and php5-dev.

You must first run dpkg –s installation package name (for example, dpkg –s php) to check whether PHP and other components are installed. If the components are not installed, run `apt -get install php5 php5-dev`. (php5-cli and php5-common are automatically installed at the same time.)

4. Install and configure SASL support.

   You must first run `dpkg –s installation package name` (for example, `dpkg –s libsasl2`) to check whether libsasl2 cloog-ppl and other components are installed. If they are not installed, run the following command:

   ```
   apt-get install libsasl2-dev cloog-ppl cd /usr/local/src
   ```

5. Run the following command to install Libmemcache of the specified version:

   Note:
   Before running the command, check whether the specified package (including the source code package) is installed. If yes, skip this step.

   ```
   wget https://launchpad.net/libmemcached/1.0/1.0.18/+download/
   libmemcached-1.0.18.tar.gz tar -zxvf libmemcached-1.0.18.tar.gz cd
   libmemcached-1.0.18 ./configure --prefix=/usr/local/libmemcached
   make make install cd ..
   ```

6. Run the following command to install Memcached of the specified version:

   Note:
   Check whether the Memcached client package (including the source code package) has been installed. If yes, no installation is required. However, you need to recompile the package to add the -enable-memcached-sasl extension.

   ```
   wget http://pecl.php.net/get/memcached-2.2.0.tgz tar zxvf memcached
   -2.2.0.tgz cd memcached-2.2.0 phpize5 ./configure --with-libmemcach
   ed-dir=/usr/local/libmemcached --enable-memcached-sasl make make
   install
   ```

7. Configure PHP to support Memcached and run a test.

   ```
   echo "extension=memcached.so" >>/etc/php5/conf.d/pdo.ini echo "
   memcached.use_sasl = 1" >>/etc/php5/conf.d/pdo.ini php -m |grep mem
   memcached
   ```

   If this component is displayed, the installation and configuration are completed.

PHP sample code

Example 1: Connect to ApsaraDB for Memcache and perform the Set and Get operations

```
<?php $connect = new Memcached; //Declare a new Memcached connection
 $connect->setOption(Memcached::OPT_COMPRESSION, false); //Disable
 the compression function $connect->setOption(Memcached::OPT_BINARY
_PROTOCOL, true); //Use the binary protocol $connect->setOption(
Memcached::OPT_TCP_NODELAY, true); //Important: PHP Memcached has
a bug that causes a fixed latency of 40 ms when the Get value does
 not exist. Enabling this parameter can avoid this bug $connect->
addServer(aaaaaaaaaa.m.yyyyyyyyyy.ocs.aliyuncs.com, 11211); //Add
the address and port number of the ApsaraDB for Memcache instance $
connect->setSaslAuthData(aaaaaaaaaa, password); //Set the ApsaraDB for
 Memcache account and password for authentication. Skip this step if
the password-free feature is enabled $connect->set("hello", "world");
echo hello: ,$connect->get("hello"); $connect->quit(); ?>
```

Example 2: Cache an array in ApsaraDB for Memcache

```
<?php $connect= new Memcached; //Declare a new Memcached connection
 $connect->setOption(Memcached::OPT_COMPRESSION, false); //Disable
 the compression function $connect->setOption(Memcached::OPT_BINARY
_PROTOCOL, true);//Use the binary protocol $connect->setOption(
Memcached::OPT_TCP_NODELAY, true); //Important: PHP Memcached has a
bug that causes a fixed latency of 40 ms when the Get value does not
exist. Enabling this parameter can avoid this bug $connect->addServer
(xxxxxxxx.m.yyyyyyyy.ocs.aliyuncs.com, 11211);//Add the address and
port number of the ApsaraDB for Memcache instance $connect->setSaslAut
hData(xxxxxxxx, bbbbbbbb);//Set the ApsaraDB for Memcache account
 and password for authentication. Skip this step if the password-
free feature is enabled $user = array( "name" => "ocs", "age" =>
1, "sex" => "male" ); //Declare an array $expire = 60; //Set the
expiration time test($connect->set(your_name,$user,$expire), true, Set
 cache failed); if($connect->get(your_name)){ $result =$connect->get
(your_name); }else{ echo "Return code:", $connect->getResultCode();
echo "Retucn Message:", $connect->getResultMessage (); //If an error
 is returned, parse the return code $result=" "; } print_r($result);
 $connect->quit(); function test($val, $expect, $msg) { if($val!= $
expect) throw new Exception($msg); } ?>
```

Example 3: Use ApsaraDB for Memcache together with the MySQL database

```
<?php $connect = new Memcached; //Declare a new Memcached connection
 $connect->setOption(Memcached::OPT_COMPRESSION, false);//Disable
the compression function $connect->setOption(Memcached::OPT_BINARY
_PROTOCOL, true);//Use the binary protocol $connect->setOption(
Memcached::OPT_TCP_NODELAY, true); //Important: PHP Memcached has a
bug that causes a fixed latency of 40 ms when the Get value does not
exist. Enabling this parameter can avoid this bug $connect->addServer(
xxxxxx.m.yyyyyyyy.ocs.aliyuncs.com, 11211);//Add the instance address
and port number $connect->setSaslAuthData(xxxxxx, my_passwd);//Set the
 ApsaraDB for Memcache account and password for authentication. Skip
 this step if the password-free feature is enabled $user = array( "
name" => "ocs", "age" => 1, "sex" => "male" ); //Define an array if($
connect->get(your_name)) { $result =$connect->get(your_name); print_r
($result); echo "Found in OCS, get data from OCS"; //If the value is
 obtained, print the value source as ApsaraDB for Memcache exit; }
else { echo "Return code:", $connect->getResultCode(); echo "Retucn
```

```
 Message:", $connect->getResultMessage ();//Throw the return code $
db_host=zzzzzz.mysql.rds.aliyuncs.com; //Database address $db_name=
my_db; //database name $db_username=db_user; //Database user name $
db_password=db_passwd;//Database password $connection=mysql_connect($
db_host,$db_username,$db_password); if (!mysql_select_db($db_name, $
connection)) { echo Could not select database; //An error is thrown
if database connection fails exit; } $sql = "SELECT name,age,sex FROM
 test1 WHERE name = ocs"; $result = mysql_query($sql, $connection);
while ($row = mysql_fetch_assoc($result)) { $user = array( "name" =>
 $row["name"], "age" => $row["age"], "sex" => $row["sex"], ); $expire
 = 5; //Set the value expiration time in the cache test($connect->
set(your_name,$user,$expire), true, Set cache failed); //Write the
value to the ApsaraDB for Memcache cache } mysql_free_result($result
); mysql_close($connection); } print_r($connect->get(your_name)); //
Print the value obtained echo "Not Found in OCS,get data from MySQL
"; //Confirm the value obtained from the database $connect->quit();
function test($val, $expect, $msg) { if($val!= $expect) throw new
Exception($msg); } ?>
```

18.2.4.4 Python

Download a client

Client download URL

About the client

Client version

Environment configuration

Depends on Bmemcached (SASL extensions supported). To download Bmemcached, click *Here*.

Python sample code

```
#!/usr/bin/env python import bmemcached client = bmemcached.Client((ip
:port), user, passwd) print client.set(key, value11111111111) print
client.get(key)
```

18.2.4.5 C#/.NET: EnyimMemcached

Download a client

Client download URL

About the client

Client version

C#/.NET sample code

```
using System.Net; using Enyim.Caching; using Enyim.Caching.Configurat
ion; using Enyim.Caching.Memcached; namespace OCS.Memcached { public
 sealed class MemCached { private static MemcachedClient MemClient
; static readonly object padlock = new object(); //Thread-safe
single instance mode public static MemcachedClient getInstance() {
```

```
if (MemClient == null) { lock (padlock) { if (MemClient == null) {
MemClientInit(); } } } return MemClient; } static void MemClientInit
() { //Initialize the cache MemcachedClientConfiguration memConfig =
new MemcachedClientConfiguration(); IPAddress newaddress = IPAddress
.Parse(Dns.GetHostEntry ("your_ocs_host").AddressList[0].ToString
()) ; //Replace your_ocs_host with the ApsaraDB for Memcache intranet
address IPEndPoint ipEndPoint = new IPEndPoint(newaddress, 11211); //
Configuration file - IP address memConfig.Servers.Add(ipEndPoint); //
Configuration file - protocol memConfig.Protocol = MemcachedProtocol
.Binary; // Configuration file - permission memConfig.Authentication
.Type = typeof(PlainTextAuthenticator); memConfig.Authentication.
Parameters["zone"] = ""; memConfig.Authentication.Parameters["userName
"] = "username"; memConfig.Authentication.Parameters["password"] =
 "password"; //Complete the following settings based on the maximum
 connections of the instance memConfig.SocketPool.MinPoolSize = 5;
memConfig.SocketPool.MaxPoolSize = 200; MemClient=new MemcachedClient(
memConfig); } } }
```

Dependency

Code:

```
MemcachedClient MemClient = MemCached.getInstance();
```

18.2.4.6 C

Download a client

Client download URL

About the client

Client version

Environment configuration

1. Download, compile, and install the C++ client.

   https://launchpad.net/libmemcached/1.0/1.0.18/+download/libmemcached-1.0.18.tar.gz

2. Run the following command:

   ```
   tar -xvf libmemcached-1.0.18.tar.gz
   ```

   ```
   cd libmemcached-1.0.18
   ```

   ```
   ./configure
   ```

   ```
   sudo make install
   ```

C++ sample code

1. Download *ocs_test.tar.gz*.
2. Run the following command:

```
tar -xvf ocs_test.tar.gz

cd ocs_test

vim ocs_test_sample1.cpp
```

3. Set TARGET_HOST to the internal network address of the ApsaraDB for Memcache instance, USERNAME to the user name of your instance, and PASSWORD to the password you set.
4. Run `build.sh` to generate ocs_test. Run `./ocs_test`. A key is written to the ApsaraDB for Memcache instance. Get the key from the ApsaraDB for Memcache instance and delete it from the instance.

   The code of ocs_test_sample1.cpp is as follows:

```
#include <iostream> #include <string> #include <libmemcached/
memcached.h> using namespace std; #define TARGET_HOST "" #define
USERNAME "" #define PASSWORD "" int main(int argc, char *argv[]) {
memcached_st *memc = NULL; memcached_return rc; memcached_server_st
 *server; memc = memcached_create(NULL); server = memcached_
server_list_append(NULL, TARGET_HOST, 11211,&rc); /* SASL */
sasl_client_init(NULL); rc = memcached_set_sasl_auth_data(memc
, USERNAME, PASSWORD); if(rc != MEMCACHED_SUCCESS) { cout<<"Set
SASL err:"<< endl; } rc = memcached_behavior_set(memc,MEMCACHED_
BEHAVIOR_BINARY_PROTOCOL,1); if(rc != MEMCACHED_SUCCESS) { cout<<"
Binary Set err:"<<endl; } /* SASL */ rc = memcached_server_push(
memc,server); if(rc != MEMCACHED_SUCCESS) { cout <<"Connect Mem err
:"<< rc << endl; } memcached_server_list_free(server); string key =
 "TestKey"; string value = "TestValue"; size_t value_length = value
.length(); size_t key_length = key.length(); int expiration = 0;
uint32_t flags = 0; //Save data rc = memcached_set(memc,key.c_str(),
key.length(),value.c_str(),value.length(),expiration,flags); if (rc
 != MEMCACHED_SUCCESS){ cout <<"Save data failed: " << rc << endl;
return -1; } cout <<"Save data succeed, key: " << key << " value: "
 << value << endl; cout << "Start get key:" << key << endl; char*
result = memcached_get(memc,key.c_str(),key_length,&value_length,&
flags,&rc); cout << "Get value:" << result << endl; //Delete data
cout << "Start delete key:" << key << endl; rc = memcached_delete(
memc,key.c_str(),key_length,expiration); if (rc != MEMCACHED_SUCCESS
) { cout << "Delete key failed: " << rc << endl; } cout << "Delete
key succeed: " << rc << endl; //free memcached_free(memc); return 0
; }
```

The following is a sample code of using ApsaraDB for Memcache with another C++ program, where the ApsaraDB for Memcache cache and MySQL database are combined. You can follow the steps in the preceding example to compile and install the C++ client.

1. Create the sample database and table in the MySQL database.

```
mysql -h host -uUSER -pPASSSWORD

create database testdb;
```

```
create table user_info (user_id int, user_name char(32) not null,
password char(32) not null, is_online int, primary key(user_id) );
```

2. Download *ocs_test_2.tar.gz* and run the following command:

```
tar -xvf ocs_test_2.tar.gz

cd ocs_test

vim ocs_test_sample2.cpp
```

> **Note:**
> Set OCS_TARGET_HOST to the internal network address of the ApsaraDB for Memcache instance, OCS_USERNAME to the ApsaraDB for Memcache instance name, OCS_PASSWORD to the password you set, MYSQL_HOST to the MySQL database address, MYSQL_USERNAME to the database user name, and MYSQL_PASSWORD to the database password.

3. Run `build.sh` to generate ocs_test and run `/ocs_test`.

   The ocs_test_sample2.cpp code is as follows:

```
#include <iostream> #include <string> #include <sstream> #include <
libmemcached/memcached.h> #include <mysql/mysql.h> using namespace
 std; #define OCS_TARGET_HOST "xxxxxxxxxx.m.yyyyyyyyy.ocs.aliyuncs
.com" #define OCS_USERNAME "your_user_name" #define OCS_PASSWORD
 "your_password" #define MYSQL_HOST "zzzzzzzzzz.mysql.rds.aliyuncs
.com" #define MYSQL_USERNAME "db_user" #define MYSQL_PASSWORD "
db_paswd" #define MYSQL_DBNAME "testdb" #define TEST_USER_ID "100
" MYSQL *mysql = NULL; memcached_st *memc = NULL; memcached_return
rc; int InitMysql() { mysql = mysql_init(0); if (mysql_real_connect
(mysql, MYSQL_HOST, MYSQL_USERNAME, MYSQL_PASSWORD, MYSQL_DBNA
ME, MYSQL_PORT, NULL, CLIENT_FOUND_ROWS) == NULL ) { cout << "
connect mysql failure!" << endl; return EXIT_FAILURE; } cout << "
connect mysql success!" << endl; return 0; } bool InitMemcached() {
memcached_server_st *server; memc = memcached_create(NULL); server
 = memcached_server_list_append(NULL, OCS_TARGET_HOST, 11211,&rc); /
* SASL */ sasl_client_init(NULL); rc = memcached_set_sasl_auth_data
(memc, OCS_USERNAME, OCS_PASSWORD); if (rc != MEMCACHED_SUCCESS)
 { cout<<"Set SASL err:"<< endl; return false; } rc = memcached_
behavior_set(memc,MEMCACHED_BEHAVIOR_BINARY_PROTOCOL,1); if (rc !
= MEMCACHED_SUCCESS) { cout<<"Binary Set err:"<<endl; return false
; } /* SASL */ rc = memcached_server_push(memc,server); if (rc !
= MEMCACHED_SUCCESS) { cout <<"Connect Mem err:"<< rc << endl;
return false; } memcached_server_list_free(server); return true; }
struct UserInfo { int user_id; char user_name[32]; char password[
32]; int is_online; }; bool SaveToCache(string &key, string &value
, int expiration) { size_t value_length = value.length(); size_t
 key_length = key.length(); uint32_t flags = 0; //Save data rc =
memcached_set( memc,key.c_str(), key.length(), value.c_str(), value.
length(), expiration, flags); if (rc != MEMCACHED_SUCCESS){ cout <<"
Save data to cache failed: " << rc << endl; return false; } cout <<"
Save data to cache succeed, key: " << key << " value: " << value <<
```

```
endl; return true; } UserInfo *GetUserInfo(int user_id) { UserInfo
 *user_info = NULL; //get from cache string key; stringstream out
; out << user_id; key = out.str(); cout << "Start get key:" <<
key << endl; size_t value_length; uint32_t flags; char* result =
memcached_get(memc, key.c_str(), key.size(), &value_length, &flags
, &rc); if (rc != MEMCACHED_SUCCESS) { cout << "Get Cache Failed
, start get from mysql."<< endl; int status; char select_sql[1024
]; memset(select_sql, 0x0, sizeof(select_sql)); sprintf(select_sql
, "select * from user_info where user_id = %d", user_id); status
 = mysql_query(mysql, select_sql); if (status !=0 ) { cout << "
query from mysql failure!" << endl; return NULL; } cout << "the
status is :" << status << endl; MYSQL_RES *mysql_result = mysql_stor
e_result(mysql); user_info = new UserInfo; MYSQL_ROW row; while (
row = mysql_fetch_row(mysql_result)) { user_info->user_id = atoi
(row[0]); strncpy(user_info->user_name, row[1], strlen(row[1]));
strncpy(user_info->password, row[2], strlen(row[2])); user_info->
is_online = atoi(row[3]); } mysql_free_result(mysql_result); return
 user_info; } cout << "Get from cache succeed" << endl; user_info
 = new UserInfo; memcpy(user_info, result, value_length); return
 user_info; } bool DeleteCache(string &key, int expiration) { rc
 = memcached_delete(memc, key.c_str(), key.length(), expiration);
if (rc != MEMCACHED_SUCCESS) { cout << "Delete key failed: " <<
rc << endl; return false; } cout << "Delete key succeed: " << rc
 << endl; return true; } void PrintUserInfo(UserInfo *user_info
) { cout << "user_id: " << user_info->user_id << " " << " name:
 " << user_info->user_name << endl; } bool SaveMysql(UserInfo *
user_info) { char insert_sql[1024]; memset(insert_sql, 0x0, sizeof
(insert_sql)); sprintf(insert_sql, "insert into user_info(user_id,
user_name, password, is_online) values(%d, %s, %s, %d)", user_info
->user_id, user_info->user_name, user_info->password, user_info->
is_online); int status = mysql_query(mysql, insert_sql); if (status
 != 0) { cout << "insert failed" << endl; return false; } cout <<
 "insert user_info" << endl; //insert mysql return true; } int main
(int argc, char *argv[]) { if (InitMysql() != 0) { return -1; } if
 (!InitMemcached()) { return -1; } //generate user_info UserInfo
user_info; user_info.user_id = atoi(TEST_USER_ID); strcpy(user_info
.user_name, "James"); strcpy(user_info.password, "12345678");
user_info.is_online = 1; //save to mysql if (!SaveMysql(&user_info))
 { //return -1; } string user_str; user_str.assign((char*)&user_info
, sizeof(UserInfo)); //save to memcached string key_str = TEST_USER_
ID; SaveToCache(key_str, user_str, 10); //start get, exist in
memcahced UserInfo *get_user_info = GetUserInfo(user_info.user_id);
PrintUserInfo(get_user_info); //wait 10 secons sleep(2); //delete
memcached or expired DeleteCache(key_str, 0); //start get, exist in
 mysql delete get_user_info; get_user_info = GetUserInfo(user_info.
user_id); PrintUserInfo(get_user_info); delete get_user_info; //free
 memcached_free(memc); mysql_close(mysql); return 0; }
```

18.2.5 Connect to an instance over the Internet

ECS Windows

Currently, ApsaraDB for Memcache is accessible through the ECS internal network. To locally access ApsaraDB for Memcache through the Internet, create a port mapping through netsh on the ECS Windows server.

1. Log on to the ECS Windows server and run the following command in CMD:

`netsh interface portproxy add v4tov4 listenaddress=`*`IP address of the ECS instance`* `listenport=11211 connectaddress=`*`ApsaraDB for Memcache connection address`* `connectport=11211`

Note:

- `netsh interface portproxy delete v4tov4 listenaddress=`*`IP address of the ECS instance`* `listenport=11211 //`You can delete unnecessary mappings.
- `netsh interface portproxy show all //`You can check the mappings on the current server.

2. Perform a verification test after the configuration is complete.

   Connect to the ECS Windows server locally over Telnet, write data, and perform query and verification. For example, if the IP address of the ECS Windows server is 1.1.1.1, then `telnet 1.1.1.1 11211`.

After performing the preceding steps, you can use a local PC or server to connect to port 11211 of the ECS Windows server over the Internet and access ApsaraDB for Memcache.

Note:
Portproxy is provided by Microsoft rather than open source software. See the netsh official documentation on portproxy or consult with Microsoft engineers if you have any problems in configuration and use. Alternatively, use another scheme. For example, use portmap to configure proxy mappings.

ECS Linux

Currently, ApsaraDB for Memcache is accessible through the ECS internal network. To locally access ApsaraDB for Memcache through the Internet, install rinetd on the ECS Linux server for forwarding.

1. Install rinetd on the ECS Linux server.

   ```
   wget http://www.boutell.com/rinetd/http/rinetd.tar.gz&&tar -xvf rinetd.tar.gz&&cd rinetd
   ```

   ```
   sed -i s/65536/65535/g rinetd.c (Modify the port range. Otherwise, an error will be reported.)
   ```

   ```
   mkdir /usr/man&&make&&make install
   ```

Note:

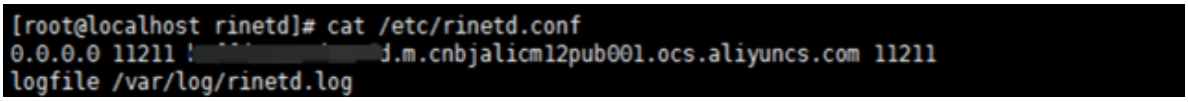
The rinetd installation package obtained from the download URL may be unavailable. You can download the rinetd installation package from other sources.

2. Create the configuration file.

   ```
   vi /etc/rinetd.conf
   ```

3. Enter the following information:

   ```
   0.0.0.0 11211 Connection address of the ApsaraDB for Memcache instance 11211 logfile /var/log/rinetd.log
   ```



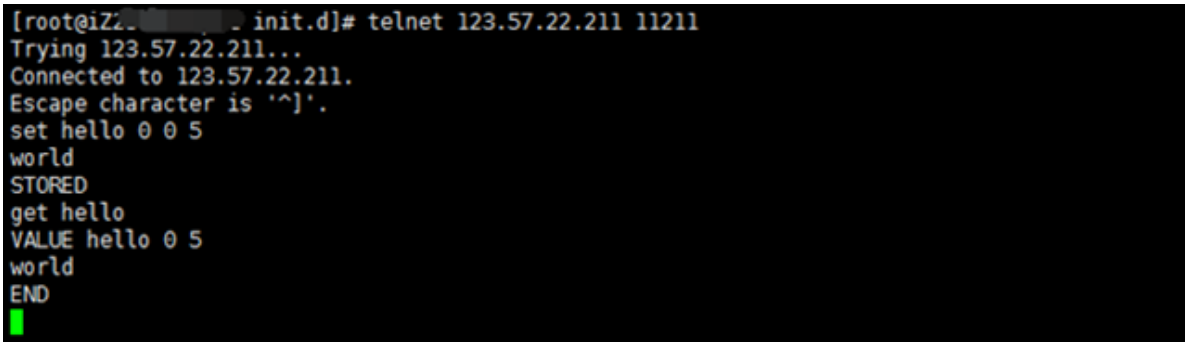
4. Run `rinetd` to start rinetd.

Note:

Run `echo rinetd >>/etc/rc.local` to make rinetd autorun. Run `pkill rinetd` to kill the rinetd process.

5. Perform a verification test.

   Connect to the ECS Linux server locally over Telnet, write data, and perform query and verification. For example, if the IP address of the server with rinetd installed is 1.1.1.1, then `telnet 1.1.1.1 11211`.



After performing the preceding steps, you can use a local PC or server to connect to port 11211 of the ECS Linux server through the Internet and access ApsaraDB for Memcache.

Note:

Rinetd is open source software. Read the rinetd official documentation or consult with rinetd engineers if you have any problems.

18.3 Manage instances

18.3.1 Modify the password of an instance

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click **Reset Password** in the upper right corner.
4. On the **Reset Password** page, enter a new password and click **Submit**.

18.3.2 View details of an instance

After creating an instance, you can view details of the instance on Apsara Stack console.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details**. On the **Instance Information** page, view the instance details.

   The **Instance Information** page displays **Basic Information**, **Specifications**, and **Connection Information**. The following table lists information in each area.

   Table 18-2: Instance Information

Region	Item
Basic Information	• Instance ID • Name • Status • Region • Department • Project • Created At: • Zone • Network Type
Specifications	• Instance Specifications • Maximum Connections • Maximum Internal Network Bandwidth

Region	Item
	• Maintenance Time • Whitelist
Connection Information	• Connection Address • Port Number

18.3.3 Modify an instance name

After creating an instance, you can modify the instance name on Apsara Stack console.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, locate the target instance and choose > **Edit**.
3. In the **Edit Instance Information** window, set **Name** for the instance and click **Confirm**.

18.3.4 Modify configurations

ApsaraDB for MemcacheApsaraDB for MemcacheAllows you to modify configurations.

Context

Note:
The instance experiences intermittent interruption for several seconds during configuration modification. Perform upgrade during off-peak hours if possible.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, locate the target instance and click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click **Change Instance** in the upper right corner. In the **Change Instance** window, set **Instance Specifications** and click **Confirm**.

   You will be notified after the configuration is successfully changed.

18.3.5 Set the maintenance period

Context

To ensure stability of ApsaraDB for Memcache instances on the Alibaba Cloud platform, the backend system irregularly maintains instances and machines.

Before the official maintenance, ApsaraDB for Memcache sends SMS messages and emails to contacts configured in your Alibaba Cloud account.

To guarantee the stability of the maintenance process, instances will enter the **Being Maintained** state before the preset O&M time on the day of maintenance. When an instance is in this state, access to data in the database is not affected. However, modification-related functions (for example, configuration modification) are temporarily unavailable for this instance on the console, whereas query functions such as performance monitoring are still available.

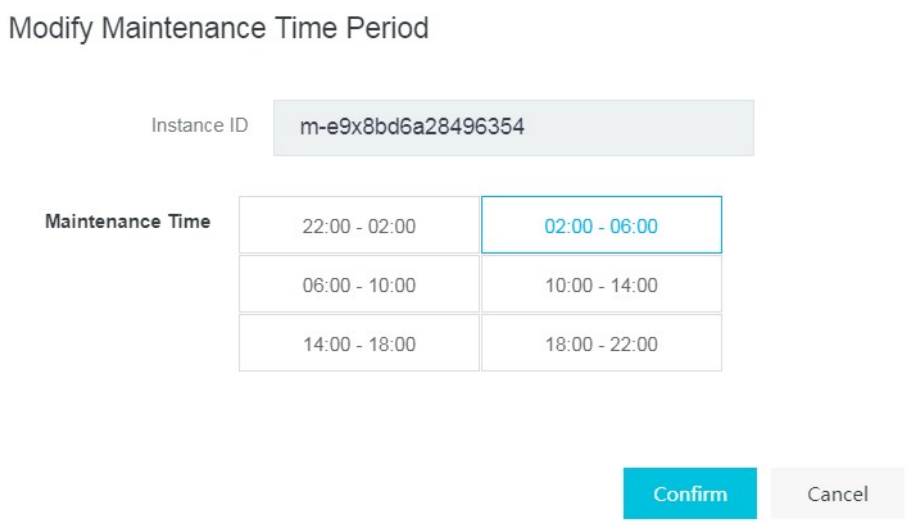
Note:
After the preset O&M time is reached, instances may be interrupted intermittently during maintenance. We recommend that you maintain instances during off-peak hours.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, click the **Instance ID** or choose > **Details** to go to the **Instance Information** page.
3. Click **Modify Maintenance Time Period** in the upper right corner. The page shown in the following figure is displayed.

   The default maintenance period for the ApsaraDB for Memcache instance is from 02:00 to 06:00.

   Figure 18-2: Set the maintenance time period



4. Select a maintenance time period and click **OK**.

The maintenance time period is in Beijing time.

18.3.6 Clear an instance

Context

> **Note:**
> Be cautious when clearing the instance as this operation erases all data of the instance and the erased data cannot be recovered.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, locate the target instance and click > **Clear**.
3. In the **Clear Instance** window, click **Confirm**.

18.3.7 Release an instance

You can release your ApsaraDB for Memcache instance as needed.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, locate the target instance and choose > **Release**.
3. In the **Delete Instance** window, click **Confirm**.

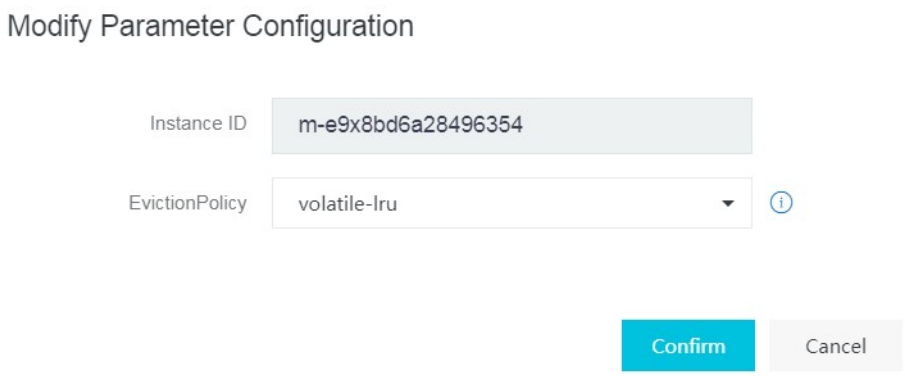
18.4 Set parameters

ApsaraDB for MemcacheSupports six data eviction policies. You can set the maxmemory-policy parameter on the console to configure a data eviction policy as needed.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click the **Parameters** tab.
4. Choose **Action** > **Edit** next to the parameter **`EvictionPolicy`**.
5. Select a data eviction policy and click **Confirm**.

Figure 18-3: Set parameters



18.5 Backup and recovery

18.5.1 Automatic backup (backup policy setting)

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, click the **Instance ID** or choose > **Details** to go to the **Instance Information** page.
3. Click the **Backup and Restore** tab.
4. Click **Backup Settings**.
5. Click **Set** and set **Backup Period** and **Backup Time**.

   Backup data is retained for seven days by default. This setting cannot be modified.
6. Click **Confirm** to complete automatic backup setting.

18.5.2 Manual backup (instant backup)

In addition to the general backup settings, you can initiate a manual backup request on the console at any time.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, click the **Instance ID** or choose > **Details** to go to the **Instance Information** page.
3. Click the **Backup and Restore** tab.
4. Click **Create Backup** in the upper right corner.

5. Click **Confirm** to instantly back up the instance.

   On the **Backup Data** page, you can select time ranges and query historical backup data. By default, backup data is retained for seven days. You can query historical backup data of the past seven days.

18.5.3 Restore data

The data restore function minimizes the damage caused by database misoperation. Currently, ApsaraDB for Memcache supports data restore from backup.

Procedure

1. *Log on to the ApsaraDB for Memcache console*.
2. In the instance list, click the instance ID or choose > **Details** to go to the **Instance Information** page.
3. Click the **Backup and Restore**tab.
4. On the **Backup and Restore** page, click the **Backup Data** tab.
5. Locate the backup file to be recovered and choose **Action** > **Data Restore**.
6. In the **Data Restore**window, click **Confirm** to restore the data of the original instance.

   As the data restore operation is highly risky, we recommend that you clone the instance if time permits. Create an instance based on the backup dataset to be restored and perform data restore after verifying that the data is correct.

18.6 Supported protocols and commands

Any clients compatible with the Memcached protocol can access ApsaraDB for Memcache. You can select any Memcached client supporting SASL or Memcached Binary Protocol based on the application features.

Protocol

- Memcached Binary Protocol (binary)
- SASL authentication protocol

Operation

ApsaraDB for Memcache supports the following command operations.

Operation code	Operation command	Remarks
0x00	Get	-

Operation code	Operation command	Remarks
0x01	Set	-
0x02	Add	-
0x03	Replace	-
0x04	Delete	-
0x05	Increment	-
0x06	Decrement	-
0x07	Quit	-
0x08	Flush	ApsaraDB for Memcache supports the second-level time accuracy.
0x09	GetQ	-
0x0a	No-op	-
0x0b	Version	-
0x0c	GetK	-
0x0d	GetKQ	-
0x0e	Append	-
0x0f	Prepend	-
0x10	Stat	Not supported
0x11	SetQ	-
0x12	AddQ	-
0x13	ReplaceQ	-
0x14	DeleteQ	-
0x15	IncrementQ	-
0x16	DecrementQ	-
0x17	QuitQ	-
0x18	FlushQ	-
0x19	AppendQ	-
0x1a	PrependQ	-
0x1b	Verbosity	Not supported
0x1c	Touch	-

Operation code	Operation command	Remarks
0x1d	GAT	-
0x1e	GATQ	-
0x20	SASL list mechs	-
0x21	SASL Auth	-
0x22	SASL Auth	-

18.7 Restrictions

Project	Restrictions
Data type	ApsaraDB for Memcache only supports the Key-Value data format. Complex data types such as Array, Map, and List are not supported.
Data reliability	ApsaraDB for Memcache stores data in the memory. This service does not guarantee that the cached data will not be lost. ApsaraDB for Memcache is not suitable for storing data that requires high consistency.
Data volume	ApsaraDB for Memcache supports a maximum of 1 KB in key size and 1 MB in value size for a single piece of cached data. ApsaraDB for Memcache is not suitable for storing oversized data.
Transaction support	ApsaraDB for Memcache does not support transactions. ApsaraDB for Memcache is not suitable for storing data with transaction requirements. Data with transaction requirements needs to be directly written into the database.
Scenarios	When data access traffic is evenly distributed and there is no obvious hotspot or less popular data, many access requests cannot hit the cached data in ApsaraDB for Memcache. Therefore, the effect of ApsaraDB for Memcache as a database cache is not significant. You must consider the data access requirements of the business model when selecting a database cache.
Data deletion policy	ApsaraDB for Memcache expiration mechanism: Each key expires at a custom time. After expiration, the key becomes inaccessible. The space occupied by the expired key is not recycled immediately after expiration, but is recycled at 02:00 Beijing Time (UTC+8) every day.

Project	Restrictions
Data expiration policy	Like open source Memcached, ApsaraDB for Memcache adopts the LRU algorithm to determine whether the data expires. Expired data is not deleted and the space occupied by the expired data is not recycled immediately after expiration. The space is recycled by a background program periodically.
Connection processing	ApsaraDB for Memcache server does not proactively close client connections in idle state.
Data expiration	We recommend that you control and manage the key expiration time.

19 ApsaraDB for MongoDB

19.1 What is ApsaraDB for MongoDB

ApsaraDB for MongoDB is fully compatible with the MongoDB protocol and provides stable, reliable, and automatically scalable database services. It offers you a full range of database solutions including disaster recovery, backup, restore, monitoring, and alarms.

ApsaraDB for MongoDB offers the following basic features:

- Automatically creates a three-node MongoDB replica set, Supports advanced functions such as disaster recovery, switchover and failover. All functions are completely transparent to the users .
- Supports quick database backup and restore. Allows users to perform quick conventional database backup and rollback.
- Provides more than 20 performance metrics for monitoring and alarm functions. Gives you a full view of database performance.
- Provides visual data management tools to facilitate your operation and maintenance work.

19.2 Restrictions

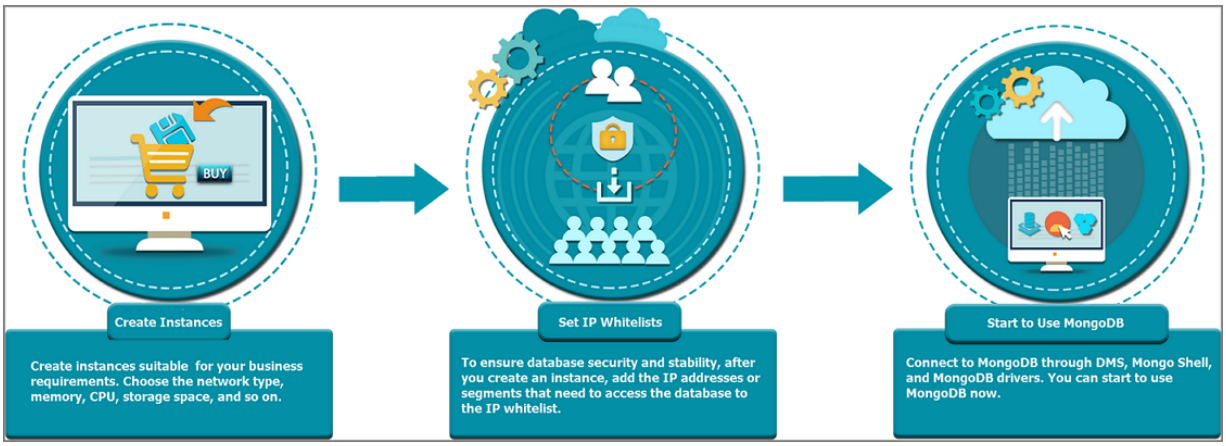
You can easily migrate self-built MongoDB databases to ApsaraDB for MongoDB instances. However, note that ApsaraDB for MongoDB has certain restrictions.

Table 19-1: Restrictions of ApsaraDB for MongoDB

Operation	Restriction
Copy database	• The system automatically creates a three-node replica set. • Of the three nodes, the primary and secondary nodes are provided for users, whereas the standby node is invisible to users. • Currently, the secondary nodes cannot be created manually.
Restart database	Instances must be restarted on the console.

19.3 Procedure

Before you use Alibaba Cloud ApsaraDB for MongoDB for the first time, read *Restrictions*. Before using the new instance that you bought, you must complete the following operations:



19.4 Quick start

19.4.1 Log on to the ApsaraDB for MongoDB console

Take the Chrome browser as an example to describe how to log on to the ApsaraDB for MongoDB console through the Apsara Stack console as cloud product users.

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

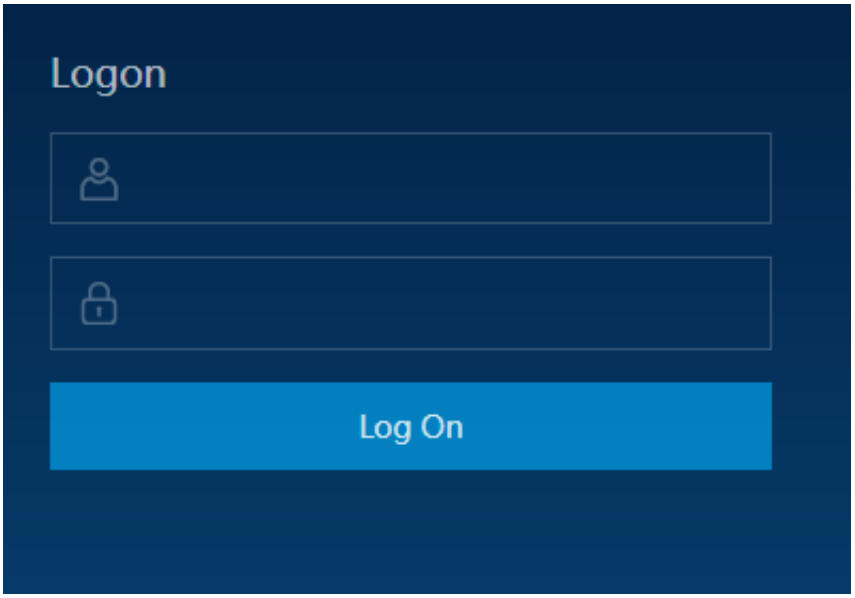
  The access address of the Apsara Stack console is http://`x.x.x.x`/manage, where `x.x.x.x` represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*`IP address or domain name address of the Apsara Stack console`*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 19-1: Log on to the Apsara Stack console*.

Figure 19-1: Log on to the Apsara Stack console



3. Enter the correct username and password.
   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **Log On** to go to the **Home** page.
5. In the menu bar, choose **Console** > **Database** > **ApsaraDB for MongoDB**.

19.4.2 Create an instance

Prerequisites

Make sure that you have the account required for logging on to Apsara Stack console.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click **Create Instance** in the upper right corner of the page to go to the **MongoDB Instance Creation** page.

3. Set **`Configurations`**, **`Network Type`**, **`Specification Configuration`**, and other parameters of the instance, as described in *Table 19-2: New instance configurations*.

Table 19-2: New instance configurations

Configurations	Description
Department	Select a department for the instance.
Project	Select a project for the instance.
Region	Select a region for the instance. Products from different network regions are not interoperable. Changing region is not supported after choosing this.
Zone	Select the zone where the project for the instance is located.
Network Type	Virtual Private Cloud (VPC): VPC helps you build an isolated network environment in Alibaba Cloud. The user can customize the route table , IP address range, and gateway in VPC.
Specification Configuration	Select node specifications and storage space.
Password Settings	Set the password used for the first database logon. You can also select **Configure After Creation** and set the password later by referring to *Reset Password*.
Instance Name	Enter a name for the instance.

4. Click **Create** to create the instance.

19.4.3 Set a whitelist

To guarantee database security and stability, you need to add IP addresses or IP address segments used for database access to the whitelist of the target instance before using an ApsaraDB for MongoDB instance. Proper use of the whitelist provides ApsaraDB for MongoDB with high-level access protection. We recommend that you regularly maintain the whitelist. This document describes how to set a whitelist.

Context

Note:

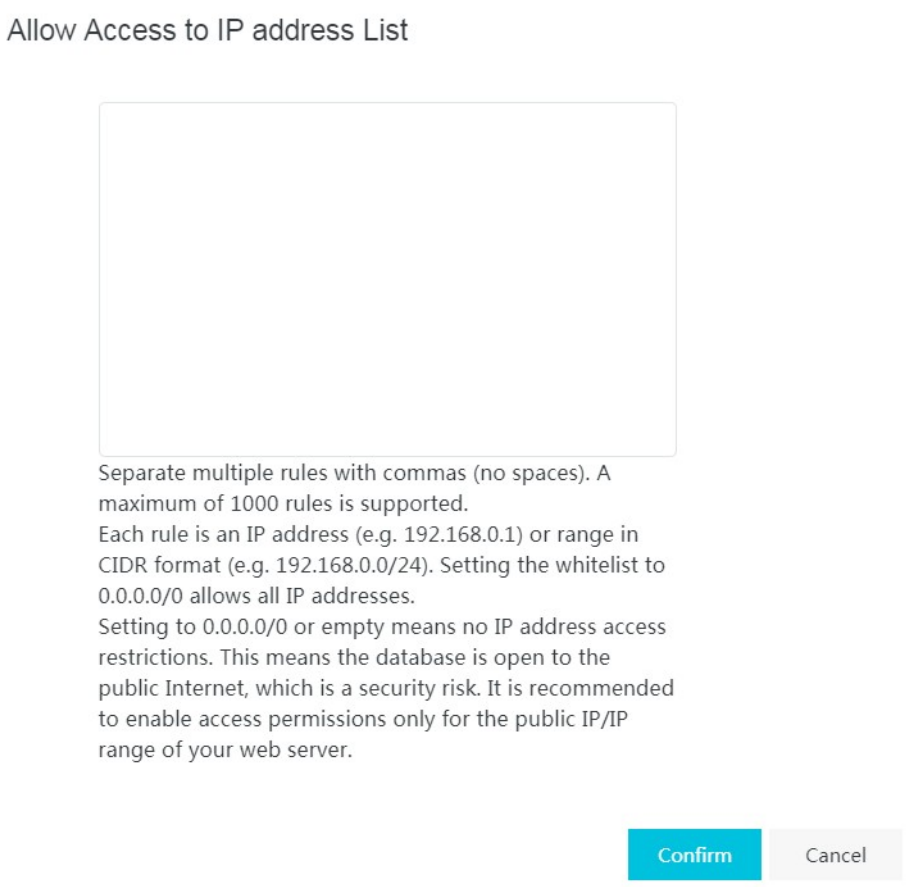
- The system creates a **default** whitelist group for each instance. This whitelist group can be modified but cannot be deleted.

- After you create an ApsaraDB for MongoDB instance, the IP address 0.0.0.0/0 is automatically added to the **default** whitelist group, which allows access to the ApsaraDB for MongoDB instance using any IP address. This configuration may greatly reduce database security, and is not recommended unless necessary. Therefore, when configuring the whitelist, delete 0.0.0.0/0 and add IP addresses or IP address segments allowed to access the ApsaraDB for MongoDB instance.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, choose **Security Control** > **Whitelist Settings** to go to the **Whitelist** page.
4. Click **Modify Whitelist** on the right. The **Allow Access to IP List** dialog box is displayed, as shown in *Figure 19-2: Allow Access to IP List*.

Figure 19-2: Allow Access to IP List



5. Add IP addresses or IP address segments used for accessing the ApsaraDB for MongoDB instance to the whitelist and click **Confirm**.

19.4.4 Obtain the seven elements required to connect to an instance

Context

The initial ApsaraDB for MongoDB three-node replica set provides the connection addresses of two data nodes to be used for access. Before connecting to an ApsaraDB for MongoDB instance, obtain the following elements:

- Instance user name
- Password
- Replica set name
- Domain name addresses and port numbers of the two nodes

This document describes how to obtain the seven elements required to connect to an instance.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, select **Database Connection** to go to the **Network Information** page. This page displays the six elements except for the logon password, as shown in *Figure 19-3: Network Information*.

Figure 19-3: Network Information

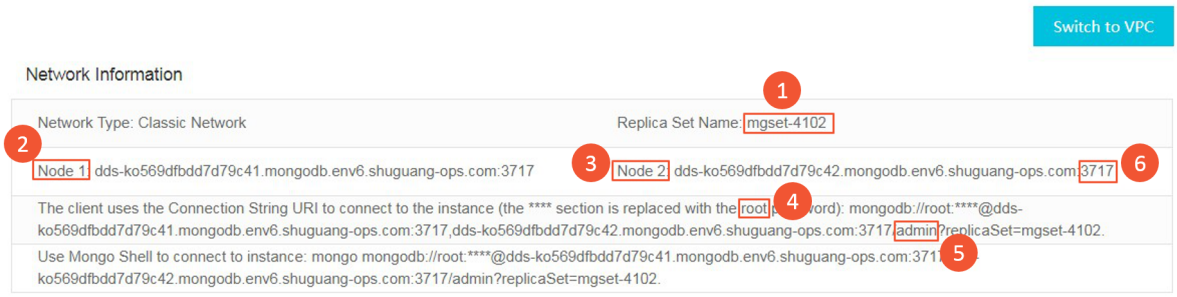


Table 19-3: Six elements required to connect to an instance

Element	Description
Replica Set Name	See 1 in the figure.

Element	Description
Name of node 1	See 2 in the figure.
Name of node 2	See 3 in the figure.
Default account for initial database logon: root	See 4 in the figure.
Name of the default database: admin	See 5 in the figure.
Port to connect to the database: 3717	See 6 in the figure.

The password to connect to the database is the password you set when you created the instance. You can change the password by referring to *Reset a password*.

19.4.5 Connect to Mongo shell

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, select **Database Connection** to go to the **Network Information** page. This page displays the six elements except for the logon password.

   For more information about how to obtain the elements, see:*Obtain the seven elements required to connect to an instance*.
4. In the ECS instance, run mongo to establish a connection. A command example is as follows:

```
mongo --host dds-xxxx.mongodb.rds.aliyuncs.com:3717 -u root -p
123456 --authenticationDatabase admin
```

19.5 Manage instances

19.5.1 Query details

You can query the details of an instance to get its basic information, running status, and configurations. This chapter describes how to query the details of an instance.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. You can go to the Instance Details page in either of the following ways:
   - Click the ID of the target instance to go to the **Basic Information** page.
   - In the Action column of the target instance, choose > **View Details**.

19.5.2 Restart an instance

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. In the Action column of the target instance, choose  > **Restart Instance**.

19.5.3 Modify configurations

You can modify the configurations of your instance, such as memory and storage space, if the configurations are too high or cannot meet the performance requirement of an application.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. In the Action column of the target instance, choose  > **Change Configurations**. The **Change MongoDB Specifications** window is displayed.
3. On the configuration page, select the required configurations.
4. Click **OK**.

19.5.4 Release an instance

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. In the Action column of the target instance, choose  > **Delete Instance**.
3. Click **OK**.

19.5.5 Edit an instance name

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. In the Action column of the target instance, choose  > **Edit Instance Name**. The **Edit Instance Name** window is displayed.
3. In the **`Instance Name`** text box, enter a new name for the instance.
4. Click **Confirm**.

19.6 Switch to VPC

Context

ApsaraDB supports both the classic network and Virtual Private Cloud (VPC). On Alibaba Cloud platform, a classic network and a VPC instance are different in the following aspects:

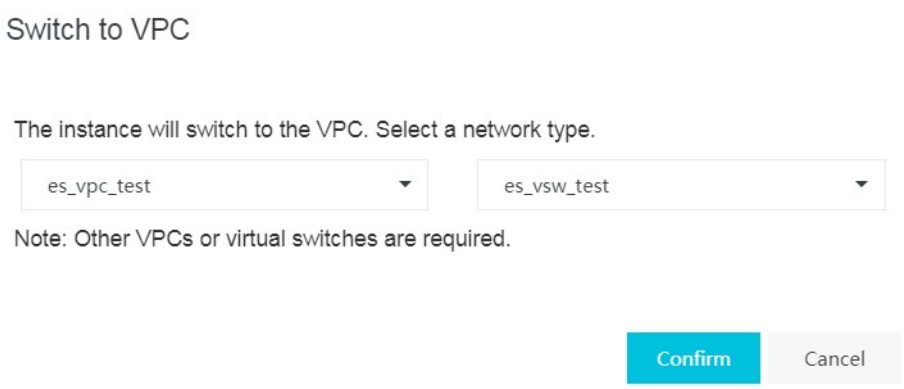
- Classic network: The cloud services in a classic network are not isolated, and unauthorized access to a cloud service is blocked only by the security group or whitelist policy of the service.
- Virtual Private Cloud (VPC): VPC helps you build an isolated network environment in Alibaba Cloud. You can customize the route table, IP address range, and gateway in VPC. In addition, you can combine your data center and cloud resources in Alibaba Cloud VPC into a virtual data center through a leased line or VPN to migrate applications to the cloud smoothly.

To use VPC to create an ApsaraDB for MongoDB instance, make sure that the ApsaraDB for MongoDB and VPC instances are in the same region.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, select **Database Connection**.
4. Click **Switch to VPC**. The *Switch to VPC* page is displayed, as shown in the following figure.

   Figure 19-4: Switch to VPC



5. Select the expected VPC type and switch, and click **Confirm** to create a VPC instance.

19.7 Reset a password

If you forget the password of your database account when using ApsaraDB for MongoDB, you can reset the password on the ApsaraDB for MongoDB console.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. On the **Basic Information** page, click **Reset Password** in the upper right corner. The **Reset Password** window is displayed.
4. Enter a new logon password.
5. Click **Submit**.

> **Note:**
> To ensure the security of your account, we recommend that you change your password every three months.

19.8 Backup and recovery

19.8.1 Set backup conditions

You can set backup conditions to enable ApsaraDB for MongoDB to automatically back up data at your selected time points.

Procedure

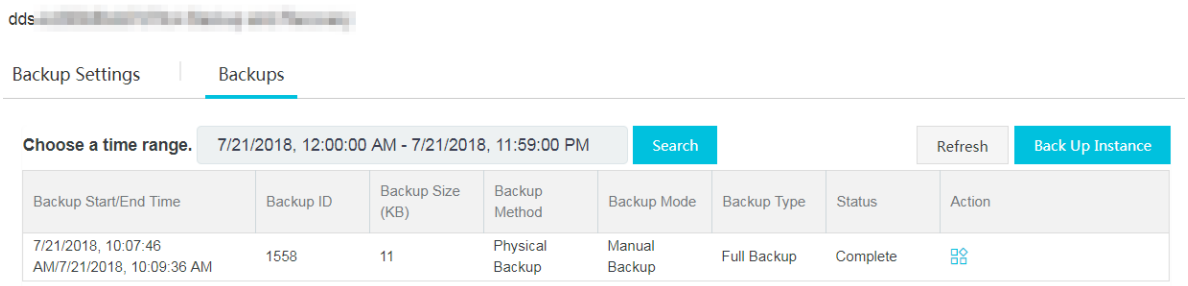
1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, choose **Backup and Restore** > **Backup Management**. The **Backup and Restore** page is displayed.
4. On the **Backup and Restore** page, click **Set** in the upper right corner to set the following backup conditions:
   - Backup Retention Period (days): By default, backup data is retained for seven days. This setting cannot be modified.
   - Backup Period: Set the period of data backup.
   - Backup Time: Time when the backup starts.
5. After setting the backup conditions, click **Confirm**.

19.8.2 Search a backup list

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, choose **Backup and Restore** > **Backup Management**. The **Backup and Restore** page is displayed.
4. Click **Backups**, as shown in *Figure 19-5: Backup list details*.

   Figure 19-5: Backup list details



5. Select a time range and click **Search**. The system searches the backups that are generated in the selected time range.

   You can click the icon to perform the following operations:

   - Download: Download the backup files that are generated in the specified time range. For more information, see *Download backup data*.
   - Restore data: Restore data from the backup files that are generated in the specified time range. For more information, see *Restore data*.
   - Create an instance from a backup point: Create an instance from the specified backup point. For information, see *Download an instance from a backup point*. You can refer to *Create an instance* to create an instance.

19.8.3 Create an instance from backup point

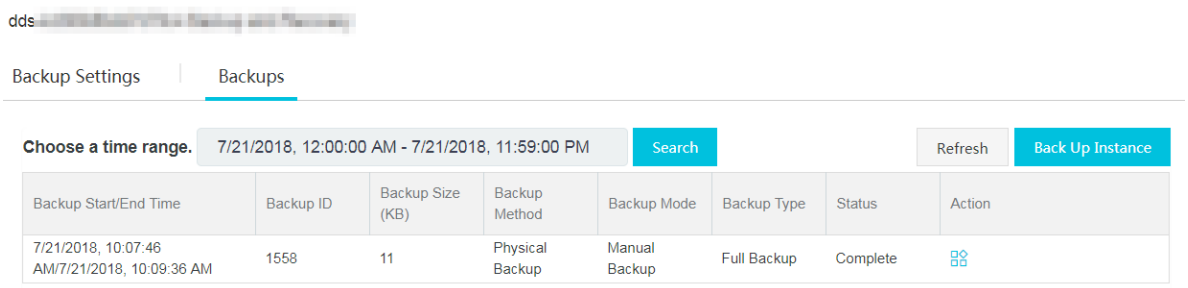
You can use any backup time point to clone a new instance.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.

3. In the left-side navigation menu, choose **Backup and Restore** > **Backup Management**. The **Backup and Restore** page is displayed.
4. Click **Backups**, as shown in *Figure 19-6: Backup list details*.

Figure 19-6: Backup list details



5. Click **Create Instances from Backup Point**.
6. In the displayed **Select a time** dialog box, set the time and click **Confirm**.
7. On the displayed *Create an instance* page, create an instance.

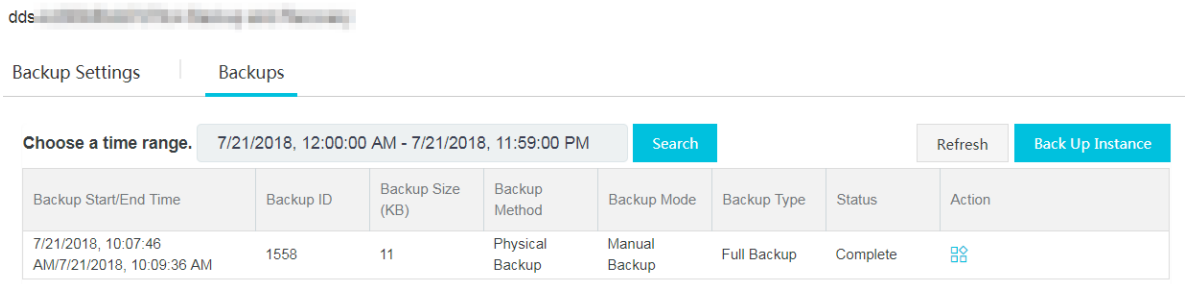
> **Note:**
> The storage space of the new instance must be equal to or greater than that of the source instance.

19.8.4 Back up an instance

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, choose **Backup and Restore** > **Backup Management**. The **Backup and Restore** page is displayed.
4. Click **Backups**, as shown in *Figure 19-7: Backup list details*.

Figure 19-7: Backup list details



5. Click **Back Up Instance** to manually back up your instance.

19.8.5 Restore data

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, choose **Backup and Restore** > **Backup Management**. The **Backup and Restore** page is displayed.
4. Switch to the **Backups** tab.
5. In the Action column of the target backup list, choose > **Restore Data**.
6. In the displayed **Restore Data** dialog box, click **OK**.

19.8.6 Download backup data

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, choose **Backup and Restore** > **Backup Management**. The **Backup and Restore** page is displayed.
4. Switch to the **Backups** tab.
5. In the Action column of the target backup list, choose > **Download**.
6. In the displayed **Download** dialog box, click **Confirm** to download the backup data to the local machine.

19.8.7 Download an instance from a backup point

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, choose **Backup and Restore** > **Backup Management**. The**Backup and Restore** page is displayed.
4. Switch to the **Backups** tab.
5. In the Action column of the target backup list, choose > **Create Instances from Backup Point** and click **Confirm** to go to the **MongoDB Instance Creation** page. Set the related information to create an instance. The created instance is displayed on the instance page.

19.9 Audit logs

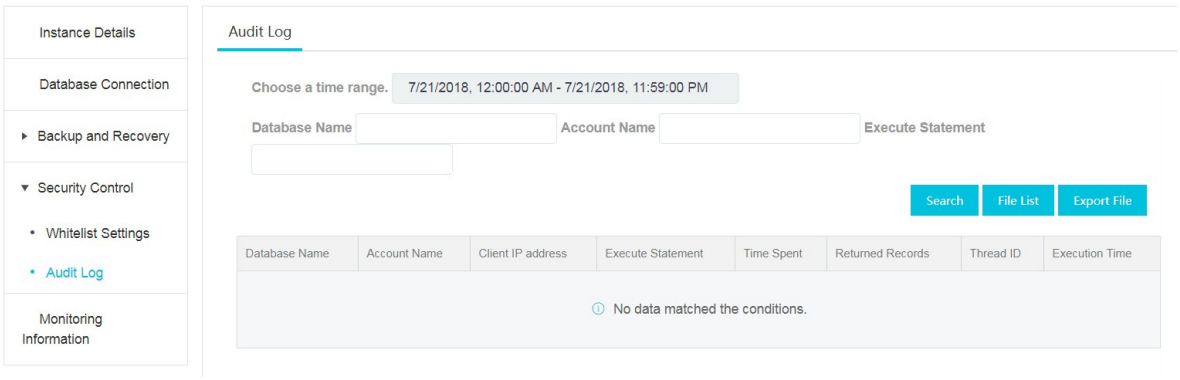
Context

Audit logs record all operations that a client performs on the connected database for subsequent fault analysis, behavior analysis, and security audits. Audit logs help you effectively obtain information about data execution for analysis. The recording of audit logs has become a supervision requirement for core business scenarios such as AntCloud.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation menu, choose **Security Control** > **Audit Log** to go to the **Audit Log** page, as shown in *Figure 19-8: Audit logs*.

Figure 19-8: Audit logs



4. On the **Audit Log** page, you can search and export files.
   - Search: Enter keywords such as the start time and end time of audit logs, database name, and database account to search logs by condition.
   - File List: Click to list audit log files.
   - Export File: Click to export audit log files.

19.10 Performance and monitoring

The ApsaraDB for MongoDB console provides abundant performance metrics for you to conveniently check and master the running status of instances. You can check instance monitoring data on the ApsaraDB for MongoDB console.

Procedure

1. *Log on to the ApsaraDB for MongoDB console*.
2. Click the ID of the target instance to go to the **Basic Information** page.
3. In the left-side navigation pane, select **Monitoring Information** to go to the **Resource Monitoring** page.

   You can select a time range to query historical metrics. *Table 19-4: Metric Information* lists various metrics.

   Table 19-4: Metric Information

Metrics	Description	Monitoring Frequency	Monitoring Period
CPU Utilization	Instance CPU Utilization	300s/time	30 days
Memory Utilization	Memory utilization by instance	300s/time	30 days
IOPS Utilization	IOPS used by the instance, including: • Data disk IOPS • Log disk IOPS	300s/time	30 days
IOPS Utilization	Percentage of the IOPS volume used by the instance to the maximum available IOPS volume	300s/time	30 days
Disk Space Utilization	Disk space occupied by the instance, including • Total used space	300s/time	30 days

Metrics	Description	Monitoring Frequency	Monitoring Period
	• Occupied data disk space • Occupied log disk space		
Disk Space Utilization	Percentage of the total space used by the instance to the maximum available space permitted by specifications	300s/time	30 days
opcounters	Operation QPS of the instance, including: • Number of insert operations • Number of query operations • Number of delete operations • Number of update operations • Number of getmore operations • Number of command operations	300s/time	30 days
connections	Current connections of the instance	300s/time	30 days
cursors	Number of cursors used by the instance currently, including: • Number of currently opened cursors • Number of expired cursors	300s/time	30 days
network	Network traffic of the instance, including: • Incoming traffic • Outgoing traffic • Number of processed requests	300s/time	30 days
globalLock	Length of the instance queue for the global lock, including: • Length of the instance queue waiting to read the global lock • Length of the instance queue waiting to write the global lock • Length of the instance queue waiting to perform all types of operations on the global lock	300s/time	30 days
wiredTiger	Cache indicators of the instances WiredTiger engine, including: • Volume of data read to the cache	300s/time	30 days

Metrics	Description	Monitoring Frequency	Monitoring Period
	• Capacity of the disk with data written from the cache • Maximum available disk capacity that is configured		

20 Server Load Balancer (SLB)

20.1 Introduction to Server Load Balancer

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

SLB virtualizes the added ECS instances into a high-performance and highly available application service pool by setting virtual service addresses, and distributes the requests from clients to ECS instances in the backend server pool based on forwarding rules.

Server Load Balancer provides the following features:

- **Protocol support**: Both Layer-4 (TCP and UDP) Server Load Balancer and Layer-7 (HTTP and HTTPS) Server Load Balancer are provided.
- **Health check**: Server Load Balancer checks the health status of backend ECS instances and automatically blocks abnormal ECS instances and distributes requests to them when they become normal.
- **Session persistence**: Server Load Balancer provides the session persistence feature. Requests from the same client are forwarded to the same backend ECS instance in a session lifecycle.
- **Scheduling algorithm**: Server Load Balancer supports the following scheduling algorithms:
  - Round robin: Requests are sequentially distributed across backend ECS instances.
  - Weighted least connections (WLC): A backend ECS instance with a smaller number of connections receives a larger percentage of live connections.
- **Domain name/URL-based forwarding**: For the layer-7 (HTTP and HTTPS) protocols, Server Load Balancer forwards requests to different server groups based on domain names or URLs.
- **Certificate management**: Server Load Balancer provides unified certificate management for the HTTPS protocol. You do not need to upload certificates to backend ECS instances. Deciphering is performed on Server Load Balancer to reduce the CPU overheads of backend ECS instances.

20.2 Quick start

This tutorial guides you to create an Internet Server Load Balancer instance to forward clients requests from to two backend ECS instances.

In this tutorial, two ECS instances on which an Apache Web application is deployed are added as backend servers to receive requests forwarded by Server Load Balancer.

Note:
If you want to create multiple listeners to forward different requests to different ECS instances, you must create a VServer group before adding the listeners. For more information, see *Add a VServer group*.

20.2.1 Planning and preparation

Before creating an SLB instance, you must plan the deployment of backend ECS instances, the network type (intranet/Internet), and the listening protocol to be configured.

Make the following preparations before creating an SLB instance:

- Create ECS instances

  ECS instances are used to receive and process requests forwarded by the SLB listeners. Before using Server Load Balancer, create ECS instances and deploy applications on them. Make sure that the department of the ECS instances is the same as that of the SLB instance, and the security rules of the ECS instances allow HTTP/HTTPS access on port 80/443.
- Plan the network type

  Server Load Balancer supports creating Internet instances and intranet instances. Different service IP address are allocated based on the network type. Select a network type of an SLB instance based on your business needs:

  — Internet: An Internet SLB instance distributes requests from the Internet only. After you create an Internet SLB instance, the system allocates a public IP address to the instance. You can bind your domain name to the public IP address to provide external services.
  — Intranet: An Internet SLB instance distributes requests from the intranet only. For intranet SLB instances, you have to further select to use the classic network or VPC network:

    ■ If you select the classic network, the system allocates an intranet IP address to the instance, which is managed by Alibaba Cloud. The SLB instance can be accessed only by the ECS instances in the classic network.

- If you select the VPC network, the system allocates a private IP address to the instance, which is an unused IP address of the VSwitch that the SLB instance belongs to. The SLB instance can be accessed only by ECS instances in the same VPC.

- Plan the listener protocol

  Alibaba Cloud provides both Layer-4 (TCP/UDP) Server Load Balancer and Layer-7 (HTTP /HTTPS) Server Load Balancer. You can configure different listeners based on different scenarios.

  Compared with Layer-4 listeners, Layer-7 listeners require an extra link of Tengine processing , therefore, the performance of Layer-7 listeners is less efficient to that of the Layer-4 listener. In addition, the performance of Layer-7 listeners may not be good because of insufficient client ports and too many connections to backend servers. We recommend that you use Layer-4 listeners if you have high requirements on the performance.

20.2.2 Create an SLB instance

An SLB instance is a running entity of Server Load Balancer. You can add multiple listeners and backend servers to an SLB instance.

Prerequisites

- Create ECS instances and deploy applications on them.
- Make sure that the department of the ECS instances is the same as that of the SLB instance, and the security rules of the ECS instances allow HTTP/HTTPS access over port 80/443.

Procedure

1. *Log on to the Apsara Stack console*.
2. Click **Create Instance**.
3. From the **Department** list, select a department.

   Note:
   Make sure that the department of the SLB instance is the same as that of the backend ECS instance.
4. From the **Project** list, select a project for the SLB instance.
5. In the **Name** field, enter the instance name.
6. In the **Network Type** area, select the instance type and network type.

In this tutorial, **Internet** is selected and leave the **IP Address** field as blank to use the system allocated IP address.

7. Click **Create**.

What's next

Add a listener

20.2.3 Add a listener

You must add at least one listener to the SLB instance to forward front-end requests to the backend servers. In addition to the forwarding configurations, you can also configure the health check settings when adding a listener.

Prerequisites

Create an SLB instance

Procedure

1. *Log on to the Apsara Stack console*.
2. On **Instances** page, click the ID of the target instance.
3. On the **SLB Instance Details** page, click the **Listener** tab.
4. On the **Listener** page, click **Add**.
5. In the **Add Listener** dialog box, configure the listener.

   In this tutorial, the listener configurations are as follows:

   - **SLB Protocol [Port]**: HTTP 80
   - **Backend Protocol [Port]**: HTTP 80
   - **Scheduling Algorithm**: Round robin
   - **Set Peak Bandwidth**: 100 Mbps
   - **Session Persistence**: Disabled
   - **Health Check Settings**: Use the default health check configurations
6. Click **Confirm**.

What's next

Add backend servers

20.2.4 Add backend servers

After configuring listeners, you must add backend servers to the SLB instance to receive and process requests forwarded by the listeners.

Procedure

1. *Log on to the Apsara Stack console*.
2. On **Instances** page, click the ID of the target instance.
3. On the **SLB Instance Details** page, click **Backend Server**.
4. Click **Add Backend Server**.
5. In the displayed dialog box, select the ECS instances to be added as the backend servers, and click **Add**.

   In this tutorial, two ECS instances with Apache applications deployed are added as the backend servers. An ECS instance with a higher weight will receive more requests. You can set the weight based on the external service capability of the backend ECS instance. In this tutorial, the default weight is used.

20.3 Manage SLB instances

On the Server Load Balancer console, you can edit, delete, start, or stop an SLB instance.

An SLB instance is a running entity of Server Load Balancer. To use Server Load Balancer, you must create an SLB instance and add listeners and backend servers to the instance.

20.3.1 Create an SLB instance

Before using Server Load Balancer, you must create an SLB instance. You can add multiple listeners and backend servers to an SLB instance.

Procedure

1. *Log on to the Apsara Stack console*.
2. Click **Create Instance**.
3. On the **Create SLB Instance** page, configure the SLB instance according to the following information, and then click **Create**.

   Table 20-1: Server Load Balancer configurations

Configuration	Description
Region	Displays the region of the instance.

Configuration	Description
Department	Select the department to which the SLB instance belongs.
Project	Select the project to which the SLB instance belongs.
Name	Enter a name for the SLB instance. The name can contain 1 to 63 characters. It must start with a letter and can contain numbers, letters, hyphens (-), and underscores (_).
Instance Type	Select an instance type of the SLB instance: • **Internal**: Select this type if the SLB instance only distributes client requests from the internal access. For intranet SLB instances, you must further select the network type (the classic network or the VPC network). If the VPC network is used, select the VPC and the VSwitch to which the instance belongs. • **External**: Select this type if the SLB instance only distributes client requests from the Internet.
Network Type	Select the network type of the SLB instance: • **Classic Network**: If the classic network is selected, an internal IP address will be allocated to the SLB instance. • **VPC**: If the VPC network is selected, a private IP address from the VSwitch that the SLB instance belongs to will be allocated to the SLB instance. **Note:** This configuration is required only when the instance type of the SLB instance is intranet.
IP Address	Enter the IP address of the SLB instance. If no IP address is specified, the system automatically allocates an IP address to the instance based on the network type.
Quantity	Enter the number of the instances to be created. The system can create SLB instances with the same configuration in batches.

20.3.2 Start or stop an instance

You can start or stop your SLB instance at any time.

Procedure

1. *Log on to the Apsara Stack console*.

2. On the **Instances** page, find the target instance.
3. Click the icon, and then click **Start** or **Stop** to start or stop the instance.
4. In the displayed dialog box, click **Confirm**.

20.3.3 View instance details

You can get the information such as the instance type, the department and project that the instance belongs to on the SLB details page.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, find the target instance.
3. Click the instance ID, or click the icon and then click **View Details** to view the instance details.

20.3.4 Modify attributes of an instance

You can edit the name and description of an SLB instance.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, find the target instance.
3. Click the icon and then click **Edit**.
4. In the displayed dialog box, edit the name and description, and then click **Confirm**.

20.3.5 Change the ownership of an instance

You can change the department and project that an SLB instance belongs to.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, find the target instance.
3. Click the icon and then click **Change Ownership**.
4. In the displayed dialog box, select the department and project to which the SLB instance belongs, and click **Confirm**.

20.3.6 Delete an instance

Delete an SLB instance if you no longer need it.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, find the target instance.
3. Click the icon and then click **Delete**.
4. In the displayed dialog box, click **Confirm**.

20.4 Configure listeners

The SLB listeners receive and forward client requests to backend ECS instances based on the forwarding rules.

Server Load Balancer provides Layer-4 (TCP and UDP) and Layer-7 (HTTP and HTTPS) listeners. Select the protocol of the listener based on your business needs.

Protocol	Feature	Scenario
TCP	• A connection-oriented protocol. A reliable connection must be established with the peer end before data can be sent and received. • Source address-based session persistence. • The source address is available at the network layer. • Fast data transmission.	Applicable to scenarios with high requirements on reliability and data accuracy, but with tolerance for low speeds, such as file transmission, sending or receiving e-mails, and remote logon.
UDP	• A non-connection-oriented protocol. Before sending data, UDP directly performs data packet transmission instead of making three handshakes with the other party. It does not provide error recovery and data retransmission. • Fast data transmission, however, the reliability is relatively low.	Applicable to scenarios with preference of real-time content over reliability, such as video chats and pushes of real-time financial quotations.
HTTP	• An application layer protocol mainly used to package data • Cookie-based session persistence • Obtain the source address using X-Forward-For	Applicable to applications that need to recognize data content, such as web applications and small-size mobile games.

Protocol	Feature	Scenario
HTTPS	• Encrypted data transmission to prevent unauthorized access • Unified certificate management service. You can upload certificates to Server Load Balancer. The decryption operations are completed directly on Server Load Balancer.	Applicable to applications requiring encrypted transmission.

20.4.1 Health check concepts

Server Load Balancer checks the service availability of the backend servers (ECS instances) by performing health checks on them. Health check improves the overall availability of the front-end service, and avoids the impact of service availability caused by the exceptions of backend ECS instances.

With health check enabled, when Server Load Balancer discovers that an instance is unhealthy, it stops distributing requests to that instance and only resumes distributing requests to the instance when it changes to the healthy status.

If your service is load-sensitive, frequent health checks may affect normal service access. You can reduce the health check frequency, increase the health check interval, or change Layer-7 health check to Layer-4 health check to reduce the impact on your service. However, to ensure continuous service availability, we do recommend that you disable the health check.

Note:
The health check configuration is included in the listener configuration. For more information, see *Add a Layer-4 listener* and *Add a Layer-7 listener*.

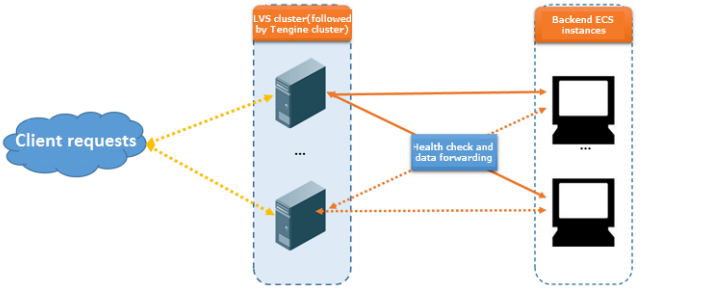
Health check process

Server Load Balancer is deployed in clusters. Data forwarding and health checks are handled at the same time by the node servers in the LVS cluster and Tengine cluster.

Servers in the LVS cluster perform data forwarding and health checks independently and in parallel based on the configured forwarding rules. If the health check done by an LVS node server on a backend ECS instance fails, the LVS node server no longer sends new client requests to the abnormal ECS instance. All servers in the LVS cluster perform this operation simultaneously without affecting each other.

The IP address ranges used to perform the health check are the IP address ranges of Server Load Balancer, including 100.64.0.0/10, 10.18.0.0/16, 10.19.0.0/16, and 10.49.0.0/16. If the backend ECS instances enable IP table forwarding or other access controls, authorize access of these IP address ranges on the intranet NIC.

Figure 20-1: Health check process



Health check

The health check process varies by the listener protocols as follows:

- **Health check of HTTP/HTTPS listeners**

  For Layer-7 (HTTP or HTTPS) listeners, Server Load Balancer detects the status of backend servers by sending HTTP head requests.

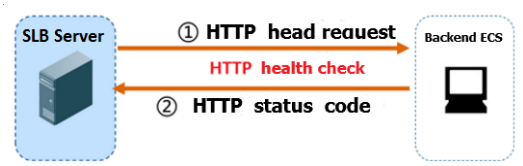
  > **Note:**
  > For HTTPS listeners, certificates are managed in Server Load Balancer. Data (including health check data and service interaction data) between Server Load Balancer and backend ECS instances is not transmitted over HTTPS to improve the system performance.

  The health check process of a Layer-7 listener is as follows:

  1. The Tengine node server sends an HTTP head request to the backend server at *ECS internal IP:health check port* based on the health check configuration of the listener.
  2. After receiving the request, the backend ECS instance returns an HTTP status code that indicates the service running status.
  3. If the Tengine node server does not receive the response from the backend ECS instance within the specified response timeout, then the ECS instance is considered unhealthy.
  4. If the Tengine node server receives a response from the backend ECS instance within the specified response timeout, the server compares the response with the configured status

code. If the response matches the status code, then the ECS instance is considered healthy , otherwise, the ECS instance is considered unhealthy.

Figure 20-2: Health check of HTTP/HTTPS listeners



- **Health check of TCP listeners**

  For Layer-4 TCP listeners, Server Load Balancer detects the status of backend servers by sending TCP detections

  The health check process of a TCP listener is as follows:

  1. The LVS node server sends a TCP SYN data packet to the backend ECS instance at *ECS internal IP:health check port* based on the health check configuration of the listener.
  2. After receiving the request, the backend ECS instance returns an SYN+ACK packet if the corresponding port is listening.
  3. If the LVS node server does not receive the packet from the backend ECS instance within the configured response timeout period, the ECS instance is considered unhealthy. Then, the LVS node server sends an RST data packet to the backend ECS server to terminate the TCP connection.
  4. If the LVS node server receives the data packet from the backend ECS instance within the configured response timeout period, the ECS instance is considered healthy. Then, the LVS node server sends an RST data packet to the backend ECS server to terminate the TCP connection.

  Note:

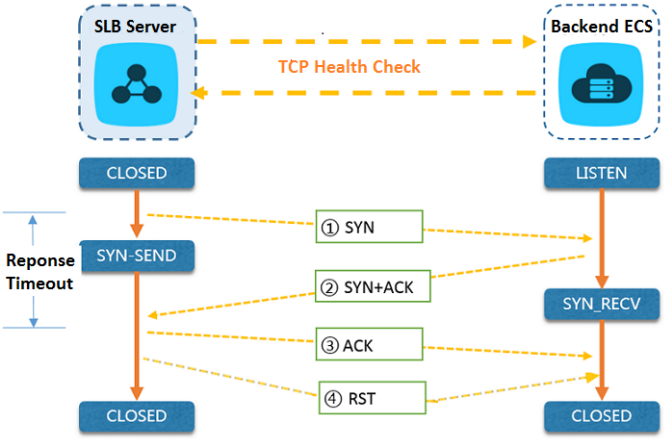
  In general, TCP three-way handshakes are conducted to establish a TCP connection. After the LVS node server receives an SYN + ACK data packet from the backend ECS instance, the LVS node server sends an ACK data packet, and then immediately sends an RST data packet to terminate the TCP connection.

  This mechanism may cause backend ECS instances to think an error occurred in the TCP connection, such as an abnormal exit, and then throw a corresponding error message, such as `Connection reset by peer`.

Resolution:

- Use the HTTP health check.
- After enabling real IP obtaining, ignore the connection errors caused by the health check IPs.

Figure 20-3: Health check of TCP listeners



- **Health check of UDP listeners**

  For layer-4 UDP listeners, Server Load Balancer detects the status of the backend servers through UDP packet detection.

  The health check process of a UDP listener is as follows:

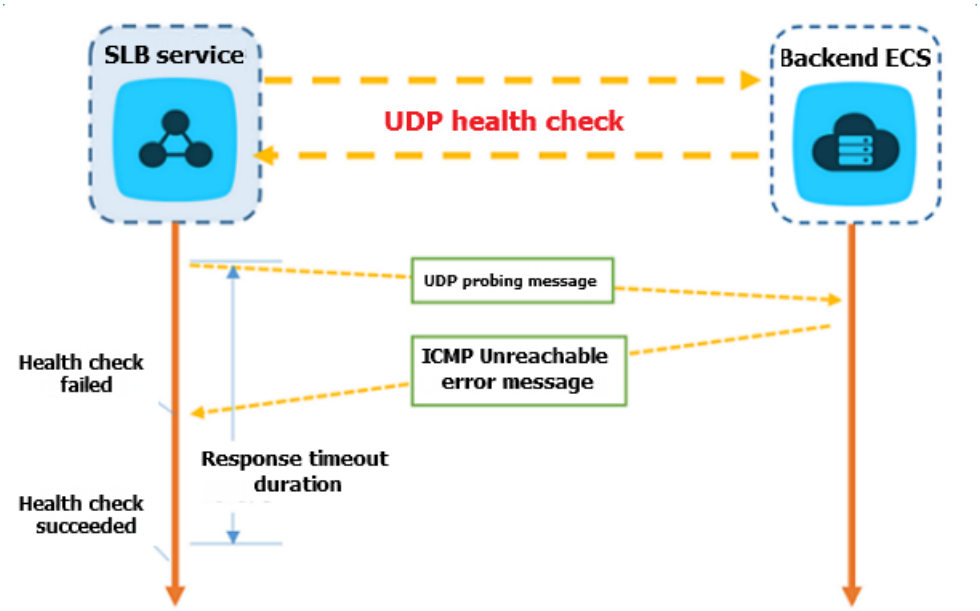
  1. The LVS node server sends a UDP packet to the ECS instance at *ECS internal IP:health check port* based on the health check configuration of the listener.
  2. If the corresponding port of the backend ECS instance is not listening normally, the system returns an ICMP error message such as `port XX unreachable`. Otherwise, no message is returned.
  3. If the LVS node server receives the preceding error message from the backend ECS instance within the specified response time, the ECS instance is considered unhealthy and the health check fails.
  4. If the LVS node server does not receive any message from the backend ECS instance within the specified response time, the ECS instance is considered healthy and the health check succeeds.

Note:

The actual health status may be different from the result of the health check performed by the UDP listeners.

If the backend ECS instance is a Linux server, in high-concurrency scenarios, the anti-ICMP attack protection of Linux limits the speed of the server in sending ICMP messages. In this case, even if a server exception occurs, the error message `port XX unreachable` cannot be returned to the front end, thereby Server Load Balancer may consider that the health check succeeds because it does not receive the ICMP response. As a result, the actual service status is different from the health check result.

Figure 20-4: Health check of UDP listeners



Health check time window

Health check effectively improves service availability. However, to reduce the impact on the system availability caused by frequent ECS switches due to health check failure, Server Load Balancer declares an ECS instance healthy or unhealthy only after successive successes or failures within a specified timeframe. The following factors determine the health check time window:

- Health check interval (the time interval between two consecutive health checks)
- Response timeout (the amount of time to wait for the response from the server)
- Health check threshold (the number of consecutive successful or failed health checks)

The health check time window is calculated as follows:

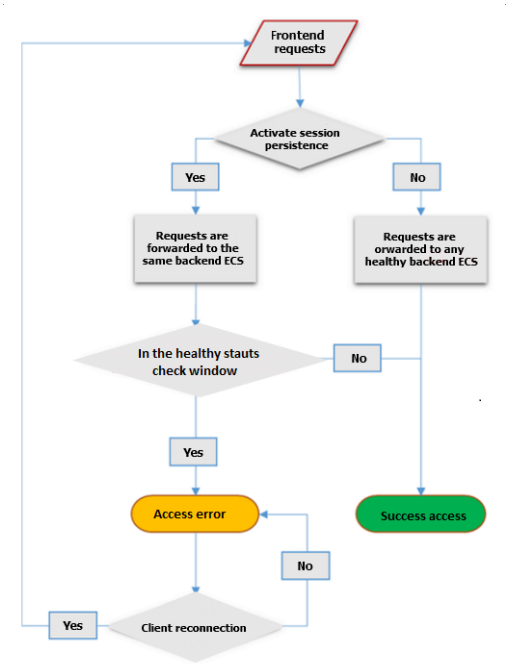
- Health check failure time window = (health check interval + response timeout) × unhealthy check threshold
- Health check success time window = health check interval × healthy check threshold

By default, the healthy status check window for the TCP/HTTP/HTTPS listeners is 6 seconds, and the unhealthy status check window for them is 21 seconds. For UDP listeners, the healthy status check window is 15 seconds, and the unhealthy status check window is 45 seconds.

The health check status has the following impact on request forwarding:

- If the health check on the target ECS instance fails, new requests are not forwarded to the ECS instance, and front-end access is not influenced.
- If the health check on the target ECS instance succeeds, new requests are forwarded to the ECS instance and front-end access is normal.
- If an exception occurs on the target ECS instance during the health check failure time window, but the health check threshold does not reach the specified value, the requests are forwarded to the ECS instance and front-end access fails.

Figure 20-5: Request forwarding



20.4.2 Add a Layer-4 listener

Alibaba Cloud supports Layer-4 (TCP and UDP) Server Load Balancer. A Layer-4 listener of Server Load Balancer forwards requests directly to the backend ECS instances without modifying the headers.

Context

For more information, see *Configure listeners*.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. On the **SLB Instance Details** page, click the **Listener** tab.
4. On the **Listener** page, click **Add**.
5. In the **Add Listener** dialog box, configure the listener as follows, and click **Confirm**.

Table 20-2: Layer-4 listener configuration

Configuration	Description
SLB Protocol [port]	Select the front-end protocol and port which receives requests and forwards the requests to backend servers. For Layer-4 listeners, select **TCP** or **UDP**.
Backend Protocol [port]	Enter the port of the application deployed on ECS instances.
Scheduling Algorithm	Select a forwarding rule: • **Round Robin**: Requests are sequentially distributed to ECS instances. • **Least Connections**: Backend servers with a higher weight will receive a larger percentage of live connections at any one time. If the weights are the same, the system directs connections to the server with the fewest established connections.
Set Peak Bandwidth	Set the peak bandwidth of the listener in Mbps.
Session Persistence	Select whether to enable session persistence. For TCP listeners, session persistence is based on IP addresses . If enabled, requests from the same IP address are forwarded to the same backend server. If session persistence is enabled, specify a session timeout value in the **Timeout** field.

Configuration	Description
Configure Idle Connection Timeout	Specify the idle connection timeout in seconds. If no request is received during the specified timeout period, Server Load Balancer will close the connection and restart the connection when the next request comes.
Use VServer Group	Select whether to use a VServer group. If VServer group is used, select the VServer group to bind with the listener. A VServer group consists of ECS instances that provide the same services. Server Load Balancer forwards client requests to the ECS instances in the specified VServer group according to the configured listening rules. If VServer group is not used, Server Load Balancer forwards client requests to the ECS instances in the server pool according to the configured listening rules.  **Note:** The VServer group cannot be modified after the listener is added.
Health Check Settings	
Check Port	Port used by the health check service to access backend ECS instances. By default, it is the backend port specified when adding the listener.
Response Timeout (seconds)	The maximum amount of time to wait for a health check response. If the backend ECS instance does not correctly respond within the specified time, then the health check fails.
Health Check Interval (Seconds)	The time interval between two consecutive health checks. All nodes in the LVS cluster perform health checks on the ECS instances at the specified interval independently and in parallel.
Unhealthy Threshold	Number of consecutive failures of the health check performed by the same LVS node server on the same backend ECS instance (from success to failure).
Healthy Threshold	Number of consecutive successes of the health check performed by the same LVS node server on the same backend ECS instance (from failure to success).

20.4.3 Add a Layer-7 listener

A Layer-7 listener uses reverse proxy implementation. After HTTP requests of clients arrive at an SLB listener, the SLB instance establishes a TCP connection with the backend ECS instances.

That is, the service uses the new TCP connection to allow HTTP protocol to access backend servers, instead of directly forwarding packets to the backend servers.

Context

For more information, see *Configure listeners*.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. On the **SLB Instance Details** page, click the **Listener** tab.
4. On the **Listener** page, click **Add**.
5. In the **Add Listener** dialog box, configure the listener as follows, and click **Confirm**.

Table 20-3: Layer-7 listener configuration

Configuration	Description
Basic Settings	
SLB Protocol [Port]	Select the front-end protocol and port used to receive requests and forward the requests to backend servers. For Layer-7 listeners, select **HTTP** or **HTTPS**.
Backend Protocol [Port]	Enter the port of the application deployed on the backend ECS instances.
Scheduling Algorithm	Select a forwarding rule: • **Round Robin**: Requests are sequentially distributed to ECS instances. • **Least Connections**: Backend servers with a higher weight will receive a larger percentage of live connections at any one time. If the weights are the same, the system directs connections to the server with the fewest established connections.
Set Peak Bandwidth	Set the peak bandwidth of the listener in Mbps.
Two-way Authentication	Select whether to enable two-way authentication. After two-way authentication is enabled, you must upload both the server certificate and the CA certificate. By default, two-way authentication is disabled for HTTPS listeners . Click **Upload Certificate** to upload the server and CA certificates. For more information, see *Upload certificates*.

Configuration	Description
	Note: This option is applicable only to HTTPS listeners.
Select Server Certificate	Select a server certificate. The server certificate is used by the client browser to check whether the certificate sent by the server is signed and issued by a trusted center. **Note:** This option is applicable only to HTTPS listeners.
Select CA Certificate	Select a CA certificate. The CA certificate is used by the server to verify a client's identity. If the verification fails, connection is denied. The CA certificate is only required when the two-way authentication is enabled. You can use a self-signed CA certificate for verification. **Note:** This option is applicable only to HTTPS listeners with two-way authentication enabled.
Session Persistence	Select whether to enable session persistence. For Layer-7 (HTTP and HTTPS) listeners, Server Load Balancer supports cookie-based session persistence.
Cookie Persistence	Select a cookie processing method: • **Cookie Insert**: When a client accesses Server Load Balancer for the first time, Server Load Balancer inserts a cookie (inserts a SERVERID string in the HTTP/HTTPS response message) in the response request. When the client accesses Server Load Balancer with this cookie next time, Server Load Balancer forwards the requests to the recorded ECS instance. If this method is used, specify a cookie timeout value in the **Timeout** field. • **Cookie Rewrite**: Server Load Balancer will overwrite the original cookie when it discovers that a new cookie is set. When a client accesses Server Load Balancer with the new cookie next time, Server Load Balancer forwards the requests to the recorded ECS instance.

Configuration	Description
	If this method is used, specify the cookie name inserted in the HTTPS/HTTP response in the **Cookie Name** field and maintain the cookie timeout value in the backend ECS instance. **Note:** This option is available only when session persistence is enabled.
Use VServer Group	Select whether to use a VServer group. If VServer group is used, select the VServer group to bind with the listener. A VServer group consists of ECS instances that provide the same services. Server Load Balancer forwards client requests to the ECS instances in the specified VServer group according to the configured listening rules. If VServer group is not used, Server Load Balancer forwards client requests to the ECS instances in the server pool according to the configured listening rules. **Note:** The VServer group cannot be modified after the listener is added.
Health Check Settings	
Enable Health Check	Select whether to enable health check. This feature is enabled by default. To ensure continuous service availability, we recommend that you enable health check.
Domain Name and Check Path	In HTTP health check, the Server Load Balancer system uses the intranet IP address of the backend ECS instance to send an HTTP head request to the default home page of the application server. • If the page for heath check is not the default home page of the application server, you must specify the domain name and the specific check path. • If you set limitation on the host field parameters of the HTTP head request, you only need to specify the check path, namely the URI of the page file used for health check.

Configuration	Description
Health Status	Set the HTTP status code indicating that the health check is normal.
Check Port	Port used by the health check service to access backend ECS instances. By default, it is the backend port specified when adding the listener.
Response Timeout (Seconds)	The maximum amount of time to wait for a health check response. If the backend ECS instance does not correctly respond within the specified time, then the health check fails.
Health Check Interval (Seconds)	The time interval between two consecutive health checks. All nodes in the LVS cluster perform health checks on the ECS instances at the specified interval independently and in parallel.
Unhealthy Threshold	Number of consecutive failures of the health check performed by the same LVS node server on the same backend ECS instance (from success to failure).
Healthy Threshold	Number of consecutive successes of the health check performed by the same LVS node server on the same backend ECS instance (from failure to success).

20.4.4 Configure forwarding rules

Layer-7 Server Load Balancer supports adding forwarding rules based on domain name or URL, so as to forward requests from different domain names or URLs to different ECS instances.

Context

You can add multiple forwarding rules in a listener. Each forwarding rule is associated with a VServer groups (A VServer group consists of multiple ECS instances.) For example, you can forward all read requests to one group of backend servers and all write requests to another VServer group to meet different business demands.

After forwarding rules are added, the system forwards client requests as follows:

- If the client request matches a forwarding rule, the request will be forwarded to the VServer group specified in the forwarding rule.
- If the client request does not match a forwarding rule, the request will be forwarded to the VServer group configured for the listener. If the listener does not configure a VServer group, the request will be forwarded to the backend ECS instances in the server pool of the Server Load Balancer instance.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. On the **SLB Instance Details** page, click the **Listener** tab.
4. On the **Listener** page, locate the target listener.

   Only HTTP and HTTPS listeners support configuring domain name and URL forwarding rules.
5. Click the icon, and then click **Configure Forwarding Rule**.
6. In the **Configure Forwarding Rule** dialog box, click **Add Forwarding Rule**.
7. Configure the forwarding rules based on the following information, and click **Save**.

   - Configure a domain name forwarding rule

     — Leave the **URL** field empty (no "/" is required) when configuring a domain name forwarding rule separately. The domain name consists of only letters, numbers, hyphens (-), or dots (.).

     — The rule can be configured in exact match or wildcard match mode. For example, `www.aliyun.com` is a precise domain name, while `*.aliyun.com` and `*.market.aliyun.com` are wildcard domain names. When a front-end request matches multiple domain name rule, the exact domain takes high priority.

     Table 20-4: Domain name match rules

Mode	Request URL	Domain name rule (√ indicates that the request matches the domain name rule, while x indicates not matching.)		
		www.aliyun.com	*.aliyun.com	*.market.aliyun.com
Exact match	www.aliyun.com	√	x	x
Wildcard domain name match	market.aliyun.com	x	x	x
	info.market.aliyun.com	x	x	√

   - Configure a URL forwarding rule

- Leave the domain name configuration item empty when configuring a URL forwarding rule separately.
- The URL contains only letters, numbers, or special characters including hyphens (-), dots (.), slashes (/), percent signs (%), question marks (?), number signs (#), and ampersands .
- The URL must start with a slash (/).

  Note:
  You cannot only enter one slash (/) in the **URL** field. If so, the URL forwarding rule is invalid.

- The URL forwarding rule supports string match and complies with sequential matching. For example, "/admin", "/bbs_", and "/ino_test".

- Add a domain name+URL forwarding rules

  If you want to forward traffic based on the same domain name, but different paths, you can configure a forwarding rule with both domain name and URL. We recommend further configuring a domain-only forwarding rule to avoid access errors due to no matching URLs.

  For example, there are two domain names: www.aaa.com and www.bbb.com. The requirement is to forward the requests from www.aaa.com/index.html to ServerGroup1 and forward other requests from xxx.html to ServerGroup2. You need to configure the following forwarding rules. If rule2 is absent, requests from www.aaa.com will get a 404 error.

20.4.5 Set access control

You can add a whitelist to allow specific IP addresses to access Server Load Balancer.

Context

When setting a whitelist, note:

- Certain business risks exist because once a whitelist is set, only the IP addresses in the whitelist can access the Server Load Balancer listener.
- If no whitelist is configured after access control is enabled, no IP address can access the Server Load Balancer listener.
- A short interruption to the Server Load Balancer listener may occur when configuring a whitelist .

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. On the **SLB Instance Details** page, click the **Listener** tab.
4. On the **Listener** page, locate the target listener.
5. Click the icon, and then click **Configure Access Control**.
6. In the displayed dialog box, enable access control, and enter the IP addresses that are allowed to access the listener.

   Separate multiple IP addresses by commas. You can add up to 300 unique IP addresses. You can also enter IP addresses in the form of CIDR block, such as 10.23.12.0/24.
7. Click **Confirm**.

20.4.6 Stop a listener

After a listener is stopped, it no longer forwards traffic.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. On the **SLB Instance Details** page, click the **Listener** tab.
4. Click the icon of the target listener, and then click and click **Stop**.

20.4.7 Start a listener

You can restart a listener that has stopped traffic forwarding.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. On the **SLB Instance Details** page, click the **Listener** tab.
4. Click the icon of the target listener, and then click and click **Start**.

20.4.8 Edit a listener

You can edit the configuration of a listener.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.

3. On the **SLB Instance Details** page, click the **Listener** tab.
4. Click the icon of the target listener, and then click and click **Edit**.

20.4.9 Delete a listener

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. On the **SLB Instance Details** page, click the **Listener** tab.
4. Click the icon of the target listener, and then click and click **Delete**.
5. In the displayed dialog box, click **Confirm**.

20.5 Configure backend servers

Before using Server Load Balancer, you must add ECS instances as the backend servers of your Server Load Balancer instance to receive and process forwarded requests .

Server Load Balancer virtualizes multiple ECS instances in the same region into an application server pool featuring high performance and high availability by setting a virtual IP address. Server Load Balancer performs health check on ECS instances in the application server pool, automatica lly blocks abnormal ECS instances, and forwards traffic to them again when they become normal . Health check improves the overall availability of the front-end service, and avoids the impact of service availability caused by the exceptions of backend ECS instances.

You can increase or decrease the number of backend ECS instances at any time. However, to ensure the stability of your external services, make sure that you have enabled the health check feature and that at least one ECS instance in the Server Load Balancer instance is running normally when you perform the preceding operations.

20.5.1 Add an ECS instance

After creating an SLB instance, you must add ECS instances to it to process requests forwarded by the SLB instance.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. Click the **Backend Server** tab.
4. On the **Backend Server** page, click **Add Backend Server**.

5. In the displayed dialog box, select the ECS instance to be added as the backend server and set the server weight.

   An ECS instance with a higher weight will receive more requests. Set the weight based on the external service capability of the backend ECS instance.

   > **Note:**
   > If the weight is set to **0**, no requests will be forwarded to the ECS instance.

6. Click **Add**.

20.5.2 Modify the weight of an ECS instance

You can modify the weight of an added ECS instance. An ECS instance with a higher weight will receive more requests. Set the weight based on the external service capability of the backend ECS instance.

Context

> **Note:**
> If you add a backend ECS instance to a VServer group, you must modify the weight of the ECS instance when editing the VServer group. For more information, see *Edit a VServer group*.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. Click the **Backend Server** tab.
4. Click the icon of the target ECS instance and then click **Edit**.
5. In the displayed dialog box, modify the weight of the ECS instance, and click **Confirm**.

   > **Note:**
   > If the weight is set to **0**, no requests will be forwarded to the ECS instance.

20.5.3 Remove an ECS instance

You can remove an ECS instance from an SLB instance. However, directly removing an ECS instance from an SLB instance may cause transient service interruption. We recommend that you

modify the weight of the ECS instance to zero and remove the ECS instance after no traffic is forwarded to instance.

Context

> **Note:**
> If you have added a backend ECS instance to a VServer group, you must remove the ECS instance when editing the VServer group. For more information, see *Edit a VServer group*.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. Click the **Backend Server** tab.
4. Click the icon of the target ECS instance and then click **Remove**.
5. In the displayed dialog box, click **Confirm**.

20.5.4 Add a VServer group

A VServer group is a group of ECS instances. By using VServer groups (configure different listeners with different VServer groups), Server Load Balancer can forward client requests to different ECS instances.

Context

Note the following when adding a VServer group:

- An ECS instance can be added to multiple VServer groups.
- A VServer group can be associated with multiple listeners.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. Click the **VServer Group** tab.
4. On the **VServer Group** page, click **Add VServer Group**.
5. On the **Create VServer Group** page, complete these steps:
   a) In the **VServer Group Name** field, enter a VServer group name.
   b) Select a query criteria and enter the corresponding value. Then, click **Search** to query the ECS instance to be added.

c) In the **Available Servers** list, click the ECS instance to be added.

d) In the **Selected Servers** list, set the port and weight of the ECS instance.

An ECS instance with a higher weight will receive more requests. Set the weight based on the external service capability of the backend ECS instance.

> **Note:**
> If the weight is set to **0**, no requests will be forwarded to the ECS instance.

e) Click **Confirm**.

20.5.5 View a VServer group

You can view the status, configured ports, and other information of the ECS instances added to a VServer group.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. Click the **VServer Group** tab.
4. Click the ID of the target VServer group, or click the icon and then click **View**.

20.5.6 Edit a VServer group

You can modify the name of a VServer group, modify the weights of the ECS instances in the VServer group, and add new ECS instances to or remove ECS instances from the VServer group.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. Click the **VServer Group** tab.
4. Click the icon of the target VServer group, and then click **Edit**.
5. In the displayed dialog box, add or remove an ECS instance or modify the weight of the ECS instance.

20.5.7 Delete a VServer group

You can delete an unnecessary VServer group.

Prerequisites

If the VServer group to be deleted is associated with a listener, you must delete the listener before deleting the VServer group.

Procedure

1. *Log on to the Apsara Stack console*.
2. On the **Instances** page, click the ID of the target instance.
3. Click the **VServer Group** tab.
4. Click the icon of the target VServer group, and then click **Delete**.
5. In the displayed dialog box, click **Yes** and then click **Confirm**.

20.6 Manage certificates

Server Load Balancer provides the certificate management function for HTTPS listeners to manage certificates.

To configure HTTPS listeners, you must upload the required certificates:

- For HTTPS two-way authentication, upload the CA and server certificates.
- For HTTPS one-way authentication, upload the server certificate.

After the certificates are uploaded to Server Load Balancer, you do not need to deploy the certificates on the backend ECS instances. Private keys uploaded to the certificate management system are encrypted.

- Server certificate: Used by the client browser to check whether the certificate sent by the server is signed and issued by a trusted center. You can buy a server certificate from Alibaba Cloud Security Certificate Service or other service providers.
- Client certificate: Proves the identity of the client user so that the client user can prove the true identity when communicating with the server. You can use a self-signed CA certificate to sign the client certificate.
- CA certificate: After receiving the client certificate sent by your browser, the server verifies the client certificate by using the CA certificate. If the verification fails, the server rejects connecting to the client.

20.6.1 Certificate format

Only PEM certificates in the Linux environment can be uploaded to Server Load Balancer.

Requirements on the certificate format

Make sure that your certificate to be uploaded meets the following requirements:

- Certificates issued by a root CA

  If the certificate is issued by a root CA, the received certificate is the only one that is required to upload to Server Load Balancer. The website that is configured with the certificate will be trusted by the web browser without configuring additional certificates.

  The certificate format must meet the following requirements:

  - The certificate content is placed between -----BEGIN CERTIFICATE-----, -----END CERTIFICATE-----. Include the header and footer when uploading the certificate.
  - Each line except the last must contain exactly 64 characters. The last line can contain 64 or fewer characters.
  - Space is not allowed in the content.
- Certificates issued by an intermediate CA organization

  If a certificate is issued by an intermediate CA, you will obtain multiple intermediate certificates . You must combine the server certificate and the immediate certificate first, and then upload it to Server Load Balancer.

  The format of the certificate chain must meet the following requirements:

  - Put the server certificate in the first place and the intermediate certificates in the second place without any space in between.
  - Space is not allowed in the content.
  - Each line except the last must contain exactly 64 characters. The last line must contain 64 or fewer characters.
  - Conform to the certificate requirements as described in the certificate description. In general , the intermediate CA will provide an instruction about the certificate format when issuing the certificate, the certificate chain must conform to the format requirements.

Requirements on the RSA private key format

When uploading a server certificate, you also need to upload the private key of the certificate. The RSA private key format must meet the following requirements:

- The key is placed between -----BEGIN RSA PRIVATE KEY-----, -----END RSA PRIVATE KEY -----. Include the header and footer when uploading the key.
- Space is not allowed in the content. Each line except the last must contain exactly 64 characters. The last line can contain 64 or fewer characters.

Note:

If your private key is encrypted. For example, the header and footer are -----BEGIN PRIVATE KEY-----, -----END PRIVATE KEY----- or -----BEGIN ENCRYPTED PRIVATE KEY-----, -----END ENCRYPTED PRIVATE KEY-----, or the private key contains Proc-Type: 4,ENCRYPTED, run the following command to convert the private key before uploading it to Server Load Balancer:

```
openssl rsa -in old_server_key.pem -out new_server_key.pem
```

20.6.2 Generate a CA certificate

When configuring HTTPS listeners, you can use self-signed CA certificates. Follow the instructions in this document to generate a CA certificate using OpenSSL.

Procedure

1. Run the following commands to create a *ca* folder under the */root* directory, and create four subfolders under the *ca* folder.

```
$ sudo mkdir ca
$ cd ca
$ sudo mkdir newcerts private conf server
```

   - The *newcerts* folder is used to store the digit certificate signed by a CA certificate.
   - The *private* folder is used to store the private key of the CA certificate.
   - The *conf* folder is used to store the configuration files.
   - The *server* folder is used to store the server certificate.

2. Create an *OpenSSL. conf* file with the following configurations in the *conf* directory.:

```
[ ca ]
default_ca = foo

[ foo ]
dir = /root/ca
database = /root/ca/index.txt
new_certs_dir = /root/ca/newcerts

certificate = /root/ca/private/ca.crt
serial = /root/ca/serial
private_key = /root/ca/private/ca.key
RANDFILE = /root/ca/private/.rand

default_days = 365
default_crl_days= 30
default_md = md5
unique_subject = no
policy = policy_any
```

```
[ policy_any ]
countryName = match
stateOrProvinceName = match
organizationName = match
organizationalUnitName = match
localityName = optional
commonName      = supplied
emailAddress    = optional
```

3. Run the following commands to generate a private key file.

```
$ cd /root/ca
$ sudo openssl genrsa -out private/ca.key
```

4. Run the following command and input the required information according to the prompts. Press Enter to generate a *csr* file.

```
$ sudo openssl req -new -key private/ca.key -out private/ca.csr
```

Figure 20-6: Generate a key

5. Run the following command to generate a *.crt* file.

```
$ sudo openssl x509 -req -days 365 -in private/ca.csr -signkey
private/ca.key -out private/ca.crt
```

6. Run the following command to set the start sequence number for the key, which can be any four characters.

```
$ sudo echo FACE > serial
```

7. Run the following command to create a CA key library.

```
$ sudo touch index.txt
```

8. Run the following command to create a certificate revocation list for the removed client certificates.

```
$ sudo openssl ca -gencrl -out /root/ca/private/ca.crl -crldays 7 -
config "/root/ca/conf/openssl.conf"
```

The execution results are as follows:

```
Using configuration from /root/ca/conf/openssl.conf
```

9. Run the following command to view the generated CA certificate.

```
cd private
```

```
ls
```

20.6.3 Generate a client certificate

A client certificate proves the identity of the client user so that the client user can prove the true identity when communicating with the server.

Prerequisites

A CA certificate is required to sign the client certificate. Make sure that you *Generate a CA certificate*.

Procedure

1. Run the following command to create the *users* directory in the *ca* directory to store keys.

   ```
   $ sudo mkdir users
   ```

2. Run the following command to create a key for the client.

   ```
   $ sudo openssl genrsa -des3 -out /root/ca/users/client.key 1024
   ```

   When creating the key, enter the pass phrase as the key password to prevent unauthorized use if the key leaks. Enter the same password twice.

3. Run the following command to create the certificate signing request file *csr* for the key.

   ```
   $ sudo openssl req -new -key /root/ca/users/client.key -out /root/ca
   /users/client.csr
   ```

   Enter the pass phrase stored in the preceding step as prompted, press Enter, and enter the required information as prompted.

   Note:

   A challenge password is the password of the client certificate (which must be separated from the password of the *client.key* file). It can be the same as the password of the server or root certificate.

4. Run the following command to use the CA key to sign the client key.

   ```
   $ sudo openssl ca -in /root/ca/users/client.csr -cert /root/ca/
   private/ca.crt -keyfile /root/ca/private/ca.key -out /root/ca/users/
   client.crt -config "/root/ca/conf/openssl.conf"
   ```

   When you are prompted to confirm the signature, enter y.

5. Run the following command to convert the certificate to the *PKCS12* file that can be identified by most browsers.

```
$ sudo openssl pkcs12 -export -clcerts -in /root/ca/users/client.crt
 -inkey /root/ca/users/client.key -out /root/ca/users/client.p12
```

   Enter the pass phrase of the client key as prompted and press Enter, and enter the password for exporting the certificate as prompted. This password is used to protect the client certificate, which is required when the client certificate is installed.

6. Run the following command to view the generated client certificate.

```
cd users
ls
```

20.6.4 Upload certificates

Server Load Balancer provides the certificate management function to support data transfer encryption and authentication over HTTPS. You can store certificates in the Server Load Balancer certificate management system without deploying certificates on backend ECS instances.

Note:
Each account can create up to 100 certificates.

Prerequisites

You have generated a server certificate or CA certificate to be uploaded.

Procedure

1. *Log on to the Apsara Stack console*.
2. Click the **Certificate Management** tab.
3. Click **Upload Certificate**.
4. In the **Upload Certificate** dialog box, provide the following information, and click **Confirm**.

Table 20-5: Upload certificate configuration

Configuration	Description
Region	Select the region where the certificate is uploaded. Server Load Balancer manages certificates by regions. To use a certificate in multiple regions, upload the certificate in each region.
Department	Select the department to use the certificate.
Project	Select the project to use the certificate.

Configuration	Description
Certificate Type	Select a type for the certificate to be uploaded. • Server certificate: For one-way authentication, only server certificate is required. The client uses it to check whether the certificate sent by the server is issued by a trusted center. • CA certificate: For two-way authentication, a CA certificate is required in addition to a server certificate. The server uses the CA certificate to authenticate the CA signature on the client certificate, as part of the authorization before launching a secure connection.
Certificate Name	Enter a certificate name.
Certificate Contents	Enter the certificate content. The certificate must be in the PEM format. You can click **Examples** to view the example format. For more information, see *Certificate format*.
Private Key	Enter the private key of the server certificate. The private key must meet the format requirement of Server Load Balancer. You can click **Examples** to view the example format.

20.6.5 Convert certificate formats

Server Load Balancer supports PEM certificates only. Certificates in other formats must be converted to the PEM format before they can be uploaded to Server Load Balancer.

Context

We recommend that you use OpenSSL for format conversion. The following describes how to convert popular certificate formats to PEM:

- DER: This format is usually used on a Java platform.
- P7B: This format is usually used in a Windows server or Tomcat.
- PFX: This format is usually used in a Windows server.

Convert DER to PEM

1. Run the following command to convert the certificate format.

   ```
   openssl x509 -inform der -in certificate.cer -out certificate.pem
   ```

2. Run the following command to convert the private key.

   ```
   openssl rsa -inform DER -outform PEM -in privatekey.der -out
   privatekey.pem
   ```

Convert P7B to PEM

1. Run the following command to convert the certificate format.

   ```
   openssl pkcs7 -print_certs -in incertificat.p7b -out outcertificate.cer
   ```

2. Obtain the content of `[——-BEGIN CERTIFICATE——-, ——-END CERTIFICATE——-]` in *`outcertificat.cer`* and upload the content as a certificate.

Convert PFX to PEM

1. Run the following command to extract the private key.

   ```
   openssl pkcs12 -in certname.pfx -nocerts -out key.pem -nodes
   ```

2. Run the following command to extract the certificate.

   ```
   openssl pkcs12 -in certname.pfx -nokeys -out cert.pem
   ```

20.6.6 Replace a certificate

When your certificate expires, you can generate and upload a new certificate, and delete the existing one.

Procedure

1. Create and upload a new certificate.

   For more information, see *Generate a certificate* and *Upload a certificate*.

2. Configure the new certificate in HTTPS listener configuration.

   For more information, see *Add a Layer-7 listener*.

3. On the **Certificate Management** page, click the icon of the target certificate and click **Delete Certificate** to delete the certificate.

21 Virtual Private Cloud (VPC)

21.1 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.

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.

21.2 Plan and design your network

When creating a VPC and VSwitches, you must specify the private IP address range for your VPC in the form of CIDR block.

CIDR is a bit- and prefix-based standard used to interpret IP addresses. It combines multiple address blocks to a route entry to facilitate routing. Such address blocks are CIDR blocks.

VPC CIDR block

When planning the CIDR block for a VPC, follow these guidelines:

- You can use the standard private CIDR blocks listed in the following table and their subnets as the VPC CIDR block. Only one CIDR block can be specified for each VPC. The available CIDR blocks are specified in the **`vpc_customer_private_cidr`** parameter in the global configuration of the VPC when planning the configurations of Apsara Stack.

CIDR block	Number of available private IPs
192.168.0.0/16	65,532
172.16.0.0/12	1,048,572
10.0.0.0/8	16,777,212

- When a VPC is created using APIs, the mask length of the VPC CIDR block is between 8 and 24 bits.
- After a VPC is created, its CIDR block cannot be modified.

VSwitch CIDR block

When planning the CIDR block for a VSwitch, follow these guidelines:

- You must specify the IP address range of a VSwitch in form of CIDR block. The mask of a VSwitch CIDR block is 16 to 29 bits long, providing 8 to 65,536 IP addresses.
- The CIDR block of a VSwitch must be a subset of the CIDR block of the home VPC.

Note:

If the CIDR blocks of your VSwitch and VPC are the same, you can create only this VSwitch.

- The first and the last three IP addresses of each VSwitch are reserved by the system. For example, if the CIDR block of a VSwitch is 192.168.1.0/24, 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved by the system.
- In the same VPC, the CIDR block of a VSwitch cannot be the same as the destination CIDR block of a route entry, but can be a subset of it.
- After a VSwitch is created, its CIDR block cannot be modified.

21.3 Quick start

This tutorial guides you quickly build a VPC and create an ECS instance in the VPC.

21.3.1 Log on to the VPC console

Follow the instructions in the section to log on to the VPC console.

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

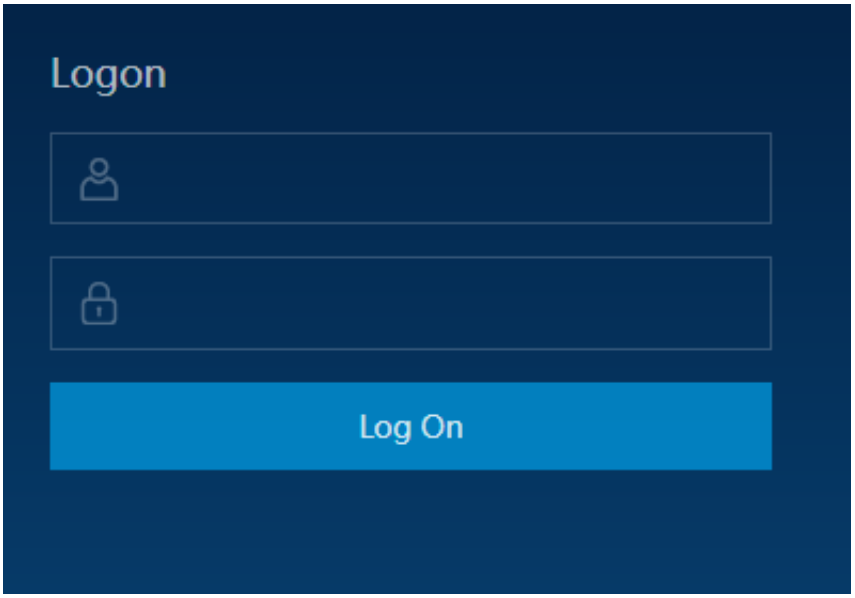
  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 21-1: Log on to the Apsara Stack console*.

Figure 21-1: Log on to the Apsara Stack console



3. Enter the correct username and password.
   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **Log On** to go to the **Home** page.
5. From the top menu, click **Console** > **Virtual Private Cloud**.

21.3.2 Create a VPC and VSwitch

To use cloud products in a VPC, you must create a VPC and a VSwitch.

Context

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** page, click **Create**.

3. Configure the VPC according to the following information and click **Confirm**.

Table 21-1: VPC configuration

Configuration	Description
Name	Enter a name for the VPC. The name can contain 2-128 characters, and must start with an English letter or Chinese character. It cannot contain special characters such as the at sign (@), backslash (/), colon (:), angle brackets (<>), braces ({}), square brackets ([]), and space.
Description	Add a description for the VPC.
Region	Select a region for the VPC.
Department	Select a department for the VPC.
CIDR Block	Select a CIDR block for the VPC. After the VPC is created, its CIDR block cannot be modified.

4. Click **Next**.
5. In the **Create VSwitch** dialog box, configure the VSwitch according to the following information and click **Confirm**.

Table 21-2: VSwitch configuration

Configuration	Description
Zone	Select a zone for the VSwitch. In a VPC, a VSwitch can be in only one zone and cannot span multiple zones. You can deploy cloud resources in VSwitches in different zones to achieve cross-zone disaster tolerance. **Note:** A cloud resource can only be located in one VSwitch.
Name	Enter a name for the VSwitch.
CIDR Block	Enter the CIDR block of the VSwitch. Follow these guidelines when specifying the CIDR block: • You must specify the IP address range of a VSwitch in form of CIDR block. The mask of a VSwitch CIDR block is 16 to 29 bits long, providing 8 to 65,536 IP addresses. • The CIDR block of a VSwitch must be a subset of the CIDR block of the home VPC.

Configuration	Description
	Note: If the CIDR blocks of your VSwitch and VPC are the same, you can create only this VSwitch. • The first and the last three IP addresses of each VSwitch are reserved by the system. For example, if the CIDR block of a VSwitch is 192.168.1.0/24, 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved by the system. • In the same VPC, the CIDR block of a VSwitch cannot be the same as the destination CIDR block of a route entry, but can be a subset of it. • After a VSwitch is created, its CIDR block cannot be modified.
Description	Enter the description of the VSwitch.

21.3.3 Create a security group

Before creating an ECS instance in a VPC, you must first create a security group. The security group is an important means for network security isolation and can be used to set network access control for one or more ECS instances.

Procedure

1. Log on to the ECS console.
2. Click the **Security Groups** tab. Then, click **Create Security Group**.
3. On the **Create Security Group** page, configure the security group according to the following information and click **Confirm**.

Table 21-3: Security group configuration

Configuration	Description
Security Group Name	Enter a name for the security group.
Region	Select a region for the security group. Make sure that the region for the security group is the same as that for the VPC.
Department	Select a department for the security group. Make sure that the department for the security group is the same as that for the VPC.
Description	Add a description for the security group.

Configuration	Description
Project	Select a project for the security group. Make sure that the project for the security group is the same as that for the VPC.
Network Type	Select **VPC**.
VPC	Select the created VPC for the security group.

21.3.4 Create an ECS instance

After creating a security group and VSwitch, you can create an ECS instance in the VSwitch to deploy your applications.

Procedure

1. Log on to the ECS console.
2. Click the **Instances** tab. Then, click **Create Instance**.
3. On the **Elastic Compute Service (ECS)** page, configure the ECS instance and click **Create**.

> **Note:**
> Set the network type to **VPC**, and then select the created VPC and VSwitch.

21.4 VPC

Virtual Private Cloud (VPC) is an isolated network environment built based on Apsara Cloud. You have full control over your own VPC instance, including choosing the IP address range and configuring the route tables and gateways. You can also use Alibaba Cloud resources, such as ECS, ApsaraDB for RDS, and Server Load Balancer instances, in your VPC instance.

21.4.1 Create a VPC

You must create a VPC and a VSwitch to use cloud products.

Context

Note the following before creating a VPC:

- You can specify only one CIDR block for each VPC. For more information, see *Plan and design your network*.
- After a VPC is created, a VRouter and a route table are created automatically. Each VPC contains only one router and one route table.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** page, click **Create**.
3. Configure the VPC according to the following information and click **Confirm**.

   Table 21-4: VPC configuration

Configuration	Description
Name	Enter a name for the VPC. The name can contain 2-128 characters, and must start with an English letter or Chinese character. It cannot contain special characters such as the at sign (@), backslash (/), colon (:), angle brackets (<>), braces ({}), square brackets ([]), and space.
Description	Add a description for the VPC.
Region	Select a region for the VPC.
Department	Select a department for the VPC.
CIDR Block	Select a CIDR block for the VPC. After the VPC is created, its CIDR block cannot be modified.

4. Click **Next** to create a VSwitch.

   For more information, see *Create a VSwitch*.

21.4.2 View a VPC

On the VPC console, you get the detailed information about a VPC such as the IP address range of the VPC, the VSwitch and VRouter of the VPC, and cloud resources deployed in the VPC, and so on.

Procedure

1. *Log on to the VPC console*.
2. Locate the target VPC, click its ID or the icon, and click **Details** to view the VPC details.

21.4.3 Modify a VPC

After creating a VPC, you can modify the name and description of this VPC.

Procedure

1. *Log on to the VPC console*.
2. Locate the target VPC instance, click the icon, and then click **Edit**.

3. In the **Modify VPC** dialog box, enter the new name and description of the VPC instance, and then click **Confirm**.

21.4.4 Delete a VPC

You can delete a VPC when you no longer need it.

Prerequisites

Before deleting a VPC, you must release or move all resources, including VSwitches, from the VPC.

After the VPC is deleted:

- The security group under this VPC instance is deleted as well.
- Related data cannot be restored.

Procedure

1. *Log on to the VPC console*.
2. Locate the target VPC instance, click the icon, and then click **Delete**.
3. In the displayed dialog box, click **Confirm**.

21.5 VSwitch

A VSwitch is a basic network device of a VPC. After a VPC is created, you can divide the VPC into several subnets by adding VSwitches. A maximum of 24 VSwitches can be configured for a VPC.

In a VPC, a VSwitch can be in only one zone and cannot span multiple zones. You can deploy cloud resources in VSwitches in different zones to achieve cross-zone disaster tolerance.

Note:
VSwitches do not support multicasting or broadcasting.

21.5.1 Create a VSwitch

A VSwitch is a basic network device of a VPC. After a VPC is created, you can divide the VPC into several subnets by adding VSwitches and then deploy cloud resources in the VSwitches.

Context

When creating a VSwitch, note:

- You can create a maximum of 24 VSwitches for a VPC.

- After a VSwitch is created, the system automatically adds a system route entry that uses the VSwitch CIDR block as the destination CIDR block.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** page, click the ID of the target VPC.
3. On the page of VPC details, click **VSwitch**.
4. Click **Create** in the upper right corner.
5. In the **Create VSwitch** dialog box, provide the following information and click **Confirm**.

Table 21-5: VSwitch configuration

Configuration	Description
Zone	Select a zone for the VSwitch. In a VPC, a VSwitch can be located in only one zone and cannot span across several zones. You can deploy the cloud product instances in VSwitches in different zones to implement cross-zone disaster recovery. **Note:** A cloud product instance can be added to only one VSwitch.
Name	Enter the name of the VSwitch. The VPC name is a string of 2 to 128 English characters. It must start with an uppercase/lowercase letter and can contain numbers, underlines (_), and hyphens (-).
CIDR Block	Enter the CIDR block of the VSwitch. • You must specify the IP address range of a VSwitch in form of CIDR block. The mask of a VSwitch CIDR block is 16 to 29 bits long, providing 8 to 65,536 IP addresses. • The CIDR block of a VSwitch must be a subset of the CIDR block of the home VPC. **Note:** If the CIDR blocks of your VSwitch and VPC are the same, you can create only this VSwitch. • The first and the last three IP addresses of each VSwitch are reserved by the system. For example, if the CIDR block of a VSwitch is 192.168.1.0/24, 192.168.1.0, 192.168.1.253, 192.168.1.254, and 192.168.1.255 are reserved by the system.

Configuration	Description
	• In the same VPC, the CIDR block of a VSwitch cannot be the same as the destination CIDR block of a route entry, but can be a subset of it. • After a VSwitch is created, its CIDR block cannot be modified.
Description	Enter the description of the VSwitch.

21.5.2 View a VSwitch

On the VPC console, you get the detailed information about a VSwitch such as the IP address range, the status and the zone of the VSwitch.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** page, click the ID of the target VPC.
3. On the VPC details page, click the **VSwitch** tab to view VSwitches.

21.5.3 Modify a VSwitch

After creating a VSwitch, you can modify the name and description of this VSwitch.

Procedure

1. *Log on to the VPC console*.
2. Click the ID of the target VPC.
3. On the VPC details page, click **VSwitch**.
4. Locate the target VSwitch, click the icon, and then click **Edit**.
5. In the displayed dialog box, modify the name and description of the VSwitch, and click **Confirm**.

21.5.4 Delete a VSwitch

You can delete a created VSwitch.

Prerequisites

Before deleting a VSwitch, you must release or move cloud products from this VSwitch.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** page, click the ID of the target VPC.

3. On the VPC details page, click **VSwitch**.
4. Locate the target VSwitch, click the icon, and then click **Delete**.
5. In the displayed dialog box, click **Confirm**.

21.6 VRouter and route table

A VRouter is the hub in a VPC. As an important component of a VPC, a VRouter can connect all VSwitches in the VPC. It also serves as a gateway that connects the VPC to other networks.

A VRouter is automatically created after a VPC is created. When the VPC is deleted, the VRouter is also deleted. A VRouter cannot be created or deleted directly. Each VRouter includes a route table. A VRouter forwards network traffic according to the route entries in the route table.

A route table refers to a list of route entries in the VRouter. A route table is automatically created for a VPC when the VPC is created. When the VPC is deleted, the route table is also deleted. A route table cannot be created or deleted directly.

Each item in a route table is a route entry, which defines the next hop address for the network traffic to be routed to the specified destination CIDR block. The route entries include the system routes and custom routes:

- System route entries

  After you create a VPC, a system route entry destined for 100.64.0.0/10 is automatically created, which is used for the intercommunication of the cloud products in the VPC. Additional ly, a system route entry destined for the CIDR block of the VSwitch is also automatically created after you create a VSwitch.
- Custom route entries

  You can forward the traffic to a specified next hop by adding a custom route entry. For more information, see *Add a custom route entry*.

21.6.1 View a VRouter

On the VPC console, you get the detailed information about a VRouter such as the ID of the VRouter and the route entries in the route table.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** page, click the ID of the target VPC.
3. On the VPC details page, click **VRouter** to view the route table.

21.6.2 Add a custom route entry

Each item in the route table is a route entry. A route entry defines the next hop address for the network traffic pointing to the destination CIDR block. You can add a custom route entry to route the traffic to a specified destination.

Procedure

1. *Log on to the VPC console*.
2. On the **VPC** page, click the ID of the target VPC.
3. On the VPC details page, click **VRouter**.
4. Click **Create** in the **Route Table** area.
5. In the displayed dialog box, configure the route entry according to the following information and click **Confirm**.

Table 21-6: Route entry configuration

Configuration	Description
Destination CIDR Block	Enter the destination CIDR block of the route entry. • The destination CIDR block of a route entry cannot be same as the CIDR block of a VSwitch in the VPC or a subset of the CIDR block of the VSwitch. • The destination CIDR block of a route entry cannot be 100.64.0.0/10 or a subset of 100.64.0.0/10. • The destination CIDR blocks of route entries in the same route table cannot be the same. • If the specified destination CIDR block is an IP address, the default subnet mask 32 is used.
Next Hop Type	Select the next hop type of the route entry. • **ECS Instance**: Forward requests from the destination CIDR block to an ECS instance. • **Router Interface**: Forward requests from the destination CIDR block to a specified router interface. Requests will be routed to the peer router interface through this router interface.
Next hop instance ID	If the next hop type is set to **ECS Instance**, select the ECS instance that receives the forwarded traffic. Note the following: • The next-hop ECS instance specified in a route entry must belong to the same VPC as the route table. • Multiple route entries can be directed to the same ECS instance.

Configuration	Description
Router Interface	If the next hop type is set to **Router Interface**, select the router interface that receives the forwarded traffic.

22 Log Service (Log)

22.1 What is Log Service?

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

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

22.2 Log on to the Log Service Console

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

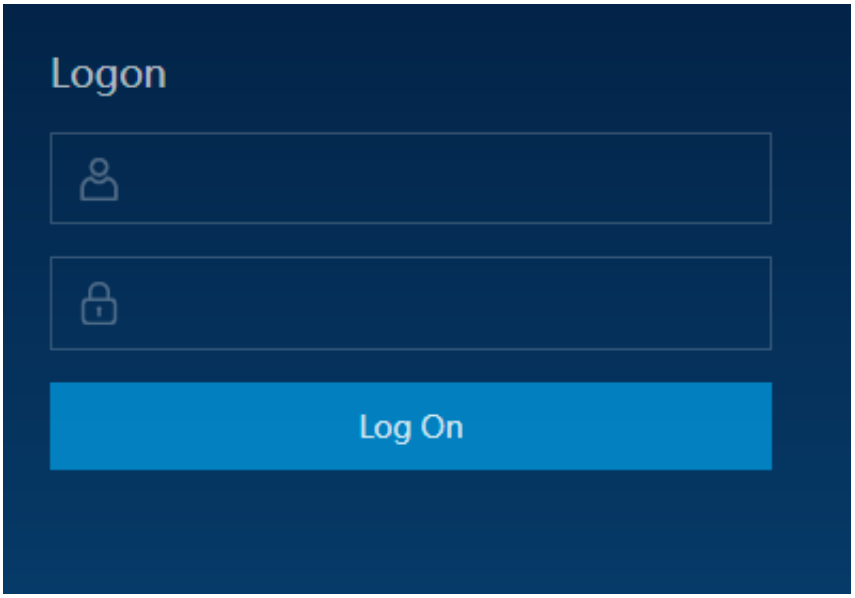
  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 22-1: Log on to the Apsara Stack console*.

Figure 22-1: Log on to the Apsara Stack console



3. Enter the correct username and password.
   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **Log On** to go to the **Home** page.
5. On the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Log Service**.
6. Set **Region** and **Department**, and then click **SLS** to log on to the Log Service console.

22.3 Preparations

22.3.1 Preparation

Log Service provides multiple log collection methods. You can use Log Service to collect Elastic Compute Service (ECS) logs, local server logs, IoT device logs, and other cloud product logs.

Procedure

1. View Access Key.

   Access Keys are essential for you to use the Log Service. Ensure that your account has a pair of access keys. View Access Key information of your account. See *View the key pair*.

2. Create a project.

   You can also click **Create Project** in the top right corner on the homepage to create a project.

   You can modify project annotations and delete projects. For more details, see *Operate on projects*.

3. Create a Logstore.

   When the project is created, the system prompts you to create a Logstore. You can also open the project and click **Create** in the top right corner to create a Logstore.

   You can also modify and delete Logstores. For more details, see *Operate on Logstores*.

4. Manage shards (optional)

   When creating a Logstore, you can select the number of shards based on the volume and generation speed of your logs. You can also change the number of shards by splitting or merging shards when modifying the Logstore.

   For more information about how to merge and split shards, see *Operate on shards*.

22.3.2 View the key pair

AccessKey is a requirement for Log Service operations through APIs/SKDs.

Procedure

1. *Log on to the Log Service Console*.
2. On the top navigation bar, click the user name and select **Personal Information**.
3. In the left-side navigation pane, select **AccessKey**.
4. In the dialog box that appears, click **Confirm**.

   Now you can view the AccessKey ID and AccessKey Secret of the current account.

22.3.3 Operate on projects

You can create and delete projects on the Log Service console.

Create a project

> **Note:**
>
> - The name of a project must be globally unique. If the project name you select is already in use, the message **Project XXX already exists** is displayed. Then use another name.
> - Each department can have up to 10 projects.

1. *Log on to the Log Service Console*.
2. Click **Create Project** in the upper-right corner.
3. Enter **Project Name** and **Region**, and click **Confirm**.

Configuration	Description
Project Name	A project name can contain only lowercase letters, numbers, and hyphens (-). It must start and end with a lowercase letter or number and must be 3 to 63 bytes in length. **Note:** The name of a project cannot be modified after the project is created.
Description	Enter the brief description about the project, which will be displayed on the **Projects** page. If you want to modify the description after the project is created, go to the **Projects** page and click **Modify Description**.
Region	Select a region based on the Apsara Stack deployment. The region cannot be changed after a project is created.

Delete a project

In some cases (for example, to disable a log service instance or destroy all logs in a project), you may need to delete the entire project. Log Service allows you to conveniently delete a project on the console.

> **Note:**

The deletion of a project permanently releases all its logs and configurations. The data is irrecoverable. To avoid data loss, make sure you want to delete the project before the operation.

1. Locate the project you want to delete on the **Projects** page.
2. Click **Delete** on the right.
3. In the dialog box that appears, click **Confirm**.

22.3.4 Operate on Logstores

You can use the Log Service console or APIs to create a Logstore. For more information about how to create a Logstore through APIs, see **Create Logstores** in *Cite LeftLog Service API ReferenceCite Right*.

A Logstore is a set of resources created under a project. All the data in a Logstore comes from the same source. A Logstore is a unit for querying, analyzing, and shipping the collected logs.

Create a Logstore

Note:

- A Logstore must be created under a certain project.
- Each project can have up to 100 Logstores.
- The name of a Logstore must be unique in the project to which it belongs.
- The retention period of logs can be modified after a Logstore is created. On the Logstores page, click **Modify** in the **Actions** column. Modify **Data Retention Period**, click **Modify**, and close the dialog box.

1. *Log on to the Log Service Console*.
2. Click a project name to go to the **Logstores** page. Click **Create** to create a Logstore.
3. Enter configurations about the Logstore and click **Confirm**.

Configuration	Description
Logstore Name	A Logstore name can only contain lowercase letters, numbers, hyphens (-), and underscores (_). It must start and end with a lowercase letter or number and must be 3 to 63 bytes in length. A LogStore name must be unique in the project to which it belongs. **Note:** A Logstore name cannot be modified after creation.

Configuration	Description
Data Retention Period	Specifies the number of days for which the data will be retained in the Logstore. Value range: 1 to 365. Data not within the time period will be deleted.
Shards	Specifies the number of shards in the Logstore. You can create 1 to 10 shards for each Logstore. You can create up to 100 shards for each project.

Modify Logstore configurations

After a Logstore is created, you can modify the Logstore configurations as needed.

1. *Log on to the Log Service Console*.
2. Click the name of a project.
3. On the **Logstores** page, select the expected Logstore and click **Modify** in the **Actions** column.
4. In the dialog box that appears, modify the Logstore configurations. Close the dialog box.

   For the adjustment of shards, see *Operate on shards*.

Delete a Logstore

You may need to delete a Logstore in certain cases, for example, to discard the Logstore. Log Service allows you to delete Logstores on the console.

Note:

- After a Logstore is deleted, its logs will be lost permanently and cannot be recovered, and a Logstore with the same name cannot be created. Exercise caution when performing this operation.
- Before deleting a Logstore, you must delete all its Logtail configurations.

1. *Log on to the Log Service Console*.
2. Click the name of a project.
3. On the **LogStores** page, select the LogStore you want to delete and click **Delete** on the right.
4. In the dialog box that appears, click **Confirm**.

22.3.5 Operate on shards

Logstore read/write logs must be saved in a certain shard. Each Logstore has several shards. When creating a Logstore, you must specify its number of shards. After the Logstore is created, you can split or merge shards to increase or reduce shards.

Split shards

Each shard can write data at 5 MB/s and read data at 10 MB/s. When the data traffic exceeds the shard service capability, we recommend that you add shards immediately. A shard can be scaled up through the split operation.

To perform the split operation, specify a shard ID in the read/write state and a MD5. The MD5 must be greater than the shard BeginKey and smaller than the shard EndKey.

Another two shards can be split from a shard. That is, the number of shards is increased by 2 after split. After split, the original shard transitions from the read/write state to the read-only state. Data in the shard can still be consumed, but new data cannot be written to the shard. The two new shards are in the read/write state and follow the original shard. The MD5 range of the new shards covers that of the original shard.

1. *Log on to the Log Service Console*.
2. Click the name of a project.
3. On the **Logstores** page, select a Logstore and click **Modify** in the **Actions** column.
4. Select the shard to be split, and click **Split** on the right.
5. Click **Confirm** and close the dialog box.

   After split, the original shard enters the read-only state, and the MD5 range of the new shards covers that of the original shard.

Merge shards

A shard can be scaled down through the merge operation. The merge operation combines the ranges of a specified shard and its following shard and assigns the combined range to a new shard in the read/write state. The original two shards enter the read-only state.

You must specify a shard (except the last one) in the read/write state. The server finds the shard that follows the specified shard and merges the ranges of the two shards. After merging, the specified shard and the following shard enter the read-only state. Data in the two shards can still be consumed, but new data cannot be written to them. A shard in the read/write state is generated , and its MD5 range covers the MD5 ranges of the original two shards.

Procedure

1. *Log on to the Log Service Console*.
2. Click the name of a project.
3. On the **Logstores** page, select the expected Logstore and click **Modify** in the **Actions** column.

4. Select the shard to be merged and click **Merge** on the right. Then, close the dialog box.

   After the merge operation is completed, the specified shard and the following shard enter the read-only state, and the data can still be consumed. A shard in the read/write state is generated, and its MD5 range covers the MD5 ranges of the original two shards.

22.4 Data collection

22.4.1 Producer Library

LogHub Producer Library is a write LogHub class library for Java applications with high concurrency. Both Producer Library and Consumer Library package LogHub read and write requests to lower the threshold for data collection and consumption.

Features

- Provides an asynchronous sending interface, thus ensuring thread security.
- Adds configurations for multiple projects.
- Configures the number of network IO threads used for sending.
- Configures the number and size of logs merged into a package.
- Controllable memory usage. In other words, when the memory usage reaches the threshold you configured, the send interface of Producer will be blocked until idle memory is available.

Benefits

- Logs collected by agents not flushed into a disk: Data is directly sent to the server through the network after the data is generated.
- High-concurrency write operations on the agent: For example, there are more than one hundred write operations within one second.
- Agent computing logically separated from I/O: Logging does not affect the computing time used .

In the above scenarios, Producer Library helps you reduce program development costs, aggregate multiple write requests, and asynchronously send them to the LogHub server. During the process, you can configure parameters for batch aggregation, the server exception processing logic, and so on.

Comparison of the above access methods:

Access method	Advantage/Disadvantage	Target scenario
SDK direct transmission	Logs are not flushed into a disk, but are directly sent to the server. Switching between the network I/O and program I/O needs to be properly processed.	Logs are not flushed into a disk.
Producer Library	Logs are not flushed into a disk, but are combined and asynchronously sent to the server at a large throughput.	Logs are not flushed into a disk. The QPS is high on the agent.

Procedure

- *Java Producer*
- *Log4J1.XAppender (based on Java Producer)*
- *Log4J2.XAppender (based on Java Producer)*
- *LogBack Appender (based on Java Producer)*
- *C Producer*
- *C Producer Lite*

22.4.2 Use LogStash to collect logs

Log Service supports collection of server logs through LogStash and upload of data to Log Service through a plug-in.

At present, Log Service supports collection of logs through APIs, SDKs and LogStash. As an open source log management tool, LogStash can quickly collect and define distributed and diversified logs, and transmit them to a specified location, for example, a server or file. You can install LogStash and plug-ins on ECS instances and IDC machines or virtual machines of other cloud vendors and perform a simple configuration to easily migrate server logs to the cloud.

After you install LogStash and related plug-ins on the machines, and configure file directories and Log Service projects or Logstores, LogStash automatically traces changes of log files, collects log files in real time, parses the log files, and sends them to Log Service.

Log Service supports entry of data through LogStash. It provides the following features:

- Collect logs of various types on machines and support data sources such as files, TCP, and Syslog.
- Access the account security system and support data signature transmission and access permission control through secret key pairs.

- Support batch transmission of logs to reduce TPS costs arising out of entry into Log Service.
- Compress log data and enter the data to Log Service to reduce the occupied network egress bandwidth.

22.4.2.1 Quick installation

You can select a default mode to quickly install LogStash on your server.

Context

Log Service provides an installation package based on LogStash 2.2.2, which integrates JRE 1.8, Log Service write plug-in, and NSSM 2.24. The deployment procedure using this package is simpler than *Custom Installation*. If you have complex requirements, you can select custom installation.

Procedure

1. Download the *installation package* and decompress it on drive C.
2. Check that the program start path of LogStash is `C:\logstash-2.2.2-win\bin\logstash.bat`.

22.4.2.2 Custom installation

You can install LogStash by custom and configure installation items as required during installation.

Context

When you have other requirements for installation and configuration of LogStash, you can select custom installation to modify the default installation and configuration.

Procedure

1. Install Java.
   1. Download the installation package.

      Go to *Java website*, download the JDK, and double-click the JDK to install it.
   2. Set environment variables.

      Add or modify environment variables in advanced system settings.

      - **PATH:** `C:\Program Files\Java\jdk1.8.0_73\bin`
      - **CLASSPATH:** `C:\Program Files\Java\jdk1.8.0_73\lib;C:\Program Files\Java\jdk1.8.0_73\lib\tools.jar`
      - **JAVA_HOME:** `C:\Program Files\Java\jdk1.8.0_73`
   3. Perform verification.

Run `PowerShell` or `cmd.exe` for verification.

```
PS C:\Users\Administrator> java -version
java version "1.8.0_73"
Java(TM) SE Runtime Environment (build 1.8.0_73-b02)
Java HotSpot(TM) 64-Bit Server VM (build 25.73-b02, mixed mode)
PS C:\Users\Administrator> javac -version
javac 1.8.0_73
```

2. Install LogStash.

   1. Download the installation package.

      Download LogStash from the official website: *On the LogStash homepage*, select LogStash 2.2 or a later version.

   2. Install LogStash.

      Decompress *logstash-2.2.2.zip* to the *C:\logstash-2.2.2* directory.

      Check that the program start path of LogStash is *C:\logstash-2.2.2\bin\logstash.bat*.

3. Install the plug-in used by LogStash to write logs to Log Service.

   Install the plug-in online or offline based on the network environment of the machine.

   - Online installation.

     The plug-in is hosted on RubyGems. For more information, click *https://rubygems.org/gems/logstash-output-logservice*.

     Run `PowerShell` or `cmd.exe` to go to the LogStash installation directory. Run the following command to install LogStash:

     ```
     PS C:\logstash-2.2.2> .\bin\plugin install logstash-output-
     logservice
     ```

   - Offline installation.

     Download LogStash from the official website: Go to the *logstash-output-logservice page* and click **Download** in the lower right corner.

     If you cannot access the Internet on the machine on which logs are collected, copy the downloaded gem package to the *C:\logstash-2.2.2* directory on the machine. Run

PowerShell or *cmd.exe* to go to the LogStash installation directory. Run the following command to install LogStash:

```
PS C:\logstash-2.2.2> .\bin\plugin install C:\logstash-2.2.2\
logstash-output-logservice-0.2.0.gem
```

- Perform verification.

```
PS C:\logstash-2.2.2> .\bin\plugin list
```

LogStash can be found in the list of plug-ins installed on the local computer. logstash-output -logservice.

4. Install NSSM.

Download NSSM from the official website: Go to *NSSM official website* and download NSSM.

After you download the installation package to the local computer, decompress it to *C:\logstash-2.2.2\nssm-2.24*.

22.4.2.3 Set LogStash to a Windows service

For convenient automatic log collection, you can set LogStash to a Windows service to maintain running of LogStash in the background and automatic start upon startup of the computer.

Context

Start *logstash.bat* under PowerShell. The LogStash process works in the foreground. It is generally used for configuration test and collection commissioning. You are advised to set LogStash to a Windows service to maintain running of LogStash in the background and automatic start upon startup of the computer.

In addition, you can start, stop, modify, and delete a service through the command line. For more information about how to use NSSM, see *NSSM official documents*.

Add the LogStash service

This operation is generally performed when LogStash is deployed the first time. Skip this operation if LogStash is added.

Run the following commands to add the LogStash service:

- 32-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe install logstash "C:\
logstash-2.2.2-win\bin\logstash.bat" "agent -f C:\logstash-2.2.2-win
\conf"
```

- 64-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe install logstash "C:\
logstash-2.2.2-win\bin\logstash.bat" "agent -f C:\logstash-2.2.2-win
\conf"
```

Start the LogStash service

If the configuration file is updated in the *conf* directory of LogStash, stop the LogStash service first. Then start the LogStash service.

Run the following commands to start the LogStash service:

- 32-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe start logstash
```

- 64-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe start logstash
```

Stop the LogStash service

Run the following commands to stop the LogStash service:

- 32-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe stop logstash
```

- 64-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe stop logstash
```

Modify the LogStash service

Run the following commands to modify the LogStash service:

- 32-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe edit logstash
```

- 64-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe edit logstash
```

Delete the LogStash service

Run the following commands to delete the LogStash service:

- 32-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win32\nssm.exe remove logstash
```

- 64-bit system

```
C:\logstash-2.2.2-win\nssm-2.24\win64\nssm.exe remove logstash
```

22.4.2.4 Create a LogStash collection configuration

Related plug-ins

- **logstash-input-file**

  The plug-in is used to collect log files in tail mode. For more information, see *logstash-input-file*.

  Note:
  path indicates the file path, which must use Unix separators, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.

- **logstash-output-logservice**

  You can use the plug-in to collect logs to Log Service.

Parameter	Description
endpoint	Entry of Log Service, for example, *http://regionid.example.com*.
project	Name of a Log Service project.
logstore	Logstore name.
topic	Log topic name. By default, the log topic is set to null.
source	Log source. If this parameter is set to null, the IP address of the local computer is automatically used. Otherwise, the set value prevails.

Parameter	Description
access_key_id	Account secret key ID.
access_key_secret	Account secret key.
max_send_retry	Maximum number of retries when packets cannot be transmitted to Log Service due to an exception. If retry fails, packets are discarded. The retry interval is 200 ms.

Procedure

1. Create a collection configuration.

   After you add a configuration file to the `C:\logstash-2.2.2-win\conf\` directory, restart LogStash to make the configuration file take effect.

   You can create a configuration file in the format `*.conf` for each type of logs. You are advised to save the configuration file in the `C:\logstash-2.2.2-win\conf\` directory for convenient management.

   Note:
   The configuration file must be UTF-8 encoded without BOM. You can modify the file encoding format by using Notepad++.

   - IIS log

     See *Use LogStash to collect IIS logs*.

   - CSV log

     The system time when logs are collected is used as the log upload time. For details, see *Use LogStash to collect CSV logs*.

   - Default log time

     Take the format of CSV logs as an example. The time in log content is used as the log upload time. For details, see *Use LogStash to collect CSV logs*.

   - General log

     By default, the system time when logs are collected is used as the uploaded log time. Log fields are not parsed. Single-line logs and multiline logs are supported. For details, see *Use LogStash to collect other logs*.

2. Verify configuration syntax.

1. Run `PowerShell` or `cmd.exe` to go to the LogStash installation directory. Run the following command to verify the configuration:

   ```
   PS C:\logstash-2.2.2-win\bin> .\logstash.bat agent --configtest --
   config C:\logstash-2.2.2-win\conf\iis_log.conf
   ```

2. Modify the collection configuration file. Add the temporary configuration item rubydebug in the output phase to output collected results to the console. During configuration, set the type field.

   ```
   output { if [type] == "***" { stdout { codec => rubydebug }
   logservice { ... } } }
   ```

3. Start PowerShell or cmd.exe, change the current directory to the installation directory of Logstash, and run `PowerShell` or `cmd.exe`. Run the following command:

   ```
   PS C:\logstash-2.2.2-win\bin> .\logstash.bat agent -f C:\logstash-
   2.2.2-win\conf
   ```

After verification is complete, end the *`logstash.bat`* process and delete the temporary configuration item rubydebug.

What's next

Start *`logstash.bat`* under PowerShell. The Logstash process works in the foreground. It is generally used for configuration test and collection commissioning. You are advised to set LogStash to a Windows service so as to maintain running of LogStash in the background and automatic start upon startup of the computer. For more information about how to set LogStash to a Windows service, see *Set LogStash to a Windows service*.

22.4.2.5 Advanced functions

LogStash provides <xref href="https://www.elastic.co/guide/en/logstash/current/index.html"format ="html" scope="external">lots of plug-ins</xref> to meet personalized requirements, for example:

- *grok*: Parse log content into multiple fields through the structure of regular expression.
- <xref href="https://www.elastic.co/guide/en/logstash/current/plugins-codecs-json_lines.html"format="html" scope="external">json_lines</xref>, *json*: Support structured parsing of JSON logs.
- *date*: Support parsing and conversion of fields related to date and time.
- *multiline*: Define more complex multi-line logs.
- *kv*: Support structured parsing of Key-Value logs.

22.4.2.6 LogStash error handling

After LogStash is configured to collect logs, if an error occurs during log collection, select a handling method based on the error type.

If the following collection errors occur when LogStash collects logs, handle the errors according to corresponding suggestions:

- Garbled characters are found in Log Service.

  By default, LogStash supports UTF-8 encoding. Check whether input files are correctly encoded.

- Errors displayed on the console

  When the following error is displayed on the console: `io/console not supported; tty will not be manipulated`, if product functions are not affected, ignore the error.

  For other errors, you are advised to refer to Google or LogStash forum for help.

22.4.3 Log4j Appender

Loghub Log4j Appender

Apache Log4j is an open source project which allows you to set the log output destination to the console, file, GUI component, socket server, NT event recorder, and Unix Syslog daemon. You can set the output format and level of each log to control log generation at a smaller granularity. Configurations can be performed using a configuration file without the need to modify the code of applications.

Alibaba Cloud Log Log4j Appender enables you to set the log output destination to Log Service. For the download address and usage, see *Github*.

22.4.4 C Producer Library

LogHub not only supports the Java version of Producer Library, but also supports the C version of Producer Library and Producer Lite Library, providing you with a concise, high-performance, and low-resource-consumption one-stop log collection solution across platforms.

For the project address on GitHub and more details, see

- *C Producer Library (recommended for the server)*
- *C Producer Lite Library (recommended for IoT and smart devices)*

22.4.5 Common log formats

22.4.5.1 Apache logs

The Apache log format and directory are specified in the */etc/apache2/httpd.conf* configuration file.

Log format

The "combined" and "common" log formats are defined in the Apache log configuration file.

- Combined format:

```
LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}
i\"" combined
```

- Common format:

```
LogFormat "%h %l %u %t \"%r\" %>s %b"
```

The following statement writes logs into the specified file in the "combined" format.

```
CustomLog "/var/log/apache2/access_log" combined
```

Field description

Field format	Meaning
%a	remote_ip
%A	local_ip
%B	size
%b	size
%D	time_taken_ms
%h	remote_host
%H	protocol
%I	ident
%m	method
%p	port
%P	pid
“%q”	url_query
“%r”	request
%s	status

Field format	Meaning
%>s	status
%t	time
%T	time_taken
%u	remote_user
%U	url_stem
%v	server_name
%V	canonical_name
%I	bytes_received
%O	bytes_sent
“%{User-Agent}i”	user_agent
“%{Referer}i”	referer

Sample log

```
192.168.1.2 - - [02/Feb/2016:17:44:13 +0800] "GET /favicon.ico HTTP/1.
1" 404 209 "http://localhost/x1.html" "Mozilla/5.0 (Macintosh; Intel
Mac OS X 10_11_3) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/48.0.
2564.97 Safari/537.36"
```

Configure Logtail to collect Apache logs.

1. Create a project and a Logstore. For detailed instructions, see *Operate on projects* and *Operate on Logstores*.
2. On the **Logstores** page, click the **Data Import Wizard** icon to access the wizard.
3. Select a data type.

   Select **Text File** and click **Next**.
4. Configure the data source.

   1. Enter the configuration name and log path, and set the log collection mode to **Full Mode**.
   2. Enter a log sample and enable **Extract Field**.
   3. Select fields to generate a regular expression, and manually adjust it.

      Log Service can automatically parse the selected sample log. That is, when you select fields, it can automatically generate a regular expression. There may be possible minor changes in the log data format, so you need to click **Manually Input Regular Expression** to

adjust the automatically generated regular expression. This makes it suitable for all the log formats that may be encountered during collection.

After modifying the regular expression, click **Validate**. Extracted results are displayed if the regular expression is correct.

4. Enter keys corresponding to the log extraction results.

   Choose a descriptive field name for each extraction result. For example, choose "time" for the time field. Enable **Use System Time** and click **Next**.

After configuring Logtail, apply the configuration to the machine group to collect Apache logs.

22.4.5.2 Nginx logs

The Nginx log format and directory are specified in the *`/etc/nginx/nginx.conf`* configuration file.

Nginx log format

The configuration file defines the print format of Nginx logs, namely, the main format:

```
log_format main $remote_addr - $remote_user [$time_local] "$request"
 $request_time $request_length  $status $body_bytes_sent "$http_refer
er"  "$http_user_agent";
```

The statement uses the "main" log format and the written file name.

```
access_log /var/logs/nginx/access.log main
```

Field description

Field name	Meaning
remoteaddr	IP address of the agent.
remote_user	User name of the agent.
request	Request URL and HTTP protocol.
status	Request status.
bodybytessent	Number of bytes (not including the size of the response header) sent to the agent. The number of bytes indicated by this variable is the same as that indicated by bytes_sent in modlogconfig of the Apache module.
connection	Serial number of a connection.

Field name	Meaning
connection_requests	Number of requests obtained by one connection
msec	Time when the log is written, which is measured in seconds and accurate to milliseconds
pipe	Whether requests are sent pipelined over HTTP. If yes, the pipe value is p; otherwise, the value is a period (.).
httpreferer	Source page of the access request.
“http_user_agent”	Browser information of the agent, which should be enclosed by double quotation marks.
requestlength	Request length, which includes the request line , request header, and request body.
request_time	Time when the request is processed, which is measured in seconds and accurate to milliseconds. The time starts when the first byte is read from the agent until logs are written after the last character is sent to the agent.
[$time_local]	Local time when the general log format is applied, which must be enclosed by braces.

Sample log

```
192.168.1.2 - - [10/Jul/2015:15:51:09 +0800] "GET /ubuntu.iso HTTP/1.0
" 0.000 129 404 168 "-" "Wget/1.11.4 Red Hat modified"
```

Configure Logtail to collect Nginx logs

1. Create a project and a Logstore. For detailed instructions, see *Operate on projects* and *Operate on Logstores*.
2. On the **Logstores** page, click **Data Import Wizard** to access the wizard.
3. Select a data type.

   Select **Text File** and click **Next**.
4. Select a data type.

   Select **NGINX Access Log** and click **Next**.
5. Configure the data source.

1. Enter **Configuration Name** and **Logs Directory Path** .
2. Enter the Nginx log format.

   Enter the log configuration part of the standard Nginx configuration file. It typically starts with `log_format`. Log Service automatically reads the Nginx key.
3. Configure **Advanced Options** as required and click **Next**.

   For the description of advanced options, see *Advanced options*.

After configuring Logtail, apply the configuration to the machine group to collect Nginx logs.

22.4.5.3 Python log

Python logging module provides a general logging system for use by third-party modules or applications. The logging module provides different log levels and records logs by different methods including file, HTTP GET/POST, SMTP, and Socket. You can customize a log recording method as required. The logging module has the same mechanism as Log4j except that they have different implementation details. The logging module provides the logger, handler, filter, and formatter features.

Python log format

The formatter specifies the log output format. The formatter constructor requires two parameters for construction: message format string and message date string. The parameters are optional.

Python log format:

```
import logging import logging.handlers LOG_FILE = tst.log handler =
logging.handlers.RotatingFileHandler(LOG_FILE, maxBytes = 1024*1024,
backupCount = 5) # Instantiate handler fmt = %(asctime)s - %(filename
)s:%(lineno)s - %(name)s - %(message)s formatter = logging.Formatter
(fmt) # Instantiate formatter handler.setFormatter(formatter) # Add
formatter logger = logging.getLogger(tst) for handler # Obtain logger
logger.addHandler(handler) with name tst # Add handler logger.setLevel
(logging.DEBUG) logger.info(first info message) logger.debug(first
debug message) for logger
```

Field description

The formatter is configured in the `%(key)s` format, that is, substitution of dictionary keywords. The following keywords are provided:

Format	Meaning
%(name)s	Name of the logger that generates logs.

Format	Meaning
%(levelno)s	Numerical log levels, including debug, info, warning, error, and critical.
%(levelname)s	Text log levels, including ‘debug’, ‘info’, ‘ warning’, ‘eerror’, and ‘critical’.
%(pathname)s	Complete path (if available) to the source file that contains the statement for log output.
%(filename)s	File name.
%(module)s	Name of the module that contains the statement for log output.
%(funcName)s	Name of the function that calls the log output function.
%(lineno)d	Line of the code (if available) that contains the statement for calling the log output function.
%(created)f	Log creation time, in the Unix time format, expressed by the number of seconds passed since 1970-1-1 00:00:00 UTC.
%(relativeCreated)d	Difference (in milliseconds) between the log creation time and the loading time of the logging module.
%(asctime)s	Log creation time. The default format is like 2003-07-08 16:49:45,896. The number following the comma (,) indicates the number of milliseconds.
%(msecs)d	Log creation time, in milliseconds.
%(thread)d	Thread ID (if available).
%(threadName)s	Thread name (if available).
%(process)d	Process ID (if available).
%(message)s	Log information,

Sample log

Output sample log:

```
2015-03-04 23:21:59,682 - log_test.py:16 - tst - first info message
2015-03-04 23:21:59,682 - log_test.py:17 - tst - first debug message
```

Configure Logtail to collect Python logs

For more information about how to configure Logtail to collect Python logs, see *Apache logs*.

Set configuration options based on your network deployment and the actual situation.

The automatically generated regular expression is based on the sample log but does not cover every log type. Therefore, you need to tune the regular expression after it is generated.

Common Python logs and their regular expressions:

- Sample log:

  ```
  2016-02-19 11:03:13,410 - test.py:19 - tst - first debug message
  ```

  Regular expression:

  ```
  (\d+-\d+-\d+\s\S+)\s+-\s+([^:]+):(\d+)\s+-\s+(\w+)\s+-\s+(.*)
  ```

- Log format:

  ```
  %(asctime)s - %(filename)s:%(lineno)s - %(levelno)s %(levelname)
  s %(pathname)s %(module)s %(funcName)s %(created)f %(thread)d %(
  threadName)s %(process)d %(name)s - %(message)s
  ```

  Sample log:

  ```
  2016-02-19 11:06:52,514 - test.py:19 - 10 DEBUG test.py test <module
  > 1455851212.514271 139865996687072 MainThread 20193 tst - first
  debug message
  ```

  Regular expression:

  ```
  (\d+-\d+-\d+\s\S+)\s-\s([^:]+):(\d+)\s+-\s+(\d+)\s+(\w+)\s+(\S+)\s
  +(\w+)\s+(\S+)\s+(\S+)\s+(\d+)\s+(\w+)\s+(\d+)\s+(\w+)\s+-\s+(.*)
  ```

22.4.5.4 Log4j log

Access method

Log Service supports the following methods to collect Log4j logs:

- LogHub Log4j Appender
- Logtail

Collect Log4j logs by using Loghub Log4j Appender

For more information, see *Log4j Appender* .

Collect Log4j logs by using Logtail

Log4j logs include Log4j 1 logs and Log4j 2 logs. This document describes how to configure a regular expression for collecting Log4j 1 logs based on the default configurations. To collect Log4j 2 logs, you need to modify the default configurations and print the complete date.

```
<Configuration status="WARN"> <Appenders> <Console name="Console"
target="SYSTEM_OUT"> <PatternLayout pattern="%d{yyyy-MM-dd HH:mm:ss:
SSS zzz} [%t] %-5level %logger{36} - %msg%n"/> </Console> </Appenders
> <Loggers> <Logger name="com.foo.Bar" level="trace"> <AppenderRef ref
="Console"/> </Logger> <Root level="error"> <AppenderRef ref="Console
"/> </Root> </Loggers> </Configuration>
```

For more information about how to configure Logtail to collect Log4j logs, see *Apache logs*. Set configuration options based on your network deployment and the actual situation.

The automatically generated regular expression is based on the sample log but does not cover every log type. Therefore, you need to tune the regular expression after it is generated.

The following shows the sample log in the default format of Log4j that is printed to a file:

```
2013-12-25 19:57:06,954 [10.207.37.161] WARN impl.PermanentTairDaoImpl
 - Fail to Read Permanent Tair,key:e:470217319319741_1,result:com
.example.tair.Result@172e3ebc[rc=code=-1, msg=connection error or
timeout,value=,flag=0]
```

Match the starting line of a multiline log (the beginning of a line is expressed by IP address information):

```
\d+-\d+-\d+\s.*
```

Regular expression used to extract log information:

```
(\d+-\d+-\d+\s\d+:\d+:\d+,\d+)\s\[([^\]]*)\]\s(\S+)\s+(\S+)\s-\s(.*)
```

Time conversion format:

```
%Y-%m-%d %H:%M:%S
```

Sample log extraction result:

Key	Value
time	2013-12-25 19:57:06,954
ip	10.207.37.161

Key	Value
level	WARN
class	impl.PermanentTairDaoImpl
message	Fail to Read Permanent Tair,key:e:470217319319741_1,result:com.example.tair.Result@172e3ebc[rc=code=-1, msg=connection error or timeout,value=,flag=0]

22.4.5.5 Node.js log

By default, Node.js logs are printed to the console, which makes data collection and troubleshooting inconvenient. The Log4js package is provided to print logs to files and customize log formats, which makes data collection and sorting easy.

```
var log4js = require(log4js); log4js.configure({ appenders: [ { type
: file, //Output filename: logs/access.log, maxLogSize: 1024, backups
:3, category: normal } ] }); var logger = log4js.getLogger(normal);
logger.setLevel(INFO); logger.info("this is a info msg"); logger.error
("this is a err msg");
```

Log format

The Log4js log that is output to a text file is shown below:

```
[2016-02-24 17:42:38.946] [INFO] normal - this is a info msg [2016-02-
24 17:42:38.951] [ERROR] normal - this is a err msg
```

Log4js has six output levels: trace, debug, info, warn, error, and fatal, in ascending order of severity.

Use Logtail to collect Node.js logs

For more information about how to configure Logtail to collect Node.js logs, see *Apache logs*. Set configuration options based on your network deployment and the actual situation.

The automatically generated regular expression is based on the sample log but does not cover every log type. Therefore, you need to tune the regular expression after it is generated. You can refer to the following Node.js sample logs to write a correct and complete regular expression for logs.

Common Node.js logs and their regular expressions:

- Node.js sample log 1

- Sample log:

```
[2016-02-24 17:42:38.946] [INFO] normal - this is a info msg
```

- Regular expression:

```
\[([^]]+)]\s\[([^\]]+)]\s(\w+)\s-(.*)
```

- Extracted fields:

  `time`, `level`, `loggerName`, and `message`

- Node.js sample log 2:

  - Sample log:

```
[2016-01-31 12:02:25.844] [INFO] access - 42.120.73.203 - - "GET
 /user/projects/ali_sls_log?ignoreError=true HTTP/1.1" 304 - "http
:// aliyun.com/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_3)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/48.0.2564.97 Safari/
537.36"
```

  - Regular expression:

```
\[([^]]+)]\s\[(\w+)]\s(\w+)\s-\s(\S+)\s-\s-\s"([^"]+)"\s(\d+)
[^"]+("[^"]+)"\s"([^"]+).*
```

  - Extracted fields:

    `time`, `level`, `loggerName`, `ip`、`request`, `status`, `referer`, and `user_agent`

22.4.5.6 WordPress log

Default WordPress log format

Raw sample log:

```
172.64.0.2 - - [07/Jan/2016:21:06:39 +0800] "GET /wp-admin/js/password
-strength-meter.min.js?ver=4.4 HTTP/1.0" 200 776 "http://wordpress
.c4a1a0aecdb1943169555231dcc4adfb7.cn-hangzhou.alicontainer.com/wp-
admin/install.php" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5)
```

```
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537
.36"
```

Match the starting line of a multiline log (the beginning of a line is expressed by IP information):

```
\d+\.\d+\.\d+\.\d+\s-\s.*
```

Regular expression used to extract log information:

```
(\S+) - - \[([^\]]*)] "(\S+) ([^"]+)" (\S+) (\S+) "([^"]+)" "([^"]+)"
```

Time conversion format:

```
%d/%b/%Y:%H:%M:%S
```

Sample log extraction result:

Key	Value
ip	10.10.10.1
time	07/Jan/2016:21:06:39 +0800
method	GET
url	/wp-admin/js/password-strength-meter.min.js?ver=4.4 HTTP/1.0
status	200
length	776
ref	http://wordpress.c4a1a0aecdb1943169555231dcc4adfb7.cn-hangzhou.alicontainer.com/wp-admin/install.php
user-agent	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537.36

22.4.5.7 Delimiter log

Log overview

Delimiter logs use linefeeds as the boundary. Each natural line is a log. The fields of each log are connected by fixed separators, such as tabs (/t), spaces, vertical bars (|), commas (,), semicolons (;), and other single characters. Fields that contain separators are enclosed by a pair of double quotes.

Common delimiter logs include CSV and TSV formatted logs.

Log format

A delimiter log is divided into several fields by separators. The **single-character** mode and **multi-character** mode are supported.

- **Single-character mode**

  In single-character mode, a delimiter log is divided into several fields by single characters, such as tabs (`\t`), spaces, vertical bars (`|`), commas (`,`), and semicolons (`;`).

  > **Note:**
  > Double quotation marks (") are used as quotes for single-character separators, but not as separators.

  When single-character separators are used, fields that contain separators are enclosed by a pair of double quotation marks (") to avoid incorrect field division. Use an escape character preceding a pair of double quotation marks `""` which is not used to enclose a field. Double quotation marks (") are either used separately as quotes on the boundary of a field or used in pair (`""`) as data within a field. For other cases not compliant with the delimiter log format definition, parse fields in other modes, such as the simple mode and regular mode.

  — **Double quotation marks used as quotes**

    When double quotation marks (") are used as quotes, fields that contain separators must be enclosed by a pair of double quotes. Quotes must be located adjacent to the separators. Modify the format if there are spaces, tabs, and other characters between them.

    For example, when commas (`,`) are used as separators and double quotation marks are used as quotes, the log format is:`1997,Ford,E350,"ac, abs, moon",3000.00`. The log can be parsed into five fields: `1997`, `Ford`, `E350`, `ac, abs, moon`, and `3000.00` The quoted `ac, abs, moon` is a complete field.

  — **Double quotation marks as a part of a field**

    When double quotation marks are a part of a field, they are escaped into `""` rather than used as quotes. Restored when the field is parsed, that is, `""` is restored to `"`.

    For example, when commas are used as separators and double quotation marks and commas are a part of a field, enclose the field with a pair of quotes and escape the double quotation marks into `""`. The log format after processing is:

`1999,Chevy,"Venture ""Extended Edition, Very Large""",",5000.00`. The log can be parsed into five fields: `1999`, `Chevy`, `Venture "Extended Edition, Very Large"`, blank field, and `5000.00`

- **Multi-character mode**

  In multi-character mode, a separator contains two or three characters, such as (`||`), (`&&&`), and (`^_^`). Logs are parsed solely based on separators, without the use of quotes for enclosing.

  Note:

  Avoid full match of separators in the fields of a log; otherwise, the log is divided incorrectly.

  For example, when the separator is set to `&&`, the log is `1997&&Ford&&E350&&ac&abs&moon&&3000.00` and parsed into five fields, `1997`, `Ford`, `E350`, `ac&abs&moon`, and `3000.00`.

Sample log

- **Log with single-character separators**

```
05/May/2016:13:30:28,10.10.10.1,"POST /PutData?Category=YunOsAccou
ntOpLog&AccessKeyId=U0Ujpek**********&Date=Fri%2C%2028%20Jun%202013
%2006%3A53%3A30%20GMT&Topic=raw&Signature=pD12XYLmGxKQ%2Bmkd6x7hA
gQ7b1c%3D HTTP/1.1",200,18204,aliyun-sdk-java 05/May/2016:13:31:23
,10.10.10.2,"POST /PutData?Category=YunOsAccountOpLog&AccessKeyId=
U0Ujpek**********&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT
&Topic=raw&Signature=pD12XYLmGxKQ%2Bmkd6x7hAgQ7b1c%3D HTTP/1.1",401,
23472,aliyun-sdk-java
```

- **Log with multi-character separators**

```
05/May/2016:13:30:28&&10.200.98.220&&POST /PutData?Category=
YunOsAccountOpLog&AccessKeyId=U0UjpekFQOVJW45A&Date=Fri%2C%2028%
20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=pD12XYLmGxKQ
%2Bmkd6x7hAgQ7b1c%3D HTTP/1.1&&200&&18204&&aliyun-sdk-java 05/May
/2016:13:31:23&&10.200.98.221&&POST /PutData?Category=YunOsAccou
ntOpLog&AccessKeyId=U0UjpekFQOVJW45A&Date=Fri%2C%2028%20Jun%202013
%2006%3A53%3A30%20GMT&Topic=raw&Signature=pD12XYLmGxKQ%2Bmkd6x7hA
gQ7b1c%3D HTTP/1.1&&401&&23472&&aliyun-sdk-java
```

Configure Logtail to collect delimiter logs

For more information about how to configure Logtail to collect Python logs, see *Apache logs*. Set configuration options based on your network deployment and the actual situation.

1. Create a project and a Logstore. For more information about how to create a project and a Logstore, see *Operate on projects* and *Operate on Logstores*.
2. On the **Logstores** page, click the **Data Import Wizard** icon to start the wizard.

3. Select a data type.

   Select **Text File** and click **Next**.

4. Configure the data source.

   a. Enter the configuration name and log path, and set the log collection mode to **Separator Mode**.

   b. Enter a sample log and select a separator.

      Select a proper separator based the log format. Otherwise, parsing may fail.

   c. Specify the key in the log extraction result.

      After you enter a sample log and select a separator, Log Service extracts fields of the log based on the separator and defines the fields as values. You need to specify keys for the values.

      The preceding sample log uses commas (`,`) as separators and contains six fields. The key values are: time, ip, url, status, latency, and user-agent.

   d. Specify the log time.

      You can use the system time as the time of a log or use a column (for example, the time field 05/May/2016:13:30:29) of the log as the time. For date format setting, see *Text - Configure a time format*.

   e. After the configuration is applied to the machine group, preview logs on the console to check whether logs are successfully collected.

22.4.5.8 JSON log

A JSON-formatted log can be written in two types of structure:

- Object:a collection of name-value pairs
- Array: an ordered list of values

Logtail supports JSON logs of the object type. Logtail automatically extracts the keys and values at the first layer of an object as the names and values of fields respectively. (The field value belongs to the object, array, or basic type, for example, a string or number.)

Logtail does not support automatic parsing of non-object data (for example, JSON arrays). You can use regular expressions for field extraction or use the simple mode for log collection by line.

Sample log

```
{"url": "POST /PutData?Category=YunOsAccountOpLog&AccessKeyId=U0Ujpek
********&Date=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw
&Signature=pD12XYLmGxKQ%2Bmkd6x7hAgQ7b1c%3D HTTP/1.1", "ip": "10.200
.98.220", "user-agent": "aliyun-sdk-java", "request": {"status": "200
", "latency": "18204"}, "time": "05/May/2016:13:30:28"} {"url": "POST
 /PutData?Category=YunOsAccountOpLog&AccessKeyId=U0Ujpek********&Date
=Fri%2C%2028%20Jun%202013%2006%3A53%3A30%20GMT&Topic=raw&Signature=
pD12XYLmGxKQ%2Bmkd6x7hAgQ7b1c%3D HTTP/1.1", "ip": "10.200.98.210", "
user-agent": "aliyun-sdk-java", "request": {"status": "200", "latency
": "10204"}, "time": "05/May/2016:13:30:29"}
```

Configure Logtail to collect JSON logs

For more information about how to configure Logtail to collect JSON logs, see *Apache logs*. Set configuration options based on your network deployment and the actual situation.

1. Create a project and a Logstore. For more information about how to create a project and a Logstore, see *Operate on projects* and *Operate on Logstores*.
2. On the **Logstores** page, click the **Data Import Wizard** icon to start the wizard.
3. Select a data type.

   Select **Text File** and click **Next**.
4. Configure the data source.

   a. Enter the configuration name and log path, and set the log collection mode to **JSON Mode**.

   b. Determine whether to use the system time as the log time based on your requirements. You can choose to enable or disable the **Use system time** feature.

   - Enable **Use system time**.

     The time field in a log is not extracted, and the log time is the time when Log Service collects the log.
   - Disable **Use system time**.

     The time field in a log is extracted to indicate the log time.

     If you disable **Use system time**, you must define a key for the time field to be extracted and define the time conversion format. For example, the time field `time` (05/May/2016:13:30:29) in a JSON object can be extracted to indicate the log time. For more information about how to configure the date format, see *Text - Configure a time format*.
5. After the configuration is applied to the machine group, preview logs on the console to check whether logs are successfully collected.

22.4.5.9 ThinkPHP log

ThinkPHP is a Web application development framework based on the PHP language.

ThinkPHP log format

The log print format in ThinkPHP is as follows:

Sample log

```
[ 2016-05-11T21:03:05+08:00 ] 10.10.10.1 /index.php
INFO: [ app_init ] --START--
INFO: Run Behavior\BuildLiteBehavior [ RunTime:0.000014s ]
INFO: [ app_init ] --END-- [ RunTime:0.000091s ]
INFO: [ app_begin ] --START--
INFO: Run Behavior\ReadHtmlCacheBehavior [ RunTime:0.000038s ]
INFO: [ app_begin ] --END-- [ RunTime:0.000076s ]
INFO: [ view_parse ] --START--
INFO: Run Behavior\ParseTemplateBehavior [ RunTime:0.000068s ]
INFO: [ view_parse ] --END-- [ RunTime:0.000104s ]
INFO: [ view_filter ] --START--
INFO: Run Behavior\WriteHtmlCacheBehavior [ RunTime:0.000032s ]
INFO: [ view_filter ] --END-- [ RunTime:0.000062s ]
INFO: [ app_end ] --START--
INFO: Run Behavior\ShowPageTraceBehavior [ RunTime:0.000032s ]
INFO: [ app_end ] --END-- [ RunTime:0.000070s ]
ERR: D model class not found for method instantiation
```

The log that is printed using this method is as follows:

Configure Logtail to collect ThinkPHP logs

For how to configure Logtail to collect Python logs, see `<xref href="LogService_user_guide_0047.dita"format="dita"/>`. Select configurations based on your network deployment and the actual situation.

The automatically generated regular expression is based on the sample log but does not cover every log type. Therefore, you need to tune the regular expression after it is generated.

ThinkPHP logs are multiline logs in varying modes, and the following fields can be extracted from these logs: time, IP address of the visitor, accessed URL, and printed message. Because the message mode is not fixed, the message is packaged into a field that contains multiple lines of information.

Parameters for configuring Logtail to collect ThinkPHP logs:

Regular expression at the beginning of the line

```
\[\s\d+-\d+-\w+:\d+:\d+\+\d+:\d+\s.*
```

Regular expression:

```
\[\s(\d+-\d+-\w+:\d+:\d+)[^:]+:\d+\s]\s+(\S+)\s(\S+)\s+(.*)
```

Time expression:

```
%Y-%m-%dT%H:%M:%S
```

22.4.5.10 Use LogStash to collect IIS logs

Before using LogStash to collect IIS logs, modify the configuration file to parse IIS log fields.

Sample log

View IIS log configurations, select the W3C format (default field setting), and save the format to put it into effect.

```
2016-02-25 01:27:04 112.74.74.124 GET /goods/list/0/1.html - 80 - 66.
249.65.102 Mozilla/5.0+(compatible;+Googlebot/2.1;++http://www.google.
com/bot.html) 404 0 2 703
```

Collection configuration

```
input { file { type => "iis_log_1" path => ["C:/inetpub/logs/LogFiles
/W3SVC1/*.log"] start_position => "beginning" } } filter { if [type
] == "iis_log_1" { #ignore log comments if [message] =~ "^#" { drop
 {} } grok { # check that fields match your IIS log settings match =>
 ["message", "%{TIMESTAMP_ISO8601:log_timestamp} %{IPORHOST:site} %{
WORD:method} %{URIPATH:page} %{NOTSPACE:querystring} %{NUMBER:port
} %{NOTSPACE:username} %{IPORHOST:clienthost} %{NOTSPACE:useragent}
 %{NUMBER:response} %{NUMBER:subresponse} %{NUMBER:scstatus} %{NUMBER
:time_taken}"] } date { match => [ "log_timestamp", "YYYY-MM-dd HH
:mm:ss" ] timezone => "Etc/UTC" } useragent { source=> "useragent
" prefix=> "browser" } mutate { remove_field => [ "log_timestamp
"] } } } output { if [type] == "iis_log_1" { logservice { codec => "
json" endpoint => "***" project => "***" logstore => "***" topic =>
 "" source => "" access_key_id => "***" access_key_secret => "***"
max_send_retry => 10 } } }
```

Note:

- The configuration file must be UTF-8 encoded without BOM. You can modify the file encoding format by using Notepad++.
- *path* indicates the file path, which must use Unix separators, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.

- The *type* field must be modified and saved in the file. If multiple Logstash configuration files exist on one computer, ensure that the setting of type is unique in the configuration files. Otherwise, data may not be properly processed.

Related plug-ins: *file* and *grok*.

Restart LogStash to make the configuration take effect.

Create a configuration file in the *conf* directory. Restart LogStash to make the setting take effect. For more information about how to restart LogStash, see *Set LogStash to a Windows service*.

22.4.5.11 Use LogStash to collect CSV logs

Before using LogStash to collect CSV logs, modify the configuration file to parse CSV log fields.

Context

The system time when CSV logs are collected or the time in log content can be used as the log upload time. Based on different definitions of log time, configure LogStash to collect CSV logs in two modes.

Use the system time as the log upload time.

- **Sample log**

```
10.116.14.201,-,2/25/2016,11:53:17,W3SVC7,2132,200,0,GET,project/
shenzhen-test/logstore/logstash/detail,C:\test\csv\test_csv.log
```

- **Collection configuration**

```
input { file { type => "csv_log_1" path => ["C:/test/csv/*.log"]
start_position => "beginning" } } filter { if [type] == "csv_log_1
" { csv { separator => "," columns => ["ip", "a", "date", "time",
 "b", "latency", "status", "size", "method", "url", "file"] } } }
output { if [type] == "csv_log_1" { logservice { codec => "json"
endpoint => "***" project => "***" logstore => "***" topic => ""
source => "" access_key_id => "***" access_key_secret => "***"
max_send_retry => 10 } } }
```

Note:

- The configuration file must be UTF-8 encoded without BOM. You can download Notepad++ to modify the file encoding format.
- *path* indicates the file path, which must use Unix separators, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.

- The *type* field must be modified and saved in the file. If multiple LogStash configuration files exist on one computer, ensure that the setting of *type* is unique in the configuration files. Otherwise, data may not be properly processed.

Related plug-ins: *file* and *csv*.

- **Restart LogStash to make the modification take effect.**

Create a configuration file in the *conf* directory. See *Set LogStash to a Windows service*.

Restart LogStash to make the setting take effect.

Use the time in log content as the log upload time.

- **Sample log**

```
10.116.14.201,-,Feb 25 2016 14:03:44,W3SVC7,1332,200,0,GET,project/
shenzhen-test/logstore/logstash/detail,C:\test\csv\test_csv_withtime
.log
```

- **Collection configuration**

```
input { file { type => "csv_log_2" path => ["C:/test/csv_withtime
/*.log"] start_position => "beginning" } } filter { if [type] == "
csv_log_2" { csv { separator => "," columns => ["ip", "a", "datetime
", "b", "latency", "status", "size", "method", "url", "file"] } date
 { match => [ "datetime" , "MMM dd YYYY HH:mm:ss" ] } } } output {
if [type] == "csv_log_2" { logservice { codec => "json" endpoint =>
 "***" project => "***" logstore => "***" topic => "" source => ""
access_key_id => "***" access_key_secret => "***" max_send_retry =>
10 } } }
```

Note:

- The configuration file must be UTF-8 encoded without BOM. You can download Notepad++ to modify the file encoding format.
- *path* indicates the file path, which must use Unix separators, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.
- The *type* field must be modified and saved in the file. If multiple LogStash configuration files exist on one computer, ensure that the setting of *type* is unique in the configuration files. Otherwise, data may not be properly processed.

Related plug-ins: *file* and *csv*.

- **Restart LogStash to make the configuration take effect.**

Create a configuration file in the *conf* directory. Restart LogStash make the configuration take effect. For more information about how to restart LogStash, see *Set LogStash to a Windows service*.

22.4.5.12 Use LogStash to collect other logs

Before using LogStash to collect logs, modify the configuration file to parse log fields.

Use the system time as the log upload time.

- **Sample log**

```
2016-02-25 15:37:01 [main] INFO com.aliyun.sls.test_log4j - single
line log 2016-02-25 15:37:11 [main] ERROR com.aliyun.sls.test_log4j
 - catch exception ! java.lang.ArithmeticException: / by zero at com
.aliyun.sls.test_log4j.divide(test_log4j.java:23) ~[bin/:?] at com.
aliyun.sls.test_log4j.main(test_log4j.java:13) [bin/:?] 2016-02-25
15:38:02 [main] INFO com.aliyun.sls.test_log4j - normal log
```

- **Collection configuration**

```
input { file { type => "common_log_1" path => ["C:/test/multiline
/*.log"] start_position => "beginning" codec => multiline {
pattern => "^\d{4}-\d{2}-\d{2} \d{2}:\d{2}:\d{2}" negate => true
auto_flush_interval => 3 what => previous } } } output { if [type
] == "common_log_1" { logservice { codec => "json" endpoint =>
 "***" project => "***" logstore => "***" topic => "" source => ""
access_key_id => "***" access_key_secret => "***" max_send_retry =>
10 } } }
```

Note:

- The configuration file must be UTF-8 encoded without BOM. You can download Notepad++ to modify the file encoding format.
- *path* indicates the file path, which must use Unix separators, for example, *C:/test/multiline/*.log*. Otherwise, fuzzy match is not supported.
- The *type* field must be modified and saved in the file. If multiple LogStash configuration files exist on one computer, ensure that the setting of *type* is unique in the configuration files. Otherwise, data may not be properly processed.

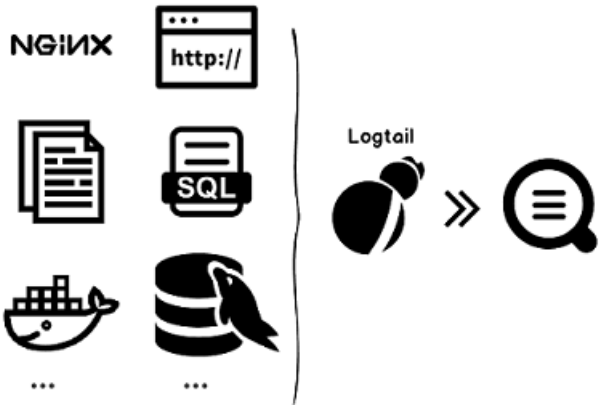
Related plug-ins: *file* and *multiline* (For a single-line log file, remove the codec => multiline line.)

- **Restart LogStash to make the configuration take effect.**

Create a configuration file in the *conf* directory. Restart LogStash make the configuration take effect. For more information about how to restart LogStash, *Set LogStash to a Windows service*.

22.5 Logtail-based collection

Log Service provides a log collection agent: Logtail access service. It allows you to collect logs from ECS servers and other servers on the console in real time. Currently, Log Service of Apsara Stack only allows you to install Logtail on a Linux server. To collect logs from a Windows server, use Logstash.



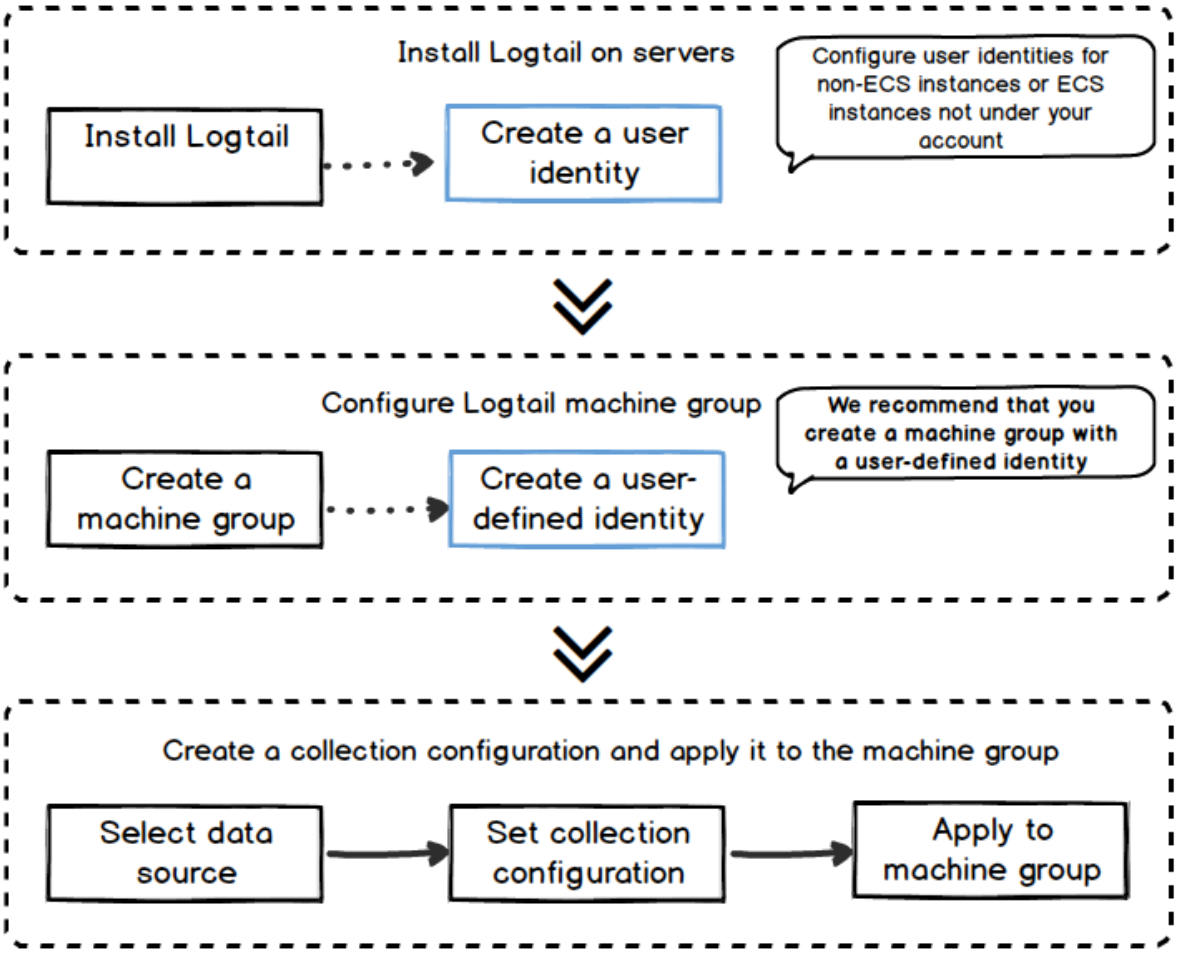
Benefits

- Non-invasive log collection based on log files. You do not need to modify any application code, and log collection does not affect the operating logic of your applications.
- Exception handling in a stable manner during the log collection process. Logtail takes data security measures such as proactive retry and local caching when the network or Log Service has an exception or when user data exceeds the reserved write bandwidth.
- Centralized management capability based on Log Service. After installing Logtail, you only need to configure the devices from which the logs you want to collect and the collection method on the server, rather than logging on to the servers one by one. For Logtail installation, see *Install Logtail on Linux*.
- Comprehensive self-protection mechanism. To ensure that the collection agent running on the client machine does not significantly affect the performance of your services, Logtail provides a protective mechanism and strictly limits its use of CPU, memory, and network resources.

Processing capability and constraints

See *Limits*.

Configuration process



Use Logtail to collect server logs as follows:

1. Install Logtail. Install Logtail on the server from which you want to collect logs. For more information, see *Install Logtail on Linux*.
2. *Create a machine group*. Log Service uses machine groups to manage all servers from which you want to collect logs with Logtail. Log Service allows you to define machine groups through IP addresses or custom identifiers. You can create a machine group as instructed when applying Logtail configurations to machine groups.
3. Create Logtail configurations and apply them to a machine group. You can use **Data Import Wizard** to create Logtail configurations to collect *text files* and *syslogs*, and apply the Logtail configurations to a machine group.

After the preceding process is completed, logs of a specific type on the ECS servers are collected and sent to the selected Logstore. Historical logs are not collected. You can use the Log Service

console or SDKs and APIs to query these logs. You can also view the Logtail collection status on each ECS server, for example, whether the collection is normal and whether any error occurs.

For the complete Logtail access service operations on the Log Service console, see *Collect logs by Logtail*.

Docker

- Container service: See **Integrated Log Service** in *Cite LeftContainer Service User GuideCite Right*.
- Built-in Docker of ECS/IDC: Mount the log directories from containers to the host server.
  - *Install Logtail on Linux*.
  - Mount the log directories from containers to the host server.
    - Method 1: Run commands. For example, if the directory on the host server is `/log/webapp` and the directory in a container is `/opt/webapp/log`, run the following command:

      ```
      docker run -d -P --name web -v /src/webapp:/opt/webapp training
      /webapp python app.py
      ```

    - Method 2: Use an orchestration template.

Note:

We recommend that you modify the Logtail startup parameters, change the checkpoint save path of Logtail, and mount the log directories to the host server. It prevents repeated collection due to checkpoint information loss when containers are released.

Core concepts

- **Machine group**A machine group contains one or more machines on which logs of a specific type are collected. You can bind Logtail configurations to a machine group. Log Service then collects logs from all the servers in the machine group based on the same Logtail configurations. The Log Service console allows you to manage machine groups conveniently, including operations to create, delete, add, and remove servers.
- **Logtail agent**: Logtail is an agent running on a server to collect logs. For more information, see *Install Logtail on Linux*. After Logtail is installed on a server, configure Logtail and apply the configurations to a machine group.

- In **Linux**, Logtail is installed in the `/usr/local/ilogtail` directory and starts two independent processes (a collection process and a daemon) with their names beginning with ilogtail. The program running log is `/usr/local/ilogtail/ilogtail.LOG`.

- **Logtail configuration**: A set of log collection policies of Logtail. You can configure the data source, collection mode, and other parameters for Logtail to create custom collection policies for all the servers in a machine group. Logtail configurations describe how to collect a specific type of logs on servers, parse the collected logs, and send the logs to the specified Logstore of Log Service. Using the console, you can add Logtail configurations for each Logstore. It enables the Logstore to accept logs collected through the Logtail configurations.

Basic functions

The Logtail access service provides the following functions:

- **Real-time log collection:** It dynamically monitors log files, reads them in real time, and parses incremental logs. There is a delay of less than 3s between log generation and log transfer to Log Service.

   **Note:**
  The Logtail access service does not support the collection of historical data. If the time when a log is read is more than 5 minutes after the time of creation, the log is discarded.

- **Automatic log rotation processing:** Many applications rotate log files based on the file size or generation date. During the rotation process, the original log file is renamed and a new empty log file is created. For example, rotation of `app.LOG` generates `app.LOG.1` and `app.LOG.2`. You can specify the file (for example, `app.LOG`) to which collected logs are written. Logtail automatically detects the log rotation process and ensures that no logs are lost during this process.

   **Note:**
  If log files are rotated multiple times within several seconds, data loss may occur.

- **Automatic handling of collection exceptions:** If data sending fails due to the errors such as server abnormality, improper network measures, and quota exceeding, Logtail will proactively retry the operation based on the scenario. If retry fails, Logtail writes the data to the local cache and resends the data later.

   **Note:**

The local cache is on your server disk. If the local data is not accepted by the server within 24 hours, it will be discarded and deleted from the cache.

- **Flexible collection policy configuration:** Using Logtail configurations, you can flexibly specify how logs are collected on an ECS server. Specifically, you can select log directories and files by means of exact match or fuzzy match using wildcards based on the actual scenario. You can customize an extraction method for log collection and set the names of extracted fields. Log extraction by regular expression is supported. Because the log data models of Log Service require that each log have precise timestamp information, Logtail provides custom log time formats. It allows you to extract the required timestamp information from logs of different formats.
- **Automatic synchronization of collection configurations:** After you create or update configurations on the Log Service console, Logtail automatically accepts and applies the changes within three minutes. Collected data is not lost during the configuration update process.
- **Automatic agent upgrade:** After you manually install Logtail on a server, Log Service automatically maintains the upgrade of Logtail without manual intervention. No log data is lost during the Logtail upgrade process.
- **Self status monitoring:** To prevent the Logtail agent from consuming excessive resources and thus affecting services, Logtail monitors its resource (CPU and memory) consumption in real time. The Logtail agent automatically restarts when the resource usage limit is exceeded to avoid any impact on the ongoing operations on the server. The agent takes proactive measures of network traffic limitation to prevent excessive bandwidth consumption.

  Note:

  - Logs may be lost when the Logtail agent restarts.
  - If the Logtail agent exits due to an exception of its processing logic, the corresponding protective mechanism is triggered and the agent is restarted to continue log collection. However, logs may be lost before restart.
- **Transferred data signature:** To prevent data tampering during the transfer process, the Logtail agent proactively obtains your AccessKey to sign all log data packets before they are sent.

  Note:

  The Logtail agent uses an HTTPS channel to obtain your AccessKey to ensure its security.

22.5.1 Installation

Before using Logtail of Log Service to collect server logs, install a Logtail agent on the server and set startup parameters as needed.

22.5.1.1 Install Logtail (for Linux)

Applicable systems:

Linux x86-64 (64-bit) servers in the following versions:

- Aliyun Linux
- Ubuntu
- Debian
- CentOS
- OpenSUSE

Procedure

1. Download the Logtail installation script.

   Run the following command to download Logtail:

   ```
   logtail.your Log Service endpoint/logtail.sh
   ```

2. Execute the installation script.

   Start the shell terminal and run the following command as an administrator to install Logtail:

   ```
   sh logtail.sh
   ```

   Note:

   Logtail installation uses the overwrite mode. If you have installed Logtail before, the installer uninstalls and deletes the `/usr/local/ilogtail` directory, and then reinstalls it.

What's next

View the Logtail version.

The following information shows that the running Logtail version is 0.9.4:

```
$ls /usr/local/ilogtail/ilogtail -lh lrwxrwxrwx 1 root root 34 Nov 3
12:00 /usr/local/ilogtail/ilogtail -> /usr/local/ilogtail/ilogtail_0.9
.4
```

Uninstall Logtail.

Download**logtail.sh** by referring to *Install Logtail*. Run the following command as an administrator in shell mode:

```
wget http://{sls data endpoint}/logtail.sh chmod 755 logtail.sh sh
logtail.sh uninstall
```

22.5.1.2 Configure startup parameters

This topic describes how to configure Logtail startup parameters. You can use this topic as parameter setting reference.

Context

The configuration of Logtail startup parameters is applicable to the following scenarios:

- If many log files are collected, excessive memory space is occupied. The metadata of each file must be maintained in memory, including the file signature, collection location, and file name.
- CPU utilization is high due to heavy log data traffic.
- A high volume of log data leads to heavy traffic sent to Log Service.
- Syslogs and TCP data streams need to be collected.

Startup configuration

- File path:

  ```
  /usr/local/ilogtail/ilogtail_config.json
  ```

- File format:

  JSON

- File sample (only partial configuration items are shown)

  ```
  { ... "cpu_usage_limit" : 0.4, "mem_usage_limit" : 100, "max_bytes_
  per_sec" : 2097152, "process_thread_count" : 1, "send_reque
  st_concurrency" : 4, "streamlog_open" : false, "streamlog_pool_size_
  in_mb" : 50, "streamlog_rcv_size_each_call" : 1024, "streamlog_
  formats":[], "streamlog_tcp_port" : 11111, "buffer_file_num" : 25, "
  buffer_file_size" : 20971520, "buffer_file_path" : "", ... }
  ```

Common configuration parameters

Parameter name	Parameter value	Parameter description
cpu_usage_limit	CPU usage threshold, double type, calculated per core.	For example, the value 0.4 indicates that the CPU utilization of Logtail is limited to 40% of single-core capacity. Logtail restarts automatically when the threshold is exceeded. In many cases, the single-

Parameter name	Parameter value	Parameter description
		core CPU processing capability is about 24 MB/s in easy mode and about 12 MB/s in full mode.
mem_usage_limit	In-memory usage threshold, int type, measured in MB.	For example, the value 100 indicates that the memory usage of Logtail is restricted to 100 MB. Logtail restarts automatically when the threshold is exceeded. If you need to collect more than 1000 distinct files, increase the threshold value properly.
max_bytes_per_sec	Traffic limit on the raw data sent by Logtail, int type, measured in bytes per second.	For example, the value 2097152 indicates that the data transfer rate of Logtail is restricted to 2 MB/s.
process_thread_count	Number of threads Logtail uses to write data to log files.	The default value is 1, which supports a write speed of 24 MB/s in easy mode and 12 MB/s in complete regex mode. Adjust the threshold only when necessary.
send_request_concurrency	By default, Logtail sends data packets asynchronously. You can set a larger asynchronous concurrency value if the write TPS is large.	By default, four asynchronous concurrencies are available. You can calculate the proper concurrency quantity based on the condition that one concurrency supports 0.5 MB/s to 1 MB/s network throughout. The actual concurrency quantity varies with the network delay.
streamlog_open	Syslog reception switch, bool type.	False indicates that syslog reception is disabled and true indicates that syslog reception is enabled.
streamlog_pool_size_in_mb	Size of syslog memory pool used to receive logs. The memory is used to cache syslog data. The unit is MB.	Logtail requests memory when it starts. Set the pool size based on the machine memory size and your needs.
streamlog_rcv_size_each_call	The cache size used each time Logtail calls the Linux socket rcv interface. The value ranges from 1024 to 8192, in bytes.	You can increase the value in the case of heavy syslog traffic.
streamlog_formats	Method of parsing received syslogs.	--

Parameter name	Parameter value	Parameter description
streamlog_tcp_addr	The binding address Logtail uses to receive syslogs. The default value is 0.0.0.0.	For details, see *Collect syslogs through Logtail*.
streamlog_tcp_port	The TCP port through which Logtail receives syslogs.	The default value is 11111.
buffer_file_num	When a network exception occurs or the write quota is exceeded, Logtail writes the logs that are parsed in real time to a local file (in the installation directory) and then tries to resend the logs to Log Service after recovery. This parameter indicates the maximum number of cached files.	The default version is 25.
buffer_file_size	Maximum number of bytes of each buffered file. **`buffer_file_num * buffer_file_size`** indicates the maximum disk space available for cached files.	The default value is 20,971,520 bytes (20 MB).
buffer_file_path	Directory that stores cached files. After you modify this parameter, manually move the files named in the format of *`logtail\_buffer\_file_*`* in the old cache directory to the new directory so that Logtail can read the cached files and delete them after sending.	The default value is null, indicating that cached files are stored in the Logtail installation directory */usr/local/ilogtail*.
bind_interface	Name of the NIC bound to the local machine, for example, eth1. Only Linux version is supported.	By default, the available NICs are bound automatically. If this parameter is configured, Logtail uses only the specified NIC to upload logs.

Parameter name	Parameter value	Parameter description
check_point_filename	Full path of the checkpoint files. The parameter is used to customize the storage path of the checkpoint files of Logtail.	By default, the files are stored in */tmp/logtail_check_point*. We recommend that Docker users modify the checkpoint file storage path, and mount the file path to the host server to prevent repeated collection due to checkpoint information loss when containers are released. For example, set check_point_filename in Docker to `/data/logtail/check_point.dat` and add `-v /data/docker1/logtail:/data/logtail` to Docker startup commands. In addition, mount the */data/docker1/logtail* directory on the host server to the */data/logtail* directory on the Docker.

Note:

- The preceding table only lists the common startup parameters. If *ilogtail_config.json* has parameters not listed above, the default values are used.
- Add or modify the values of configuration parameters as required. You do not need to add unused configuration parameters to *ilogtail_config.json*.

Modify configuration

1. Configure *ilogtail_config.json* as required.

   Check that the modified configurations are JSON compatible.

2. Restart Logtail to apply the configurations.

```
/etc/init.d/ilogtaild stop /etc/init.d/ilogtaild start /etc/init.d/
ilogtaild status
```

22.5.2 Data sources

Log Service collects logs from multiple data sources, such as text logs and syslogs.

22.5.2.1 Text log

The Logtail agent helps you easily collect logs from ECS instances through the console.

Context

After a Logstore is created, the system prompts you to go to the data import wizard. In the dialog box that appears, click **Confirm** to create a Logtail configuration. Alternatively, you can click **Data Import Wizard** on the **Logstores** page to create a Logtail configuration.

Prerequisites

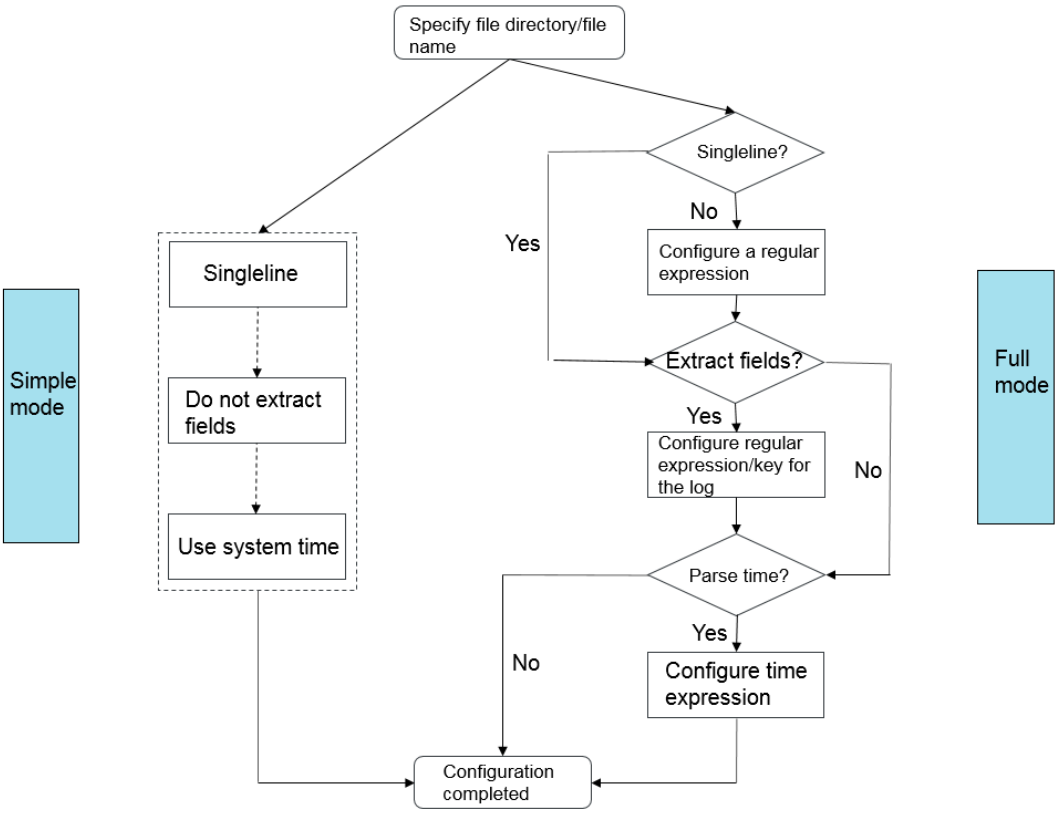
You must install Logtail before using it to collect logs. Apsara Stack Log Service allows you to install Logtail on a Linux operating system. For the installation method, see *Install Logstail on Linux*.

Restrictions

- A single file can only be collected with one configuration. Use a soft link to collect multiple copies of a log file. For example, if you need two copies of the logs under */home/log/nginx/log*, configure the original path for one copy. For the other copy, configure a soft link path (`ln -s /home/log/nginx/log /home/log/nginx/link_log`).
- For the operating systems supported by the Logtail agent, see *Install Logstail on Linux*.

Logtail collection configuration procedure

You can configure Logtail to collect text logs on the console. Logtail supports various log collection methods such as easy mode, separator mode, JSON mode, and full mode. The following describes how to configure Logtail in easy mode and full mode:



Procedure

1. *Log on to the Log Service Console*.
2. Create a project and a Logstore. For detailed instructions, see *Operate on projects* and *Operate on Logstores*.
3. On the Log Service console, click the project to go to the **Logstores** page.
4. Select the Logstore and click the **Data Import Wizard** icon next to it to start the data import configuration.
5. Select a data type.

   Choose**Custom DataText File**. Click **Next** to go to the **Configure Data Source** page.
6. Set the data source.

   1. Specify the configuration name.

      The configuration name can only contain lowercase letters, numbers, hyphens (-), and underlines (_). It must start and end with a lowercase letter or number and must be 3 to 63 bytes in length.

      Note:

After the configuration name is specified, it cannot be modified.

2. Specify the log directory and file name.

   The directory structure supports both the complete path mode and the wildcard mode.

   Note:

   - The directory wildcards can only be asterisks (*) or question marks (?).
   - A single file can only be collected with one configuration.

   Both the complete file name and the wildcard can be used as the log file name. For file naming rules, see *Wildcard matching*.

   Multi-level directory matching is set as the log search mode. This indicates that all files with compliant file names under the folder can be monitored (including all subdirectories).

   - For example, `/apsara/nuwa/ … /*.log` indicates the files suffixed with `.log` in the `/apsara/nuwa` directory (including the recursive subdirectories).
   - For example, `/var/logs/app_* … /*.log*` indicates the files whose names containing `.log` in the directories (including the recursive subdirectories) in the `app_*` mode under the `/var/logs` directory.

3. Specify the log collection mode.

   Currently, Log Service allows you to parse logs in **NGINX Configuration**, **Easy Mode**, **Separator Mode**, **JSON Mode**, and **Full Mode**. In this example, the easy mode and full mode are used to introduce the collection mode settings.

   - Easy mode

     The easy mode refers to the single-line mode. In single-line mode, one line is counted as one log by default. Two logs in a log file are separated by a linefeed. In single-line mode, no log field is extracted (the default regular expression is (.*)), and the system time of the current server is recorded as the log generation time. If you want to use more detailed settings later, you can change the configuration to full mode and configure all the settings.

     In easy mode, you only need to specify the file directory and file name. Logtail collects logs line by line. It does not extract fields from the log content. In addition, the log time is set to the system time of the server when the log is captured.

   - Full mode

If you need to customize the field extraction settings (for example, cross-line logs and field extraction), select **Full Mode**.

1. Enter **Log Sample**.

   The purpose of providing a log sample is facilitating the Log Service console to automatically extract the full mode in logs. Be sure to use a log from the actual environment.

2. Disable **Singleline**.

   The single-line mode is the default option. This means that a log contains only a line. If you need to collect cross-line logs (such as Java program logs), you must disable **Singleline** and then set **Regular Expression**.

3. Set **Regular Expression**.

   This option provides two functions: automatic generation and manual input. After entering the log sample, click **Auto Generate**. Then the system automatically generates a regular expression. If the regular expression is not generated, switch to the manual mode and enter a regular expression for verification.

4. Set **Extract Field**.

   If you need to analyze and process fields one by one in the log content, use the **Extract Field** function to convert the specified field into a key-value pair before sending it to the server. Therefore, you need to specify a method for parsing the log content (specifically, a regular expression).

   The Log Service console allows you to specify a regular expression for parsing in two ways. The first way is to automatically generate a regular expression through simple interactions. You can select fields from the log sample. Then, the Log Service console automatically generates a regular expression.

   Although this method is more convenient, but the automatically generated regular expressions are not optimal in most cases. Therefore, the Log Service console also allows you to manually enter regular expressions. You can click **Manually Input Regular Expression** to switch to manual entry mode. After entering the regular expression manually, click **Validate** next to it to check whether the regular expression can be used to parse and extract the sample log.

Regardless of the automatic generation or manual input method, you must name each extracted field, that is, set the key for the field.

4. Set **Use System Time**.

   The **Use System Time** option is enabled by default. If it is disabled, you must specify the **Value** field as the time field during field extraction and name this field `time` (as shown above). After selecting the `time` field, click **Auto Generate** in **Time Format** to generate a method to parse this field. For more information about log time formats, see *Text - Configure a time format*.

5. Configure**Advanced Options** as appropriate.

   Based on actual demands, configure **Local Cache**, **Topic Generation Mode**, **Log File Encoding**, **Maximum Watch Directory Depth**, **Timeout Attribute**, and **Filter Configuration**. Remain the default configurations unless otherwise required.

Configuration	Description
Local Cache	Choose whether to enable the **Local Cache** function. When Log Service is unavailable, logs can be cached to a local directory and then uploaded after the service is recovered. The default maximum size of logs that can be cached is 1 GB.
Topic Generation Mode	• **Null - no topic**: This is the default option, meaning that the topic is a null string and you can query a log without entering the topic. • **Machine Group Topic Attribute**: If this mode is selected, logs generated by different front-end servers can be distinguished. • **File Path Regular**: When this mode is selected, you must enter a **Custom Regular** below to extract a part of the path as the topic. This mode is used to distinguish the log data generated by a user or an instance.
Custom Regular	If you choose to generate a topic in **File Path Regular** mode, enter a custom regular expression here.
Log File Encoding	• utf8: UTF-8 encoding. • gbk: GBK encoding.
Maximum Watch Directory depth	Specifies the maximum depth of the monitoring directory when logs are collected from the log source, that is, up to what levels logs are collected. The maximum monitoring directory depth ranges from 0 to 1000, of which 0 means only the directory at the current level is monitored.

Configuration	Description
Timeout Attribute	If a log file is not updated within the specified period of time, the system considers that the file has timed out. You can configure **Timeout Attribute** as follows: • Never timed out: All log files are continuously monitored and never time out. • 30 minute timeout: If a log file is not updated in 30 minutes, the system considers that the log file has timed out and no longer monitors the file.
Filter Configuration	Only logs that completely meet the filter conditions are collected. For example, if `Key:level Regex:WARNING|ERROR` is configured, only the logs of warning or error level are collected. To filter out the logs that do not meet certain conditions, you can also use this method. For example, `Key:level Regex :^(?!.*(INFO|DEBUG))` indicates that the logs of info or debug level are not collected. For similar examples, see *regex-exclude-word* and *regex-exclude-pattern*.

6. After setting, click **Next**.
7. Select the expected machine group and click **Apply to Machine Group** to apply the configuration to the specified machine group.

   If you have not created a machine group, you must create one first. For details about how to create a machine group, see *Create Machine Group*.

   Note:

   - It takes up to three minutes for the Logtail configuration to come into effect after being pushed, please be patient.
   - After creating Logtail configurations, you can view the Logtail configuration list, modify the Logtail configurations, or delete Logtail configurations. For details, see *Logtail configurations*.

Logtail configuration items

You need to enter the configuration items when configuring Logtail. The configuration items and their restrictions are listed in the table below:

Configuration	Description
Log Path	The directory structure supports both the complete path mode and the wildcard mode. The wildcard mode means multi-level directory matching. This indicates that all files with compliant file names under the folder are monitored (including all subdirectories).
Log File Name	Specifies the name of the collected log file. The name is case sensitive and can contain wildcards, for example, *.log. In Linux operating systems, the file name can contain wildcards ***, [...]**, and **?**.
Local Storage	Indicates whether to enable the local cache to temporarily store logs that cannot be sent due to short-term network interruptions.
First-line Log Header	Indicates the starting line of a multiline log by means of a regular expression. Line feeds cannot be used to separate individual logs in multiline log collection mode (for example, collecting application logs with stack information). You need to specify a starting line to delimit multiline logs. Because the starting line (for example, timestamp) of each log may be different, you need to specify a starting line match rule. A regular expression is used as a match rule here.
Log Parsing Expression	Indicates how to extract a piece of log information and convert it into a log format supported by Log Service. You must specify a regular expression to extract the required log field information and define the name of each field to be extracted.
Log Time Format	Defines how to parse the time format of the timestamp string in log data. For details, see *Logtail log time format*.

What's next

After the configuration is made, Log Service can collect logs. You can view the collected logs.

In addition to using Logtail to collect logs, Log Service also provides APIs and SDKs for you to write logs conveniently.

Write logs with APIs.

Log Service provides RESTful APIs to help you write logs. You can use the PostLogstoreLogs API to write data. For a complete API reference, see *Cite LeftLog Service Development GuideCite Right*.

Write logs with SDKs.

In addition to APIs, Log Service also provides SDKs for a variety of languages (Java, .NET, PHP, and Python) that allow you to easily write logs. For a complete SDK reference, see *Cite LeftLog Service Development GuideCite Right*.

22.5.2.2 Text - Configuration parsing

Specify the way for separating log lines

A complete access log (for example, Nginx access log) occupies a line. Individual logs are separated by linefeeds. Two samples of access logs are shown below.

```
10.1.1.1 - - [13/Mar/2016:10:00:10 +0800] "GET / HTTP/1.1" 0.011 180
404 570 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; 360se
)" 10.1.1.1 - - [13/Mar/2016:10:00:11 +0800] "GET / HTTP/1.1" 0.011
180 404 570 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1;
360se)"
```

For Java applications, a program log spans several lines. The characteristic of the log beginning is used to distinguish the beginning of each log. Java program logs are shown below.

```
[2016-03-18T14:16:16,000] [INFO] [SessionTracker] [SessionTrackerImpl.
java:148] Expiring sessions 0x152436b9a12aecf, 50000 0x152436b9a12aed2
, 50000 0x152436b9a12aed1, 50000 0x152436b9a12aed0, 50000
```

The above Java program logs are all started with time and the regular expression at the beginning of the line is: [\d+-\d+-\w+:\d+:\d+,\d+]\s.*

Extract log fields

According to the data models of log service, a log contains one or more key–value pairs. To extract specified fields for analysis, you need to configure a regular expression. If log content does not need to be processed, the whole log can be considered as a key–value pair.

For the above access log, you can choose whether to extract fields.

- Fields are extracted:

  When the regular expression is `(\S+)\s-\s-\s\[(\S+)\s[^]]+]\s"(\w+).*`, the content to be extracted is: `10.1.1.1`, `13/Mar/2016:10:00`, and `GET`.
- Fields are not extracted:

When the regular expression is `(.*)`, the content to be extracted is:`10.1.1.1 - - [13/Mar/2016:10:00:10 +0800] "GET / HTTP/1.1" 0.011 180 404 570 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; 360se)"`.

Specify the log time

According to the data models of Log Service, each log must have a time field in the Unix timestamp format. Currently, the log time can be set to the system time (when Logtail captures the log) or the time in the log content.

In the above access log:

- If you need to extract the time field in the log, the time is `13/Mar/2016:10:00:10`, and the time expression is `%d/%b/%Y:%H:%M:%S`.
- If the system time is used, the time is the timestamp when the log is captured.

22.5.2.3 Text - Configure a time format

Each log in Log Service must contain a timestamp. When collecting logs from users log files, the Logtail access service must extract the timestamp string in a log record and parse it as a timestamp. Therefore, you need to specify a timestamp format for parsing.

Logtail in Linux supports all time formats provided by the strftime function. Logtail only parses and uses the timestamp strings that can be expressed in the log formats defined by the strftime function.

The timestamp strings of logs have diverse formats. To make configuration easier, the following table lists the common log time formats supported by Logtail:

Supported format	Description	Example
%a	Abbreviation of week.	Fri
%A	Full name of week.	Friday
%b	Abbreviation of month.	Jan
%B	Full name of month.	January
%d	The Nth day of a month, ranging from 01 to 31 in decimal format.	07, 31
%h	Abbreviation of month, which is the same as `%b`.	Jan
%H	Hour, in 24-hour format.	22

Supported format	Description	Example
%I	Hour, in 12-hour format.	11
%m	Month, in decimal format.	08
%M	Minute, ranging from 00 to 59 in decimal format.	59
%n	Linefeed.	Linefeed
%p	Local time in am or pm format.	AM/PM
%r	Time combination in 12-hour format, which is the same as `%I:%M:%S %p`.	11:59:59 AM
%R	Combination of hours and minutes, which is the same as `%H:%M`.	23:59
%S	Seconds, ranging from 00 to 59 in decimal format.	59
%t	Tab character.	Tab character
%y	Year (excluding century), ranging from 00 to 99 in decimal format.	04; 98
%Y	Year, in decimal format.	2004; 1998
%z	Time zone or abbreviation.	-07:00, +0800
%C	Century, ranging from 00 to 99 in decimal format.	16
%e	The Nth day of a month, ranging from 1 to 31 in decimal format. Prefix a blank to a single-digit number.	7, 31
%j	The Nth day of a year, ranging from 00 to 366 in decimal format.	365
%u	Week in decimal format. It ranges from 1 to 7, and the value 1 indicates Monday.	2
%U	The Nth week of a year. The first day of a week is Sunday. It ranges from 00 to 53.	23

Supported format	Description	Example
%V	The Nth week of a year. The first day of a week is Monday . If the first week of a month contains four or more days , this is considered the first week. Otherwise, the next week is considered the first week. It ranges from 01 to 53.	24
%w	Week in decimal format. It ranges from 0 to 6, and the value 0 indicates Sunday.	5
%W	The Nth week of a year. The first day of a week is Monday. It ranges from 00 to 53.	23
%c	Standard date and time.	If you want to specify a long or short date, use the above formats for more accurate expression.
%x	Standard date.	If you want to specify a long or short date, use the above formats for more accurate expression.
%X	Standard time.	If you want to specify a long or short date, use the above formats for more accurate expression.
%s	Unit timestamp.	1476187251

22.5.2.4 Text - Import historical log files

Logtail collects only incremental log files by default. To import historical log files, use the historical log file importing function in Logtail.

Prerequisites

- The Logtail version must be `0.16.6` or later.
- Historical files to be collected must be in the configured collection range and have never been collected by Logtail.

- The last modification time of the historical files must be earlier than the Logtail configuration time.
- The maximum latency for local event importing is one minute.
- Because local configuration loading is a special behavior, Logtail sends `LOAD_LOCAL_EVENT_ALARM` to the server to notify the user of such events.

Context

Logtail collects files based on events, which are generated during monitoring or periodical file polling. Besides, Logtail can load events from local files to drive log collection. Historical file collection is a function implemented based on local event loading.

Procedure

1. Create a collection configuration.

   Create collection configurations according to *Text log* and apply the configurations to the machine group. Ensure that the files are within the configured range.

2. Obtain the unique configuration ID.

   You can obtain the unique ID of the collection configuration from `/usr/local/ilogtail/user_log_config.json` as follows:

   ```
   grep "##" /usr/local/ilogtail/user_log_config.json | awk {print $1
   } "##1.0##log-config-test$multi" "##1.0##log-config-test$ecs-test"
    "##1.0##log-config-test$metric_system_test" "##1.0##log-config-test
   $redis-status"
   ```

3. Add a local event.

   The storage path of local events is `/usr/local/ilogtail/local_event.json`. The files are of the standard JSON type, in the following format:

   ```
   [ { "config" : "${your_config_unique_id}", "dir" : "${your_log_dir
   }", "name" : "${your_log_file_name}" }, { ... } ... ]
   ```

   - **Configuration item**

Configurations	Description	Example
config	Unique configuration ID obtained in Step 2.	##1.0##log-config-test$ecs-test
dir	Folder of the file. **Note:** A folder must not end with `/`.	`/data/logs`

Configurations	Description	Example
name	Log file name	`access.log.2018-08-08`

Note:
To prevent loading of invalid JSON data on Logtail, we recommend that you save local event configurations to a temporary file and copy the configurations to `/usr/local/ilogtail/local_event.json` after editing.

- **Sample configuration**

```
$ cat /usr/local/ilogtail/local_event.json
                                    [
                                    {
                                    "config": "##1.0##log-config-test$
ecs-test",
                                    "dir": "/data/log/",
                                    "name": "access.log.2017-08-08"
                                    },
                                    {
                                    "config": "##1.0##log-config-test$
ecs-test",
                                    "dir": "/tmp",
                                    "name": "access.log.2017-08-09"
                                    }
                                    ]
```

- **Check whether Logtail has loaded the configuration.**

  Logtail typically loads the local configuration file to the memory and clears the content of `local_event.json` within one minute after `local_event.json` is saved locally.

  You can check whether Logtail has read events using any of the following methods:

  — If the content of `local_event.json` is cleared, Logtail has read the event information.
  — Check whether the file `/usr/local/ilogtail/ilogtail.LOG` contains the keyword `process local event`. If the content of `local_event.json` is cleared but the keyword is not found in the file, your local configuration file may be filtered out due to invalid content.
  — *Query error diagnostics* to check for `LOAD_LOCAL_EVENT_ALARM` notification.

- **Check whether the configuration is loaded but no data is collected.**

  If Logtail has loaded the configuration but no data is collected, the possible causes can be:

  — The configuration is invalid.
  — The local `config` does not exist.

- No log files exist in the path specified in Logtail collection configuration.
- The log file has already been collected by Logtail.

- **Determine how to re-collect data that has been collected before.**

  To collect data that has been collected before, perform the following steps:

  1. Run the `/etc/init.d/ilogtaild stop` command to stop Logtail.
  2. Search for the corresponding log file path in the `/tmp/logtail_check_point` file.
  3. Delete checkpoint(JSON object) from the log file and save the file.
  4. Add a local event as described in step 3.
  5. Run the `/etc/init.d/ilogtaild start` command to start Logtail.

22.5.2.5 Text - generate a topic

A topic is a custom field used to mark a batch of logs. Logs in one Logstore can be grouped by log topics. You can specify a topic when writing a log or querying logs.

Log is the minimum data unit processed in Log Service. It is defined in semi-structured data mode. The specific data model consists of topic, time, content, and source. For details, see *Cite LeftLog Service OverviewCite Right*.

A topic is a custom field used to mark a batch of logs. Logs in one Logstore can be grouped by log topics. You can specify a topic when writing a log or querying logs. For example, access logs are marked by sites, and platform users can use user IDs as the log topics and write them into logs. In this way, users can view only their own logs based on log topics. If there is no need to group logs in a Logstore, one log topic can be used for all logs. The default value of this field is a null string, which is also a valid topic.

Note:

You cannot set a topic for syslogs.

You can set or change the topic in the Log Service console.

Topic generation mode

You can set topics when collecting logs by using Logtail or when uploading data by using APIs or SDKs. At present, the following topic generation modes are supported in the Log Service console: **Null - no topic**, **Machine Group Topic Attribute**, and **File Path Regular**.

- **Null - no topic**

When you configure Logtail to collect text files in the Log Service console, the default log topic generation mode is **Null - no topic**. That is, the topic is a null string, and logs can be directly queried without a topic.

- **Machine Group Topic Attribute**

  The **Machine Group Topic Attribute** mode is used to differentiate log data generated by different servers. If log data of different servers is stored in the same file path and the same file, you can divide machines into different machine groups when you want to differentiate the log data of different servers by topic. That is, set **Group Topic** differently for different machine groups when creating machine groups and set **Topic Generation Mode** to **Machine Group Topic Attribute**. Apply the previously created Logtail configuration to the machine groups to complete the configuration.

  If **Machine Group Topic Attribute** is selected, Logtail uploads the topic attribute of the machine group to which the current machine belongs as the topic name to Log Service, when reporting data. When you perform a query by using the **LogSearch/Analytics** function, you need to specify a topic (namely the topic attribute of the target machine group) as the query condition.

- **File Path Regular**

  The **File Path Regular** mode is used to differentiate the log data generated by a user or an instance. If service logs are stored in different directories by user or instance, Log Service cannot distinguish which user or instance generates the logs when collecting log files so long as subdirectories are different and log file names are the same. In this case, you can set **Topic Generation Mode** to **File Path Regular**, enter the regular expression of the file path, and set the topic to the instance name.

  When **File Path Regular** is selected as the topic generation mode, Logtail uploads the instance name as the topic name to Log Service when reporting data. The topic generated varies with your directory structure and configuration. You need to specify the topic name as the instance name when you performa query by using the **LogSearch/Analytics** function.

Set a log topic

Procedure

1. Configure Logtail in the Log Service console by referring to *Text log*.

   If you want to set the topic generation mode to **Machine Group Topic Attribute**, set **Group Topic** in the **Create Machine Group** or **Modify Machine Group** dialog box first.

2. In Logtail configurations, expand **Advanced Options**, and set **Topic Generation Mode**.

22.5.2.6 Syslog

Logtail supports local configuration of TCP ports to receive the syslog data transferred by syslog agents via the TCP protocol. The received data will be parsed by Logtail and forwarded to LogHub.

Prerequisites

You must install Logtail before using it to collect logs. Apsara Stack Log Service allows you to install Logtail on a Linux operating system. For the installation method, see *Install Logstail on Linux*.

Step 1: Create a Logtail syslog configuration.

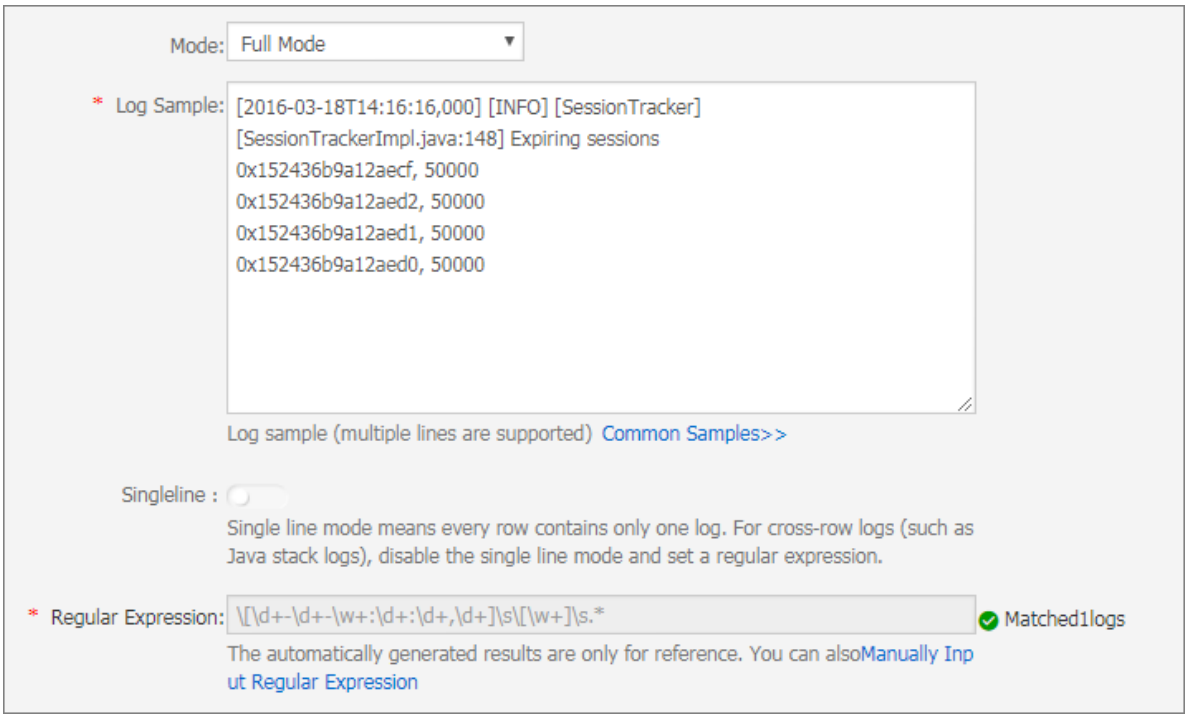
1. *Log on to the Log Service Console*.
2. Create a project and a Logstore. For detailed instructions, see *Operate on projects* and *Operate on Logstores*.
3. On the log service console, click the project to go to the **Logstores** page.
4. Select the Logstore and click **Data Import Wizard** next to it to start the data import configuration.
5. Select the data source type.

   Choose **Custom Data** > **Syslog** and click **Next**.

6. Specify **`Configuration Name`**.

   The configuration name can only contain lowercase letters, numbers, hyphens (-), and underlines (_). It must start and end with a lowercase letter or number and must be 3 to 63 bytes in length.

   Note:
   After the configuration name is specified, it cannot be modified.



7. Specify **Tag Settings**.

   For details about how to set the tag, see *Syslog collection reference*.

8. Configure**Advanced Options** as appropriate.

   Choose whether to enable the **Local Cache** function. When Log Service is unavailable, logs can be cached to a local directory and then uploaded after the service is recovered. Local cache is enabled by default, and the maximum size of logs that can be cached is 1 GB.

9. Apply the Logtail configuration to a machine group as prompted.

   Select the desired machine group and click **Apply to Machine Group** to apply the configuration to the machine group.

   If you have not created a machine group, you must create one first. For details about how to create a machine group, see *Create Machine Group*.

Step 2: Configure the Logtail protocol

Locate `ilogtail_config.json` under the Logtail installation directory `/usr/local/ilogtail/`, and modify the syslog-related configurations as required.

Procedure

1. Check whether Syslog is enabled.

   True indicates Syslog is enabled while False indicates Syslog is disabled.

   ```
   "streamlog_open" : true
   ```

2. Configure the size of the Syslog memory pool for storing received logs.

   Logtail requests memory of the specified size at one time when launched. Set the pool size according to the machine memory size and your needs. The unit is MB.

   ```
   "streamlog_pool_size_in_mb" : 50
   ```

3. Configure the buffer size.

   Configure the buffer size used each time Logtail calls the socket io rcv interface. The unit is byte.

   ```
   "streamlog_rcv_size_each_call" : 1024
   ```

4. Configure the syslog format.

   ```
   "streamlog_formats":[]
   ```

5. Configure the TCP port.

   Configure the TCP port used by Logtail to receive syslogs. The port 11111 is used by default.

   ```
   "streamlog_tcp_port" : 11111
   ```

6. After configuration, restart Logtail.

   To restart Logtail, run the following command to stop and then start the Logtail agent.

   ```
   sudo /etc/init.d/ilogtaild stop sudo /etc/init.d/ilogtaild start
   ```

Step 3: Install rsyslog and modify its configuration

If rsyslog is already installed on the machine, skip this step.

7. Install rsyslog.

   Visit the following descriptions for more information on installing rsyslog:

   - *Ubuntu installation method*

- *Debian installation method*
- *RHEL/CENTOS installation method*

8. Modify configuration.

   Modify the configurations in `/etc/rsyslog.conf` as required, for example:

```
$WorkDirectory /var/spool/rsyslog # where to place spool files $
ActionQueueFileName fwdRule1 # unique name prefix for spool files
 $ActionQueueMaxDiskSpace 1g # 1gb space limit (use as much as
possible) $ActionQueueSaveOnShutdown on # save messages to disk
 on shutdown $ActionQueueType LinkedList # run asynchronously $
ActionResumeRetryCount -1 # infinite retries if host is down #
Define the fields in logs. $template ALI_LOG_FMT,"0.1 sys_tag %
timegenerated:::date-unixtimestamp% %fromhost-ip% %hostname% %pri
-text% %protocol-version% %app-name% %procid% %msgid% %msg:::drop-
last-lf%\n" *.* @@10.101.166.173:11111;ALI_LOG_FMT
```

   Note:

   In the template `ALI_LOG_FMT`, the value of the second field is `sys_tag`. This value must be consistent with the value created in step 1. This configuration indicates that all (`\*.\*`) syslogs received by this machine are formatted according to `ALI_LOG_FMT`, and sent to 10.101.166.173:11111 via TCP. The machine 10.101.166.173 must be in the machine group in step 1 and configured according to step 2.

9. Start rsyslog.

```
sudo /etc/init.d/rsyslog restart
```

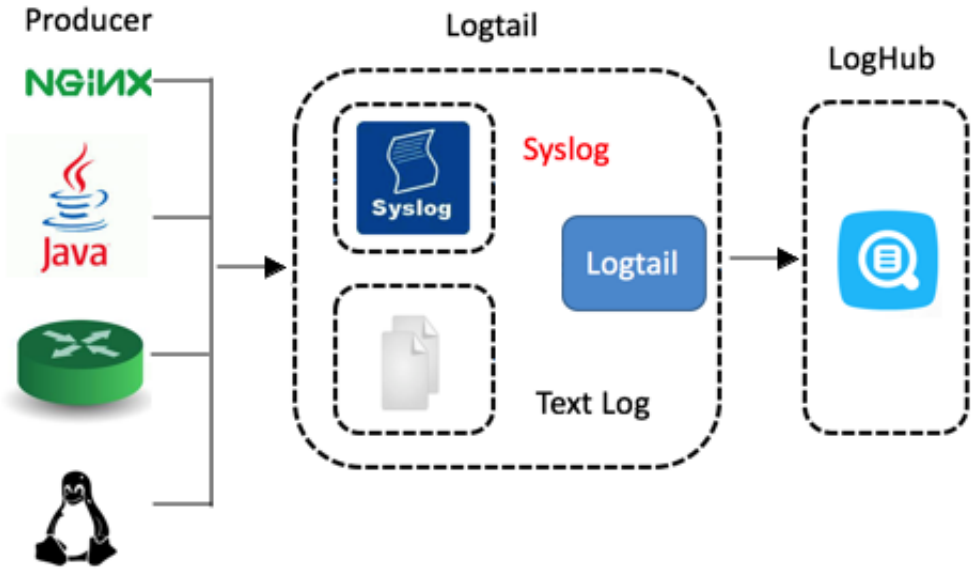
   Before starting rsyslog, check whether another syslog agent is installed on the machine, such as syslogd, sysklogd, or syslog-ng. If yes, disable the installed rsyslog.

   After completing the three steps above, you can collect syslogs on the machine to Log Service.

   For more information about syslog collection and syslog formatting, see *Syslog collection reference*.

22.5.2.7 Syslog collection reference

Currently, Logtail supports collection of syslogs and text files, as shown in the following figure.



Logtail collects syslogs based on TCP. For more information about how to configure Logtail to collect syslogs, see *Collect syslogs through Logtail*.

Benefits

For the syslog concept, see *Syslog*.

Compared with text files, syslogs can be directly collected to LogHub without being flushed into disks. It provides enhanced confidentiality and removes the need for parsing. A single machine delivers 80 MB/s of throughput.

How it works

Logtail supports local TCP port configurations and receives the logs forwarded by syslog agents. The following figure shows the relationship among Logtail, syslogs, and LogHub. Logtail with the TCP port enabled receives the syslogs forwarded by rsyslog or other syslog agents over TCP, parses the received logs, and forwards the logs to LogHub.

Syslog format

Logtail receives data as streams through the TCP port. If you want to parse individual logs from the data streams, ensure that the log format meets the following requirements:

- Logs are separated by linefeeds (`\n`), but do not contain linefeeds.
- Only the message body of a log can contain spaces; other fields cannot contain spaces.

The syslog format is shown as follows:

```
$version $tag $unixtimestamp $ip [$user-defined-field-1 $user-defined-
field-2 $user-defined-field-n] $msg\n"
```

Meanings of the fields:

Log field	Meaning
version	Specifies the version of the log format. Logtail uses the version to parse user-defined-field.
tag	Specifies the data tag used to locate the project or Logstore. It cannot contain spaces or linefeeds.
unixtimestamp	Specifies the timestamp of the log.
ip	Specifies the IP address of the machine that corresponds to the log. If the IP address is 127.0.0.1, it is replaced with the peer address of the TCP socket when the log is sent to the server.
user-defined-field	Zero or multiple custom fields can be set. The fields cannot contain spaces or linefeeds. The braces indicate that the fields are optional.
msg	Specifies the message body of the log. The `\n` symbol appended to the message is a linefeed, but the message cannot contain linefeeds.

The following example is a log that meets the above format requirements:

```
2.1 streamlog_tag 1455776661 10.101.166.127 ERROR com.alibaba.
streamlog.App.main(App.java:17) connection refused, retry
```

In addition to syslogs, Logtail can also collect other types of logs that meet the following requirements:

- The formatted logs can meet the requirements.
- Logs can be appended to the remote server over TCP.

Rules for Logtail to parse syslogs

You should add a configuration in Logtail for parsing syslogs, for example:

```
"streamlog_formats": [ {"version": "2.1", "fields": ["level", "method
"]}, {"version": "2.2", "fields": []}, {"version": "2.3", "fields": ["
pri-text", "app-name", "syslogtag"]} ]
```

Logtail identifies the corresponding user-defined-field format in streamlog_formats based on the version field. According to the sample configuration, the preceding sample log with the version

field 2.1 contains two user-defined fields: level and method. The sample log is parsed in the following format:

```
{ "source": "10.101.166.127", "time": 1455776661, "level": "ERROR",
 "method": "com.alibaba.streamlog.App.main(App.java:17)", "msg": "
connection refused, retry" }
```

The version field is used to parse the user-defined-field. The tag is used to search for the project or Logstore to which the data is to be sent. The two fields are not sent to the Log Service instance as the log content. In addition, Logtail has preset some log formats in which the version field starts with "0." or "1.", for example, 0.1 or 1.1. Therefore, custom version fields cannot start with "0." or "1.".

Common Logtail syslog collection tools

- **log4j**

  — Introduce the Log4j library.

  ```
  <dependency> <groupId>org.apache.logging.log4j</groupId> <
  artifactId>log4j-api</artifactId> <version>2.5</version> </
  dependency> <dependency> <groupId>org.apache.logging.log4j</
  groupId> <artifactId>log4j-core</artifactId> <version>2.5</version
  > </dependency>
  ```

  — Introduce the Log4j configuration file `log4j_aliyun.xml`.

  ```
   <?xml version="1.0" encoding="UTF-8"?> <configuration status="OFF
  "> <appenders> <Socket name="StreamLog" protocol="TCP" host="10.
  101.166.173" port="11111"> <PatternLayout pattern="%X{version} %X
  {tag} %d{UNIX} %X{ip} %-5p %l %enc{%m}%n" /> </Socket> </appenders
  > <loggers> <root level="trace"> <appender-ref ref="StreamLog" />
   </root> </loggers> </configuration>
  ```

  10.101.166.173:11111 is the address of the machine where Logtail is located.

  — Set ThreadContext in programs.

  ```
  package com.alibaba.streamlog; import org.apache.logging.log4j.
  LogManager; import org.apache.logging.log4j.Logger; import org.
  apache.logging.log4j.ThreadContext; public class App { private
  static Logger logger = LogManager.getLogger(App.class); public
  static void main( String[] args ) throws InterruptedException {
  ThreadContext.put("version", "2.1"); ThreadContext.put("tag", "
  streamlog_tag"); ThreadContext.put("ip", "127.0.0.1"); while(true)
   { logger.error("hello world"); Thread.sleep(1000); } //ThreadCont
  ext.clearAll(); } }
  ```

- **Tengine**

  Tengine can collect syslogs by ilogtail.

Tengine uses the ngx_http_log_module for logging to the local syslog agent, which forwards the logs to rsyslog.

For syslog configuration on Tengine, see *Configure syslog in Tengine.*

Example:

Send INFO-level access logs of the user type to Unix dgram(/dev/log) of the local machine and set the application tag to Nginx.

```
access_log syslog:user:info:/var/log/nginx.sock:nginx
```

Rsyslog configuration:

```
module(load="imuxsock") # needs to be done just once input(type="
imuxsock" Socket="/var/log/nginx.sock" CreatePath="on") $template
ALI_LOG_FMT,"2.3 streamlog_tag %timegenerated:::date-unixtimestamp%
 %fromhost-ip% %pri-text% %app-name% %syslogtag% %msg:::drop-last-lf
%\n" if $syslogtag == nginx then @@10.101.166.173:11111;ALI_LOG_FMT
```

- **Nginx**

  The collection of Nginx access logs is used as an example.

  Access log configuration:

  ```
  access_log syslog:server=unix:/var/log/nginx.sock,nohostname,tag=
  nginx;
  ```

  Rsyslog configuration:

  ```
  module(load="imuxsock") # needs to be done just once input(type="
  imuxsock" Socket="/var/log/nginx.sock" CreatePath="on") $template
  ALI_LOG_FMT,"2.3 streamlog_tag %timegenerated:::date-unixtimestamp%
   %fromhost-ip% %pri-text% %app-name% %syslogtag% %msg:::drop-last-lf
  %\n" if $syslogtag == nginx then @@10.101.166.173:11111;ALI_LOG_FMT
  ```

  For more information, visit *http://nginx.org/en/docs/syslog.html*.

- **Python syslog**

  Example:

  ```
  import logging import logging.handlers logger = logging.getLogger(
  myLogger) logger.setLevel(logging.INFO) #add handler to the logger
   using unix domain socket /dev/log handler = logging.handlers.
  SysLogHandler(/dev/log) #add formatter to the handler formatter =
  logging.Formatter(Python: { "loggerName":"%(name)s", "asciTime":"%(
  asctime)s", "pathName":"%(pathname)s", "logRecordCreationTime":"%(
  created)f", "functionName":"%(funcName)s", "levelNo":"%(levelno)s",
   "lineNo":"%(lineno)d", "time":"%(msecs)d", "levelName":"%(levelname
  ```

```
)s", "message":"%(message)s"}) handler.formatter = formatter logger.
addHandler(handler) logger.info("Test Message")
```

22.5.3 Machine group

22.5.3.1 Create a machine group

Log Service manages all ECS servers whose logs need to be collected using Logtail in machine groups.

After creating Logtail configurations, you can create a machine group on the **Machine Groups** page of a project in the Log Service project list. You can also go to the **Apply to Machine Group** page and click **Create Machine Group** to create a machine group.

A machine group is defined by either of the following two items:

- IP address: defines the name of the machine group and adds the internal IP addresses of servers to the group.

  You can add internal IP addresses of ECS servers to a machine group so that multiple ECS servers are directly added to the group, and centrally configure Logtail for the ECS servers.
- ID: indicates membership of the machine group, and is associated with the IDs configured on corresponding machines.

  The system consists of multiple modules, and each component of each module can be horizontally expanded separately. One module can contain multiple machines and the machine group is created for each module separately to collect logs by type. Therefore, you must set a custom ID for each module separately and configure an ID for the server of each module. For example, a common website generally consists of an HTTP request processing module, a cache module, a logic processing module, and a storage module, which can be identified as http_module, cache_module, logic_module, and store_module, respectively.

Procedure

1. *Log on to the Log Service Console*.
2. Create a project and a Logstore. For detailed instructions, see *Operate on projects* and *Operate on Logstores*.
3. On the **Projects** page on the Log Service console, click a project name to go to the **Logstores** page.
4. In the left-side navigation pane, click **Logtail Machine Group** to go to the **Machine Groups** page. Click **Create Machine group**.

Alternatively, after creating collection configurations in the data import wizard, click **Create Machine Group** on the **Apply to Machine Group** page.

5. Enter a name in **Group Name**.

   The machine group name can only contain lowercase letters, numbers, hyphens (-), and underlines (_). It must start and end with a lowercase letter or number and must be 3 to 128 bytes in length.

6. Select **Machine Group Identification**.

   - **IPs**

     After this option is selected, you need to enter the ECS servers internal IP address in the **IPs** field.

     Note:

     - Make sure that the ECS servers you enter belong to your department.
     - Make sure that the ECS servers you enter are in the same department as the current Log Service project.
     - Make sure that you use the ECS servers internal IP address (not public IP address) and use linefeeds to separate multiple IP addresses.

   - **User Defined Identity**

     After this option is selected, you must enter your custom ID in **User Defined Identity**.

     Before entering a custom ID, you must have created the custom ID on the server collecting logs. For more information about how to use the user defined identity, see *Use user-defined identity*.

     When scaling up a module, for example, adding servers to a front-end module, you just need to install Logtail and create the configuration file with the custom ID `http_module` on the added servers. Then configurations of different machine groups are automatically synchronized. After the configuration, you can click **Machine Status** to view the new servers.

7. Enter a topic in **Group Topic**.
8. click **Confirm**.

   You can now view the machine group you just created on the machine group list.

What's next

After the machine group is created, you can view the machine group list, modify machine groups, view machine group status, manage machine group configurations, and delete machine groups.

22.5.3.2 Manage a machine group

Log Service manages all ECS servers whose logs need to be collected using Logtail in machine groups. You can go to the **Machine Groups** page by selecting a project from the Log Service project list. Log Service allows you to create, modify, and delete machine groups, view the machine group list and status, manage configurations, and apply machine group IDs.

Create a machine group

A machine group is defined by either of the following two items:

- IP address: defines the name of the machine group and adds the internal IP addresses of servers to the group.
- ID: defines the ID of the machine group and configures the IDs of the machines in the group for association.

You can refer to *Create a machine group* to see how to create a machine group.

View the machine group list

1. *Log on to the Log Service Console*.
2. Create a project and a Logstore. For detailed instructions, see *Operate on projects* and *Operate on Logstores*.
3. Click the name of a project to go to the **Logstores** page. In the left-side navigation pane, click **Logtail Machine Group** to go to the **Machine Groups** page.

   You can view all machine groups under the project.

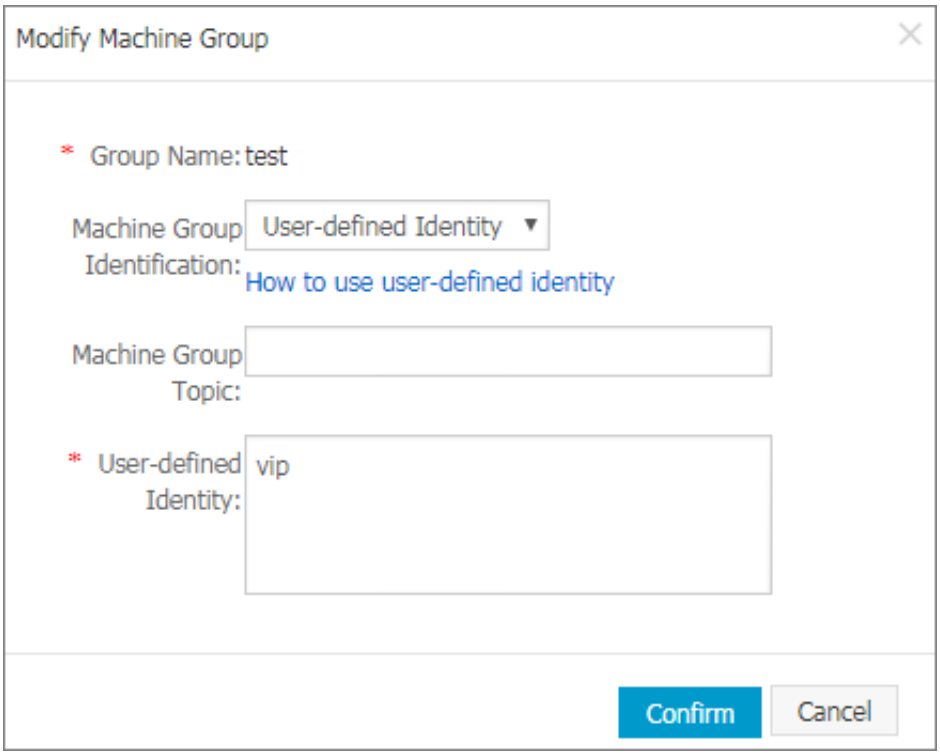
Modify a machine group

After a machine group is created, you can adjust the ECS server list in the machine group as required.

Note:
The machine group name cannot be modified after the machine group is created.

1. *Log on to the Log Service Console*.
2. Create a project and a Logstore. For detailed instructions, see *Operate on projects* and *Operate on Logstores*.
3. Click the name of a project to go to the **Logstores** page. In the left-side navigation pane, click **Logtail Machine Group** to go to the **Machine Groups** page.

   You can view all machine groups under the project.
4. Select the machine group you want to modify and click **Modify**.
5. Modify the configurations of the machine group and then click **Confirm**.

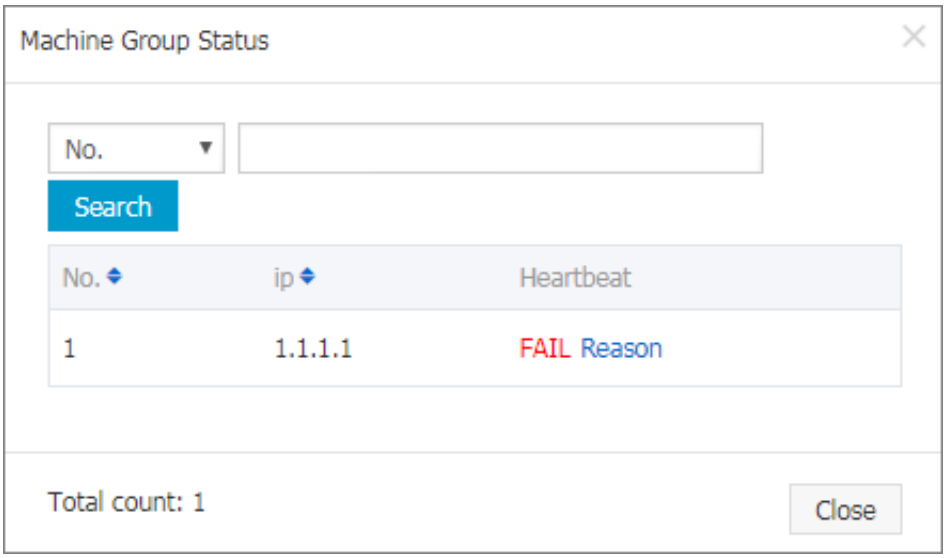


View the machine group status

To verify that the Logtail agent is successfully installed on all ECS servers in a machine group, you can view the heartbeat information of the Logtail agent.

1. *Log on to the Log Service Console*.
2. Create a project and a Logstore. For detailed instructions, see *Operate on projects* and *Operate on Logstores*.
3. Click the name of a project to go to the **Logstores** page. In the left-side navigation pane, click **Logtail Machine Group** to go to the **Machine Groups** page.
4. Select a machine group and click **Machine Status**.

If the Logtail agent is successfully installed on all ECS servers, the heartbeat status of all ECS servers is **OK**. If the heartbeat status of an ECS server is**FAIL**, perform self-check as prompted. If the problem persists, submit a ticket for help.



Note:

- Heartbeat **OK** means that Logtail is successfully connected to Log Service. After a machine is added to a machine group, a latency of several minutes exists before the heartbeat **OK** status is displayed. Please wait patiently.
- If the heartbeat status of an ECS server is always **FAIL**, perform troubleshooting by referring to *Install Logtail on Linux*.

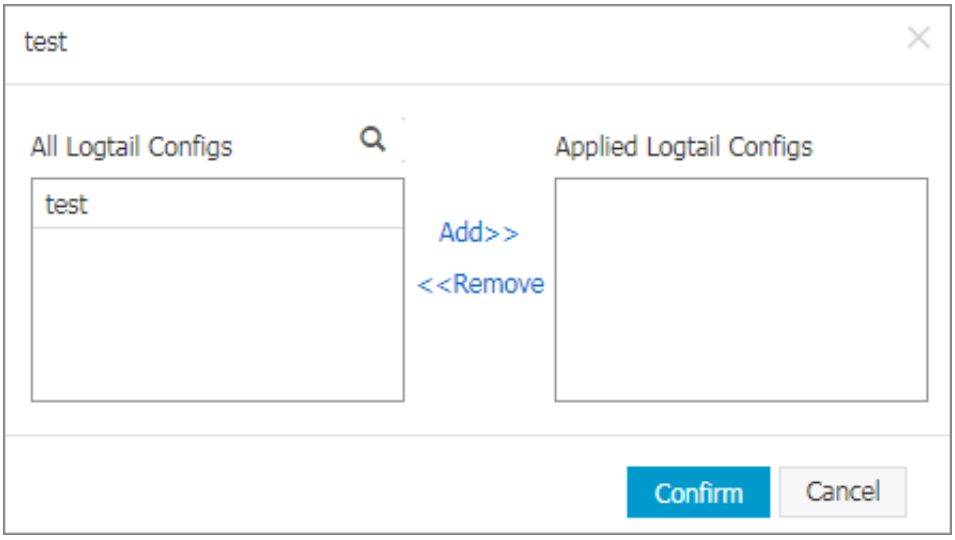
Management configurations

Log Service manages all ECS servers whose logs need to be collected in the form of machine groups. One important management item is the collection configurations of the Logtail agent. For details, see *Use Logtail to collect text files* and *Use Logtail to collect syslogs*. You can apply or delete the Logtail configurations of a machine group to decide on each ECS server, what logs are collected, how the logs are parsed, and to which LogStore the logs are sent by the Logtail agent.

1. *Log on to the Log Service Console*.
2. Create a project and a Logstore. For detailed instructions, see *Operate on projects* and *Operate on Logstores*.
3. Click the name of a project to go to the **Logstores** page. In the left-side navigation pane, click **Logtail Machine Group** to go to the **Machine Groups** page.
4. Select a machine group and click **Config**.

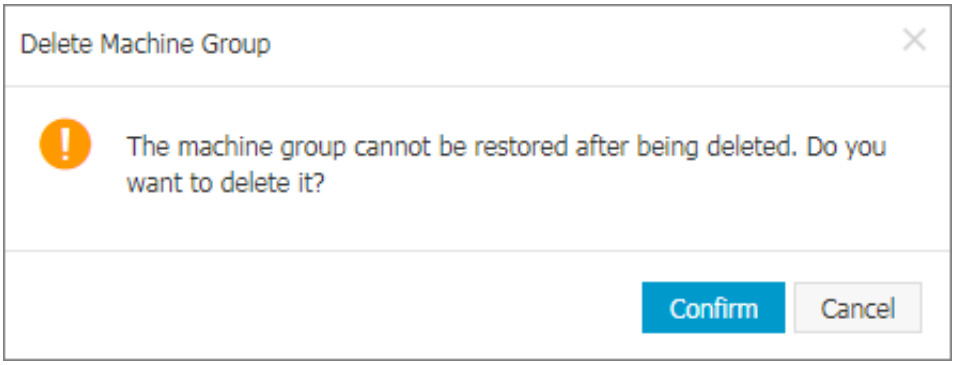
5. Select the Logtail configuration and click **Add** or **Delete** to modify the Logtail configuration applied to the machine group.

   After a Logtail configuration is created, the configuration is issued to the Logtail agent on each ECS server in the machine group. After a Logtail configuration is deleted, the configuration is also removed from the Logtail agent.



Delete a machine group

1. *Log on to the Log Service Console*.
2. Create a project and a Logstore. For detailed instructions, see *Operate on projects* and *Operate on Logstores*.
3. Click the name of a project to go to the **Logstores** page. In the left-side navigation pane, click **Logtail Machine Group** to go to the **Machine Groups** page.
4. Select a machine group and click **Delete**.
5. Click **Confirm** in the confirmation dialog box.



22.5.3.3 Configure a user-defined identity for a machine group

Logtail reports machine IDs to the Log Service agent after it starts. Logtail operates properly only when the IDs reported by the agent are the same as the IDs of machines in the machine group.

Besides IP addresses, you can use the label user-defined ID to dynamically define the machine group.

User-defined identity is advantageous in following scenarios:

- In a custom network environment such as Virtual Private Cloud (VPC), the IP addresses of different machines may conflict with each other, which makes Log Service fail to manage Logtail. User-defined ID helps to avoid such situation.
- Multiple machines use the same label to implement the auto scaling of the machine group. You must only configure the same user-defined ID for the newly added machine. Log Service can automatically identify it, and add to the machine group.

Procedure

To use the user-defined identity to dynamically define the machine group, procedure is as follows:

Step 1 Enable user-defined ID

Linux Logtail

Set the user-defined ID by using the `/etc/ilogtail/user_defined_id` file.

For example, set a user-defined machine ID as follows:

```
#cat /etc/ilogtail/user_defined_id
```

Step 2 Create a machine group.

1. *Log on to the Log Service Console*.
2. Click the name of a project to go to the **Logstores** page.
3. On the **Machine Groups** page, click **Create Machine Group** on the upper-right corner.
4. Complete the configurations for the machine group.
   - Group Name: Enter a name for the machine group.
   - Machine Group Identification: Select **User-defined Identity**.
   - User-defined Identity: Enter the user-defined ID configured in step 1.
5. Click **Confirm** to create the machine group. To expand machines, complete step 1 on the server to be added.

Step 3 View machine group status.

On the **Machine Groups** page, click **Machine Status** at the right of the machine group to view the list of machines that use the same user-defined identity and their heartbeat status.

Other operations

Disable user-defined ID

To use IP address as the machine group identification, delete the user_defined_id file. The configuration takes effect in one minute.

- **Linux operating system**

```
rm -f /etc/ilogtail/user_defined_id
```

Effective time

After you add, delete, or modify the user_defined_id file, the latest configuration takes effect in one minute by default.

Example

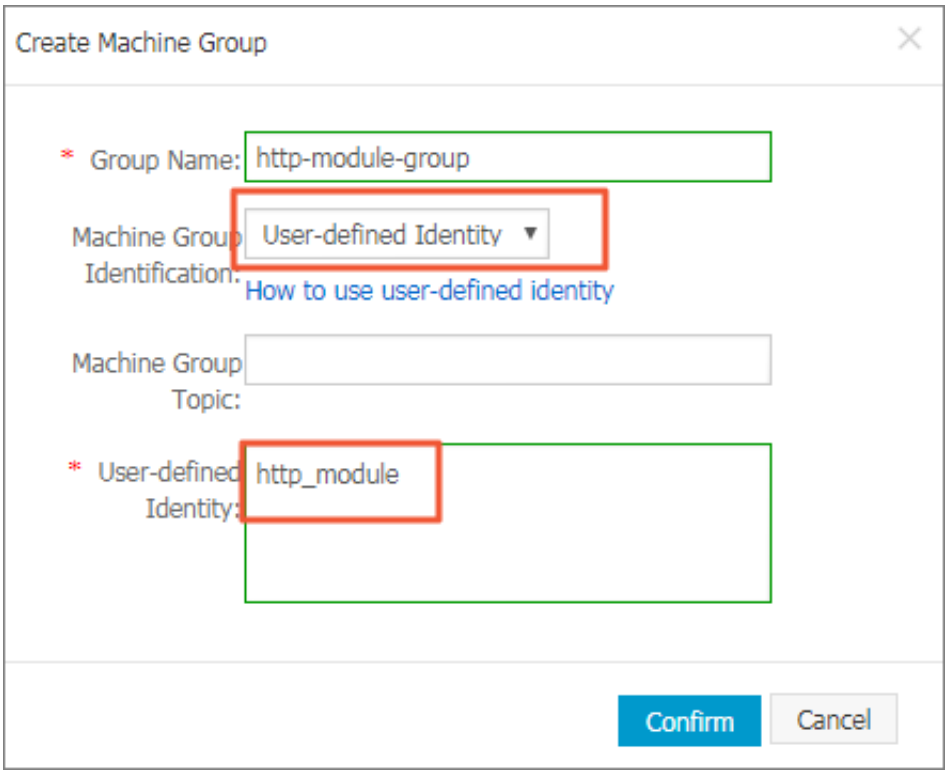
Generally, the system is composed of multiple modules. Each module can contain multiple machines, for example, a common website is composed of frontend HTTP request processing module, cache module, logic processing module, and storage module. Each part is horizontally scalable. Therefore, logs must be collected in real time when machines are being added.

1. Create a user-defined identity.

   After installing the Logtail client, enable the user-defined ID for the server. For the modules in the preceding example, the user-defined identities can be defined as http_module, cache_module, logic_module, and store_module.

2. Create a machine group.

   Enter the corresponding user-defined identify of the machine group in the **User Defined Identity**field when creating the machine group. See the following configurations of the http_module machine group. The http_module Machine Group is shown in the following figure:



3. Click **Machine Status** at the right of the machine group to view the list of machines that use the same user-defined identity and their heartbeat status.

4. If the frontend module has a machine 10.1.1.3 added, complete step 1 on the newly added machine. After the successful operation, you can view the added machine in the **Machine Group Status** dialog box.

22.5.3.4 Manage collection configurations

The Logtail agent provides an easy method to collect logs from ECS through the Log Service console. You must create log collection configurations for the Logtail agent after installing it. For more information about Logtail installation, see *Install Logtail on Linux*. You can create and modify the Logtail configurations of LogStores in the LogStore list.

Create Logtail configurations

For more information about how to create Logtail configurations on the Log Service console, see *Use Logtail to collect text files* and *Use Logtail to collect syslogs*.

View the Logtail configuration list

1. *Log on to the Log Service Console*.
2. Click the name of a project to go to the **Logstores** page.
3. On the **LogStores** page, click Logtail Config **(Manage)** in the **Log Collection Mode** column to go to the **Logtail Configurations** page.

> **Note:**
> A single file can only be collected with one configuration.

Modify Logtail configurations

1. *Log on to the Log Service Console*.
2. Click the name of a project.
3. On the **Logstores** page, click **Manage** to go to the **Logtail Configurations** page.
4. Click the name of the Logtail configuration to be modified.

   You can modify the log collection mode and specify the machine group to which the modified mode is applied. The configuration modification process is the same as the configuration creation process.

Delete Logtail configurations

1. *Log on to the Log Service Console*.
2. Click the name of a project.
3. On the **LogStores** page, click Logtail Config **(Manage)** in the **Log Collection Mode** column to go to the **Logtail Configurations** page.
4. Select the Logtail configuration you want to delete and click **Delete** on the right.

   After the configuration is deleted successfully, it is unbound from the machine group and Logtail no longer collects the log files that match the deleted configuration.

> **Note:**
> Before deleting a Logstore, you must delete all of its Logtail configurations.

22.5.4 Troubleshooting

22.5.4.1 View the local log collection status

Overview

Logtail can be used to view the health status and log collection progress. This helps you to check log collection issues and customize status monitoring for log collection.

Instruction

If a Logtail agent that supports status query is installed, you can view the local log collection status by running commands on the agent. For more information about how to install Logtail, see *Install Logtail (for Linux)*.

Run the `/etc/init.d/ilogtaild -h` command on the agent to check whether the agent allows you to view the local log collection status. If the `logtail insight, version` keywords are returned, status query is supported on the Logtail agent.

```
/etc/init.d/ilogtaild -h Usage: ./ilogtaild { start | stop (graceful,
flush data and save checkpoints) | force-stop | status | -h for help
}$ logtail insight, version : 0.1.0 commond list : status all [index
] get logtail running status status active [--logstore | --logfile]
index [project] [logstore] list all active logstore | logfile. if use
 --logfile, please add project and logstore. default --logstore status
 logstore [--format=line | json] index project logstore get logstore
 status with line or json style. default --format=line status logfile
 [--format=line | json] index project logstore fileFullPath get log
 file status with line or json style. default --format=line status
 history beginIndex endIndex project logstore [fileFullPath] query
logstore | logfile history status. index : from 1 to 60. in all, it
means last $(index) minutes; in active/logstore/logfile/history, it
means last $(index)*10 minutes
```

Currently, Logtail supports the following query commands, command functions, query time intervals, and time windows for result statistics:

Command	Feature	Query time interval	Time window for statistics
all	Query the running status of Logtail	Past 60 minutes	1 minutes
active	Query currently active Logstores or log files (with data collected)	Past 600 minutes	10 minutes
logstore	Query the collection status of a Logstore	Past 600 minutes	10 minutes

Command	Feature	Query time interval	Time window for statistics
logfile	Query the log file collection status	Past 600 minutes	10 minutes
history	Query the collection status of a Logstore or log file over a period of time	Past 600 minutes	10 minutes

Note:

- The `index` parameter in the command indicates the index value of the time window, which is counted from the current time. Its valid range is from 1 to 60. If the time window for statistics is 1 minute, windows in the last `(index, index-1]` minutes are queried. If the time window for statistics is 10 minutes, windows in the last `(10*index, 10*(index-1)]` minutes are queried.
- All query commands belong to status subcommands, so the main command is **status**.

all command

Command format

```
/etc/init.d/ilogtaild status all [ index ]
```

Note:

The **all** command is used to view the running status of Logtail. The index parameter is optional. If left blank, the default value is 1.

Example

```
/etc/init.d/ilogtaild status all 1 ok /etc/init.d/ilogtaild status all
 10 busy
```

Output description

Item	Description	Priority	Solution
ok	The current status is normal.	None.	No action needed.
busy	The current collection speed is high, and	None.	No action needed.

Item	Description	Priority	Solution
	Logtail is running properly.		
many_log_files	The large number of log files are being collected.	Low	Check for any files that need not be collected in the configuration.
process_block	Current log parsing is blocked.	Low	Check whether logs are generated too quickly. If you still get this output, change *Configure startup parameters* as needed to modify the maximum CPU usage or the highest number of concurrent network transmissions.
send_block	Current sending is blocked.	Relatively high	Check whether logs are generated too quickly or the network status remains normal. If you still get this output, change *Configure startup parameters* as needed to modify the maximum CPU usage or the highest number of concurrent network transmissions.
send_error	Failed to upload log data.	High	To troubleshoot the issue, see *Query error diagnostics*.

active command

Command format

```
/etc/init.d/ilogtaild status active [--logstore] index /etc/init.d/
ilogtaild status active --logfile index project-name logstore-name
```

Note:

- The `active [--logstore] index` command is used to query currently active Logstores. The `--logstore` parameter can be removed without changing the meaning of the command.
- The `active --logfile index project-name logstore-name` command is used to view all active log files in a Logstore for a project.
- The **active** command is used to view active log files level by level. We recommend that you first locate the currently active Logstore and then query active log files in this Logstore.

Example

```
/etc/init.d/ilogtaild status active 1 sls-zc-test : release-test sls-
zc-test : release-test-ant-rpc-3 sls-zc-test : release-test-same-regex
-3 /etc/init.d/ilogtaild status active --logfile 1 sls-zc-test release
-test /disk2/test/normal/access.log
```

Output description

- If you run the `active --logstore index` command, all currently active Logstores are returned in the format of `project-name : logstore-name`. If you run the `active --logfile index project-name logstore-name` command, the complete paths of active log files are returned.
- A Logstore or log file with no log collection activity in the current query window does not appear in the output.

logstore command

Command format

```
/etc/init.d/ilogtaild status logstore [--format={line|json}] index
project-name logstore-name
```

Note:

- The **logstore** command is used to output the collection status of a specified project and Logstore in LINE or JSON format.
- If the `--format=` parameter is not set, `--format=line` is selected by default. The echo information is returned in LINE format. *NOTE*: The `--format` parameter must be placed after `logstore`.
- If this Logstore is unavailable or has no log collection activity in the current query window, you get an empty output in LINE format and a `null` value in JSON format.

Example

```
/etc/init.d/ilogtaild status logstore 1 sls-zc-test release-test-same
 time_begin_readable : 17-08-29 10:56:11 time_end_readable : 17-08-
29 11:06:11 time_begin : 1503975371 time_end : 1503975971 project :
sls-zc-test logstore : release-test-same status : ok config : ##1.0
##sls-zc-test$same read_bytes : 65033430 parse_success_lines : 230615
 parse_fail_lines : 0 last_read_time : 1503975970 read_count : 687
avg_delay_bytes : 0 max_unsend_time : 0 min_unsend_time : 0 max_send_s
uccess_time : 1503975968 send_queue_size : 0 send_network_error_c
ount : 0 send_network_quota_count : 0 send_network_discard_count : 0
send_success_count : 302 send_block_flag : false sender_valid_flag :
true /etc/init.d/ilogtaild status logstore --format=json 1 sls-zc-test
 release-test-same { "avg_delay_bytes" : 0, "config" : "##1.0##sls-zc-
test$same", "last_read_time" : 1503975970, "logstore" : "release-test
-same", "max_send_success_time" : 1503975968, "max_unsend_time" : 0,
 "min_unsend_time" : 0, "parse_fail_lines" : 0, "parse_success_lines
" : 230615, "project" : "sls-zc-test", "read_bytes" : 65033430, "
read_count" : 687, "send_block_flag" : false, "send_network_discard
_count" : 0, "send_network_error_count" : 0, "send_network_quota_c
ount" : 0, "send_queue_size" : 0, "send_success_count" : 302, "
sender_valid_flag" : true, "status" : "ok", "time_begin" : 1503975371,
 "time_begin_readable" : "17-08-29 10:56:11", "time_end" : 1503975971,
 "time_end_readable" : "17-08-29 11:06:11" }
```

Output description

Keyword	Meaning	Unit
status	Overall status of this Logstore. For specific statuses, description, and change methods, see the following table.	None.
time_begin_readable	Start time of reading.	None.
time_end_readable	End time of reading.	None.
time_begin	Start time of counting.	Unix timestamp, in seconds.
time_end	End time of counting.	Unix timestamp, in seconds.
project	Project name.	None.
logstore	Logstore name.	None.
config	Name of collection configuration (the globally unique configuration name consists of ##1.0##, project, $, and config).	None.
read_bytes	Number of logs read in the window.	byte

Keyword	Meaning	Unit
parse_success_lines	Number of successfully parsed lines in the window.	line
parse_fail_lines	Number of lines that fail to be parsed in the window.	line
last_read_time	Last read time in the window.	Unix timestamp, in seconds.
read_count	The number of times that logs are read in the window.	Count
avg_delay_bytes	Average of the differences between the current offset and the file size each time logs are read in the window.	byte
max_unsend_time	Maximum time that unsent data packets are in the send queue when the window ends. The value is 0 when the queue is empty.	Unix timestamp, in seconds.
min_unsend_time	Minimum time that unsent data packets are in the send queue when the window ends. The value is 0 when the queue is empty.	Unix timestamp, in seconds.
max_send_success_time	Maximum time that data is successfully sent in the window.	Unix timestamp, in seconds.
send_queue_size	Number of unsent data packets in the current send queue when the window ends.	
send_network_error_count	Number of unsent data packets in the window due to network errors.	
send_network_quota_count	Number of unsent data packets in the window due to quota exceeded.	
send_network_discard_count	Number of discarded data packets in the window due to data exceptions or insufficient permissions.	

Keyword	Meaning	Unit
send_success_count	Number of successfully sent data packets in the window.	
send_block_flag	Whether the send queue is blocked when the window ends .	None.
sender_valid_flag	Whether the send flag of this Logstore is valid when the window ends. TRUE indicates the flag is valid, and FALSE indicates it is disabled due to network errors or quota errors.	None.

Logstore status

Status	Meaning	Processing method
ok	The status is normal.	No action needed.
process_block	Log parsing is blocked.	Check whether logs are generated too quickly. If you still get this output, change *Configure startup parameters* as needed to modify the maximum CPU usage or the highest number of concurrent network transmissions.
parse_fail	Log parsing failed.	Check whether the log format is consistent with the log collection configuration.
send_block	Current sending is blocked.	Check whether logs are generated too quickly or the network status remains normal. If you still get this output, change *Configure startup parameters* as needed to modify the maximum CPU usage or the highest number of concurrent network transmissions.

Status	Meaning	Processing method
sender_invalid	An exception occurred when sending log data.	Check the network status. If the network is normal, see *Query error diagnostics*.

logfile command

Command format

```
/etc/init.d/ilogtaild status logfile [--format={line|json}] index
project-name logstore-name fileFullPath
```

Note:

- The **logfile** command is used to output the collection status of a specific log file in LINE or JSON format.
- If the `--format=` parameter is not set, `--format=line` is selected by default. The echo information is returned in LINE format.
- If this log file is unavailable or has no log collection activity in the current query window, you get an empty output in LINE format or a `null` value in JSON format.
- The `--format` parameter must be placed after `logfile`.
- `filefullpath` must be a full path name.

Example

```
/etc/init.d/ilogtaild status logfile 1 sls-zc-test release-test-same /
disk2/test/normal/access.log time_begin_readable : 17-08-29 11:16:11
time_end_readable : 17-08-29 11:26:11 time_begin : 1503976571 time_end
 : 1503977171 project : sls-zc-test logstore : release-test-same
status : ok config : ##1.0##sls-zc-test$same file_path : /disk2/test
/normal/access.log file_dev : 64800 file_inode : 22544456 file_size_
bytes : 17154060 file_offset_bytes : 17154060 read_bytes : 65033430
 parse_success_lines : 230615 parse_fail_lines : 0 last_read_time :
1503977170 read_count : 667 avg_delay_bytes : 0 /etc/init.d/ilogtaild
 status logfile --format=json 1 sls-zc-test release-test-same /disk2
/test/normal/access.log { "avg_delay_bytes" : 0, "config" : "##1.0
##sls-zc-test$same", "file_dev" : 64800, "file_inode" : 22544456, "
file_path" : "/disk2/test/normal/access.log", "file_size_bytes" :
17154060, "last_read_time" : 1503977170, "logstore" : "release-test
-same", "parse_fail_lines" : 0, "parse_success_lines" : 230615, "
project" : "sls-zc-test", "read_bytes" : 65033430, "read_count" :
667, "read_offset_bytes" : 17154060, "status" : "ok", "time_begin" :
1503976571, "time_begin_readable" : "17-08-29 11:16:11", "time_end" :
1503977171, "time_end_readable" : "17-08-29 11:26:11" }
```

Output description

Keyword	Meaning	Unit
status	Collection status of this log file in the current window period. See **status** of the **logstore** command.	None.
time_begin_readable	Start time of reading.	None.
time_end_readable	End time of reading.	None.
time_begin	Start time of counting.	Unix timestamp, in seconds.
time_end	End time of counting.	Unix timestamp, in seconds.
project	Project name.	None.
logstore	Logstore name.	None.
file_path	Path to the log file.	None.
file_dev	Device ID of the log file.	None.
file_inode	Inode of the log file.	None.
file_size_bytes	Size of the last scanned file in the window.	byte
read_offset_bytes	Current parsing offset of this file.	byte
config	Name of collection configuration (the globally unique configuration name consists of `##1.0##`, project, `$`, and config).	None.
read_bytes	Number of logs read in the window.	byte
parse_success_lines	Number of successfully parsed lines in the window.	line
parse_fail_lines	Number of lines that fail to be parsed in the window.	line
last_read_time	Last read time in the window.	Unix timestamp, in seconds.
read_count	The number of times that logs are read in the window.	Count
avg_delay_bytes	Average of the differences between the current offset and	byte

Keyword	Meaning	Unit
	the file size each time logs are read in the window.	

history command

Command format

```
/etc/init.d/ilogtaild status history beginIndex endIndex project-name
logstore-name [fileFullPath]
```

Note:

- The **history** command is used to query the collection status of a Logstore or log file over a period of time.
- `beginIndex` and `endIndex` indicate the start and end values for the code query window index. `beginIndex <= endIndex` is required.
- If `fileFullPath` is not entered, the collection information of the Logstore is queried. If this parameter is entered, the collection information of the log files is queried.

Example

```
/etc/init.d/ilogtaild status history 1 3 sls-zc-test release-test-same
 /disk2/test/normal/access.log begin_time status read parse_success
parse_fail last_read_time read_count avg_delay device inode file_size
 read_offset 17-08-29 11:26:11 ok 62.12MB 231000 0 17-08-29 11:36:11
 671 0B 64800 22544459 18.22MB 18.22MB 17-08-29 11:16:11 ok 62.02MB
230615 0 17-08-29 11:26:10 667 0B 64800 22544456 16.36MB 16.36MB 17
-08-29 11:06:11 ok 62.12MB 231000 0 17-08-29 11:16:11 687 0B 64800
22544452 14.46MB 14.46MB $/etc/init.d/ilogtaild status history 2 5
sls-zc-test release-test-same begin_time status read parse_success
parse_fail last_read_time read_count avg_delay send_queue network_er
ror quota_error discard_error send_success send_block send_valid
max_unsend min_unsend max_send_success 17-08-29 11:16:11 ok 62.02MB
230615 0 17-08-29 11:26:10 667 0B 0 0 0 0 300 false true 70-01-01 08:
00:00 70-01-01 08:00:00 17-08-29 11:26:08 17-08-29 11:06:11 ok 62.12MB
 231000 0 17-08-29 11:16:11 687 0B 0 0 0 0 303 false true 70-01-01 08:
00:00 70-01-01 08:00:00 17-08-29 11:16:10 17-08-29 10:56:11 ok 62.02MB
 230615 0 17-08-29 11:06:10 687 0B 0 0 0 0 302 false true 70-01-01 08:
00:00 70-01-01 08:00:00 17-08-29 11:06:08 17-08-29 10:46:11 ok 62.12MB
 231000 0 17-08-29 10:56:11 692 0B 0 0 0 0 302 false true 70-01-01 08:
00:00 70-01-01 08:00:00 17-08-29 10:56:10
```

Output description

- This command outputs the historical collection information about a Logstore or log file i
- For the description of each output field, see the `logstore` and `logfile` commands.

Return values

Normal return value

0 is returned when all command inputs are valid (**including failure to run a query on a log store or log file**), for example:

```
/etc/init.d/ilogtaild status logfile --format=json 1 error-project
 error-logstore /no/this/file null echo $? 0 /etc/init.d/ilogtaild
status all ok echo $? 0
```

Abnormal return value

A non-zero return value indicates an exception. See the following table for details.

Return value	Type	Output	Troubleshooting
10	Invalid command or missing parameters	`invalid param, use -h for help.`	Enter `-h` to view the help information.
1	The query goes beyond the 1-60 time window	`invalid query interval`	Enter `-h` to view the help information.
1	Cannot query the specified time window	`query fail, error : $(error)`. For more information, see *errno interpretation*.	This issue may occur if the startup time of Logtail is less than the query time span. For other cases, open a ticket.
1	Mismatch of query window time	`no match time interval, please check logtail status`	Check whether Logtail is running. For other cases, open a ticket.
1	No data in the query window	`invalid profile , maybe logtail restart`	Check whether Logtail is running. For other cases, open a ticket.

Example

```
/etc/init.d/ilogtaild status nothiscmd invalid param, use -h for
help. echo $? 10 /etc/init.d/ilogtaild status/all 99 invalid query
interval echo $? 1
```

Scenarios

You can use Logtail health check to view the overall status of Logtail, and perform collection progress query to obtain related metrics during collection. With the obtained information, you can monitor log collection in a customized manner.

Monitor the running status of Logtail

Monitor the running status of Logtail by using the `all` command.

How it works: The current status of Logtail is queried every minute. If Logtail is in the `process_block`, `send_block`, or `send_error` state for 5 minutes, an alarm is triggered.

You can adjust the alert duration and the range of statuses to be monitored based on the importance of log collection in specific scenarios.

Monitor the log collection progress

Monitor the collection progress of a Logstore by using the `logstore` command.

How it works: The `logstore` command is called every 10 minutes to get status information about this Logstore. If `avg_delay_bytes` is over 1 MB (1024 × 1024) or `status` is not `ok`, an alarm is triggered.

The `avg_delay_bytes` alarm threshold can be adjusted based on the log collection traffic.

Determine whether collection of a log file is complete

Determine whether collection of a log file is complete by using the `logfile` command.

How it works: After writing to the log file stops, the `logfile` command is called every 10 minutes to obtain the status information of this file. If this file shows the same value for `read_offset_bytes` and `file_size_bytes`, it means that collection of this log file is complete.

Troubleshoot log collection issues

If log collection is delayed on a server, use the `history` command to find related collection information on this server.

1. If `send_block_flag` is TRUE, it indicates that the log collection delays because of the network.

- If `send_network_quota_count` is greater than 0, you must *split the shard* of the Logstore.
- If `send_network_error_count` is greater than 0, you must check the network connectivity.
- If no related network error occurs, you must adjust the *concurrent transmission limit and traffic limit* of Logtail.

2. Sending-related parameters are normal, but the `avg_delay_bytes` value is higher than a normal one.
   - The average log parsing speed can be calculated by using `read_bytes` to determine whether log generation traffic is normal.
   - The *configuration parameters* of Logtail can be adjusted as needed.
3. `parse_fail_lines` is greater than 0.

   Check whether the parsing configurations for log collection match all logs.

22.5.4.2 Query error diagnostics

Errors may occur during log collection by Logtail, such as regular expression parsing failures, incorrect file paths, and traffic exceeding the shard service capability. Currently, the query function is provided for debugging log collection errors.

Procedure

1. Go to the error diagnostics page.

   Log on to the Log Service Console. Select the name of a project to go to the **Logstores** page. Click **Diagnose** in the **Log Collection Mode** column.
2. View log collection errors.

   The error diagnostics page lists the Logtail collection errors of the specified Logstore.
3. Query log collection errors by machine.

   To query all log collection errors occurred to a specific machine, enter the IP address of the machine in the search box on the query page. Logtail reports errors every five minutes.

   After an error is rectified, check whether the error is reported again based on the error time statistics after the service recovers. Historical errors are displayed before expiration. You can ignore these errors and check only the new errors reported after error rectification.

Diagnostics reference

Error type	Description	Processing method
LOGFILE_PERMINSSION_ALARM	Logtail has no permission to read the specified file.	Check the Logtail startup account for the server. We recommend that you use the the root account.
SPLIT_LOG_FAIL_ALARM	The regular expression at the beginning of the line does not match the beginning of the line of the log, and thus the log cannot be split into lines.	Verify the regular expression at the beginning of the line. For single-line logs, set the regular expression at the beginning of the line to `.*`.
MULTI_CONFIG_MATCH_ALARM	Only one file can be collected by a Logtail configuration at one time.	Check whether a file is collected by multiple configurations. If there are, delete unnecessary configurations.
REGEX_MATCH_ALARM	The log content does not match the regular expression in regular expression mode.	Copy the log sample from the error content for re-matching and generate a new regular expression for parsing.
PARSE_LOG_FAIL_ALARM	Log parsing using JSON or separators fails due to the nonconforming log format.	Click the error to view relevant details.
CATEGORY_CONFIG_ALARM	Logtail collection configuration is invalid.	A common error is that the file path fails to be extracted as a topic by a regular expression. For other errors, submit a ticket.
LOGTAIL_CRASH_ALARM	Logtail crashes because the the resource usage	To change the CPU utilization and memory utilization thresholds, see *Configure startup*

Error type	Description	Processing method
	threshold of the server is exceeded.	*parameters*. If you have doubts, submit a ticket.
REGISTER_INOTIFY_FAIL_ALARM	Registration of log listening in Linux fails, possibly because Logtail does not have the folder access permission or the folder has been deleted.	Check whether Logtail has the folder access permission and whether the folder exists.
DISCARD_DATA_ALARM	This error is due to insufficient CPU resources configured for Logtail or throttling on sent packets.	To modify the maximum value of CPU utilization or the maximum number of concurrently sent packets, see *Configure startup parameters*. If you have doubts, submit a ticket.
SEND_DATA_FAIL_ALARM	(1) No AccessKey is created for the primary account. (2) The Logtail agent cannot connect to Log Service, or the network link quality is poor. (3) The write quota of Log Service is insufficient.	(1) Check the AccessKey according to *View the key pair*. (2) Check the local configuration file / usr/local/ilogtail/ ilogtail_config.json, run curl <service address>, and check the return result. (3) Add the number of shards for Logstores to support writing of larger data volumes.
PARSE_TIME_FAIL_ALARM	Logtail fails to parse the time field based on the time parsing expression.	Configure the time parsing expression correctly based on the log time.
REGISTER_INOTIFY_FAIL_ALARM	Logtail fails to register inotify watcher for the log directory.	Check whether the log monitoring directory exists. If the directory exists, check the

Error type	Description	Processing method
		directory permission setting.
SEND_QUOTA _EXCEED_ALARM	Log writing exceeds the traffic limit.	*Split shards* on the console.
READ_LOG_D ELAY_ALARM	Log collection lags behind log generation . This error is typically due to insufficient CPU resources configured for Logtail or throttling on sent packets.	Modify the maximum CPU utilization or the maximum number of concurrently sent packets according to *Query error diagnostics*. If you have doubts, submit a ticket.
DROP_LOG_ALARM	Log collection lags behind log generation , and the number of unprocessed log rotations exceeds 20. This error is typically due to insufficient CPU resources configured for Logtail or throttling on sent packets.	Modify the maximum CPU utilization or the maximum number of concurrently sent packets according to *Configure startup parameters*. If you have doubts, submit a ticket.
LOGDIR_PER MINSSION_ALARM	Logtail has no permission to read the log monitoring directory.	Check whether the log monitoring directory exists. If the directory exists, check the directory permission setting.
ENCODING_C ONVERT_ALARM	Code conversion fails.	Check whether the configuration is consistent with the log encoding format.
OUTDATED_L OG_ALAR	Logs expire with a time lag of more than 12 hours. Possible cause: Log parsing lags more than 12 hours, the custom time field is incorrectly configured , or the time output of	Check whether READ_LOG_D ELAY_ALARM exists . If yes, handle the error according to the instructions of READ_LOG_D ELAY_ALARM. If not,

Error type	Description	Processing method
	the logging program is abnormal.	check the time field. If the time field is correctly configured, check whether the time output of the logging program is normal. If you have doubts, submit a ticket.
STAT_LIMIT_ALARM	The number of files in the log collection configuration directory exceeds the limit.	Check whether the log collection configuration directory contains many files and subdirectories, and properly configure the root directory of log monitoring and the maximum monitoring depth of the directory.
DROP_DATA_ALARM	Flushing logs into the local disk times out when exiting the process and the logs unflushed are discarded.	This error is typically due to serious log collection congestion. Modify the maximum CPU utilization or the maximum number of concurrently sent packets according to *Configure startup parameters*. If you have doubts, submit a ticket.
INPUT_COLLECT_ALARM	Input source collection is abnormal.	Handle the exception according to the error prompt information.
HTTP_LOAD_ADDRESS_ALARM	The address input by HTTP is invalid.	Check the validity of the address.
HTTP_COLLECT_ALARM	HTTP collection is abnormal.	Handle the exception according to the error prompt information. The error is typically due to timeout.
FILTER_INIT_ALARM	Filter initialization fails.	This error is typically due to an invalid regular expression. Handle the

Error type	Description	Processing method
		exception according as prompted.
INPUT_CANAL_ALARM	mysql binlog runs abnormally.	Handle the exception according to the error prompt information. (The canal service may restart during configuration update. You can ignore errors caused by service restart.)
CANAL_INVALID_ALARM	The internal status of mysql binlog is abnormal.	This error is typically due to meta information inconsistency caused by table schema information change during running. Check whether the table schema is being modified when the error is reported. If not, submit a ticket.
MYSQL_INIT_ALARM	MySQL initialization is abnormal.	Handle the exception according to the error prompt information.
MYSQL_CHECKPOING_ALARM	The checkpoint format in MySQL is abnormal.	Check whether the checkpoint configuration is modified. If not, submit a ticket.
MYSQL_TIMEOUT_ALARM	MySQL query times out.	Check whether MySQL server and network connection are normal.
MYSQL_PARSE_ALARM	Parsing MySQL query results fails.	Check whether the checkpoint format configured on MySQL is consistent with the format of corresponding fields.

Note:

To view all the complete log lines discarded due to a parsing failure, log on to the machine and check */usr/local/ilogtail/ilogtail.LOG*.

22.5.4.3 Log collection error troubleshooting

If an error occurred when you use Logtail to collect logs, perform the following steps for troubleshooting:

Procedure

1. Check whether the primary account is configured with an Access Key.

   On the top navigation bar, click the user name and select **Personal Information** to go to the **Personal Information** page. Click **AccessKey**. In the **Get AccessKey** dialog box, click **Confirm** to check whether the current account has an AccessKey.

2. Check whether the Logtail heartbeat of the machine group is normal.

   Log on to the Log Service Console. On the **Machine Groups** page, check the machine group status. If the heartbeat status shows OK, go to the next step; if the heartbeat status shows FAIL, continue with the troubleshooting.

   The causes of Logtail heartbeat failure include:

   - **Logtail is not installed.**

     Check the client status in Linux:

     ```
     sudo /etc/init.d/ilogtaild status
     ```

     If the Logtail agent is not installed, see *Install Logtail in Linux* to install Logtail on the server where you need to collect logs.

   - **Incorrect parameters are selected during installation.**

     As Log Service operates by region, you must specify the correct endpoint when installing the Logtail agent. Check configuration of your installed agents in the following paths:

     - Linux:*/usr/local/ilogtail/ilogtail_config.json*

     Check the following settings:

     - The network ingress connected to the client is in the same region as your Log Service project. For a list of network ingresses, see *Endpoints*.

- Check whether a correct domain name is selected based on the network environment of your machine. If an internal domain is selected for a VPC environment, client connection will fail. Run Telnet to test the domain name configuration in *ilogtail_config.json*, for example:`telnet logtail.cn-hangzhou-intranet.log.aliyuncs.com 80`

- **An incorrect IP address or user ID is configured on the server.**

  Generally, Logtail obtains the IP address of a machine in one of the following ways:

  - If host name binding in */etc/hosts* on the machine, confirm the bound IP address. You can run the `hostname` command to check the host name.
  - If host name binding is not configured, Logtail obtains the IP address of the machines first network adapter.

  View the IP addresses on the server in the following paths:

  - Linux:*/usr/local/ilogtail/app_info.json*

  If the IP addresses of the machine group at the server end are inconsistent with the IP addresses obtained by the Logtail agent, make the following changes as needed:

  - If incorrect IP addresses of the machine group are entered on the server, modify and save the correct IP addresses of the machine group. Then check the IP addresses again one minute later.
  - If the network configuration (for example, */etc/hosts*) of the machines is modified, restart Logtail to obtain new IP addresses.

  Run the following command to restart Logtail if necessary:

  - Linux:`sudo /etc/init.d/ilogtaild stop; sudo /etc/init.d/ilogtaild start`

3. Check that the collection configuration is created and applied to the machine group.

   After you confirm that the Logtail agent status is normal, check the following configuration:

   a) Check that Logtail configuration is created.

      Check that the log monitoring directory and log file name match those on machines. The directory structure supports both the complete path mode and the wildcard mode.

   b) Check that Logtail configuration is applied to the machine group.

      View the Logtail machine group, click **Config** to check that the target configuration is applied to the machine group.

4. Check for collection errors.

If Logtail is properly configured, check that new data is generated in log file in real time. As Logtail collects incremental data only, it does not read inventory files if the files are not updated. If the log file is updated but the updates cannot be queried in Log Service, diagnose the problem following the steps below:

- **Collection error diagnostic**

  See *Query Logtail collection errors*. Handle errors based on the types of errors reported by Logtail.

- **View Logtail logs**

  Client logs include key information and all warning and error logs. To query complete and real-time errors, view client logs in the following paths:

  — Linux:*/usr/local/ilogtail/ilogtail.LOG*

- **Quota exceeded**

  If many logs or file data needs to be collected, you may modify startup parameters of Logtail to achieve higher log collection throughput. Refer to *Configure startup parameters* to make adjustments.

If the problem persists, submit a ticket to Log Service engineers and provide key information collected during troubleshooting.

22.5.5 Limits

This document describes restrictions on Logtail, including restrictions on file collection, resources, and error handling.

Table 22-1: File collection restriction

Category	Limits
File encoding	UTF8/GBK encoding of log files is supported , and UTF8 encoding is recommended for better processing performance. If log files are encoded in other encoding formats, errors such as garbled characters and data losses occur.
Log file size	Not limited.
Log file rotation	Supported. Files named as `.log*` or `.log` are supported.

Category	Limits
Log collection behavior upon log parsing block	When log parsing is blocked, Logtail retains the open state of the log file descriptor (FD). If log file rotation occurs multiple times during the block, Logtail attempts to keep the log parsing sequence of each rotation. If more than 20 unparsed logs are rotated, Logtail does not process subsequent log files. For more information, see *Related technical documents*.
Soft link	Monitored directories can be soft links.
Single log size	The maximum size of a single log is 512 KB. If multiple-line logs are divided by the regular expression at the beginning of the line, the maximum size of each log is still 512 KB. If the size of a log exceeds 512 KB, the log is forcibly split into multiple parts for collection. For example, if the size of a log is 1025 KB, the first 512 KB is processed, then the 512 KB in the middle is processed, and lastly the 1 KB in the end is processed.
Regular expression	Regular expressions can be Perl-compatible regular expressions.
Multiple collection configurations for the same file	Not supported. You are advised to collect log files in a Logstore and configure multiple subscriptions. If this function is required, configure soft links for log files to bypass the restriction.
File opening behavior	Logtail retains the open state of a file to be collected. Logtail closes the file if the file does not have any modification within five minutes (in case that rotation does not occur).
First log collection behavior	Logtail collects only incremental log files. If modifications are found in a file for the first time and the file size exceeds 1 MB, Logtail collects the logs from the last 1 MB. Otherwise, Logtail collects the logs from the beginning. If a log file is not modified after the configuration is issued , Logtail does not collect this file.
Non-standard text log	If a log contains \0 lines, the log is truncated at the position of the first \0 line.

Table 22-2: Checkpoint management

Item	Capabilities and limits
Checkpoint timeout interval	If a file has not been modified for more than 30 days, the checkpoint is deleted.
Checkpoint save policy	Regular save every 15 minutes. Files are automatically saved when the program exits.
Checkpoint save path	The default save path is `/tmp/logtail_checkpoint`. To modify this path, see *Configure startup parameters*.

Table 22-3: Limits on configuration

Item	Capabilities and limits
Configuration update	Updated configuration takes effect with a delay of about 30s.
Dynamic configuration loading	Supported. The configuration update does not affect other collections.
Number of configured tasks	Theoretically unlimited. We recommend that the number of collection configurations on a server is no more than 100.
Multi-tenant isolation	Collection configurations for different tenants are isolated.

Table 22-4: Limits on resources and performance

Item	Capabilities and limits
Log processing throughput	The default traffic of raw logs is limited to 2 MB/s. (Data is uploaded after it is encoded and compressed, with a general compression ratio of 5 to 10 times.) Logs may be lost if the traffic limit is exceeded. To adjust the parameter, see *Configure startup parameters*.
Maximum performance	Single-core capability: The maximum processing capability is 100 MB/s for logs in simple mode, 20 MB/s by default for logs in regular mode (depending on the complexity of regular expressions), 40 MB/s for logs in delimiter mode, and 30 MB/s for logs in JSON mode. After multiple processing threads are enabled, the performance can be improved by 1.5 to 3 times.

Item	Capabilities and limits
Number of monitored directories	Logtail actively restricts the number of monitored directories to avoid excessive consumption of user resources. When the monitoring upper limit is reached, Logtail stops monitoring more directories and log files. Logtail monitors a maximum of 3,000 directories, including subdirectories.
Default resource restriction	By default, Logtail occupies up to 40% of CPU and 256 MB of memory. If logs are generated at a high speed, you can modify the parameters by referring to the *topic*.
Resource out-of-limit processing policy	If the resources occupied by Logtail in 3 minutes exceed the upper limit, Logtail is forced to restart, which may cause loss or duplication of data.

Table 22-5: Limits on error handling

Item	Capabilities and limits
Network error handling	If the network connection is abnormal, Logtail actively retries and automatically adjusts the retry interval.
Handling of resource quota exceeding	If the data transmission rate exceeds the maximum quota of Logstore, Logtail blocks log collection and automatically retries. *Related technical documents*.
Maximum retry period for timeout	If data transmission fails for more than 6 successive hours, Logtail discards the data.
Status self-check	Logtail automatically restarts in the case of an exception, for example, abnormal exit of a program or resource limit exceeding.

Table 22-6: Other limits

Item	Capabilities and limits
Log collection delay	Except for block status, the delay in log collection by Logtail does not exceed one second after logs are flushed to a disk.
Log upload policy	Logtail automatically aggregates logs in the same file before uploading the logs. Log uploading is triggered if the number of logs exceeds 2,000, the total size of the log file exceeds 2 MB, or the log collection duration exceeds 3s.

22.6 Index and query

Log Service provides LogSearch/Analytics for large-scale and real-time log query and analysis. You can enable indexing and field statistics to support this function.

Benefits

- Real time: Logs can be analyzed immediately after being written.
- Fast:
  - Query: Billions of data items can be processed within 1 second for a complex query statement (with five conditions).
  - Analysis: Tens of millions of data items can be aggregated and analyzed within 1 second for a complex analysis statement (with aggregation by five dimensions and the GroupBy condition).
- Flexible: Query and analysis conditions can be changed as required to obtain results in real time.
- Ecological: LogSearch/Analytics can seamlessly integrate with Grafana and Jaeger, in addition to the report, dashboard, and fast analysis features provided by the console, and support RESTful APIs and JDBC.

Basic concepts

When LogSearch/Analytics (indexing) is disabled, raw data can be consumed by shard sequentially, similar to Kafka. After LogSearch/Analytics is enabled, statistics and query of log data are also supported.

Data types

You can set the type of each key in a log. A full-text index is a special key, and the entire log is considered as a value. The following types are supported:

Category	Type	Description	Query example
Base	*Text type*	Text type. Keyword+fuzzy match and Chinese word segmentation are supported.	`uri:"login*" method:"post"`
Base	*Value type*	Numerical value. Range search is supported.	`status>200, status in [200, 500]`
Infrastructure	*Value typeJSON type*	Floating-point number type.	`price>28.95, t in [20.0, 37]`

Category	Type	Description	Query example
Combination	*JSON type*	The content consists of JSON fields, which are of the Text type by default. Nesting is supported. You can set indexes of the Text, Long, and Double types for the b elements at layer a in the a.b path format. The fields adopt the configured types.	`level0.key>29.95 level0.key2:"action"`
Combination	*Text type*	Search is performed on the entire log as text.	`error and "login fail"`

Syntax of LogSearch/Analytics

Query: It consists of Search and Analytics, which are separated using `|`.

```
$Search |$Analytics
```

- Search: It is a search criteria, which can be generated by using keywords, fuzzy match conditions, values, ranges, and combinations. If it is blank or an asterisk (*), all data is used.
- Analytics: It calculates and collects statistics on search results or full data.

Note:
The two parts are optional. If Search is blank, all the data for the specified period is not filtered and the result is calculated directly to collect statistics. If Analytics is blank, the search result is returned and no statistics are collected.

Enable indexing

1. Log on to the Log Service console. For more information about how to log on to the Log Service console, see Log on to Apsara Stack console in *Cite LeftApsara Stack Console User GuideCite Right*.
2. Click the name of a project.
3. Select a Logstore and click **Search** in the **LogSearch** column. Click **Enable** in the upper-right corner. If you have enabled indexing, select **Index Attributes** > **Modify** After the LogSearch/Analytics function is enabled, data is indexed in the background. Traffic and storage space for the index are required. If the function is not required, choose <menucascade>Index Attributes</menucascade>**Disable**.

- After the LogSearch/Analytics function is enabled, data is indexed in the background. Traffic and storage space for the index are required.
- If the function is not required, select **Index Attributes** > **Disable**.

4. Go to the **Search & Analysis** page to perform configuration.

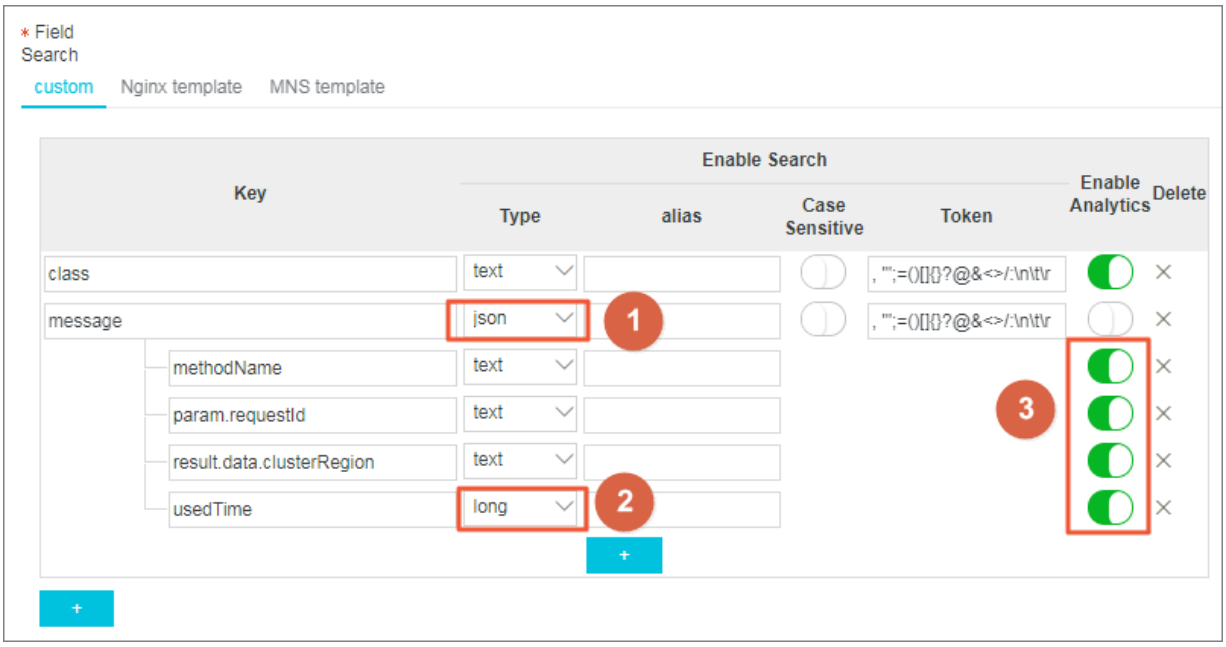
Search example

The following log includes four key values in addition to the time:

No.	Key	Type
0	time	-
1	class	text
2	status	long
3	latency	double
4	message	json

```
0. time:2018-01-01 12:00:00 1. class:central-log 2. status:200 3.
latency:68.75 4. message: { "methodName": "getProjectInfo", "success
": true, "remoteAddress": "1.1.1.1:11111", "usedTime": 48, "param":
 { "projectName": "ali-log-test-project", "requestId": "d3f0c96a-51b0
-4166-a850-f4175dde7323" }, "result": { "message": "successful", "code
": "200", "data": { "clusterRegion": "ap-southeast-1", "ProjectName
": "ali-log-test-project", "CreateTime": "2017-06-08 20:22:41" }, "
success": true } }
```

The settings are as follows:



Wherein:

- (1) indicates query of all the data of the String and Boolean types in JSON fields.
- (2) indicates query of data of the Long type.
- (3) indicates SQL analysis of configured fields.

Example 1: Query of data of the String and Boolean types

```
class : cental* message.traceInfo.requestId : 92.137_1518139699935_55
99 message.param.projectName : ali-log-test-project message.success :
true
```

Note:

- JSON fields need not be configured.
- JSON map and array expand automatically. Multi-layer nesting is supported, with layers separated by ".".

Example 2: Query of data of the Double and Long types

```
latency>40 message.usedTime > 40
```

Example 3: Combination query

```
class : cental* and message.usedTime > 40 not message.param.projectNam
e:ali-log-test-project
```

Others

If the data volume of logs to be searched for is very large (for example, the time span is long and there are more than 10 billion log entries), the data cannot be searched completely by one query request. In this case, Log Service will return the existing data to you, and inform you that the query results are not complete in the returned results.

At the same time, the server will cache query results within 15 minutes. When a portion of the query request results are cache hits, the server will continue to scan the log data that are not cache hits for this request. To reduce your workload of merging multiple query results, Log Service merges the hit query results in the cache and the new hit results of the current query and returns them to you.

Therefore, Log Service allows you to call this interface repeatedly using the same parameters to obtain the final complete results.

22.6.1 Text type

Similar to search engines, text data is queried based on term matching. Therefore, you must configure word segmentation, case sensitivity, and enable Chinese word segmentation.

Configuration instructions

Case sensitivity

Case sensitivity for raw log query. For example, if the raw log is "internalError":

- If the parameter is set to **False** (case insensitive), the sample log can be located with the keyword either "INTERNALERROR" or "internalerror".
- If the parameter is set to **True** (case sensitive), the sample log can be located only with the keyword "internalError".

Word segmentation

You can separate the contents of a raw log into several keywords by using a word segmentation.

For example, when we query the following log content:

```
/url/pic/abc.gif
```

- If no word segmentation is set, the string is considered as an individual word`/url/pic/abc.gif`. You can only query this log by using the complete string or fuzzy match such as`/url/pic/*`.
- If `/`is set as the word segmentation, the raw log is separated into three words: `url`, `pic`, and `abc.gif`. You can find the log by query or fuzzy query with any of the keywords `url`, `abc.gif`, and `pi*`, or with `/url/pic/abc.gif` (segmented into `url, pic, and abc.gif` in query).
- If the word segmentation is set to `/.`, the raw log is separated into four words: `url`, `pic`, `abc`, and `gif`.

Note:

You can extend the query range by setting appropriate word segmentations.

Contains Chinese characters

If the log contains Chinese characters, enable Chinese word segmentation. For example, for the following log content:

```
buyer:用户小李飞刀lee
```

With the default word segmentation ":", the raw log content is segmented into two words: `buyer` and `用户小李飞刀lee`. If you search for `用户`, `Lee` will not be returned. If you enable the option of **Chinese character included**, the Log Service analyzer analyzes the meaning of the Chinese words and segments the log content into five words: `buyer`, `用户`, `小李`, `飞刀`, and `lee`. You can locate the log with either the keyword `飞刀` or `小李飞刀` (resolved into: `小李 and 飞刀`).

Note:
The function of Chinese word segmentation somehow compromises the write speed. Set the option with caution based on your need.

Full text index

By default, full text query (index) considers all the fields and keys of a log, except the time field, as text data, and does not need to specify keys. For example, the following log is composed of four fields (time/status/level/message):

```
[20180102 12:00:00] 200,error,some thing is error in this field
```

- time:2018-01-02 12:00:00
- level:"error"
- status:200
- message:"some thing is error in this field"

After enabling full text index, the following text data is assembled in the "key:value + space" mode. For example:

```
status:200 level:error message:"some thing is error in this field"
```

Note:

- Prefix is not required for full text query. Enter error as the keyword, both level field and message field meet the query condition.

- You must set a word segmentation for the full text query. If a space is set as the word segmentation, status:200 is considered as a phrase. If ":" is set as the word segmentation, status and 200 are considered as two independent phrases.
- Numbers are processed as texts. For example, you can use the keyword 200 to query this log. The time field is not processed as a text.
- You can query this log if you enter a key such as "status".

22.6.2 Value type

When configuring indexes, you can configure a field as the value type and query the key by using a value range.

Configuration instructions

Supported types: `long`(long integer) or `double` (decimal). After configuring a field as the value type, you can only query the key by using a value range.

Example

To query the longkey whose key range is (1000 2000], use the following methods.

- Use values to query the longkey:

```
longKey > 1000 and longKey <= 2000
```

- Use an interval to query the longkey:

```
longKey in (1000 2000]
```

For more syntax, see *Query syntax*.

22.6.3 JSON type

JSON is a combined data type consisting of Text, Boolean, Number, Array, and Map.

Configuration instructions

Text-type data

JSON fields of the Text and Boolean types are automatically identified.

For example, the following JSON keys can be searched for using `jsonkey.key1:"text_value"` and `jsonkey.key2:true`.

```
jsonkey: {
   key1:text_value,
   key2:true,
```

```
   key3:3.14
}
```

Number-type data

You can search for data of the Double and Long types in non-JSON arrays by setting a type and specifying a path.

For example, the statement used to search for the jsonkey.key3 field of the Double type is as follows:

```
 jsonkey.key3 > 3
```

Non-fully valid JSON

Non-fully valid JSON data is parsed until the invalid content appears.

for example:

```
"json_string":
{
     "key_1" :  "value_1",
     "key_map" :
      {
             "key_2" : "value_2",
             "key_3" : "valu
```

The data following key_3 is truncated and lost. The log with missing data is correctly parsed until the `json_string.key_map.key_2` field.

Note

- JSON object and JSON array are not supported.
- Fields cannot appear in JSON arrays.
- Fields of the Boolean type can be converted to the text type.

Query syntax

The parent path prefix in JSON is required to search for a specified key. The query syntax for the text and numerical types is similar to other types. For more information, see *Query syntax*.

22.6.4 Query syntax

To help you query logs more effectively, Log Service provides a set of query syntaxes used to express query conditions. You can specify query conditions through the GetLogs and

GetHistograms APIs on Log Service or on the query page of the Log Service console. This section details the query condition syntax.

Index type

Log Service supports creating an index for the Logstore in two modes:

- Full-text indexing: The entire line of log is queried as a whole, without differentiating the key and value (Key, Value).
- Key value indexing: Query is performed when Key is specified, for example, *`FILE:app`*, *`Type: action`*. All the contained strings under this key will be hit.

Syntax keyword

LogSearch query conditions support the following keywords:

Name	Meaning
and	Binary operator. The format is `query1 and query2`, indicating the intersection of the query results of `query1` and `query2`. If there is no syntax keyword between words, the relation between words is `and` by default.
or	Binary operator. The format is . `query1 or query2` , indicating the intersection of the query results of `query1` and `query2`.
not	Binary operator. The format is `query1 not query2`, indicating a result that meets `query1` and does not meet `query2`, that is, `query1-query2`. If only `not query1` exists, it indicates that logs that do not contain the query results of `query1` are selected.
(,)	The left and right brackets are used to merge one or multiple sub -queries into one query to increase the priority of query in the brackets.
:	Used to query the key-value pair. `term1:term2` forms a key-value pair. If the key or value contains reserved characters such as spaces and colons (:)`:`, quotation marks `""` are required to enclose the entire key or value.
"	Convert a keyword into a common query character. Any term in the left and right quotation marks will be queried and will not be used as a syntax keyword. Or all the terms in the left and right quotation marks are regarded as a whole in the key-value query.

Name	Meaning
\	Escape character. Used to escape quotation marks. The quotation marks after escaping indicate the symbols themselves and are not considered as escape characters, for example, `"\""`.
|	Pipeline operator, indicating more computing based on the previous computing, for example, query1 | timeslice 1h | count.
timeslice	Time slice operator indicates the length of time during which the data is regarded as a whole for computing, and the use methods are timeslice 1h, timeslice 1m, and timeslice 1s, which respectively indicate 1 hour, 1 minute, and 1 second as a whole. For example, query1 | timeslice 1h | count indicates querying the query condition, and the total number of times with 1 hour as the time slice is returned.
count	Count operator, indicating the number of logs.
*	Fuzzy query keyword, used to replace zero or more characters. For example, if `que*` is used in a query, all the hit words starting with `que` will be returned. NOTE: Up to 100 results meeting the keyword are returned for the query.
?	Fuzzy query keyword, used to replace one character. For example, if `qu?ry` is used in a query, all the hit words starting with `qu`, ending with `ry`, and with a character in the middle are returned.
`__topic__`	Query the data under a certain topic. Under the new syntax, the data of zero or more topics can be queried in the query, for example, `__topic__:mytopicname`.
`__tag__`	Query a tag value under a tag key, for example, `__tag__:tagkey:tagvalue`.
source	Query data of an IP address, for example, `source:127.0.0.1`.
>	Query the logs with the value of a field greater than a specific number, for example, `latency > 100`.
>=	Query the logs with the value of a field greater than or equal to a specific number, for example, `latency >= 100`.
<	Query the logs with the value of a field smaller than a specific number, for example, `latency < 100`.
<=	Query the logs with the value of a field smaller than or equal to a specific number, for example, `latency <= 100`.

Name	Meaning
=	Query the logs with the value of a field equal to a specific number, for example, `latency = 100`.
in	Query the logs with a field falling in a specific range. Brackets ([]) are used to indicate closed intervals and parentheses (()) are used to indicate open intervals, with two numbers enclosed and separated by spaces. For example, `latency in [100 200]` or `latency in (100 200]`.

Note:

- Syntax keywords are case-insensitive.
- Priorities of syntax keywords are sorted in the descending order as `: > " > ( ) > and not > or`.
- Log Service also reserves the right to use the following keywords. If you need to use these keywords, enclose the keywords with double quotation marks: `sort asc desc group by avg sum min max limit`.
- If the full text index and key value index have different word segmentation characters when they are configured, data cannot be queried using the full text query method.
- To perform a numeric query, set the data type of the queried column to double or long. If no data type is set or the syntax used for the numeric range query is incorrect, Log Service translates the query condition into a full text index, which may lead to an unexpected result.
- If you change the data type of a column from text to numeric, only the = query is supported for the data prior to this change.

Query example

1. Logs that contains a and b at the same time: `a and b` or `a b`
2. Logs that contain a or b: `a or b`
3. Logs that contain a but no b: `a not b`
4. Those in all the logs that contain no a: `not a`
5. Query the logs that contain a and b, but no c: `a and b not c`
6. Logs that contain a or b and must contain c: `(a or b ) and c`
7. Logs that contain a or b, but no c: `(a or b ) not c`
8. Logs that contain a and b and may contain c: `a and b or c`
9. Logs with the FILE field containing apsara: `FILE:apsara`

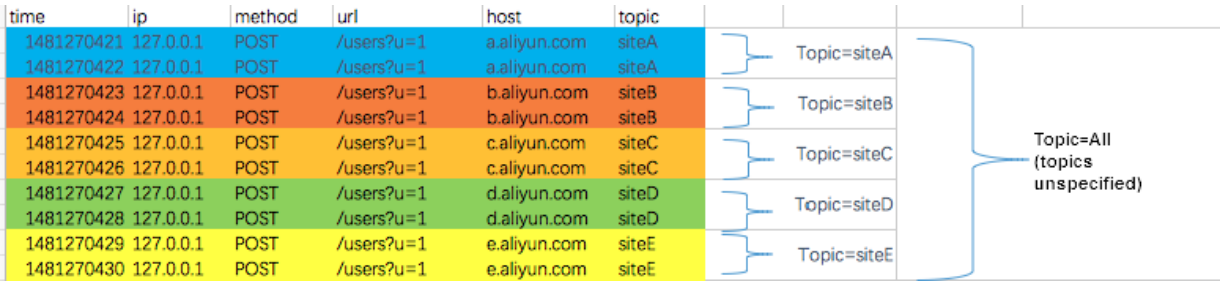
10. Logs whose FILE field contains apsara and shennong: `FILE:"apsara shennong"`, `FILE: apsara FILE: shennong`, or `FILE:apsara and FILE:shennong`
11. Logs containing and: `and`
12. Logs with the FILE field containing apsara or shennong: `FILE:apsara or FILE:shennong`
13. Logs with the file info field containing apsara: `"file info":apsara`
14. Logs that contain quotation marks: `\"`
15. Logs starting with shen: `shen*`
16. Query all the logs starting with shen under the FILE field: `FILE:shen*`
17. Query the logs starting with shen, ending with ong and with a character in the middle: `shen? ong`
18. Query all the logs starting with shen and aps: `shen* and aps*`
19. Query the distribution of logs starting with shen, with the time slice of 20 minutes: `shen* | timeslice 20m | count`
20. Query all the data under topic1 and topic2: `__topic__:topic1 or __topic__ : topic2`
21. Query all the data of tagvalue2 under tagkey1: `__tag__ : tagkey1 : tagvalue2`
22. A query for all the data with a latency greater than or equal to 100 and less than 200 can be written in either of the following ways: `latency >=100 and latency < 200` or `latency in [100 200)`.
23. A query for all the requests with a latency greater than 100 must be written in the following way: `latency > 100`.
24. Query logs that do not contain crawler and logs with http_referer not containing opx: <codeph> latency > 100</codeph>.
25. Query logs with the cdnIP field being null: <codeph>latency > 100</codeph>.
26. Query logs without the cdnIP field: `not cdnIP:*`.
27. Query logs with the cdnIP field: `cdnIP:*`.

Specified or cross-topic query

Each Logstore can be divided into one or more subspaces according to the topic. During query, the query range can be limited for the specified topic to increase the speed. Therefore, the user with the level-2 classification requirement for the LogStore is recommended to use topic to divide the LogStore.

When one or more topics are specified to perform query, query is implemented in the topic meeting the conditions only. However, if no topic is entered, the data under all the topics is queried by default.

For example, topics are used to classify logs under different domain names:



Topic query syntax:

- The data under all topics can be queried. The data of all topics is queried if no topic is specified in the query syntax and parameters.
- Topic can be queried in the query. The query syntax is `__topic__:topicName`. The old mode is still supported at the same time. The topic is specified in the URL parameter.
- Multiple topics can be queried, for example, `__topic__:topic1 or __topic__:topic2` indicates the union of data under topic1 and topic2.

22.6.5 Context query

When you expand a log file, each log records an event, and they do not exist independently. Several consecutive logs can be used to review the occurrence process of the whole event sequence.

Log context query specifies the log source (machine + files) and a log in it, and searches a number of records (the text above) before this log and a number of logs (the text below) after this log in the original file. Particularly, it is an effective way for clarifying the problem cause and effect under the DevOps scenario.

The Log Service console provides a dedicated page for query. You can use a browser to view the context information in the original file of the specified log, in a way similar to turning the pages of the original log file. By viewing the context information of a specified log, you can quickly identify related fault information during service troubleshooting, and locate problems with ease.

Scenarios

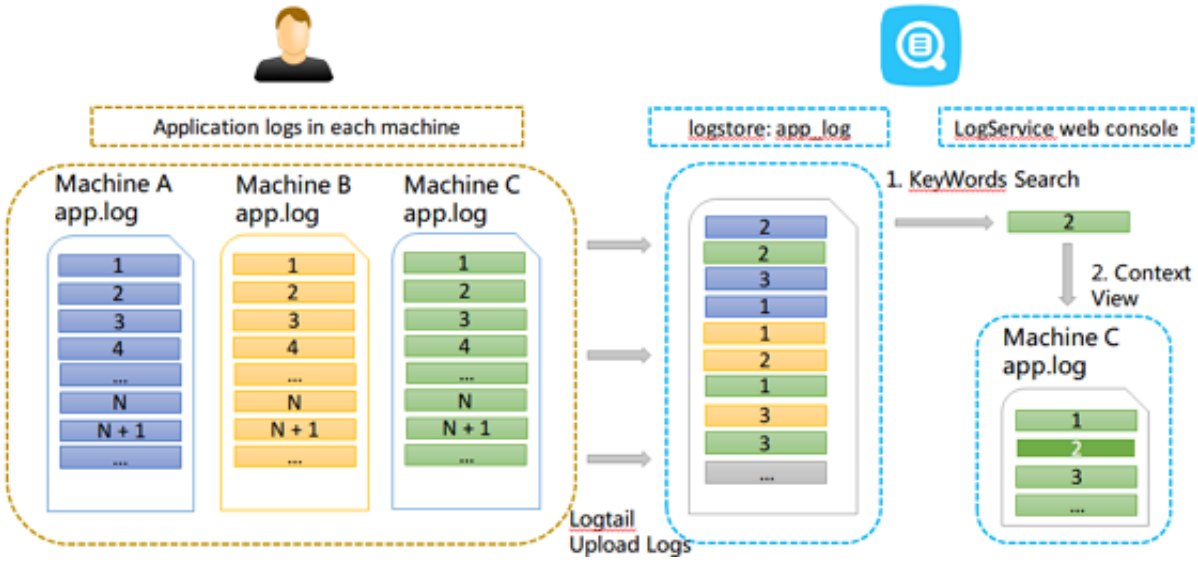
For example, the O2O take-out website will record the transaction track of an order in the program log on the server:

User Logon > **Browse Items** > **Click Items** > **Add to Shopping Cart** > **Place an Order** > **Pay for the Order** > **Deduct Payment** > **Generate Order**

If you fail to place the order, the O&M personnel need to identify the cause quickly. For a conventional context query, the administrator adds the machine logon permission to related members, and then the investigator logs on to each machine where applications are deployed in sequence and uses the order ID as the keyword to search application log files and identify the cause of a failed order.

Log Service allows you to troubleshoot in the following approach:

1. Install the log collection client Logtail on the server, and add the machine group and log collection configuration on the console. Then, Logtail starts to upload the incremental logs.
2. Access the console log query page of Log Service, specify the time range, and find the order failure log according to the order ID.
3. Page up with the found error log as benchmark till other related log information is found (for example: credit card deduction fails).



Benefits

- There is no intrusion into the application, and no need to change the log file format.

- The specified log context information of any machine and file can be viewed in the Log Service console, avoiding the trouble of logging on to each machine to view the log file.
- In combination with the time of event occurrence, if the context query is performed after the suspicious log is located quickly in the specified time range of the Log Service console, you can always get twice the result with half the effort.
- You do not need to worry about data loss caused by insufficient server storage or log file rotation, and can view historical data in the Log Service console at any time.

Prerequisites

- *Use Logtail to collect logs* To upload data to the Logstore, only machine group creation and collection configuration are required. You can also use producer-related SDKs for uploading, such as Producer Library,

  Log4J, LogBack, and C-Producer Library.
- Enable indexing.

Note:

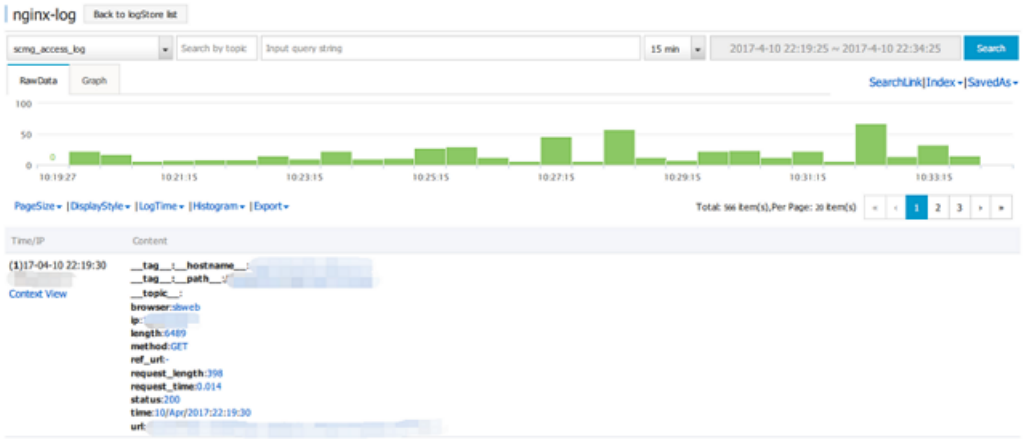
Currently, context query does not support syslogs.

Procedure

1. *Log on to the Log Service Console*.
2. Click the name of a project.
3. On the **Logstores** page, select a Logstore, and click **Search** in the **LogSearch** column to go to the search page.

   If there is a **Context View** link to the left of a log returned on the query results page, the log supports context query.
4. Enter your search and analytic syntax, select a time range, and click **Search**.

   If there is a **Context View** link to the left of a log returned on the query results page, the log supports context query.



5. Select a log and click **Context View**. On the page that appears on the right, view the context log of the target log.
6. Scroll with the mouse on the page to view the context information of the selected log. To view the historic or current information, click **Earlier** or **Later**.

22.6.6 Other functions

In addition to the statement-based query capability, the query and analysis function of Log Service provides the following extended functions for the query optimization:

Raw logs

After the index is enabled, enter the keywords in the search box and select the search time range. Then, click**Search** to view the histogram of the log quantity, the raw logs, and the statistical graph.

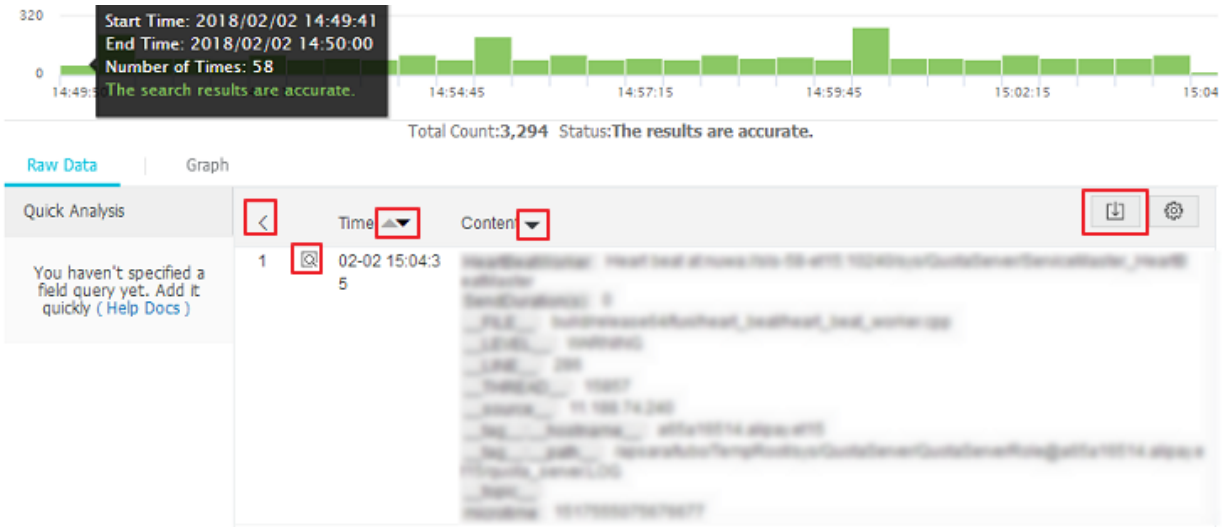
The histogram of the log quantity displays the time-based distribution of log search hit counts . With the histogram, you can view the log quantity changes over a certain period of time. By clicking the rectangular area to narrow down the time range, you can view the information about the log hits within the specified time range to refine the display of the log search results.

On the Raw Log tab, you can view the content of hit logs in time order.

- Click the triangle symbol beside the **Time** column to switch between **chronological** order and **reverse chronological** order.
- Click the triangle symbol beside the **Content**, you can switch between **Display with Line Breaks** or **Display in One Line**.
- Click the value **Keyword** in the log content to view all the log content which contains the keyword.

- On the Raw Log tab, click **Download** in the upper-right corner to download search results in CSV format. Click **Set** to add columns named after fields to the raw log results so that you can view the target fields of each raw log in the new columns.
- Click **Context** to view the 15 logs preceding the log and the 15 logs following the log. For more information, see *Context query*.

> **Note:**
> Currently, the context query feature is only applicable to the data uploaded by Logtail.



Statistical chart

After you enable indexing and enter a search analysis statement, you can view log statistics on the **Aggregation** tab.

- Data can be displayed in the following ways: tables, line charts, column charts, bar charts, pie charts, numeric values, area charts, and maps.

  You can select an appropriate statistical graph type based on the actual statistical analysis needs.
- You can adjust the display content of axes X and Y to obtain the display results that meet your needs.
- Add the analysis results to**Dashboard**. For more information, see *Dashboard*.

Context query

The Log Service console provides a dedicated page for query. You can use a browser to view the context information in the original file of the specified log, in a way similar to turning the pages of the original log file. By viewing the context information of a specified log, you can quickly identify related fault information during service troubleshooting, and locate problems with ease. For more information, see *Context query*.

Quick analysis

The quick analysis function of Log Service supports a quick interactive query. This service allows you to quickly analyze the distribution of a field over a period of time and reduce the cost of indexing key data. For more information, see*Quick analysis*.

Saved Search

On the query page, click **Saved Search** in the upper right corner to save your current query action as a quick query. Next time you can initiate the query action on the **Saved Search** tab on the left without entering the query statement manually.

The quick query condition can be used by alarm policies. If the quick query is added to the **Tab**, it can be accessed directly on the tab.

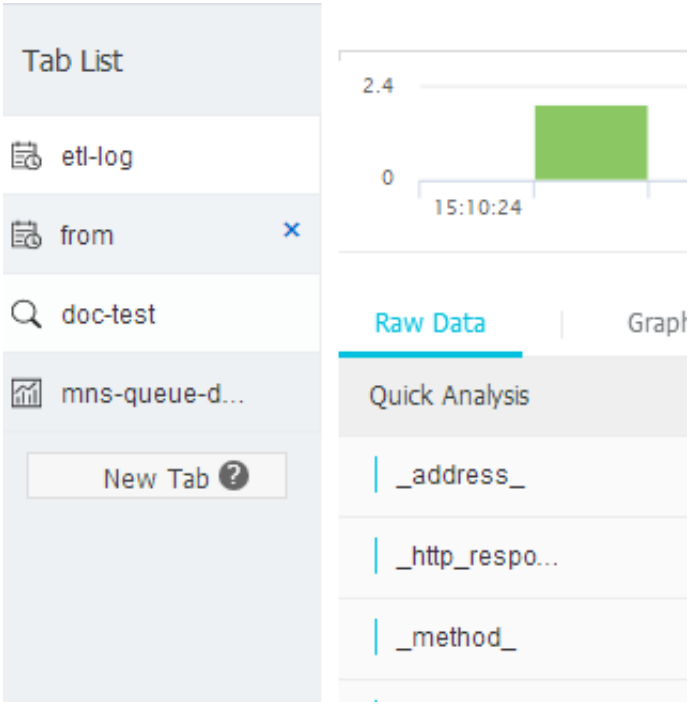
Tag

Log Service provides a tag list on the left side of the query page. You can add the following three data pages to the tag list:

- Logstore
- Saved Search

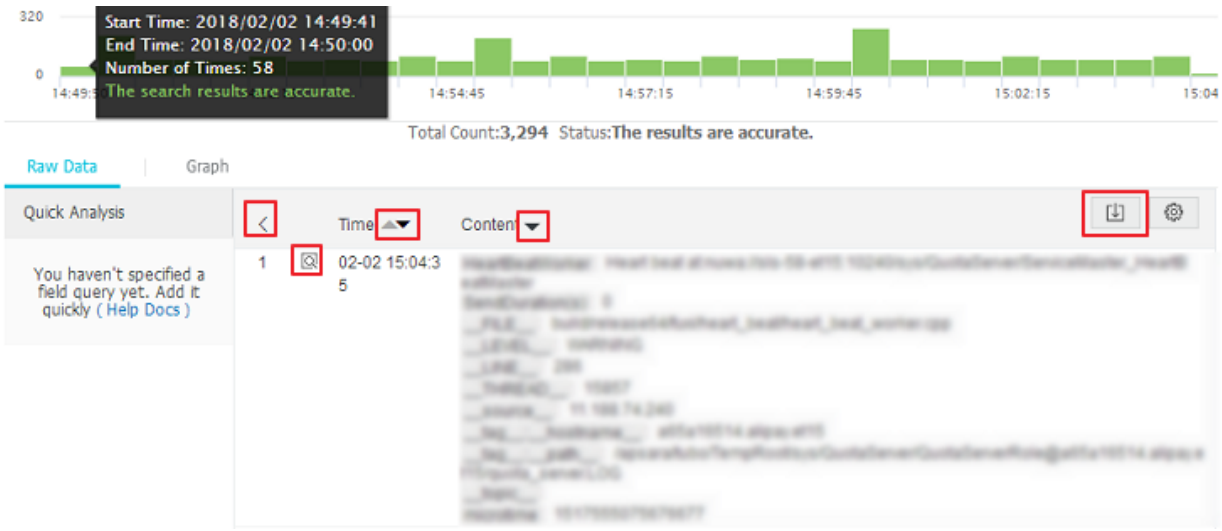
- Dashboard

You can access Logstores, saved quick queries, and dashboards in the tag list with ease. Click **Add Tab** in the label list. In the menu that appears on the right, select the Logstore, quick query, or dashboard to be added as a tag. To delete a tagl, click the X symbol next to it in the tag list.



Dashboard

Log Service provides the dashboard feature to visually display search analysis statements. For more information, see *Dashboard*.



Save as alarm

Log Service allows you to configure alerting based on your **LogSearch Results**. You can configure the alarm rules so that specific alarm content can be sent to you by using in-site notifications or DingTalk messages.

Configuration process:

1. Configure Savedsearch.
2. Configure the alarm rules.
3. Configure notification type.
4. View alarm records.

For more information, see *Configure alarming*.

22.7 Real-time analysis

Log Service supports aggregate functions. This service combines the query function with SQL compute to calculate the query result.

Syntax example:

```
status>200 |select avg(latency),max(latency) ,count(1) as c GROUP BY
method ORDER BY c DESC LIMIT 20
```

Basic syntax:

```
[search query] | [sql query]
```

The search and compute conditions, separated by `|`, filter required logs by using search queries and perform SQL query calculations for these logs. The search query syntax is specific to Log Service. For more information, see*Query syntax*.

Prerequisites

Click **Enable Analysis** under SQL fields in **Search & Analysis** before using the analysis function. See *Overview* for more information.

- If analysis is not enabled, computing capability of a maximum of 10,000 data entries is provided for each shard by default, with relatively higher latency.
- Quick analysis within seconds is available when the analysis function is enabled.
- After being enabled, the function takes effect only on new data.
- Enabling analysis will not generate additional fees.

Supported SQL syntax

Log Service supports the subsequent SQL syntax. For details, click the corresponding link.

- Aggregate functions in the SELECT statement:
  - *General aggregate functions*
  - *Map functions*
  - *Estimating functions*
  - *Mathematical statistical functions*
  - *Mathematical functions*
  - *String functions*
  - *Date and time functions*
  - *URL functions*
  - *Regular expression functions*
  - *JSON functions*
  - *Type conversion functions*
  - *Arrays*
  - *Binary string functions*
  - *Bit operation*
  - *Comparison functions and operators*
  - *Lambda function*
  - *Logical function*
  - *Geospatial functions*
- *GROUP BY syntax*
- *Window functions*
- *HAVING syntax*
- *ORDER BY syntax*
- *LIMIT syntax*
- *CASE WHEN syntax*
- *Column alias*
- *Nested subquery*

Syntax structure

The SQL syntax structure is as follows:

- You do not need to specify the FROM clause and WHERE clause for an SQL statement. The default FROM clause specifies the current Logstore where data is queried and the default WHERE clause defines the condition as search query.
- Supported clauses include SELECT, GROUP BY, ORDER BY [ASC, DESC], LIMIT, and HAVING.
- By default, only the first 10 results are returned. To return more results, append limit n, for example, `* | select count(1) as c, ip group by ip order by c desc limit 100`.

Predefined macros

Log Service provides predefined macros for statistics analysis. When a user configures an effective column, the predefined macros are added automatically.

Macro Name	Type	Meaning
`__time__`	bigint	Log time.
`__source__`	varchar	IP address of the log source. Note that the field is source in a search. The field starts and ends with an underline only in SQL.
`__topic__`	varchar	Log topic.

Restrictions

1. The maximum number of concurrencies is five for each project.
2. The maximum length of a single varchar column is 512. The exceeded content will be dropped.

Example

Calculates PV, UV, and user requests with the maximum latency every hour, and the top ten rows with the highest latency:

```
*|select date_trunc(hour,from_unixtime(__time__)) as time, count(1) as
 pv, approx_distinct(userid) as uv, max_by(url,latency) as top_latenc
y_url, max(latency,10) as top_10_latency group by 1 order by time
```

22.7.1 Analysis syntax and functions

22.7.1.1 General aggregate functions

The search analysis feature of Log Service supports log analysis using general aggregate functions. The detailed statement and meaning are as follows:

Statement	Meaning	Example
`arbitrary(x)`	Returns a value in column x randomly.	`latency > 100 | select arbitrary(method)`
`avg(x)`	Calculates the arithmetic mean of all the values in column x.	`latency > 100 | select avg(latency)`
`checksum(x)`	Calculates the checksum of all the values in a column and returns the base64-encoded value.	`latency > 100 | select checksum(method)`
`count(*)`	Calculates the number of rows in a column.	-
`count(x)`	Calculates the number of non-null values in a column.	`latency > 100 | count(method)`
`count_if(x)`	Calculates the number of x = true.	`latency > 100 | count(url like '%abc')`
`geometric_mean(x)`	Calculates the geometric mean of all the values in a column.	`latency > 100 | select geometric_mean(latency)`
`max_by(x,y)`	Returns the value of column x when column y has the maximum value.	The method for the maximum latency: `latency>100 | select max_by(method, latency)`
`max_by(x,y,n)`	Returns the values of column x corresponding to n rows with maximum values in column y.	The method for the top 3 rows with maximum latency: `latency > 100 | select`

Statement	Meaning	Example
		`max_by(method,latency,3)`
`min_by(x,y)`	Returns the value of column x when column y has the minimum value.	The method for the minimum latency:`* | select min_by(x,y)`
`min_by(x,y,n)`	Returns the values of column x corresponding to n rows with minimum values in column y.	Search the methods(x) for the minimum 3(n) latency(y) values: `* | select max_by(method,latency,3)`
`max(x)`	Returns the maximum value.	`latency > 100| select max(inflow)`
`min(x)`	Returns the minimum value.	`latency > 100| select min(inflow)`
`sum(x)`	Returns the sum of all the values in column x.	`latency > 10 | select sum(inflow)`
`bitwise_and_agg(x)`	Do the AND calculation to all the values in a column.	-
`bitwise_or_agg(x)`	Do the OR calculation to all the values in a column.	-

22.7.1.2 Map functions

The search analysis feature of Log Service supports log analysis using map functions. The detailed statement and meaning are as follows:

Statements	Meaning	Example
Subscript operator []	Obtains the results of a key in a map.	-
histogram(x)	Calculates the count grouped by the value of column x. Performs GROUP BY according to each value of column x and calculates the count. The syntax is equivalent to `select count group by x`.	`latency > 10 | histogram(status)` is equivalent to `latency > 10 | select count(1) group by status`.

map_agg(Key,Value)	Returns a map of key, value, and shows the random latency of each method.	`latency > 100 | select map_agg(method,latency)`
multimap_agg(Key,Value)	Returns a multi-value map of key, value, and returns all the latency for each method.	`latency > 100 | select multimap_agg(method, latency)`
cardinality(x) → bigint	Obtains the size of the map.	-
element_at(map<K, V>, key) → V	Obtains the value corresponding to the key.	-
map() → map<unknown, unknown>	Returns an empty map.	-
map(array<K>, array<V>) → map<K,V>	Converts two arrays into 1-to-1 maps.	`SELECT map(ARRAY[1,3], ARRAY[2,4]); – {1 -> 2, 3 -> 4}`
map_from_entries(array<row<K, V>>) → map<K,V>	Converts a multidimensional array into a map.	`SELECT map_from_entries (ARRAY[(1, 'x'), (2, 'y')]); – {1 -> 'x', 2 -> 'y'}`
map_entries(map<K, V>) → array<row<K,V>>	Converts an element in a map into an array.	`SELECT map_entries(MAP (ARRAY[1, 2], ARRAY['x', 'y'])); – [ROW(1, 'x'), ROW(2, 'y')]`
map_concat(map1<K, V>, map2<K, V>, …, mapN<K, V>) → map<K,V>	The Union of multiple maps is required, if a key exists in multiple maps, take the first one.	-
map_filter(map<K, V>, function) → map<K,V>	Refer to the lambda map_filter function.	-
transform_keys(map<K1, V>, function) → MAP<K2,V>	Refer to the lambda transform_keys function.	-
transform_values(map<K, V1>, function) → MAP<K,V2>	Refer to the lambda transform_values function.	-
map_keys(x<K, V>) → array<K>	Obtains all the keys in the map and returns an array.	-
map_values(x<K, V>) → array<V>	Obtains all values in the map and returns an array.	-

map_zip_with(map<K, V1>, map<K, V2>, function<K, V1, V2, V3>) → map<K,V3>	Refer to power functions in Lambda.	-

22.7.1.3 Estimating functions

The query and analysis function of Log Service supports analyzing logs by using estimating functions. The specific statements and meanings are as follows.

Statements	Meaning	Example
`approx_distinct(x)`	Estimates the number of unique values in column x.	-
`approx_percentile(x, percentage)`	Sorts the column x and returns the value approximately at the given percentage position.	Returns the value at the half position: `approx_percentile(x,0.5)`
`approx_percentile(x, percentages)`	It is similar to the preceding statement, but you can specify multiple percentages to return the values at each specified percentage position.	`approx_percentile(x, array(0.1,0.2))`
`numeric_histogram( buckets, Value)`	Makes statistics on the value column in different buckets. Divides the value column into buckets number of buckets and returns the key and count of each bucket, which is equivalent to `select count group by`.	For post requests, divides the delay into 10 barrels, and returns the size of each bucket: method: `method:POST | select numeric_histogram(10, latency)`

22.7.1.4 Mathematical statistical functions

The query and analysis function of Log Service supports analyzing logs by using mathematical statistical functions. The specific statements and meanings are as follows.

Statements	Meaning	Example
`corr(y, x)`	Returns the correlation coefficient of two columns. The result is from 0 to 1.	`latency>100| select corr(latency,request_size)`
`covar_pop(y, x)`	Calculates the population covariance.	`latency>100| select covar_pop(request_size, latency)`

Statements	Meaning	Example
`covar_samp(y, x)`	Calculates the sample covariance.	`latency>100| select covar_samp(request_size,latency)`
`regr_intercept(y, x)`	Returns the linear regression intercept of input values. y is the dependent value, and x is the independent value.	`latency>100| select regr_intercept(request_size,latency)`
`regr_slope(y,x)`	Returns the linear regression slope of input values. y is the dependent value, and x is the independent value.	`latency>100| select regr_slope(request_size,latency)`
`stddev(x)` or `stddev_samp(x)`	Returns the sample standard deviation of column x.	`latency>100| select stddev(latency)`
`stddev_pop(x)`	Returns the population standard deviation of column x.	`latency>100| select stddev_pop(latency)`
`variance(x)` or `var_samp(x)`	Calculates the sample variance of column x.	`latency>100| select variance(latency)`
`var_pop(x)`	Calculates the population variance of column x.	`latency>100| select variance(latency)`

22.7.1.5 Mathematical functions

The query and analysis function of Log Service supports analyzing logs by using mathematical functions. By combining query statements with mathematical functions, you can perform mathematical calculation to the log query results.

Mathematical operators

Mathematical operators such as the plus sign (+), minus sign (-), asterisk (*), slash (/), and percent sign (%) are supported. They can be used in the SELECT clause.

Example:

```
*|select avg(latency)/100 , sum(latency)/count(1)
```

Description of mathematical functions

Log Service supports the following mathematical functions:

Function Name	Meaning
`abs(x)`	Returns the absolute value of column x.
`cbrt(x)`	Returns the cube root of column x.
`ceiling(x)`	Returns the number rounded up to the nearest integer of column x.
`cosine_similarity(x,y)`	Returns the cosine similarity between the sparse vectors x and y.
`degrees`	Converts radians to degrees.
`e()`	Returns the constant Euler's number.
`exp(x)`	Returns Euler's number raised to the power of x.
`floor(x)`	Returns x rounded down to the nearest integer.
`from_base(string,radix)`	Returns the value of string interpreted as a base-radix number.
`ln(x)`	Returns the natural logarithm of x.
`log2(x)`	Returns the base-2 logarithm of x.
`log10(x)`	Returns the base-10 logarithm of x
`log(x,b)`	Returns the base-b logarithm of x.
`pi()`	Returns π.
`pow(x,b)`	Returns x raised to the power of b.
`radians(x)`	Converts angle x in degrees to radians.
`rand()`	Returns a pseudo-random value in the range 0 .0 <= x < 1.0.
`random(0,n)`	Returns a pseudo-random number between 0 and n (exclusive).
`round(x)`	Returns x rounded to the nearest integer.
`round(x, y)`	Returns x rounded to y decimal places. For example, round(1.012345,2) = 1.01.
`sqrt(x)`	Returns the square root of x.
`to_base(x, radix)`	Returns the base-radix representation of x.
`truncate(x)`	Returns x rounded to integer by dropping digits after decimal point.
`acos(x)`	Returns the arc cosine of x.

Function Name	Meaning
`asin(x)`	Returns the arc sine of x.
`atan(x)`	Returns the arc tangent of x.
`atan2(y,x)`	Returns the arc tangent of y/x.
`cos(x)`	Returns the cosine of x.
`sin(x)`	Returns the sine of x.
`cosh(x)`	Returns the hyperbolic cosine of x.
`tan(x)`	Returns the tangent of x.
`tanh(x)`	Returns the hyperbolic tangent of x.
`infinity()`	Returns the constant representing positive infinity.
`is_infinity(x)`	Determine if x is finite.
`is_finity(x)`	Determine if x is infinite.
`is_nan(x)`	Determine if x is not-a-number.

22.7.1.6 String functions

The search analysis feature of Log Service supports log analysis by using string functions. The detailed statement and meaning are as follows:

Function Name	Meaning
`length(x)`	Field length.
`levenshtein_distance(string1, string2)`	Returns the minimum Levenshtein distance between two strings.
`lower(string)`	Converts a string to lowercase characters.
`ltrim(string)`	Removes the leading whitespace.
`replace(string, search)`	Removes search from the string.
`replace(string, search,rep)`	Replaces search with rep in string.
`reverse(string)`	Reverse a string.
`rtrim(string)`	Removes trailing whitespace from the string.
`split(string,delimeter,limit)`	Splits the given string into substrings given a delimiter. Generates results in an array with subscripts starting with 1.

Function Name	Meaning
`split_part(string,delimeter,offset)`	Splits the string into substrings and returns the substrings in an array. The offset-th string will be returned. Generates results in an array with subscripts starting with 1.
`strpos(string, substring)`	Returns the starting position of the substring within the string. The position starts with 1. If not found, 0 is returned.
`substr(string, start)`	Returns substrings of the string with subscripts starting with 1.
`substr(string, start, length)`	Returns substrings of the string with subscripts starting with 1.
`trim(string)`	Removes leading and trailing whitespace from the string.
`upper(string)`	Converts the string to uppercase.
`concat(string,string.....)`	Joins two or more strings into one.
`hamming_distance (string1,string2)`	Returns the Hamming distance between two strings.

Note:

The strings are enclosed by single quotation marks. A column name is enclosed by double quotation marks. For example: In a = abc, a = string abc; in a = abc, column a = column abc.

22.7.1.7 Date and time functions

Log Service supports time functions and date functions. You can use the date and time functions introduced in this document in the analysis syntax.

Date and time

1. unixtime: The number of seconds since January 1, 1970 in the type of int. For example, `1512374067` indicates the time `Mon Dec 4 15:54:27 CST 2017`. In Log Service, The built-in time `__time__` in each log of Log Service is of this type.
2. timestamp type: Indicates the time in the format of string. For example, `2017-11-01 13:30:00`.

Date Functions

Log Service supports the following common date functions:

Function Name	Meaning	Example
`current_date`	Returns the current date.	`latency>100| select current_date`
`current_time`	hour:minute; second, millisecond time zone	`latency>100| select current_time`
`current_timestamp`	Returns the result combined by current_date and current_time .	`latency>100| select current_timestamp`
`current_timezone()`	Returns the time zone.	`latency>100| select current_timezone()`
`from_iso8601_timestamp(string)`	Converts an ISO8601 to a timestamp with time zone.	`latency>100| select from_iso8601_timestamp(iso8601)`
`from_iso8601_date(string)`	Converts an ISO8601 to a date .	`latency>100| select from_iso8601_date(iso8601)`
`from_unixtime(unixtime)`	Converts a Unix time to a timestamp.	`latency>100| select from_unixtime(1494985275)`
`from_unixtime(unixtime, string)`	Converts a Unix time to a timestamp using the string as the time zone.	`latency>100| select from_unixtime (1494985275,Asia/Shanghai)`
`localtime`	Returns the current time.	`latency>100| select localtime`
`localtimestamp`	Returns the current timestamp.	`latency>100| select localtimestamp`
`now()`	Equivalent to `current_timestamp`.	-
`to_unixtime(timestamp)`	Converts a timestamp to a Unix time.	`*| select to_unixtime(2017-05-17 09:45:00.848 Asia/Shanghai)`

Time Functions

MySQL time formats

Log Service supports the following MySQL time formats: %a, %b, and %y.

Function Name	Meaning	Example
`date_format(timestamp, format)`	Formats timestamp into a string using format.	`latency>100| select date_format (date_parse (2017-05-17 09:45:00,%Y-%m-%d %H:%i:%S), %Y-%m-%d) group by method`
`date_parse(string, format)`	Parses the string into a timestamp using format.	`latency>100|select date_parse(2017-05-17 09:45:00,%Y-%m-%d %H:%i:%S) group by method`

Table 22-7: Format Description

Format	Description
%a	Days of the week in abbreviated form (Sun .. Sat).
%b	Months in abbreviated form (Jan .. Dec).
%c	Month, numerical type (1 .. 12) [4].
%D	The day of the month with the suffix (0th, 1st, 2nd, 3rd, …).
%d	The day of the month (01 .. 31) [4].
%e	The day of the month (1 .. 31) [4].
%H	The hour (00 .. 23).
%h	The hour (01 .. 12).
%I	The hour in 12-hour format (01 .. 12).
%i	The minute (00 .. 59).
%j	The day of the year (001 .. 366).
%k	The hour (0 .. 23).
%l	The hour (1 .. 12).
%M	The month in English (January .. Dece mber).
%m	The month in number (01 .. 12) [4].
%p	AM or PM.
%r	The time in 12-hour format. The format is hh: mm:ss followed by AM or PM.

Format	Description
%S	The second (00 .. 59).
%s	The second (00 .. 59).
%T	The time in 24-hour format (hh:mm:ss).
%U	The week of the year (00 .. 53).
%u	The week of the year (00 .. 53).
%V	The week of the year (01 .. 53).
%v	The week of the year (01 .. 53), where Monday is the first day of the week; used with %x.
%W	The name of a day in a week (Sunday .. Saturday).
%w	The day of the week (0 .. 6). Sunday is the day 0.
%Y	The year.
%y	The year. Double digit.
%%	%escape character

Time period alignment functions

Log Service supports time period alignment functions, which can be aligned according to seconds , minutes, hours, days, months, and years. Time period alignment functions are usually used when statistics are made according to time.

Function syntax:

```
date_trunc(unit, x)
```

Parameters:

The optional values for Unit are as follows (x is `2001-08-22 03:04:05.000`):

Unit	Converted result
second	2001-08-22 03:04:05.000
minute	2001-08-22 03:04:00.000
hour	2001-08-22 03:00:00.000
day	2001-08-22 00:00:00.000
week	2001-08-20 00:00:00.000

Unit	Converted result
month	2001-08-01 00:00:00.000
quarter	2001-07-01 00:00:00.000
year	2001-01-01 00:00:00.000

x can be a timestamp type or Unix time.

date_trunc is only applicable to statistics at a fixed time interval. For statistics based on flexible time dimensions, for example, every 5 minutes, perform GROUP BY according to the mathematic al modulus method.

```
* | SELECT count(1) as pv, __time__ - __time__% 300 as minute5groupby
minute5 limit 100
```

In the preceding formula, `%300` indicates to make the modulus and alignment every five minutes.

Date function example

The following is a comprehensive example using time formats:

```
*|select date_trunc(minute , __time__) as t, truncate (avg(latency
) ) , current_date group by t order by t desc limit 60
```

22.7.1.8 URL functions

URL functions support extracting fields from standard URL paths. A standard URL is as follows:

```
[protocol:][//host[:port]][path][?query][#fragment]
```

Common URL functions

Function Name	Meaning	Example
`url_extract_fragment(url)`	Extracts the fragment from a URL and the result is of varchar type.	`*| select url_extract_fragment(url)`
`url_extract_host(url)`	Extracts the host from a URL and the result is of varchar type.	`*|select url_extract_host(url)`
`url_extract_parameter(url, name)`	Extracts the value of the name parameter in the query from a URL and the result is of varchar type.	`*|select url_extract_parameter(url)`

Function Name	Meaning	Example
`url_extract_path(url)`	Extracts the path from a URL and the result is of varchar type.	`*|select url_extract_path(url)`
`url_extract_port(url)`	Extracts the port from a URL and the result is of bigint type.	`*|select url_extract_port(url)`
`url_extract_protocol(url)`	Extracts the protocol from a URL and the result is of varchar type.	`*|select url_extract_protocol(url)`
`url_extract_query(url)`	Extracts the query from a URL and the result is of varchar type.	`*|select url_extract_query(url)`
`url_encode(value)`	Encodes a URL.	`*|select url_encode(url)`
`url_decode(value)`	Decodes a URL.	`*|select url_decode(url)`

22.7.1.9 Regular expression functions

A regular expression function parses a string and returns the needed substrings.

The common regular expression functions and the meanings are as follows:

Function name	Meaning	Example
`regexp_extract_all(string, pattern)`	Returns all the substrings that match the regular expression in the string as a string array.	The result of `|*SELECT regexp_extract_all(5a 67b 890m, \d+)` is `[5,67,890]`, and the result of `*|SELECT regexp_extract_all(5a 67a 890m, (\d+)a)` is `[5a,67a]`.
`regexp_extract_all(string, pattern, group)`	Returns the part of the string that hits the regular () part of the group, returns the result as an array of strings.	The result of `*| ` SELECT regexp_extract_all('5a 67a 890m', '(\d+)a',1)` is `['5','67']`.
`regexp_extract(string, pattern)`	Returns the first substring that hits the regular expression in the string.	The result of `*|SELECT regexp_extract(5a 67b 890m, \d+)` is 5.
`regexp_extract(string, pattern,group)`	Returns the first substring in the (group)th () that hits the	The result of `*|SELECT regexp_extract(5a 67b`

Function name	Meaning	Example
	regular expression in the string .	`890m, (\d+)([a-z]+),2)` is b.
`regexp_like(string, pattern)`	Determines if the string matches the regular expression and returns a bool result. The regular expression is allowed to match part of the string.	The result of `*|SELECT regexp_like(5a 67b 890m, \d+m)` is true.
`regexp_replace(string, pattern, replacement)`	Replaces the part that matches the regular expression in the string with replacement.	The result of `*|SELECT regexp_replace(5a 67b 890m, \d+,a)` is `aa ab am`.
`regexp_replace(string, pattern)`	Removes the part that matches the regular expression in the string, which is equivalent to `regexp_replace(string,patterm,)`.	The result of `*|SELECT regexp_replace(5a 67b 890m, \d+)` is `a b m`.
`regexp_split(string, pattern)`	Splits the string to an array by using the regular expression.	The result of `*|SELECT regexp_split(5a 67b 890m, \d+)` is `[a,b,m]`.

22.7.1.10 JSON functions

JSON functions can parse a string as the JSON type and extract the fields in JSON. JSON mainly has the following two structures: map and array. If a string fails to be parsed as the JSON type, the returned value is null.

Log Service supports the following common JSON functions:

Function Name	Meaning	Example
`json_parse(string)`	Converts the string into the JSON type.	`SELECT json_parse([1, 2, 3])`The result is an array of the JSON type.
`json_format(json)`	Converts the JSON type into a string.	`SELECT json_format(json_parse([1, 2, 3]))`The result is a string.
`json_array_contains(json, value)`	Judges whether a value of the JSON type or a string (with	`SELECT json_array_contains(json_parse([1, 2, 3]), 2) or SELECT`

Function Name	Meaning	Example
	the content of a JSON array) contains a specific value.	`json_array_contains([1, 2, 3], 2)`
`json_array_get(json_array, index)`	Like `json_array_contains`, but obtains the element of a subscript of a JSON array.	`SELECT json_array_get(["a", "b", "c"], 0) returns a`
`json_array_length(json)`	returns the size of JSON array.	`SELECT json_array_length([1, 2, 3]) returns result 3.`
`json_extract(json, json_path)`	indicates to extract values from a JSON object. The JSON path syntax is similar to `$.store.book[0].title`. A JSON object is returned.	`SELECT json_extract(json, $.store.book);`
`json_extract_scalar(json, json_path)`	is similar to `json_extract`, but returns a string.	-
`json_size(json, json_path)`	Returns the size of JSON object or array.	`SELECT json_size([1, 2, 3]) returns result 3.`

22.7.1.11 Type conversion functions

Log Service supports data types such as Long, Double, and Textin the configurations. Supported types for query include Bigint, Double, Varchar, Timestamp, and Int.

The type conversion function forcibly converts a column to the specified data type:

```
try_cast(value AS type) → type
```

22.7.1.12 GROUP BY syntax

GROUP BY supports multiple columns and indicating the corresponding KEY by using the SELECT column alias.

Example:

```
method:PostLogstoreLogs |select avg(latency),projectName,date_trunc(hour,__time__) as hour group by projectName,hour
```

The alias hour indicates the third SELECT column `date_trunc(hour,__time__)`. This is very helpful for complex queries.

GROUP BY supports GROUPING SETS, CUBE, and ROLLUP.

Example:

```
method:PostLogstoreLogs |select avg(latency) group by cube(projectNam
e,logstore) method:PostLogstoreLogs |select avg(latency) group by
GROUPING SETS ( ( projectName,logstore), (projectName,method)) method
:PostLogstoreLogs |select avg(latency) group by rollup(projectName,
logstore)
```

Example

Perform GROUP BY according to time

Each log has a built-in time column `__time__`. When the statistical function of any column is activated, the statistics will be automatically made for the time column.

Use the `date_trunc` function to align the time column to hour, minute, day, month, and year. `date_trunc` accepts an aligned unit and a Unix time or timestamp type column, such as `__time__`.

- PV statistics per hour and per minute

```
* | SELECT count(1) as pv , date_trunc(hour,__time__) as hour
group by hour order by hour limit 100 * | SELECT count(1) as pv
 , date_trunc(minute,__time__) as minute group by minute order by
minute limit 100
```

Note:

limit 100 indicates that up to 100 rows can be retrieved. If the LIMIT statement is not added, up to 10 rows of data can be retrieved by default.

- date_trunc is only applicable to statistics at a fixed time interval. For statistics based on flexible time dimensions, for example, every 5 minutes, perform GROUP BY in mod.

```
* | SELECT count(1) as pv, __time__ - __time__% 300 as minute5 group
 by minute5 limit 100
```

In the preceding formula, `%300` indicates making the modulus and alignment every five minutes.

Retrieve non-agg columns in GROUP BY

In standard SQL, if the GROUP BY syntax is used during the SELECT operation, the system only selects the original content of the SELECT GROUP BY column, or performs aggregate computing on any columns. Retrieving content from non-GROUP BY columns is not allowed.

For example, the following syntax is invalid. Because b is a non-GROUP BY column, the system cannot determine which row of b to output during GROUP BY based on a.

```
*|select a, b , count(c) gropu by a
```

Instead, you can use the `arbitrary` function to output b:

```
*|select a, arbitrary(b), count(c) gropu by a
```

22.7.1.13 Window functions

Window functions are used for cross-row calculation. SQL aggregate functions calculate the results of only one row or aggregate all rows into one row for calculation. Window functions support cross-row calculation and fill the calculation results in each row.

Syntax of window functions:

```
SELECT key1, key2, value, rank() OVER (PARTITION BY key2 ORDER BY
value DESC) AS rnk FROM orders ORDER BY key1,rnk
```

Core part:

```
rank() OVER (PARTITION BY KEY1 ORDER BY KEY2 DESC)
```

rank() is an aggregate function.You can use any function in analysis syntax or the function listed in this document. PARTITION BY indicates the buckets based on which values are calculated.

Special aggregate functions used in windows

Function Name	Meaning
rank()	Sorts data based on a specific column in a window and returns the serial numbers in the window.
row_number()	Returns the row numbers in the window.
first_value(x)	Returns the first value in the window. It is typically used to obtain the maximum value after values are sorted in the window.
last_value(x)	Opposite to first_value.
nth_value(x, offset)	Number of a specific offset in the window.
lead(x,offset,defaut_value)	Value of the No. offset row after a certain row in xth column in the window. If that row does not exist, use the default_value.

Function Name	Meaning
lag(x,offset,defaut_value)	Value of the No. offset row before a certain row in xth column in the window. If that row does not exist, use the default_value.

Example

- Rank the salaries of employees in their respective departments

```
* | select department, persionId, sallary , rank() over(PARTITION
 BY department order by sallary desc) as sallary_rank order by
department,sallary_rank
```

Response results:

department	persionId	sallary	sallary_rank
dev	john	9000	1
dev	Smith	8000	2
dev	Snow	7000	3
dev	Achilles	6000	4
Marketing	Blan Stark	9000	1
Marketing	Rob Stark	8000	2
Marketing	Sansa Stark	7000	3

- Calculate the salaries of employees as percentages in their respective departments

```
* | select department, persionId, sallary *1.0 / sum(sallary) over(
PARTITION BY department ) as sallary_percentage
```

Response results:

department	persionId	sallary	sallary_percentage
dev	john	9000	0.3
dev	Smith	8000	0.26
dev	Snow	7000	0.23
dev	Achilles	6000	0.2
Marketing	Blan Stark	9000	0.375
Marketing	Rob Stark	8000	0.333
Marketing	Sansa Stark	7000	0.29

- Calculate the daily UV increase over the previous day

```
* | select day ,uv, uv *1.0 /(lag(uv,1,0) over() ) as diff_perce
ntage from ( select approx_distinct(ip) as uv, date_trunc(day,
__time__) as day from log group by day order by day asc )
```

Response results:

day	uv	diff_percentage
2017-12-01 00:00:00	100	null
2017-12-02 00:00:00	125	1.25
2017-12-03 00:00:00	150	1.2
2017-12-04 00:00:00	175	1.16
2017-12-05 00:00:00	200	1.14
2017-12-06 00:00:00	225	1.125
2017-12-07 00:00:00	250	1.11

22.7.1.14 HAVING syntax

The search analysis feature of Log Service supports the HAVING syntax of standard SQL, which is used with GROUP BY to filter GROUP BY results.

Format:

```
method :PostLogstoreLogs |select avg(latency),projectName group by
projectName HAVING avg(latency) > 100
```

Differences between HAVING and WHERE

HAVING filters the results of aggregate computing after GROUP BY, and WHERE filters raw data between aggregate computing operations.

Example

Calculate the average rainfall of each province where temperature is above 10°C, and display only the provinces with average rainfall above 100 mL in the final results:

```
* | select avg(rain) ,province where teporature > 10groupby province
having avg(rain) > 100
```

22.7.1.15 ORDER BY syntax

ORDER BY is used to sort results. Currently, you can only sort results by one column.

Syntax format:

```
orderby Column name [desc|asc]
```

Example:

```
method :PostLogstoreLogs |select avg(latency) as avg_latency,
projectName group by projectName HAVING avg(latency) > 5700000 order
by avg_latency desc
```

22.7.1.16 LIMIT syntax

LIMIT is followed by a number to restrict the maximum number of rows in the results. If no LIMIT statement is added, only 10 rows are returned by default.

Note:
LIMIT OFFSET and LINES syntaxes are not supported.

Example:

```
*| select avg(latency) as avg_latency , methodgroupbymethodorderbyavg_
latencydesclimit100
```

22.7.1.17 CASE WHEN syntax

The CASE WHEN syntax is supported for classification of continuous data. For example, you can extract information from http_user_agent and classify the information into Android and iOS types:

```
SELECT CASE WHEN http_user_agent like %android% then android WHEN
http_user_agent like %ios% then ios ELSE unknown END as http_user_
agent, count(1) as pv group by http_user_agent
```

Example

- Proportion of requests with status code 200 in all requests:

```
* | SELECT sum( CASE WHEN status =200 then 1 ELSE 0 end ) *1.0 /
count(1) as status_200_percentage
```

- Distribution of latencies:

```
* | SELECT ` CASE WHEN latency < 10 then s10 WHEN latency < 100
then s100 WHEN latency < 1000 then s1000 WHEN latency < 10000 then
```

```
 s10000 else s_large end as latency_slot, count(1) as pv group by
latency_slot
```

IF syntax

The IF syntax is logically equivalent to the CASE WHEN syntax.

```
CASE WHEN condition THEN true_value [ ELSE false_value ] END
```

- if(condition, true_value)
  - If the condition is true, the true_value column is returned; otherwise, null is returned.
- if(condition, true_value, false_value)
  - If the condition is true, the true_value column is returned; otherwise, the false_value column is returned.

COALESCE syntax

The coalesce function returns the first non-null value of multiple columns.

```
coalesce(value1, value2[,...])
```

NULLIF syntax

If value1 equals value2, null is returned; otherwise, value1 is returned.

```
nullif(value1, value2)
```

TRY syntax

The TRY syntax captures some underlying exceptions, for example, returning null for an incorrect division by zero.

```
try(expression)
```

22.7.1.18 Nested subquery

You can use a nested SQL query in complicated query scenarios where a single SQL layer does not meet the requirement.

The difference between nested subquery and non-nested query is that you need to specify the from condition in the SQL statement. You need to specify the `from log` keyword in the query to read raw data from logs.

Example:

```
* | select sum(pv) from ( select count(1) as pv from log group by
method )
```

22.7.1.19 Arrays

Statement	Meaning	Example
Subscript operator []	Obtains a certain element from the array.	-

Connection operator ||	Connects two arrays into one.	`SELECT ARRAY [1] || ARRAY [2]; — [1, 2]`
		`SELECT ARRAY [1] || 2; — [1, 2]`
		`SELECT 2 || ARRAY [1]; — [2, 1]`

array_distinct	Obtain the distinct elements in the array by means of array deduplication.	-
array_intersect(x, y)	Obtains the intersection of array x and array y.	-
array_union(x, y) → array	Obtains the union of array x and array y.	-
array_except(x, y) → array	Returns an array of elements in x but not in y , without duplicates.	-
array_join(x, delimiter, null_replacement) → varchar	Concatenates the elements of the given array using the delimiter and an optional string to replace nulls.	-
array_max(x) → x	Obtains the maximum value in array x.	
array_min(x) → x	Obtains the minimum value in array x.	-
array_position(x, element) → bigint	Returns the position of the first occurrence of the element in array x (or 0 if not found).	-
array_remove(x, element) → array	Removes all elements that equal element from array x.	-

Statement	Meaning	Example
array_sort(x) → array	Sorts and returns the array x . The elements of x must be orderable. Null elements will be placed at the end of the returned array.	-
cardinality(x) → bigint	Returns the cardinality (size) of the array x.	-
concat(array1, array2, …, arrayN) → array	Concatenates arrays.	-
contains(x, element) → boolean	Returns TRUE if the array x contains the element.	-
filter(array, function) → array	This is a Lambda function. See filter() in *Lambda function*.	-
flatten(x) → array	Flattens an array(array(T)) to an array(T) by concatenating the contained arrays.	-
reduce(array, initialState, inputFunction, outputFunction) → x	See *Lambda function*reduce.	-
reverse(x) → array	Returns an array which has the reversed order of array x.	-
sequence(start, stop) → array	Generate a sequence of integers from start to stop, incrementing by 1 if start is less than or equal to stop, otherwise -1.	-
sequence(start, stop, step) → array	Generate a sequence of dates from start to stop, incrementing by step.	-
sequence(start, stop, step) → array	Generate a sequence of timestamps from start to stop , incrementing by step. The type of step can be either INTERVAL DAY TO SECOND or INTERVAL YEAR TO MONTH.	-

Statement	Meaning	Example
shuffle(x) → array	Generate a random permutation of the given array x.	-
slice(x, start, length) → array	Subsets array x starting from index start (or starting from the end if start is negative) with a length of length.	-
transform(array, function) → array	See *Lambda function*transform().	-
zip(array1, array2[, …]) → array	Merges the given arrays, element-wise, into a single array of rows. The M-th element of the N-th argument will be the N-th field of the M-th output element. If the arguments have an uneven length, missing values are filled with NULL.	`SELECT zip(ARRAY[1, 2], ARRAY['1b', null, '3b']); — [ROW(1, '1b'), ROW(2, null), ROW(null, '3b')]`
zip_with(array1, array2, function) → array	See *Lambda function*zip_with.	-

22.7.1.20 Binary string functions

The binary string type varbinary is different from the string type varchar.

Statement	Description
Connection function ||	The result of `a || b` is `ab`.
length(binary) → bigint	Returns the length of binary in bytes.
concat(binary1, …, binaryN) → varbinary	Returns the concatenation of binary1, binary2, ..., binaryN. This function provides the same functionality as the SQL-standard concatenation operator (||).
to_base64(binary) → varchar	Encodes the binary string into a base64 string representation.
from_base64(string) → varbinary	Decodes binary data from the base64 encoded string.
to_base64url(binary) → varchar	Encodes binary data into a base64 string representation using the URL safe alphabet.

Statement	Description
from_base64url(string) → varbinary	Decodes binary data from the base64 encoded string using the URL safe alphabet.
to_hex(binary) → varchar	Encodes the binary string into a hex string representation.
from_hex(string) → varbinary	Decodes binary data from the hex encoded string.
to_big_endian_64(bigint) → varbinary	Encodes bigint in a 64-bit 2's complement big endian format.
from_big_endian_64(binary) → bigint	Decodes bigint value from a 64-bit 2's complement big endian binary.
md5(binary) → varbinary	Computes the md5 hash of the binary string.
sha1(binary) → varbinary	Computes the sha1 hash of the binary string.
sha256(binary) → varbinary	Computes the sha256 hash of the binary string.
sha512(binary) → varbinary	Computes the sha512 hash of the binary string.
xxhash64(binary) → varbinary	Computes the xxhash64 hash of the binary string.

22.7.1.21 Bit operation

Statement	Description	Example

bit_count(x, bits) → bigint	Collects the number of 1 in the binary expression of x.	`SELECT bit_count(9, 64); — 2`
		`SELECT bit_count(9, 8); — 2`
		`SELECT bit_count(-7, 64); — 62`
		`SELECT bit_count(-7, 8); — 6`

bitwise_and(x, y) → bigint	Performs the AND operation on x and y in binary.	-
bitwise_not(x) → bigint	Calculates the opposite values of all bits of x in binary.	-
bitwise_or(x, y) → bigint	Performs the OR operation on x and y in binary.	-
bitwise_xor(x, y) → bigint	Performs the XOR operation on x and y in binary.	-

22.7.1.22 Comparison functions and operators

Comparison functions and operators

A comparison operation compares the values of two parameters, which can be used for any comparable types, such as int, bigint, double, and text.

Comparison operators

A comparison operator is used to compare two parameter values. During the comparison, if the logic is true, TRUE is returned. Otherwise, FALSE is returned.

Operators	Meaning
<	Less than
>	Greater than
<=	Smaller than or equal to
>=	Greater than or equal to
=	Equal to
<>	Not equal to
!=	Not equal to

Range operator

The BETWEEN operator in WHERE clause is used to select data within a given range of values.

- If the logic is true, TRUE is returned. Otherwise, FALSE is returned.

  Example: `SELECT 3 BETWEEN 2 AND 6;`The logic is true, and TRUE is returned.

  The previous example is equivalent to `SELECT 3 >= 2 AND 3 <= 6;`.
- You can combine the BETWEEN operator with the NOT operator to find rows whose column values are not in a range of values.

  Example: `SELECT 3 NOT BETWEEN 2 AND 6;`The logic is false, and FALSE is returned.

  The previous example is equivalent to `SELECT 3 < 2 OR 3 > 6;`.
- If the value of any parameter is NULL, NULL is returned.

IS NULL and IS NOT NULL

These operators are used to determine whether a parameter value is NULL.

IS DISTINCT FROM and IS NOT DISTINCT FROM

These operators are like the comparison operators, but they can determine whether a NULL value exists.

Example:

```
SELECT NULL IS DISTINCT FROM NULL; -- false SELECT NULL IS NOT
DISTINCT FROM NULL; -- true
```

As described in the following table, DISTINCT can be used to compare parameter values under multiple conditions.

a	b	a = b	a <> b	a DISTINCT b	a NOT DISTINCT b
1	1	TRUE	FALSE	FALSE	TRUE
1	2	FALSE	TRUE	TRUE	FALSE
1	NULL	NULL	NULL	TRUE	FALSE
NULL	NULL	NULL	NULL	FALSE	TRUE

GREATEST and LEAST

These operators are used to obtain the maximum and minimum values across many columns.

Example:

```
select greatest(1,2,3) ; -- 3 is returned.
```

Comparison conditions: ALL, ANY, and SOME

Comparison conditions are used to determine whether a parameter meets the specified conditions .

- ALL is used to determine whether a parameter meets all the conditions. If the logic is true, TRUE is returned. Otherwise, FALSE is returned.
- ANY is used to determine whether a parameter meets any of the conditions. If the logic is true, TRUE is returned. Otherwise, FALSE is returned.
- Same as ANY, SOME is used to determine whether a parameter meets any of the conditions.
- ALL, ANY, and SOME must follow the comparison operators.

As described in the following table, ALL and ANY support comparison and determination under multiple conditions.

Expression	Meaning
A = ALL (…)	When A is equal to all values, TRUE is returned.
A <> ALL (…)	When A is not equal to all values, TRUE is returned.
A < ALL (…)	When A is smaller than all values, TRUE is returned.
A = ANY (…)	When A is equal to any value, TRUE is returned. It is equivalent to A IN (...).
A <> ANY (…)	When A is not equal to any value, TRUE is returned.
A < ANY (…)	When A is smaller than the greatest value, TRUE is returned.

Example:

```
SELECT hello = ANY (VALUES hello, world); -- true SELECT 21 < ALL (
VALUES 19, 20, 21); -- false SELECT 42 >= SOME (SELECT 41 UNION ALL
SELECT 42 UNION ALL SELECT 43); -- true
```

22.7.1.23 Lambda function

Lambda expressions

Lambda expressions are written with

```
->
```

.

Example:

```
x -> x + 1
(x, y) -> x + y
x -> regexp_like(x, 'a+')
x -> x[1] / x[2]
x -> IF(x > 0, x, -x)
x -> COALESCE(x, 0)
x -> CAST(x AS JSON)
x -> x + TRY(1 / 0)
```

Most MySQL expressions can be used in Lambda.

filter(array<T>, function<T, boolean>) → ARRAY<T>

Filters data from an array and obtains only elements for which the function returns TRUE.

Example:

```
SELECT filter(ARRAY [], x -> true); -- []
SELECT filter(ARRAY [5, -6, NULL, 7], x -> x > 0); -- [5, 7]
SELECT filter(ARRAY [5, NULL, 7, NULL], x -> x IS NOT NULL); -- [5, 7]
```

map_filter(map<K, V>, function<K, V, boolean>) → MAP<K,V>

Filters data from a map and obtains only element pairs for which the function returns TRUE.

Example:

```
SELECT map_filter(MAP(ARRAY[], ARRAY[]), (k, v) -> true); -- {}
SELECT map_filter(MAP(ARRAY[10, 20, 30], ARRAY['a', NULL, 'c']), (k, v
) -> v IS NOT NULL); -- {10 -> a, 30 -> c}
SELECT map_filter(MAP(ARRAY['k1', 'k2', 'k3'], ARRAY[20, 3, 15]), (k,
v) -> v > 10); -- {k1 -> 20, k3 -> 15}
```

reduce(array<T>, initialState S, inputFunction<S, T, S>, outputFunction<S, R>) → R

The reduce function retrieves each element in the array from the initial state, calculates inputFunct ion(s,t) based on the state S, and generates a new state. It finally applies outputFunction to output the final state S to result R.

1. Initial state S
2. Retrieves each element T.
3. Calculates inputFunction(S,T) to generate a new state S.
4. Repeats Steps 2 and 3 to the last element and generate a new state.
5. Uses the final state S to obtain the final output result R.

Example:

```
SELECT reduce(ARRAY [], 0, (s, x) -> s + x, s -> s); -- 0
SELECT reduce(ARRAY [5, 20, 50], 0, (s, x) -> s + x, s -> s); -- 75
SELECT reduce(ARRAY [5, 20, NULL, 50], 0, (s, x) -> s + x, s -> s);
 -- NULL
SELECT reduce(ARRAY [5, 20, NULL, 50], 0, (s, x) -> s + COALESCE(x, 0
), s -> s); -- 75
SELECT reduce(ARRAY [5, 20, NULL, 50], 0, (s, x) -> IF(x IS NULL, s, s
 + x), s -> s); -- 75
SELECT reduce(ARRAY [2147483647, 1], CAST (0 AS BIGINT), (s, x) -> s
 + x, s -> s); -- 2147483648
SELECT reduce(ARRAY [5, 6, 10, 20], -- calculates arithmetic average:
10.25
              CAST(ROW(0.0, 0) AS ROW(sum DOUBLE, count INTEGER)),
              (s, x) -> CAST(ROW(x + s.sum, s.count + 1) AS ROW(sum
DOUBLE, count INTEGER)),
```

```
                s -> IF(s.count = 0, NULL, s.sum / s.count));
```

transform(array<T>, function<T, U>) → ARRAY<U>

Calls function for each element in the array to generate the new result U.

Example:

```
SELECT transform(ARRAY [], x -> x + 1); -- []
SELECT transform(ARRAY [5, 6], x -> x + 1); -- [6, 7] --Increments
each element by 1.
SELECT transform(ARRAY [5, NULL, 6], x -> COALESCE(x, 0) + 1); -- [6,
1, 7]
SELECT transform(ARRAY ['x', 'abc', 'z'], x -> x || '0'); -- ['x0', '
abc0', 'z0']
SELECT transform(ARRAY [ARRAY [1, NULL, 2], ARRAY[3, NULL]], a ->
filter(a, x -> x IS NOT NULL)); -- [[1, 2], [3]]
```

transform_keys(map<K1, V>, function<K1, V, K2>) → MAP<K2,V>

Applies the function for each key of the map to generate a new key.

Example:

```
SELECT transform_keys(MAP(ARRAY[], ARRAY[]), (k, v) -> k + 1); -- {}
SELECT transform_keys(MAP(ARRAY [1, 2, 3], ARRAY ['a', 'b', 'c']), (k
, v) -> k + 1); -- {2 -> a, 3 -> b, 4 -> c}  --Increments each key by
1.
SELECT transform_keys(MAP(ARRAY ['a', 'b', 'c'], ARRAY [1, 2, 3]), (k
, v) -> v * v); -- {1 -> 1, 4 -> 2, 9 -> 3}
SELECT transform_keys(MAP(ARRAY ['a', 'b'], ARRAY [1, 2]), (k, v) -> k
 || CAST(v as VARCHAR)); -- {a1 -> 1, b2 -> 2}
SELECT transform_keys(MAP(ARRAY [1, 2], ARRAY [1.0, 1.4]), -- {one ->
1.0, two -> 1.4}
                      (k, v) -> MAP(ARRAY[1, 2], ARRAY['one', 'two'])[
k]);
```

transform_values(map<K, V1>, function<K, V1, V2>) → MAP<K, V2>

Applies the function for all values in the map, converts V1 to V2, and generates a new map `<K,V2>`.

```
SELECT transform_values(MAP(ARRAY[], ARRAY[]), (k, v) -> v + 1); -- {}
SELECT transform_values(MAP(ARRAY [1, 2, 3], ARRAY [10, 20, 30]), (k,
v) -> v + 1); -- {1 -> 11, 2 -> 22, 3 -> 33}
SELECT transform_values(MAP(ARRAY [1, 2, 3], ARRAY ['a', 'b', 'c']), (
k, v) -> k * k); -- {1 -> 1, 2 -> 4, 3 -> 9}
SELECT transform_values(MAP(ARRAY ['a', 'b'], ARRAY [1, 2]), (k, v) -
> k || CAST(v as VARCHAR)); -- {a -> a1, b -> b2}
SELECT transform_values(MAP(ARRAY [1, 2], ARRAY [1.0, 1.4]), -- {1 ->
one_1.0, 2 -> two_1.4}
```

```
                          (k, v) -> MAP(ARRAY[1, 2], ARRAY['one', 'two
'])[k] || '_' || CAST(v AS VARCHAR));
```

zip_with(array<T>, array<U>, function<T, U, R>) → array<R>

Merges two arrays, and specifies the elements of the new array using the function. Element T in the first array and element U in the second array are used to generate the new result R.

Example:

```
SELECT zip_with(ARRAY[1, 3, 5], ARRAY['a', 'b', 'c'], (x, y) -> (y, x
)); --Transposes the positions of the elements of the first and second
 arrays to generate a new array. Result: [ROW('a', 1), ROW('b', 3),
ROW('c', 5)]
SELECT zip_with(ARRAY[1, 2], ARRAY[3, 4], (x, y) -> x + y); -- Result:
 [4, 6]
SELECT zip_with(ARRAY['a', 'b', 'c'], ARRAY['d', 'e', 'f'], (x, y) -
> concat(x, y)); --Concatenates the elements of the first and second
arrays to generate a new string. Result: ['ad', 'be', 'cf']
```

map_zip_with(map<K, V1>, map<K, V2>, function<K, V1, V2, V3>) → map<K, V3>

Merges two maps, uses values V1 and V2 to generate V3 based on each key, and generates a new map `as follows:K,V3>`.

```
SELECT map_zip_with(MAP(ARRAY[1, 2, 3], ARRAY['a', 'b', 'c']),
                    MAP(ARRAY[1, 2, 3], ARRAY['d', 'e', 'f']),
                    (k, v1, v2) -> concat(v1, v2)); --Merges values
which have the same map keys.-- {1 -> ad, 2 -> be, 3 -> cf}
SELECT map_zip_with(MAP(ARRAY['k1', 'k2'], ARRAY[1, 2]),
                    MAP(ARRAY['k2', 'k3'], ARRAY[4, 9]),
                    (k, v1, v2) -> (v1, v2)); --Generates an array
using the two values.-- {k1 -> ROW(1, null), k2 -> ROW(2, 4), k3 ->
ROW(null, 9)}
SELECT map_zip_with(MAP(ARRAY['a', 'b', 'c'], ARRAY[1, 8, 27]),
                    MAP(ARRAY['a', 'b', 'c'], ARRAY[1, 2, 3]),
                    (k, v1, v2) -> k || CAST(v1/v2 AS VARCHAR)); --
Concatenates the key values and division results of the two values --
 {a -> a1, b -> b4, c -> c9}
```

22.7.1.24 Logical function

Logical operators

Table 22-8: Logical operators

Operator	Description	Example
AND	Returns TRUE only when both the left and right operands are TRUE.	a AND b

Operator	Description	Example
OR	Returns TRUE if either the left or right operand is TRUE.	a OR b
NOT	Returns TRUE only when the right operand is FALSE.	NOT a

NULL involved in logical operation

The following table lists the true values when the values of a and b are TRUE, FALSE, and NULL, respectively.

Table 22-9: True value table 1

a	b	a AND b	a OR b
TRUE	TRUE	TRUE	TRUE
TRUE	FALSE	FALSE	TRUE
TRUE	NULL	NULL	TRUE
FALSE	TRUE	FALSE	TRUE
FALSE	FALSE	FALSE	FALSE
FALSE	NULL	FALSE	NULL
NULL	TRUE	NULL	TRUE
NULL	FALSE	FALSE	NULL
NULL	NULL	NULL	NULL

Table 22-10: True value table 2

a	NOT a
TRUE	FALSE
FALSE	TRUE
NULL	NULL

22.7.1.25 Column alias

Context

In the SQL standard, a column name must consist of letters, digits, and underlines and start with a letter.

If you have configured a column name not conforming to the SQL standard (such as User-Agent) during log collection configuration, you need to name an alias for the column on the statistic properties configuration page for query. The alias is used for SQL statistical analysis only. The original name is used in underlying storage. Therefore, you must use the original name when you perform a search.

In addition, you can give the column an alias to replace the original name for query when the column name is long.

Table 22-11: Alias example

Original column name	Alias
User-Agent	ua
User.Agent	ua
123	col
abceefghijklmnopqrstuvw	a

22.7.1.26 Geospatial functions

Concept of geometry

Geospatial functions support geometries in the Well-Known Text (WKT) format.

Table 22-12: Geometry format

Geometries	WKT format
Point	`POINT (0 0)`
LineString	`LINESTRING (0 0, 1 1, 1 2)`
Polygon	`POLYGON ((0 0, 4 0, 4 4, 0 4, 0 0), (1 1, 2 1, 2 2, 1 2, 1 1))`
MultiPoint	`MULTIPOINT (0 0, 1 2)`
MultiLineString	`MULTILINESTRING ((0 0, 1 1, 1 2), (2 3, 3 2, 5 4))`
MultiPolygon	`MULTIPOLYGON (((0 0, 4 0, 4 4, 0 4, 0 0), (1 1, 2 1, 2 2, 1 2, 1 1)), ((-1 -1, -1 -2, -2 -2, -2 -1, -1 -1)))`

Geometries	WKT format
GeometryCollection	`GEOMETRYCOLLECTION (POINT(2 3), LINESTRING (2 3, 3 4))`

Constructors

Table 22-13: Constructor Description

Function	Description
ST_Point(double, double) → Point	Returns a geometry type point object with the given coordinate values.
ST_LineFromText(varchar) → LineString	Returns a geometry type linestring object from WKT representation.
ST_Polygon(varchar) → Polygon	Returns a geometry type polygon object from WKT representation.
ST_GeometryFromText(varchar) → Geometry	Returns a geometry type object from WKT representation.
ST_AsText(Geometry) → varchar	Returns the WKT representation of the geometry.

Operators

Function	Description
ST_Boundary(Geometry) → Geometry	Returns the closure of the combinatorial boundary of this geometry.
ST_Buffer(Geometry, distance) → Geometry	Returns the geometry that represents all points whose distance from the specified geometry is less than or equal to the specified distance.
ST_Difference(Geometry, Geometry) → Geometry	Returns the geometry value that represents the point set difference of the given geometries.
ST_Envelope(Geometry) → Geometry	Returns the bounding rectangular polygon of a geometry.
ST_ExteriorRing(Geometry) → Geometry	Returns a line string representing the exterior ring of the input polygon.
ST_Intersection(Geometry, Geometry) → Geometry	Returns the geometry value that represents the point set intersection of two geometries.

Function	Description
ST_SymDifference(Geometry, Geometry) → Geometry	Returns the geometry value that represents the point set symmetric difference of two geometries. Returns the non-intersecting parts of two geometrics.

Relationship tests

Function	Description
ST_Contains(Geometry, Geometry) → boolean	Returns TRUE if and only if no points of the second geometry lie in the exterior of the first geometry, and at least one point of the interior of the first geometry lies in the interior of the second geometry. Returns FALSE if the second geometry is on the boundaries of the first geometry.
ST_Crosses(Geometry, Geometry) → boolean	Returns TRUE if the supplied geometries have some, but not all, interior points in common.
ST_Disjoint(Geometry, Geometry) → boolean	Returns TRUE if the give geometries do not spatially intersect – if they do not share any space together.
ST_Equals(Geometry, Geometry) → boolean	Returns TRUE if the given geometries represent the same geometry.
ST_Intersects(Geometry, Geometry) → boolean	Returns TRUE if the given geometries spatially intersect in two dimensions (share any portion of space).
ST_Overlaps(Geometry, Geometry) → boolean	Returns TRUE if the given geometries share space, are of the same dimension, but are not completely contained by each other.
ST_Relate(Geometry, Geometry) → boolean	Returns TRUE if first geometry is spatially related to second geometry.
ST_Touches(Geometry, Geometry) → boolean	Returns TRUE if the given geometries have at least one point in common, but their interiors do not intersect.
ST_Within(Geometry, Geometry) → boolean	Returns TRUE if first geometry is completely inside second geometry. Returns FALSE if boundary intersection exists.

Accessors

Function	Description
ST_Area(Geometry) → double	Returns the 2D Euclidean area of a geometry.
ST_Centroid(Geometry) → Geometry	Returns the point value that is the mathematic al centroid of a geometry.
ST_CoordDim(Geometry) → bigint	Return the coordinate dimension of the geometry.
ST_Dimension(Geometry) → bigint	Returns the inherent dimension of this geometry object, which must be less than or equal to the coordinate dimension.
ST_Distance(Geometry, Geometry) → double	Returns the two-dimensional cartesian minimum distance (based on spatial ref) between two geometries in projected units.
ST_IsClosed(Geometry) → boolean	Returns TRUE if the linestring's start and end points are coincident.
ST_IsEmpty(Geometry) → boolean	Returns TRUE if this Geometry is an empty geometrycollection, polygon, point etc.
ST_IsRing(Geometry) → boolean	Returns TRUE if and only if the line is closed and simple.
ST_Length(Geometry) → double	Returns the length of a linestring or multi-linestring using Euclidean measurement on a two-dimensional plane (based on spatial ref) in projected units. Returns the length of a linestring or multi-linestring. The length is a prediction specific to a two-dimensional plane based on spatial reference using Euclidean measurement.
ST_XMax(Geometry) → double	Returns X maxima of a bounding box of a geometry.
ST_YMax(Geometry) → double	Returns Y maxima of a bounding box of a geometry.
T_XMin(Geometry) → double	Returns X minima of a bounding box of a geometry.
ST_YMin(Geometry) → double	Returns Y minima of a bounding box of a geometry.

Function	Description
ST_StartPoint(Geometry) → point	Returns the first point of a LineString geometry as a Point.
ST_EndPoint(Geometry) → point	Returns the last point of a LineString geometry as a Point.
ST_X(Point) → double	Return the X coordinate of the point.
ST_Y(Point) → double	Return the Y coordinate of the point.
ST_NumPoints(Geometry) → bigint	Returns the number of points in a geometry.
ST_NumInteriorRing(Geometry) → bigint	Returns the cardinality of the collection of interior rings of a polygon.

22.7.1.27 JOIN syntax

JOIN associates the fields of multiple tables. In Log Service, JOIN is applicable to a single Logstore, between Logstore and RDS, and between Logstores. This document describes how to use JOIN across Logstores.

Procedure

1. *Download* the latest version of Python SDK.
2. Use the GetProjectLogs API for query.

SDK example

```
#!/usr/bin/env python
#encoding: utf-8
import time,sys,os
from aliyun.log.logexception import LogException
from aliyun.log.logitem import LogItem
from aliyun.log.logclient import LogClient
from aliyun.log.getlogsrequest import GetLogsRequest
from aliyun.log.getprojectlogsrequest import GetProjectLogsRequest
from aliyun.log.putlogsrequest import PutLogsRequest
from aliyun.log.listtopicsrequest import ListTopicsRequest
from aliyun.log.listlogstoresrequest import ListLogstoresRequest
from aliyun.log.gethistogramsrequest import GetHistogramsRequest
from aliyun.log.index_config import *
from aliyun.log.logtail_config_detail import *
from aliyun.log.machine_group_detail import *
from aliyun.log.acl_config import *
if __name__=='__main__':
    token = None
    endpoint = "http://cn-hangzhou.log.aliyuncs.com"
    accessKeyId  = 'LTAIvKy7U'
    accessKey='6gXLNTLyCfdsfwrewrfhdskfdsfuiwu'
    client = LogClient(endpoint, accessKeyId, accessKey,token)
    logstore = "meta"
    # In the query statements, specify two Logstores, their respective
 time ranges, and the key for Logstore association.
```

```
    req = GetProjectLogsRequest(project,"select count(1) from
sls_operation_log  s join  meta m on s.__date__ >'2018-04-10 00:00:00
' and s.__date__ <'2018-04-11 00:00:00' and m.__date__ >'2018-04-23 00
:00:00' and m.__date__ <'2018-04-24 00:00:00' and s.projectid = cast(m
.ikey as varchar)");
    res = client.get_project_logs(req)
    res.log_print();
    exit(0)
```

22.7.2 Optimize a query

The analysis efficiency varies from one query to another. Common methods of query optimization are as follows:

1. *Avoid running Group By on string columns if possible*
2. *List fields with relatively large dictionary values on top when running Group By on multiple columns*
3. *Use estimating functions*
4. *Retrieve required columns in SQL and do not read all columns if possible*
5. *Place non-GROUP BY columns in an aggregate function if possible.*

Avoid running Group By on string columns if possible

Running Group By on strings may result in a large amount of hash calculations, which account for more than 50% of total calculations.

For example, consider the following two queries:

```
* | select count(1) as pv , date_trunc(hour,__time__) as time group by
 time * | select count(1) as pv , from_unixtime(__time__-__time__%3600
) as time group by __time__-__time__%3600
```

Both Query 1 and Query 2 calculate the log count value every hour. However, Query 1 converts time into a string, for example, `2017-12-12 00:00:00`, and then runs Group By on this string. Query 2 runs Group By on the on-the-hour time value and then converts the result into a string. Query 1 is less efficient than Query 2 because the former needs to hash strings.

List fields with relatively large dictionary values on top when running Group By on multiple columns

For example, there are 13 provinces with 100 million users.

```
Quick: * | select province,uid,count(1)groupby province,uid Slow: * |
select province,uid,count(1)groupby uid,province
```

Use estimating functions

Estimating functions provide much stronger performance than accurate calculation. In estimation, accuracy is compromised to an acceptable extent for fast calculation.

```
Quick: * |select approx_distinct(ip) Slow: * | select count(distinct(
ip))
```

Retrieve required columns in SQL and do not read all columns if possible

Use the query syntax to retrieve all columns. To speed up calculation, retrieve only the required columns in SQL if possible.

```
Quick: * |select a,b c Slow:* |select*
```

Place non-GROUP BY columns in an aggregate function if possible.

For example, a user ID exactly corresponds to a user name. Therefore, use only userid in running GROUP BY.

```
Quick: * | select userid, arbitrary(username), count(1)groupby userid
Slow: * | select userid, username, count(1)groupby userid,username
```

22.7.3 Excellent analysis cases

Case list

1. *Trigger an alarm when the error rate in the recent 5 minutes exceeds 40%.*
2. *Trigger an alarm when traffic decreases sharply*
3. *Calculate the average latency of each bucket set by data range.*
4. *Return percentage data included in GROUP BY results.*
5. *Count the number of entries that satisfy the condition.*

Trigger an alarm when the error rate in the recent 5 minutes exceeds 40%.

Count the percentage of Error 500 every minute; trigger an alarm when the percentage exceeds 40% for the past 5 minutes.

```
status:500 | select __topic__, max_by(error_count,window_time)/1.0/sum
(error_count) as error_ratio, sum(error_count) as total_error from (
select __topic__, count(*) as error_count , __time__ - __time__ % 300
 as window_time from log group by __topic__, window_time ) group by
```

```
__topic__ having max_by(error_count,window_time)/1.0/sum(error_count)
 > 0.4 and sum(error_count) > 500 order by total_error desc limit 100
```

Trigger an alarm when traffic decreases sharply

Count traffic every minute; trigger an alarm when the recent traffic decreases sharply. Data from the past one minute does not cover a full minute; therefore, divide the statistical value by (max(**time**) - min(**time**)) for normalization to calculate the average traffic per minute.

```
* | SELECT SUM(inflow) / (max(__time__) - min(__time__)) as inflow_per
_minute, date_trunc(minute,__time__) as minute group by minute
```

Calculate the average latency of each bucket set by data range.

```
* | select avg(latency) as latency , case when originSize < 5000 then
 s1 when originSize < 20000 then s2 when originSize < 500000 then s3
when originSize < 100000000 then s4 else s5 end as os group by os
```

Return percentage data included in GROUP BY results.

List the count results of different departments and related percentage data. The query combines subquery and window functions. sum(c) over() indicates the sum of values in all rows.

```
* | select department, c*1.0/ sum(c) over () from(select count(1) as c
, department from log groupby department)
```

Count the number of entries that satisfy the condition.

To perform counting by URL feature in the URL, you can use the CASE WHEN syntax, or the simpler count_if syntax.

```
* | select count_if(uri like %login) as login_num, count_if(uri like %
register) as register_num, date_format(date_trunc(minute, __time__), %
m-%d %H:%i) as time group by time order by time limit 100
```

22.7.4 Quick analysis

The quick analysis function of Log Service supports a quick interactive query. This service allows you to quickly analyze the distribution of a field over a period of time and reduce the cost of indexing key data.

Features

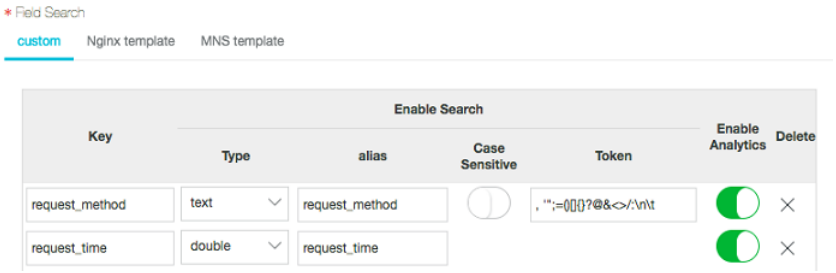
- Supports grouping statistics for the first 10 of the first 10,000 pieces of data of `Text` fields.
- Supports Quick generation of the `approx_distinct` query statement for `Text` fields.
- Supports histogram statistics for the approximate distribution of `long` or `double` fields.

- Supports quick search for the maximum value, minimum value, average, or sum for `long` or `double` fields.
- Supports generating query statements based on quick analysis and query.

The user must specify the field query properties before using the quick analysis.

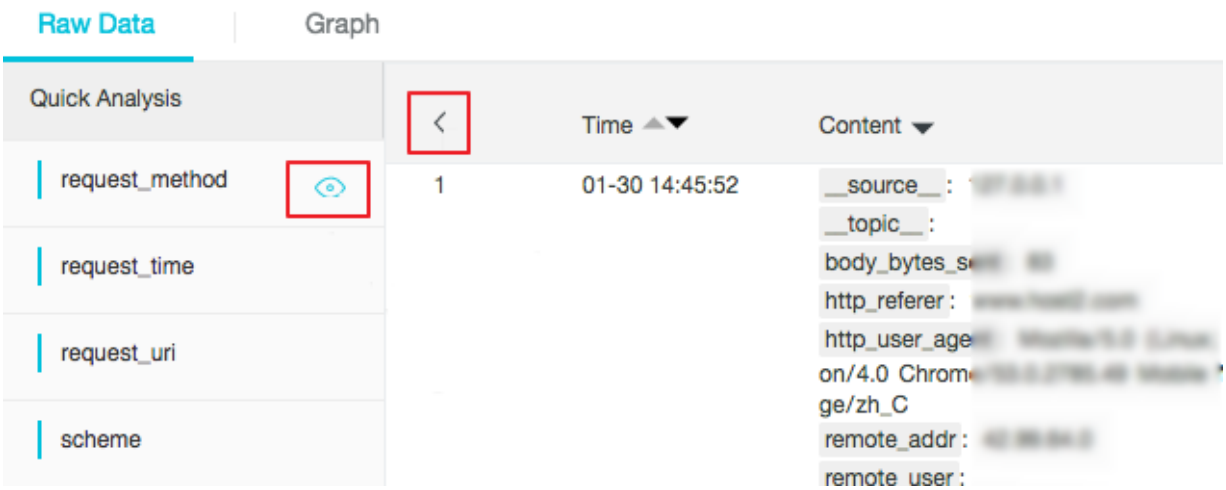
1. For specified field query, you must enable the index to activate the query and analysis function.
2. Set the `key` in the log as the field name and set the type, alias, and separator.

   If the access log contains `request_method` and `request_time` , you can configure the settings as follows:



User guide

After field query setting, go to the query page, click the **Raw Logs** tab, and view the fields in the left-side **Quick Analysis** column. Click the button above the sequence number to hide the page. Click the **eye** button to start quick analysis based on the **Current Temporal Interval** and **Current $Search** condition.



Text type

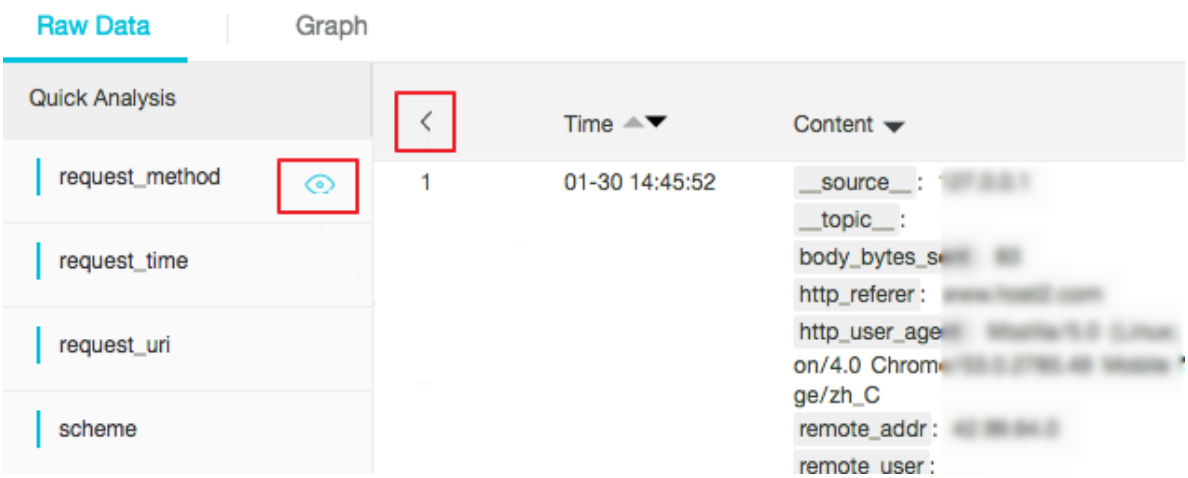
- **Group statistics for the Text type**

Click the **eye** at the right of the filed to quickly group the first 1,000 pieces of data of this **Text** field and return the ratio of the first 10 pieces.

The query statement is as follows:

```
$Search | select count(1) as pv , "${keyName}" from ( select "${
keyName}" from log limit 10000) group by "${keyName}" order by pv
desc limit 10
```

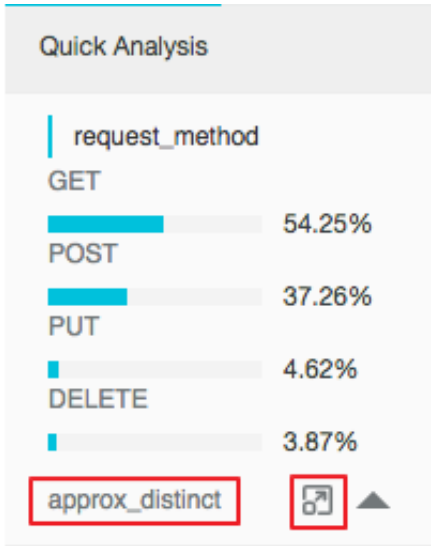
`request_method` returns the following results based on the grouping statistics, where `GET` requests are in the majority.



- **Check the number of unique entries of the field**

  Under the target fields in **Quick Analysis**, click **approx_distinct** to check the number of unique entries for `${keyName}`.

  `request_method`returns the following results based on the grouping statistics, where `GET` requests are in the majority.



- **Extend the query statement of grouping statistics to the search box**

  Click the button at the right of **approx_distinct**to extend the query statement of grouping statistics to the search box for further operations.

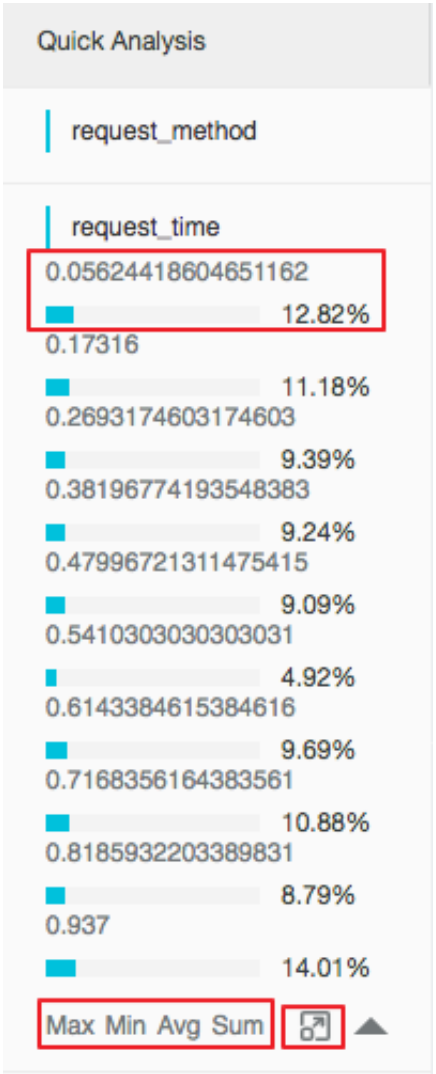
long/double

- **Histogram statistics for the approximate distribution**

  Grouping statistics is of little significance for the `long/double` fields, which have multiple type values. Therefore, histogram statistics for the approximate distribution is adopted by using 10 buckets.

```
$Search | select numeric_histogram(10, ${keyName})
```

  `request_time` returns the following result based on the histogram statistics for the approximate distribution. You can see that the request time is mostly distributed around 0.056.



- **`MaxMinAvgSum`Statement Quick Analysis**

  Click `Max`, `Min`, `Avg`, and `Sum` under the target field to quickly search for the maximum value, minimum value, average, and sum of all `Max`.

- **Extend the query statement of grouping statistics to the search box**

  Click the button at the right of `Sum` to extend the query statement of the histogram statistics for the approximate distribution to the search box for further operations.

22.7.5 JDBC protocol

In addition to RESTful APIs, you can use JDBC and standard SQL 92 for log query and analysis.

Connection parameters

Connection parameters	Example	Description
host	regionid.example.com	Access point
port	10005	10005 is the default port.
user	bq2sjzesjmo86kq	Accesskey ID
password	4fdO1fTDDuZP	Accesskey
database	sample-project	Project under an account
table	sample-logstore	Logstore under a project

The following is an example of connection through a MySQL command:

```
mysql -hcn-shanghai-intranet.log.aliyuncs.com -ubq2sjzesjmo86kq -
p4fdO1fTDDuZP -P10005 use sample-project; // Uses a specific project.
```

Prerequisites

You must use the accesskey of the primary account or a sub-account to access the JDBC interface. The sub-account must belong to the project owner and have the project-level read permission.

Syntax

NOTE

The where condition must contain `__date__` or `__time__` to limit the time range of query. The type of `__date__` is timestamp, and the type of `__time__` is bigint.

For example:

- `__date__ > 2017-08-07 00:00:00 and __date__ < 2017-08-08 00:00:00`
- `__time__ > 1502691923 and __time__ < 1502692923`

At least one of the preceding conditions must be met.

Filter syntax

The filter syntax in the WHERE condition is as follows:

Meaning	Example	Description
String search	`key = "value"`	Results after word segmentation are queried.
String fuzzy search	`key = "valu*"`	Results of fuzzy match after word segmentation are queried.
Value comparison	`num_field > 1`	The supported comparison operators include >, >=, =, <, and <=.
Logic operations	`and or not`	For example, `a = "x" and b ="y"` or `a = "x" and not b ="y"`.
Full-text search	`__line__ ="abc"`	Full-text index search requires the special key (`__line__`).

Computation syntax

For supported computation operators, see *Analysis syntax and functions*.

SQL92 syntax

The SQL92 syntax is a combination of filter and computation syntaxes.

The following query is used as an example:

```
status>200 |select avg(latency),max(latency) ,count(1) as c GROUP BY
method ORDER BY c DESC LIMIT 20
```

The filter part and time condition in the query can be combined into a new query condition based on standard SQL92 syntax.

```
select avg(latency),max(latency) ,count(1) as c from sample-logstore
 where status>200 and __time__>=1500975424 and __time__ < 1501035044
GROUP BY method ORDER BY c DESC LIMIT 20
```

Access Log Service by using JDBC protocol

Program call

Developers can use the MySQL syntax to connect to Log Service in any program that supports MySQL connector. For example, JDBC or Python MySQLdb can be used.

Example:

```
import com.mysql.jdbc.*; import java.sql.*; import java.sql.Connection
; import java.sql.ResultSetMetaData; import java.sql.Statement; public
 class testjdbc { public static void main(String args[]){ Connection
 conn = null; Statement stmt = null; try { //STEP 2: Register JDBC
 driver Class.forName("com.mysql.jdbc.Driver"); //STEP 3: Open a
 connection System.out.println("Connecting to a selected database
..."); conn = DriverManager.getConnection("jdbc:mysql://cn-shanghai-
intranet.log.aliyuncs.com:10005/sample-project","accessid","accesskey
"); System.out.println("Connected database successfully..."); //STEP
 4: Execute a query System.out.println("Creating statement..."); stmt
 = conn.createStatement(); String sql = "SELECT method,min(latency,10
) as c,max(latency,10) from sample-logstore where __time__>=1500975424
 and __time__ < 1501035044 and latency > 0 and latency < 6142629
and not (method=Postlogstorelogs or method=GetLogtailConfig) group
by method " ; String sql-example2 = "select count(1) ,max(latency),
avg(latency), histogram(method),histogram(source),histogram(status),
histogram(clientip),histogram(__source__) from test10 where __date__
 >2017-07-20 00:00:00 and __date__ <2017-08-02 00:00:00 and __line__=
abc#def and latency < 100000 and (method = getlogstorelogS or method=
Get** and method <> GetCursorOrData )"; String sql-example3 = "select
 count(1) from sample-logstore where __date__ > 2017-08-07 00:00:00
 and __date__ < 2017-08-08 00:00:00 limit 100"; ResultSet rs = stmt
.executeQuery(sql); //STEP 5: Extract data from result set while(
rs.next()){ //Retrieve by column name ResultSetMetaData data = rs.
getMetaData(); System.out.println(data.getColumnCount()); for(int i
 = 0;i < data.getColumnCount();++i) { String name = data.getColumnN
ame(i+1); System.out.print(name+":"); System.out.print(rs.getObject
(name)); } System.out.println(); } rs.close(); } catch (ClassNotFo
undException e) { e.printStackTrace(); } catch (SQLException e) { e
.printStackTrace(); } catch (Exception e) { e.printStackTrace(); }
finally { if (stmt != null) { try { stmt.close(); } catch (SQLExcepti
on e) { e.printStackTrace(); } } if (conn != null) { try { conn.close
(); } catch (SQLException e) { e.printStackTrace(); } } } } }
```

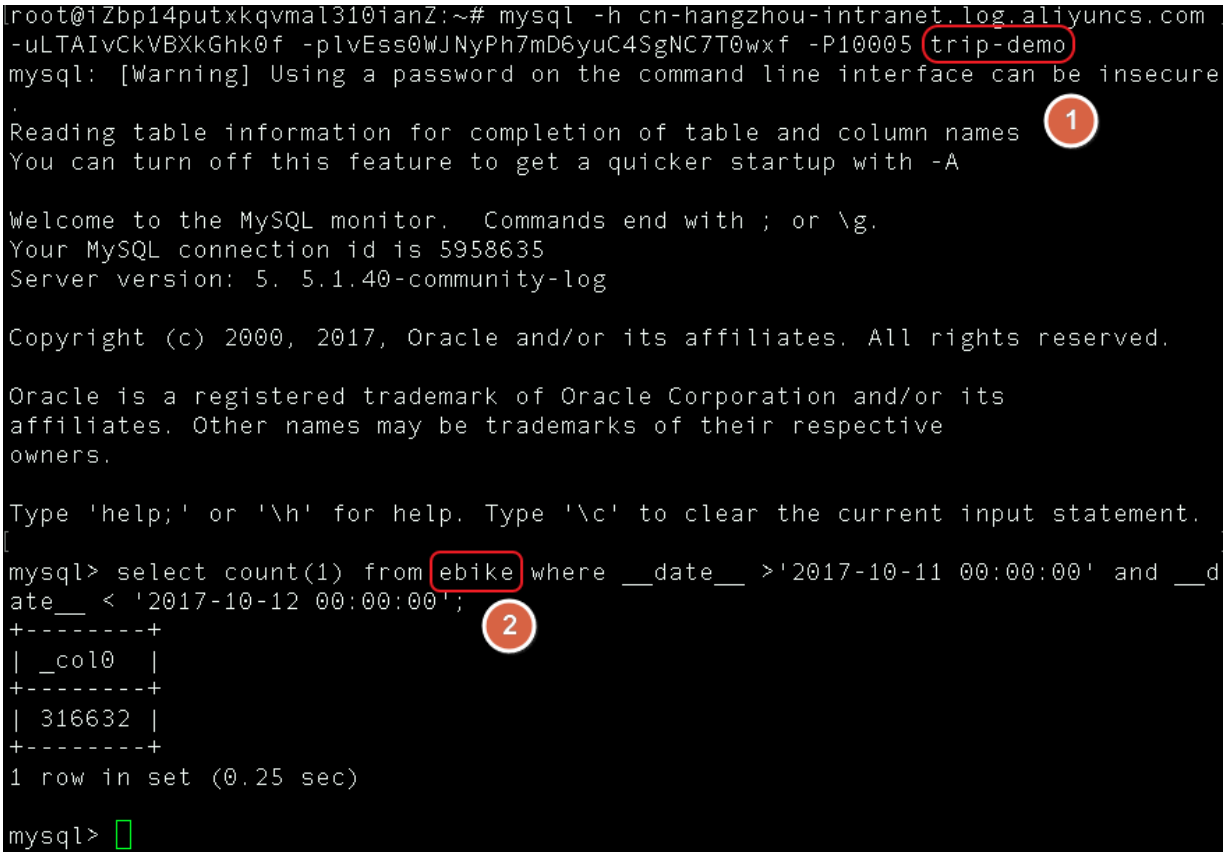
Tool call

In the classic internal network or VPC environment, use the MySQL client to connect to Log Service, as shown in *Figure 22-2: Connection example*.

Note:

1. Enter your project name at ①.
2. Enter your Logstore name at ②.

Figure 22-2: Connection example



22.8 Query and visualization

22.8.1 Analysis charts

22.8.1.1 Chart

Log Service supports aggregate functions. You can render the results of SQL aggregate computing through visual charts.

Note:

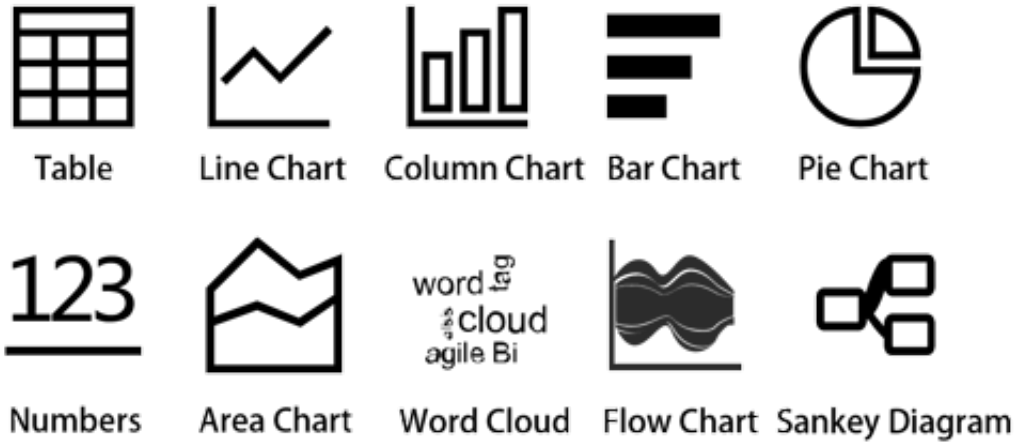
Before using visual charts, read the *Real-time analysis*.

Prerequisites

1. The log data has been successfully collected.
2. Indexes have been created and the **Analysis** feature has been enabled.
3. An analysis statement is used in query to display charts based on statistics.

Chart types

Currently, Log Service provides the following chart types:



For how to use each chart type, see the following document:

- *Table*
- *Line chart*
- *Column chart*
- *Bar chart*
- *Pie chart*
- *Number chart*
- *Area chart*
- *Flow chart*
- *Sankey diagram*
- *Word cloud*

22.8.1.2 Dashboard

After you enable the LogSearch/Analytics - Query function, in addition to entering a query condition in the search box, you can save frequently used queries to the following locations:

- Dashboard

- Saved Search
- Alarm: Created by Saved Search

Other operations

- **View**: You can perform the following operations on the existing **Dashboard**:
  - In the left-side menu of the project, go to **Search Analysis** > **Dashboard** to view or delete dashboards.
  - On the search analysis page, click **New Tag** on the left. In the pop-up menu, click the **Dashboard** tab.
- **Modify**: Click Edit on the Dashboard page to adjust the icon attributes, size, and location. You can also click Full Screen and Refresh for better effect.

Restrictions

- A maximum of five dashboards can be created for each project. A maximum of 10 analytic queries can be created for each dashboard for simultaneous display.
- Display by line charts, bar charts, pie charts, numeric values, and area charts is supported.
- You can set the positions and adjust the sizes of individual charts and save the settings.

Procedure

1. *Log on to the Log Service Console*.
2. Click the project name to go to the **Logstores**. Click the **Search** button on the right to go to the **Search** page.
3. Enter a search analysis statement in the Search box and click **Search**.
4. Add a dashboard to the search analysis view. On the search analysis page, select the corresponding view and click **Add to Dashboard**.
5. In the pop-up menu, select **New** or **Add to Existing Dashboard**, and enter a dashboard name and a table name.

22.8.1.3 Table

Context

A table is the most common data display form and the most basic data sorting method for fast reference and analysis. Log Service provides aggregate computing similar to SQL. By default, the data results obtained by a search analysis statement are displayed in tables.

Basic components

- Header
- Row
- Column

Wherein:

- The number of `SELECT` items is the number of columns.
- The number of rows is calculated based on the number of logs in the current time range. The default value is `LIMIT 100`.

Procedure

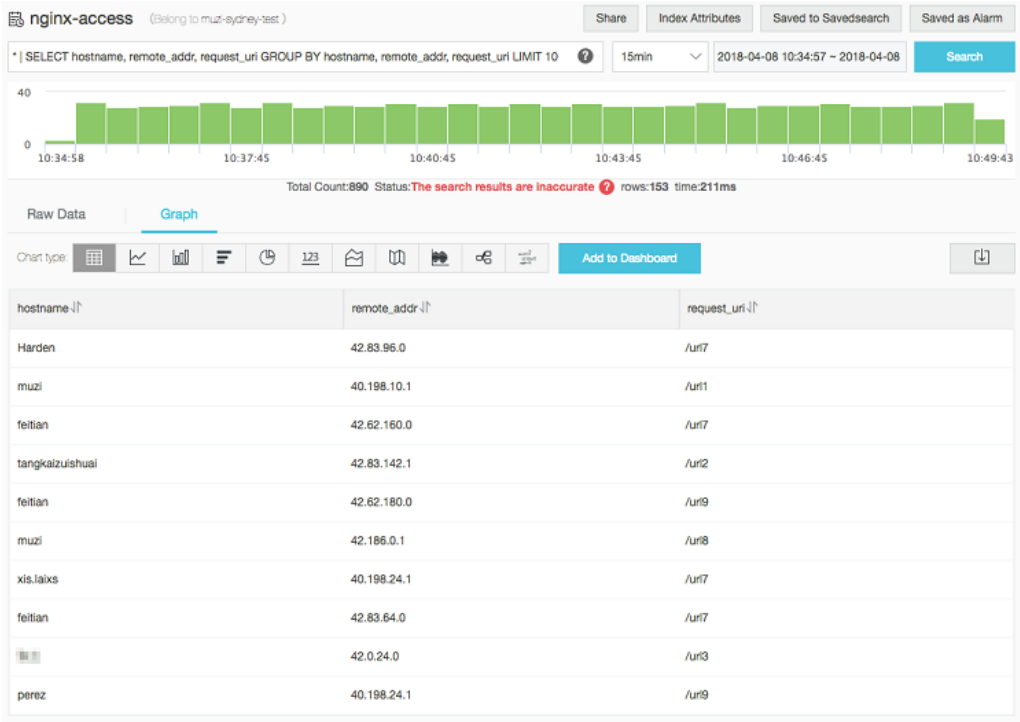
1. *Log on to the Log Service Console*.
2. Choose **Project Name** > **Search** to go to the query page.
3. Enter a search analysis statement in the Search box on the query page, select a time range, and click **Search** on the right.
4. Click the **Graph** tab. By default, results are displayed in tables, as shown in  .

Example

Filter the columns of raw log data. For example, the raw log data is as follows:

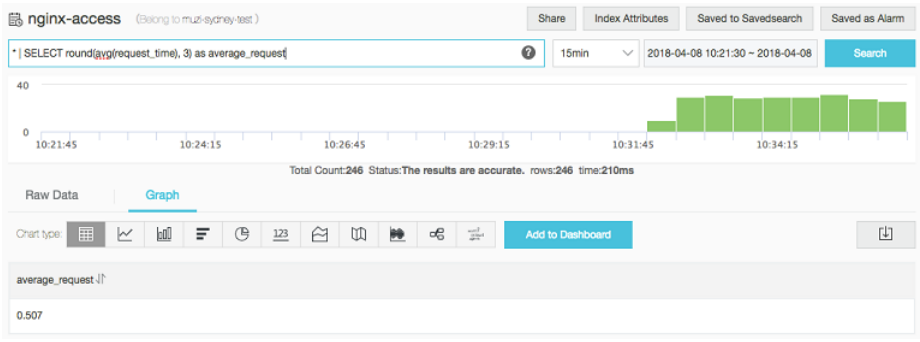
1. Filter the latest 10 logs by `hostname`, `remote_addr`, and `request_uri`.

```
* | SELECT hostname, remote_addr, request_uri GROUP BY
hostname, remote_addr, request_uri LIMIT 10
```



2. Calculate a single data item, for example, the average value of `request_time` (average request time) in the current time range, and retain three decimal digits.The statement is as follows:

```
* | SELECT round(avg(request_time), 3) as average_request
```



3. To compute grouped data, for example, the distribution of `request_method` in the current time range, and sort data in descending order.

```
* | SELECT request_method, count(*) as count GROUP BY
request_method ORDER BY count DESC
```

22.8.1.4 Line chart

A line chart is a graph for analyzing trend. It is typically used to indicate the changes of a group of data under an ordered data type (successive time intervals in most cases) for analyzing the trend of data changes intuitively.

A line chart clearly shows the changes of data over a period in the following aspects:

- Progressive increase or decrease
- Rate of increase or decrease
- Incremental or decremental pattern, for example, periodic change
- Peak and valley

A line chart is the best choice for analyzing the trend of data changes over time. You can also use multiple lines to analyze the changing trend of multiple groups of data in the same period, and then analyze the mutual effect (such as increasing or decreasing at the same time and being inversely proportional to each other) among data in different groups.

Basic components

- X axis
- Left Y axis
- (Optional) Right Y axis
- Data point
- Changing trend line
- Legend

Configuration items

Configuration item	Description
X axis	The X axis is typically an ordered data type (time series).

Configuration item	Description
Left Y axis	You can configure one or more columns of data to correspond to the value interval of the left Y axis.
Right Y axis	You can configure one or more columns of data to correspond to the value interval of the right Y axis (the layer of the right Y axis is higher than that of the left Y axis).
Column Marker	Displays a selected column in the left Y axis or right Y axis as a histogram.
Legend	Location where the legend is in the graph. You can configure the legend to the top, bottom, left, or right of the graph.
Padding	Distance between the coordinate axis and the graph boundary.

Procedure

1. *Log on to the Log Service Console*.
2. Click the name of a project, and click **Search** in the **LogSearch** column to go to the query page.
3. Enter a query statement in the search box, select the time interval, and click **Search**.
4. Click the **Graph** tab and click the  icon, that is, line chart.
5. Configure the chart properties.

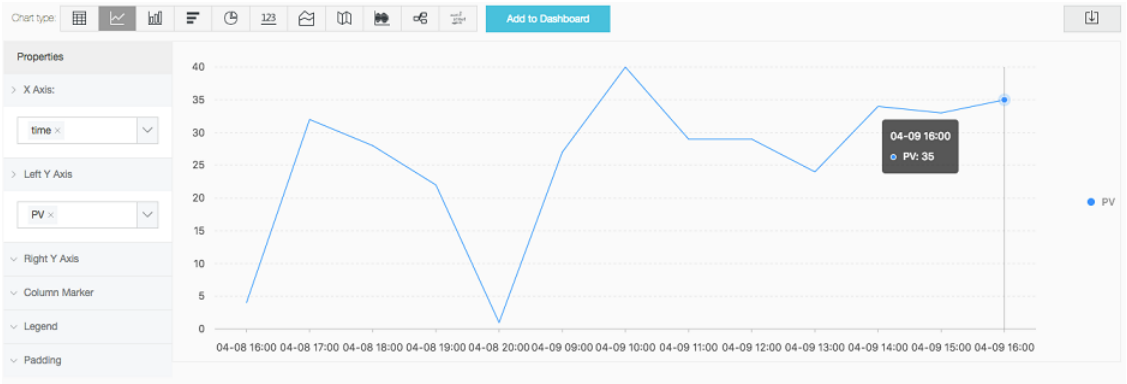
   Note:
   In a line chart, a single line must contain more than two data records to guarantee data trend analysis. We also recommend that a line chart contain no more than five lines.

   - Simple line chart

To query the access situation of the IP address `42.0.192.0` in the last day, the statement is as follows:

```
remote_addr: 42.0.192.0 | select date_format(date_trunc
(hour, __time__), %m-%d %H:%i) as time, count(1) as PV
group by time order by time limit 1000
```

Select `time` for the X axis, `PV` for the left Y axis, and **Bottom** for the legend, and adjust the padding properly.

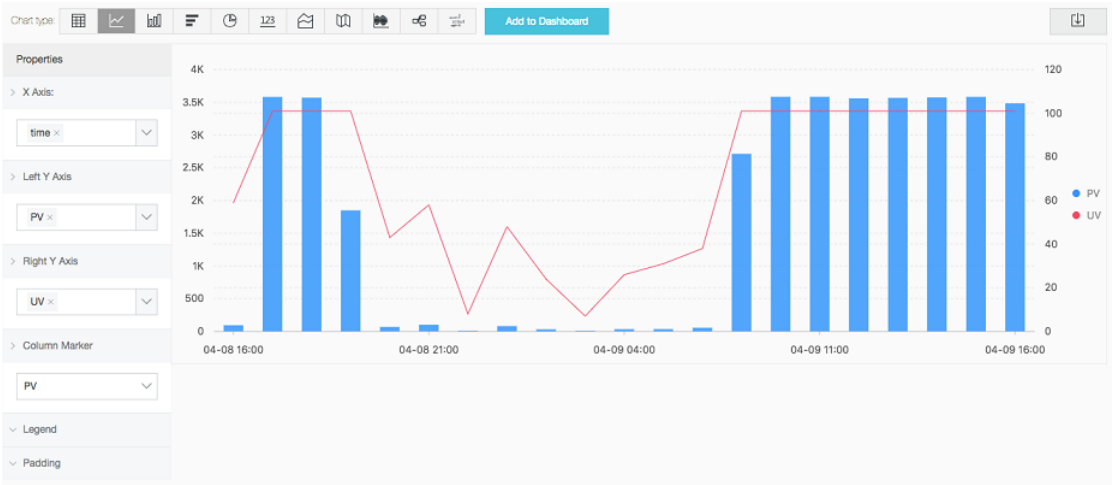


- Line chart with both left Y axis and right Y axis

  To query the access PVs and UVs in the last day, the statement is as follows:

  ```
  * | select date_format(date_trunc(hour, __time__), %m-%d %
  H:%i) as time, count(1) as PV, approx_distinct(remote_addr
  ) as UV group by time order by time limit 1000
  ```

  Select `time` for the X axis, `PV` for the left Y axis, `UV` for the right Y axis, and `PV` for the column marker.



22.8.1.5 Column chart

A column chart uses vertical or horizontal columns to show the comparison between numeric values of different data types. A line chart describes ordered data, which a column chart describes different types of data and counts the number in each data type.

You can also use multiple rectangular blocks to correspond to one type attribute in the grouping or stacked modes to analyze the differences of data types in different dimensions.

Line chart

- X axis (horizontal axis)
- Y axis (vertical axis)
- Rectangular block
- Legend

The column charts provided by Log Service use vertical columns by default. The width of a rectangle is fixed, and the height indicates a numeric value. Use the grouped column chart to display the data if multiple columns of data are mapped to the Y axis.

Configuration items

Configuration item	Description
X axis	The X axis indicates the data types.
Y axis	You can configure one or more columns of data to correspond to the value interval of the Y axis.
Legend	Location where the legend is in the graph. You can configure the legend to the top, bottom, left , or right of the graph.
Padding	Distance between the coordinate axis and the graph boundary.

Procedure

1. *Log on to the Log Service Console*.
2. Click the name of a project, and click **Search** in the **LogSearch** column to go to the query page.
3. Enter a query statement in the search box, select the time interval, and click **Search**.
4. Click the **Graph** tab and click the  icon, that is, column chart.

5. Configure the chart properties.

> **Note:**
> Use the column chart for no more than 20 pieces of data. We recommend that you use `LIMIT` to control the amount of data in case that the horizontal width is so large that the analytical comparison is not intuitive. We also recommend that you have no more than five columns of data to map to the Y axis.

- Simple column chart

  To query the number of visits for each `http_referer` in the current time range, the statement is as follows:

  ```
  * | select http_referer, count(1) as count group by
  http_referer
  ```

  Select `http_referer` for the X axis and `count` for the Y axis.

- Grouped column chart

  To query the number of visits and the average bytes for each `http_referer` in the current time range, the statement is as follows:

  ```
  * | select http_referer, count(1) as count, avg(body_bytes
  _sent) as avg group by http_referer
  ```

  Select `http_referer` for the X axis and `count` and `avg` for the Y axis.

22.8.1.6 Bar chart

The bar chart is another form of column chart, that is, the horizontal column chart. It is typically used for top N analysis and is configured in a way similar to a column chart.

Basic components

- X axis (vertical)
- Y axis (horizontal)
- Rectangular block
- Legend

A bar chart has fixed rectangular height and varying width to indicate values. When multiple data columns map to the Y-axis, the data is displayed by bars in group mode.

Properties

Table 22-14: Configuration items

Properties	Description
X Axis	Generally, the X axis indicates the data types.
Y Axis	You can configure one or more columns of data to map to the value interval of the Y axis.
Legend	The location where the legend is in the graph. You can configure the legend to the top, bottom, left, or right of the graph.
Padding	The distance between the coordinate axis and the graph boundary

Procedure

1. *Log on to the Log Service Console*.
2. Choose **Project Name** > **Search** to go to the query page.
3. Enter a query statement, select a time range, and click **Search** on the right.
4. Select the bar chart  .
5. Configure chart properties.

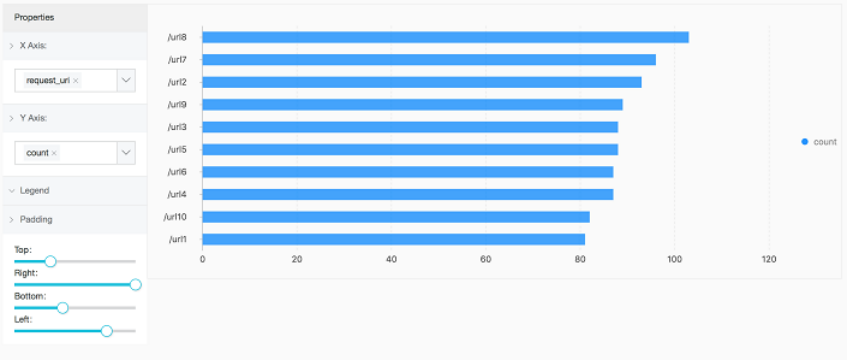
Note:

- Use the bar chart if the number of data types is no more than 20. We recommend that you use `LIMIT` to to control the number of data types in case that the vertical height is so large that the analytical comparison is not intuitive, and use the `ORDER BY` syntax for top N analysis. We also recommend that you have no more than five columns of data to map to the Y axis.
- Supports grouped bar chart, but data in all groups of the bar chart must indicate the increase or decrease at the same time.

Simple horizontal bar chart

Analyze the top 10 `request_uri` by traffic:

```
* | select request_uri, count(1) as count group by
request_uri order by count desc limit 10
```



22.8.1.7 Pie chart

The pie chart is used to indicate the ratios of different data types and compare different data types by using the radian. A pie is divided into multiple sections according to the ratios of different data types. The entire pie indicates the total amount of data, and each section (arc) indicates the ratio of a data type to the total amount of data. The sum of all the section (arc) ratios is 100%.

Basic components

- Sector
- Text percentage
- Legend

Properties

Properties	Description
Type	Data types.

Properties	Description
Value Column	The value corresponding to different types of data.
Legend	You can configure the legend to the top, bottom , left, or right of the graph.
Padding	The distance between the coordinate axis and the graph boundary.
Chart Types	Log Service provides three types of pie charts : pie chart (default), the cycle graph, and the Nightingale rose diagram.

Type

Log Service provides three types of pie charts: the default pie chart, the cycle graph, and the Nightingale rose diagram.

Cycle graph

Essentially, the cycle graph is a pie chart without the central part. Compared with the pie chart, the cycle graph has the following advantages:

- Supports displaying the total amount based on the original components, which provides you with more information.
- You may find it difficult to understand the differences between two charts simply by comparing them. Two cycle graphs can be compared by using the ring length.

Nightingale rose diagram

Essentially, the Nightingale rose diagram is not a cycle graph, but a column chart in the polar coordinate system. The data types are divided by arcs and the radius of the arc indicates the data size. Compared with the pie chart, the Nightingale rose diagram has the following advantages:

- Use the pie chart if the number of data types is no more than 10, and use the Nightingale rose diagram if the number of data types is between 11 and 30.
- Because the radius and area are of a square relationship, the Nightingale rose diagram enlarges the differences among different types of data, and is especially applicable to comparing similar values.
- A circle shows a periodic pattern. Therefore, the Nightingale rose diagram can also be used to indicate a periodic time concept, such as weeks and months.

Procedure

1. *Log on to the Log Service Console*.
2. Choose **Project Name** > **Search** to go to the query page.
3. Enter a query statement, select a time range, and click **Search** on the right.
4. Select the pie chart .
5. Configure chart properties.

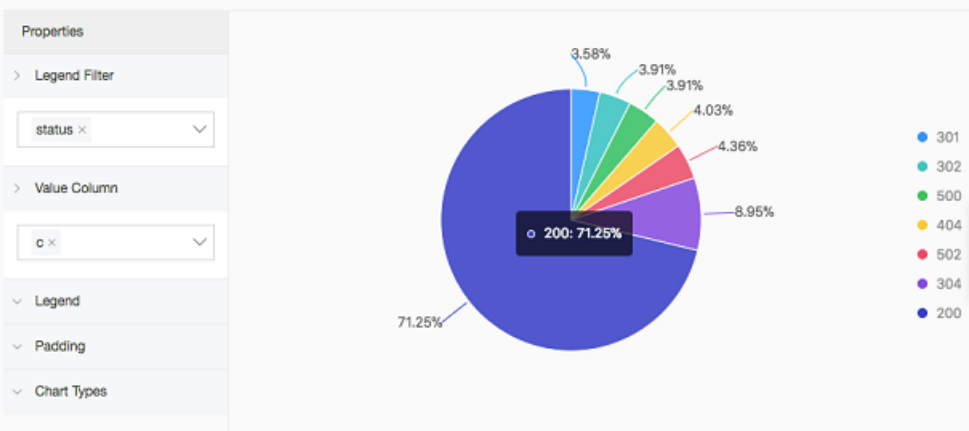
Note:

- A pie chart and a ring chart can show no more than 10 data items. We recommend that you use `LIMIT` to control the number of data items to avoid unclear analysis due to excessive planes of different colors.
- We recommend that you use the Nightingale rose diagram or a bar chart to analyze more than 10 data items.

- Pie chart

  Analyze the ratio of the access `status`:

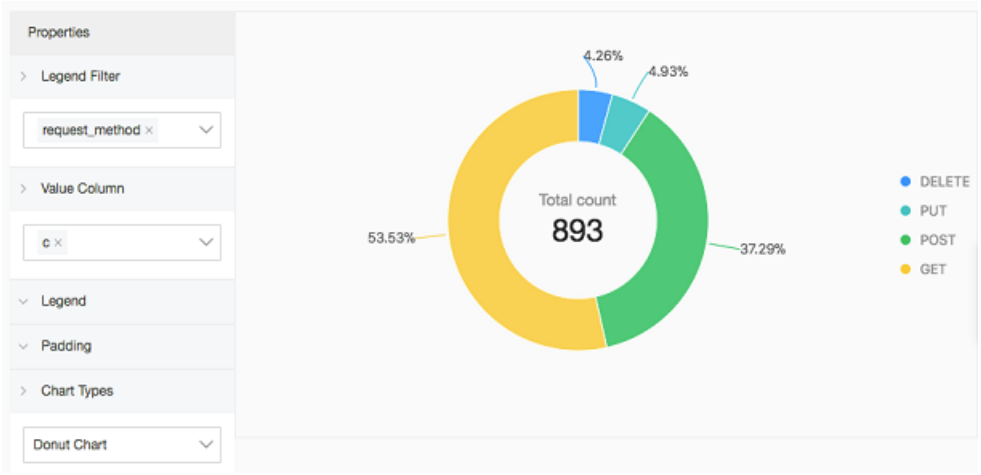
```
* | select status, count(1) as c group by status order by
c limit 10
```



- Cycle graph

Analyze the ratio of the access `request_method`:

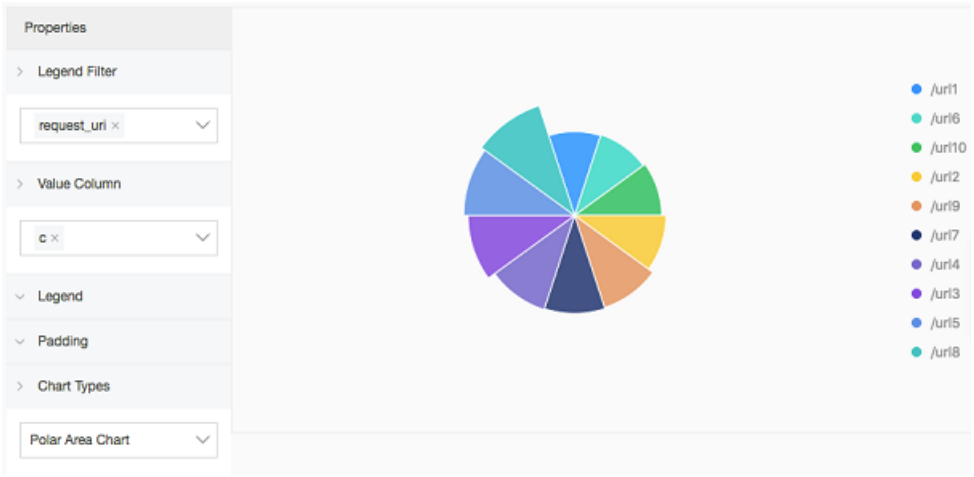
```
* | select request_method, count(1) as c group by
request_method order by c limit 10
```



- Nightingale rose diagram

  Analyze the ratio of the access `request_uri`:

```
* | select request_uri, count(1) as c group by request_uri
 order by c
```



22.8.1.8 Number chart

The number chart, as the easiest and most intuitive display type of data, clearly shows the data on a point, and is generally used to indicate the key information on a time point. Log Service number chart automatically normalizes the numeric values. For example, `230000` is processed as `230K`. You can define value formats as needed during real-time analysis. For more information, see *Mathematical functions*.

Basic components

- Main text
- (Optional) Unit
- (Optional) Description

Properties

Properties	Description
Value Column	By default, data in the first row of the specified column is displayed.
Color	The color in the number chart, including: • Font Color • Background Color
Text	The property configurations related to the text, including: • Font Size (12px–100px) • Unit • Unit Font Size (12px–100px) • Description • Description Font Size (12px–100px)

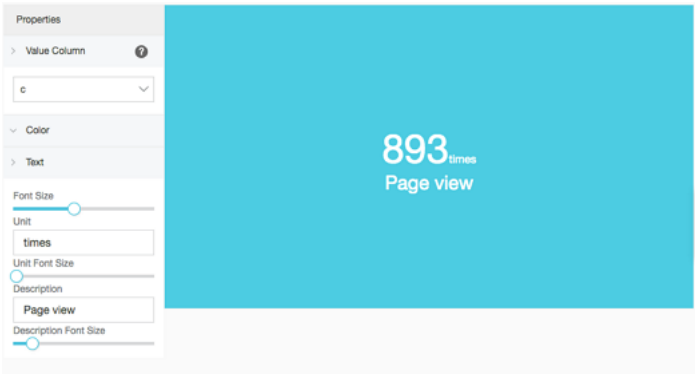
Procedure

1. *Log on to the Log Service Console*.
2. Choose **Project Name** > **Search** to go to the query page.
3. Enter a query statement, select a time range, and click **Search** on the right.
4. Select the number chart  .
5. Configure chart properties.

> **Note:**
> The number charts provided by Log Service automatically normalize values. For example, `230000` is processed as `230K`. You can define value formats as needed during real-time analysis. For more information, see *Mathematical functions*.

Execute the following query analysis statement to view the number of visits.

```
* | select count(1) as c
```



22.8.1.9 Area chart

An area chart is based on the line chart with colors filled in the section between the line and the coordinate axis. Each colored section is called an area. The colors help better outline the trend. Similar to a line chart, an area chart emphasizes the number changes over time, and is used to highlight the trend of the total number. Both the line chart and the area chart are mostly used to indicate the trend and relationship, instead of the specific values.

Basic components

- X axis (horizontal)
- Y axis (vertical)
- Area block

Properties

Properties	Description
X Axis	Generally, the X axis is an ordered data type (time series).
Y Axis	You can configure the mapping between one or more data columns and the left-axis value range.
Legend	You can configure the legend to the top, bottom , left, or right of the graph.
Padding	The distance between the coordinate axis and the graph boundary

Procedure

1. *Log on to the Log Service Console*.
2. Choose **Project Name** > **Search** to go to the query page.
3. Enter a query statement, select a time range, and click **Search** on the right.
4. Select the area chart .
5. Configure chart properties.

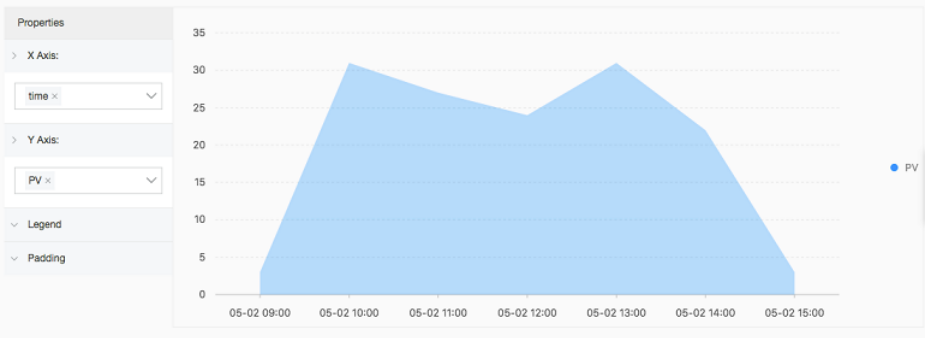
> **Note:**
> The number of data records for a single area block in the area chart must be greater than two in case that the data trend cannot be analyzed. We also recommend that you have no more than five area blocks in an area chart.

- Simple area chart

  Access details of IP address `42.0.192.0` within the last day:

  ```
  remote_addr: 42.0.192.0 | select date_format(date_trunc
  (hour, __time__), %m-%d %H:%i) as time, count(1) as PV
  group by time order by time limit 1000
  ```

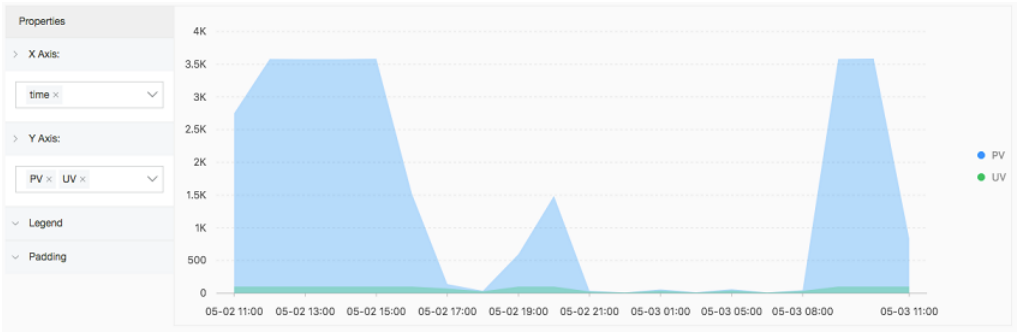
  Select `time` as the X Axis and `PV` as the Y Axis.

- Stacked area chart

  ```
  * | select date_format(date_trunc(hour, __time__), %m-%d %
  H:%i) as time, count(1) as PV, approx_distinct(remote_addr
  ) as UV group by time order by time limit 1000
  ```

  Select `time` as the X Axis and `PV` and `UV` as the Y Axis.



22.8.1.10 Flow chart

The flow chart, also known as ThemeRiver, is a stacked area chart around the central axis. The banded branches with different colors indicate different types of information. The band width indicates the corresponding numeric value. Besides, the centralized time attribute of the original data maps to the X axis, which forms a three-dimensional relationship.

A stream graph can be converted into a line chart or bar chart through the chart type function. A bar chart converted from a stream graph is in stacked mode by default. The start point of each data type is at the top of the last column.

Basic components

- X axis (horizontal)
- Y axis (vertical)
- Band

Properties

Properties	Description
X Axis	Generally, the X axis is an ordered data type (time series).
Y Axis	You can configure one or more columns of data to correspond to the value interval of the Y axis.
Aggregate Column	The information requires to be aggregated in the third dimension.
Legend	The location where the legend is in the graph. You can configure the legend to the top, bottom , left, or right of the graph.

Properties	Description
Padding	The distance between the coordinate axis and the graph boundary.
Chart Types	Area charts (default), line charts, and column charts (stacked) are available

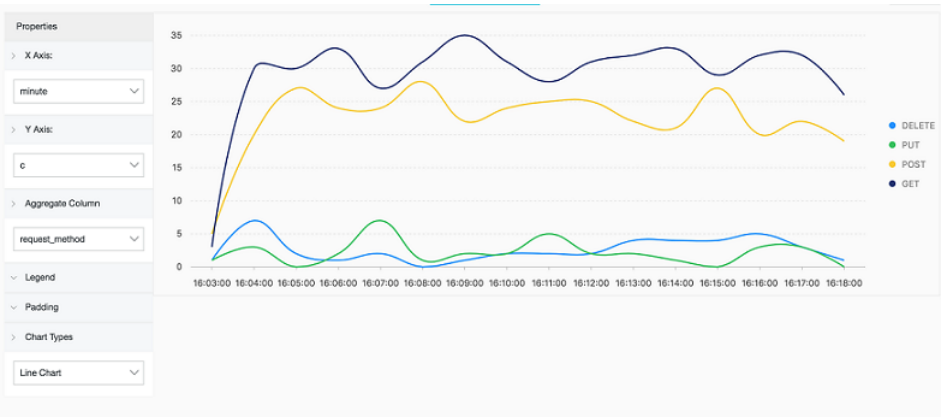
Procedure

1. *Log on to the Log Service Console*.
2. Choose **Project Name** > **Search** to go to the query page.
3. Enter a query statement, select a time range, and click **Search** on the right.
4. Select the stream graph  .
5. Configure chart properties.

   Stream graphs are suitable for displaying the three-dimensional relationship (time-type-value).

   ```
   * | select date_format(from_unixtime(__time__ - __time__% 60
   ), %H:%i:%S) as minute, count(1) as c, request_method group
   by minute, request_method order by minute asc limit 100000
   ```

   Select `minute` as the X Axis, `c` as the Y Axis, and `request_method` as the Aggregate Column.



22.8.1.11 Sankey diagram

Sankey diagram, a specific type of flow chart, is used to describe the flow from one set of values to another. This type of diagram is suitable for network traffic analysis and usually contains

three sets of values: `source`, `target`, and `value`. `source` and `target` describes the edge relationship between nodes, and `value` describes the relationship between `source` and `target`.

Features

A Sankey diagram has the following features:

- The overall width of all the main branches is the same as that of all the branches. This ensures that the inbound traffic volume equals the outbound traffic volume.
- In a Sankey diagram, different lines indicate the distribution of different flows, and the line width indicates the flow occupied by the branch proportionally.
- Different node widths indicate the flows in a particular state.

For example, the following data can be displayed in a Sankey diagram.

source	target	value
node1	node2	14
node1	node3	12
node3	node4	5
…	...	…

Basic components

- Node
- Edge

Properties

Properties	Description
Start Column	Describes the start node.
End Column	Describes the end node.
Value Column	The value that links the start node and the end node.
Padding	The distance between the coordinate axis and the graph boundary

Procedure

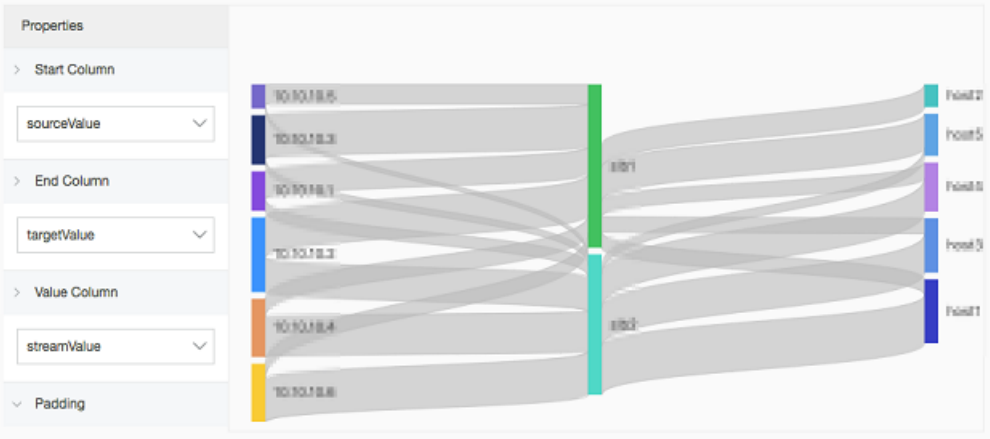
1. *Log on to the Log Service Console*.
2. Choose **Project Name** > **Search** to go to the query page.

3. Enter a query statement, select a time range, and click **Search** on the right.
4. Select the Sankey diagram .
5. Configure chart properties.

   Common Sankey diagram

   If the log field contains `source`, `target`, and `value`, each log itself is the relationship between nodes and edges, and the sum of `steamValue` can be obtained using *Nested subquery*.

```
* | select sourceValue, targetValue, sum(streamValue)
as streamValue from (select sourceValue, targetValue,
streamValue, __time__ from log group by sourceValue,
targetValue, streamValue, __time__ order by __time__ desc)
group by sourceValue, targetValue
```



22.8.1.12 Word cloud

Word cloud is the visual representation of text data, uses words to form a colorful graph similar to a cloud, and is used to display large amounts of text data. The font size or color is determined by the significance of the word, which allows you to perceive the weight of some keywords quickly.

Basic components

The word cloud shows you the words after being computed and sorted.

Properties

Properties	Description
Word column	The words to be displayed.
Value Column	The value corresponding to each word.

Properties	Description
Font Size	Adjust the font size range properly to apply to the canvas. • Maximum font size (50 px – 80 px) • Minimum font size (12 px – 24 px)
Padding	The distance between the coordinate axis and the graph boundary

Procedure

1. *Log on to the Log Service Console*.
2. Choose **Project Name** > **Search** to go to the query page.
3. Enter a query statement, select a time range, and click **Search** on the right.
4. Select the word cloud.
5. Configure chart properties.

   Analyze the distribution of the hostnames in the Nginx logs:

   ```
   * | select hostname, count(1) as count group by hostname
   order by count desc limit 1000
   ```

   Select `hostname` as the Word Column and `count` as the Value Column.

22.8.2 Interconnection with Grafana

Log Service is an end-to-end service for log data. It can be configured to automatically perform operations such as data collection, integration with various storage and computing systems, and data indexing and query. You only need to analyze the collected logs. In September 2017, Log

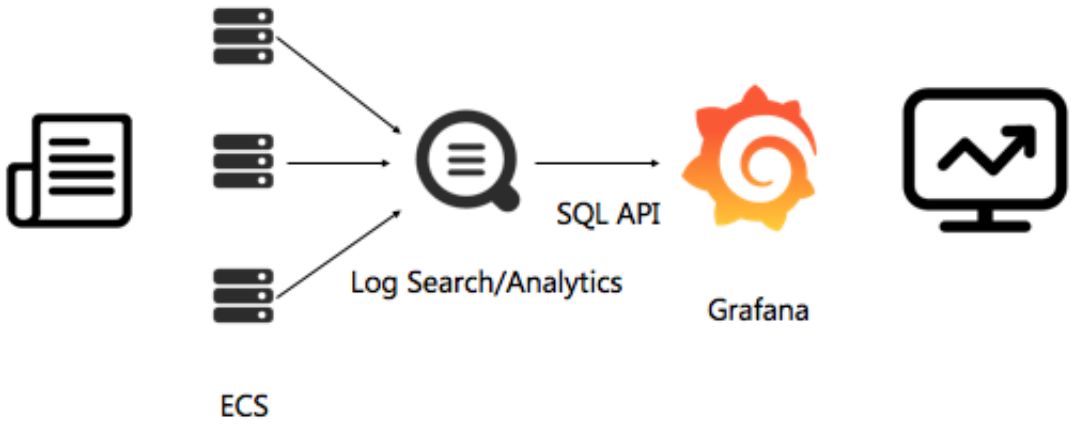
Service upgraded the LogSearch/Analytics function, which allows you to analyze logs in real time by search and SQL92 syntax.

Besides the built-in dashboard, Log Service supports the interconnection methods such as DataV , Grafana, Tableua, and QuickBI to achieve result analysis visualization. This topic shows how to use Grafana to analyze and visualize Nginx logs through Log Service.

Process structure

The process structure from log collection to analysis is as follows.

Figure 22-3: Process structure



Configuration process

1. Collect log data. For detailed instructions, see *Text log*.
2. Configure index setting and console query. For detailed instructions, see *Index and query*.
3. Install the Grafana plug-in and convert the real-time search results in the SQL database into a view.

After steps 1 and 2 are completed, you can view the collected logs on the query page.

This topic describes step 3.

Procedure

1. *Install Grafana*
2. *Install the Log Service plug-in*
3. *Configure the log data source*

4. *Add a dashboard*
    a. *Configure template variables*
    b. *Configure PV and UV*
    c. *Configure inbound and outbound bandwidth*
    d. *Analyze percentages of different HTTP methods*
    e. *Analyze percentages of different HTTP status codes*
    f. *Analyze hotspot source pages*
    g. *Analyze top latency pages*
    h. *Analyze hotspot pages*
    i. *Analyze top non-200 request pages*
    j. *Analyze the average latency of front-end and back-end*
    k. *Analyze statistics on the agent*
    l. *Save and release the dashboard*
5. *View the results*

Step 1 Install Grafana

For more information about how to install Grafana, see *Grafana official documents*.

For example, run the following command to install Ubuntu:

```
wget https://s3-us-west-2.amazonaws.com/grafana-releases/release/
grafana_4.5.2_amd64.deb sudo apt-get install -y adduser libfontconfig
sudo dpkg -i grafana_4.5.2_amd64.deb
```

If you need to use the pie chart, install the pie chart plug-in. For more information about how to install the plug-in, see *Grafana official documents*.

Run the following command:

```
grafana-cli plugins install grafana-piechart-panel
```

Step 2 Install the Log Service plug-in

Determine the location of Grafana plug-in directory. The plug-in path in Ubuntu is `/var/lib/grafana/plugins/`. Restart the Grafana server after the plug-in is installed.

For example, run the following commands to install the plug-in in Ubuntu and restart the Grafana server.

```
cd /var/lib/grafana/plugins/ git clone https://github.com/aliyun/
aliyun-log-grafana-datasource-plugin service grafana-server restart
```

Step 3 Configure the log data source

On your own computer, the log data source is installed on port 3000 by default. Enable port 3000 in the browser before the configuration.

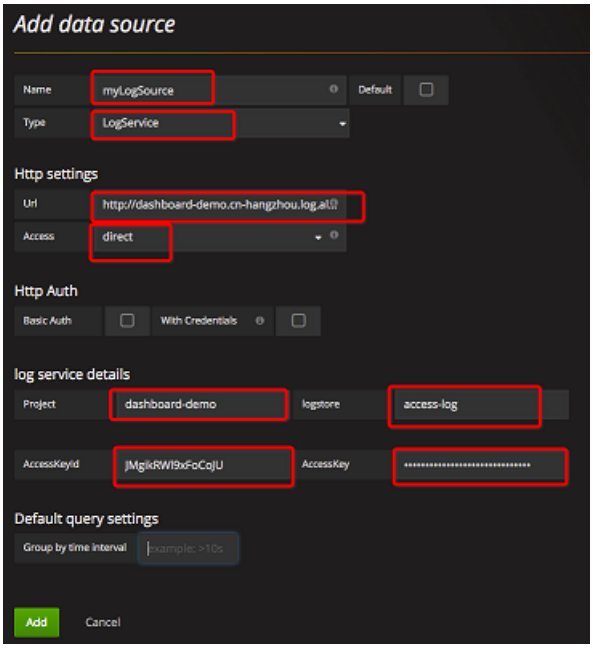
1. Click the Grafana logo in the upper-left corner and select **Data Sources** in the dialog box that appears.
2. Click **Add data source** in the upper-right corner to add a new data source. Use Grafana and Log Service for log visualized analysis.
3. Set the configuration items for the new data source.

   The configurations are as follows:

Configuration item	Content
datasource	Set the name for the new data source. Select **LogService** from the **Type** drop-down list.
Http settings	URL sample: `http://dashboard-demo.cn-hangzhou.log.aliyuncs.com`. `dashboard-demo` is the project name. `n-hangzhou.log.aliyuncs.com` is the endpoint of the region where the project is located. Use your own project and region address when configuring your own data source. Select **Direct** or **Proxy** for **Access**.
Http Auth	Retain the default values.
log service details	Detailed configurations of Log Service. Set **Project** and **Logstore**, and enter an AccessKey with the read permission. The AccessKey may belong to the primary account or sub-account.

   Sample configuration:

Figure 22-4: Sample configuration



Click **Add** to complete the data source creation.

Step 4: Add a dashboard

Click the menu in the upper-left corner, select **Dashboards** and click **New**. Add a dashboard to the menu in the upper-left corner.

4.1 Configure template variables

You can configure template variables in Grafana. In the same view, you can show different views by selecting the corresponding variables. This topic describes the configuration of time interval length and access of different domains.

1. Click the setting icon on the top of the page and click **Templating**.
2. When the configured template variables are displayed, click **New** to create a template.

   Set a time interval, of which the variable name is used in the configuration, for example, myinterval. Enter `$myinterval` in the query condition. Then the time interval you have selected is automatically used. Configure the template according to the following table.

Configuration item	Content
Name	Variable name, which can be set to myinterval .
Type	Select `Interval`.

Configuration item	Content
Label	Select `time interval`.
Internal Options	Enter `1m,10m,30m,1h,6h,12h,1d,7d,14d,30d` for the value.

3. Configure a domain name template.

   Multiple domain names can be attached to a VPS. You can view the access information of different domains. Enter `*,www.host.com,www.host0.com,www.host1.com` in the template, indicating that you can view all domains. You can also view the access information of `www.host.com`, `www.host0.com`, or `www.host1.com` respectively.

   The domain name template is configured as follows.

Configuration item	Content
Name	Variable name, which can be set to hostname.
Type	Select `Custom`.
Lable	Enter `Domain name`
Custom Options	Enter `*,www.host.com,www.host0.com,www.host1.com` for the value.

   You can view the configured template variables on the upper part of the dashboard page. Select a value from the drop-down list, such as time interval.

4.2 Configure PV and UV

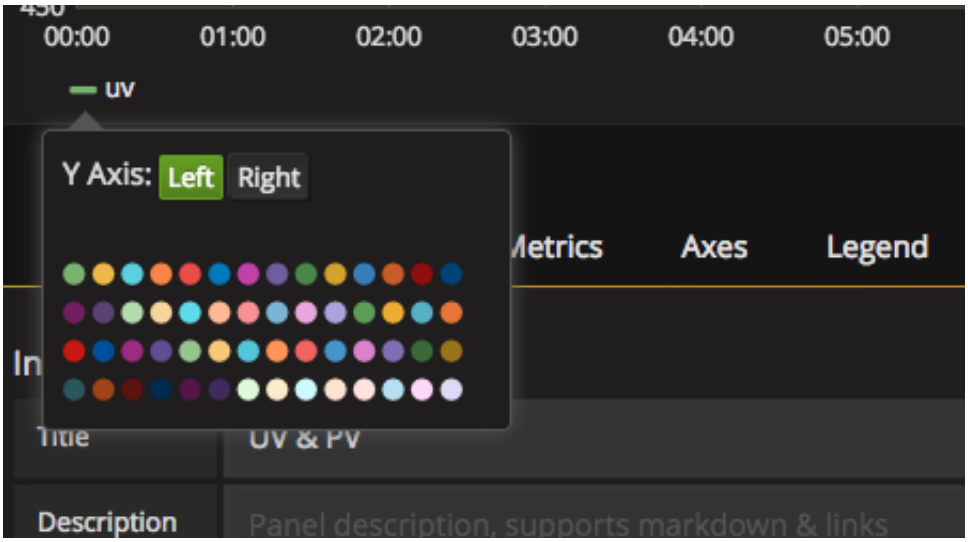
1. Click **Add Row** on the left to add a row. If a row already exists, select **Add Panel** from the pop-up menu on the left.
2. Grafana supports various views. For PV and UV data, click the **Graph** tab to create a graph view.
3. Click **Panel Title**. In the dialog box that appears, click **Edit.**
4. In Metrics configurations, select **logservice** from the **Data Source** drop-down list. Enter **Query**, **Y-column**, and **X-column**.
5. Select the configured `logservice` from the data source drop-down list.

Configuration item	Content
Query	`$hostname| select approx_distinct(remote_addr) as uv ,count(`

Configuration item	Content
	1) as pv , __time__ - __time__ % $$myinterval as time group by time order by time limit 1000 In the previous query, $hostname will be replaced with the domain name you have selected in display. $$myinterval will be replaced with the time interval. Note that there are two dollar signs ($) in front of myinterval, but only one in front of hostname.
X-column	time
Y-column	uv,pv

There is a great difference between UV and PV values, so they are displayed by two y axes. By clicking the colorful lines on the left of UV, you can show UV in the left or right y axis.

Figure 22-5: Y axis



A view needs a title, which is **Panel Title** by default. To modify the title, click the **General** tab and enter the new title in **Title**, for example, PV&UV.

4.3 Configure inbound and outbound bandwidth

You can add inbound and outbound bandwidth according to *4.2 Configure PV and UV*.

The major configuration items are as follows:

Configuration item	Content
Query	`$hostname | select sum(body_byte_sent) as net_out, sum(request_length) as net_in ,__time__ - __time__ % $$myinterval as time group by __time__ - __time__ % $$myinterval limit 10000`
X-column	Time
Y-column	net_in,net_out

4.4 Analyze percentages of different HTTP methods

You can configure the pie chart for the percentages of different HTTP methods according to *4.2 Configure PV and UV*.

Create a row, select **Pie Chart**, and enter **Query**, **Y-column**, and **X-column**.

The major configuration items are as follows:

Configuration item	Content
Query	`$hostname | select count(1) as pv ,method group by method`
X-column	pie
Y-column	method,pv

4.5 Analyze percentages of different HTTP status codes

You can configure the pie chart for the percentages of different HTTP status codes according to *4.2 Configure PV and UV*.

Create a row, and select **Pie Chart**.

The major configuration items are as follows:

Configuration item	Content
Query	`$hostname | select count(1) as pv ,status group by status`
X-column	pie
Y-column	status,pv

4.6 Analyze hotspot source pages

You can configure the pie chart for the hotspot source pages according to *4.2 Configure PV and UV*.

Create a row, and select **Pie Chart**.

The major configuration items are as follows:

Configuration item	Content
Query	`$hostname | select count(1) as pv , referer group by referer order by pv desc`
X-column	pie
Y-column	referer,pv

4.7 Analyze top latency pages

You can configure the view for the top latency pages according to *4.2 Configure PV and UV*.

If you want to show the URLs and their corresponding latencies in a table, select **Table**.

The major configuration items are as follows:

Configuration item	Content
Query	`$hostname | select url as top_latency_url ,request_time order by request_time desc limit 10`
X-column	Null
Y-column	top_latency_url,request_time

4.8 Analyze hotspot pages

You can configure the view for hotspot pages according to *4.2 Configure PV and UV*.

Create a table view. Select **logservice** from the **Data Source** drop-down list. Enter **Query**, **Y-column**, and **X-column**, as shown in the following table.

Configuration item	Content
Query	`$hostname | select count(1) as pv, split_part(url,?,1) as path`

Configuration item	Content
	group by split_part(url,?,1) order by pv desc limit 20
X-column	Null
Y-column	path,pv

4.9 Analyze top non-200 request pages

You can configure the view for the top non-200 request pages according to *4.2 Configure PV and UV*.

Create a table view. Select logservice from the Data Source drop-down list. Enter Query, Y-column, and X-column, as shown in the following table.

Configuration item	Content
Query	$hostname not status:200| select count(1) as pv , url group by url order by pv desc
X-column	Null
Y-column	url,pv

4.10 Analyze the average latency of front-end and back-end

You can configure the view for the average latency of front-end and back-end according to *4.2 Configure PV and UV*.

Create a graph view. Select logservice from the Data Source drop-down list. Enter Query, Y-column, and X-column, as shown in the following table.

Configuration item	Content
Query	$hostname | select avg(request_time) as response_time, avg(upstream_response_time) as upstream_response_time ,__time__ - __time__ % $$myinterval as time group by __time__ - __time__ % $$myinterval limit 10000
X-column	time
Y-column	upstream_response_time,response_time

4.11 Analyze statistics on the agent

You can configure the view for agent statistics according to *4.2 Configure PV and UV*.

Create a pie chart. Select logservice from the Data Source drop-down list. Enter Query, Y-column , and X-column, as shown in the following table.

Configuration item	Content
Query	`$hostname | select count(1) as pv, case when http_user_agent like %Android% then Android when http_user_agent like %iPhone% then iOS else unKnown end as http_user_agent group by case when http_user_agent like %Android% then Android when http_user_agent like %iPhone% then iOS else unKnown end order by pv desc limit 10`
X-column	pie
Y-column	http_user_agent,pv

4.12 Save and release the dashboard

Click the save button on the top to release the dashboard.

Step 5 View the results.

Open the dashboard homepage to view the effect. For the example, see *Demo*.

Set the time range for statistics on the upper part. You can also set the time interval or domains.

The dashboard for Nginx access statistics is completed, and you can find valuable information from the view.

22.9 Alerts and notifications

22.9.1 Configure alarming

Log Service can report alarms based on your **log query results**. You can configure alarm policies to receive alarms in custom WebHook notification methods.

Basic process

1. Configure Savedsearch
2. Configure the alarm rules.
3. Configure notification type.

4. View alarm records

Step 1 Configure the quick query

1. *Log on to the Log Service Console*.
2. Select the target project and click the project name.
3. On the **Logstores** page, select the required Logstore and click **Search** in the log search column.
4. After specifying the Logstore, topic, and query statement as needed, search for the specified log.
5. In the upper right corner of the page, click **Saved to Savedsearch**. Save the query statement as Savedsearch.
6. Set details of the Saved Search and click **OK**.
   - **Operation:** Select **Create Saved Search**.
   - **Saved Search Name:** Name of the Saved Search.

Step 2 Create alarm policies

After saving query parameters as a quick query, you can create alarm policies.

1. Click **Save as Alarm**.
2. Set alarm policies and click **OK**.

Policy description

- **Saved Search Name**: Currently the created Saved Search can be selected.
- **Time Range**: Time range for the data to be read when the server runs an alarm check each time. For example, 1 minute means a query on the data in the last 1 minute till the current time.

  Note:
  Currently, the server only performs sampling inspection on the first 10 data records in the time range when it checks alert policies each time.
- **Check Interval**: Time interval for the server to run an alarm check each time (currently, the minimum interval is 5 minutes).
- **Number of Triggers**: Number of consecutive checks for triggering an alarm. For example, if the interval is 5 minutes, then 2 means that an alarm is sent when two consecutive checks meet the alarm conditions (the minimum interval for generating an alarm is 10 minutes).
- **Key**: Key name used for an alarm in the log content.

- **Operator**: Supports the value type (Greater Than, Greater Than or Equal To, Less Than, and Less Than or Equal To) and character type (Include and RegEx). See the following table:

Operation	Description	Example
>	Whether the content of the column is greater than the value	$count > 0
<	Whether the content of the column is less than the value	$count<200
>=	Greater than or equal to a value	$count>=0
<=	Less than or equal to a value	$count<=0
like	Matching substring	$project like “admin”
regex	Regular expression matching string	$project regex match “^/S+$”

- **Threshold**: Alarm threshold. Specify the threshold according to the operator.

Step 3 Configure a notification method

Currently, Log Service supports the custom WebHook notification methods. When the configured alarm policies are triggered, Log Service sends you alarms using the predefined notification method.

1. In the **Action Type** area under **Action**, select **WebHook – Custom** to add the WebHook link to the WebHook address, and specify the notification content (in English only, up to 50 characters).
2. After an alarm is triggered, content like the following is sent to the WebHook address in Post mode:

```
{"uid": "13415134513","project":"ali-cn","trigger":"oplog_alert","
condition":"3413 > 3000", "message":"PV count down 30%", "context:"c
:3413"}
```

Step 4 View alarm records

You can view the specific alarm results after creating alarm policies.

1. Log on to the Log Service console. For details about how to log on to the Log Service console, see Log on to Apsara Stack console in *Cite LeftApsara Stack Management Console User GuideCite Right*.

2. Select the target project and click the project name.
3. On the **Logstores** page, choose **LogSearch/Analytics - Query** > **Alarm** on the navigation menu on the left.
4. Select an alarm policy and click **View** on the right to view specific alert results.

Alarm status:

- **Success**: The policy has been implemented successfully, and the trigger standard is displayed in the alarm trigger details.
- **Failed**: In case of a failure during query, alarm policy implementation, or notification, you can check **Trigger Details** for more information.
  - For failed query, check the syntax.
  - For failed query call, open a ticket.
  - For failed policy call, check whether the policy parameters are consistent with the returned data format.

22.10 Real-time subscription and consumption

Logs collected to LogHub of Log Service can be consumed by using the following three methods:

Method	Scenario	Real-time support	Storage duration
Real-time consumption (LogHub)	Stream computing and real-time computing	Real time (<10ms)	365 days
Index query (LogSearch)	Applicable to online query of recent hot data	Real time (1s in 99.9% cases, 3s at the longest)	365 days
Post storage (LogShipper)	Applicable to full log storage for offline analysis	5–30 minutes	Dependent on the storage system

Real-time consumption

Consumption process

After logs are written, the most common operation is to consume the logs. (Log consumption and query are essentially reading logs.) The logs in a shard are consumed as follows:

1. Obtain a cursor based on a set of criteria such as time, Begin, and End.
2. The system reads logs based on the cursor and step and returns the next cursor.
3. Move the cursor continuously to consume logs.

Consumption method

In addition to basic APIs, Log Service provides the SDKs, Storm spout, Spark client, and Web Console used for log consumption.

- Use *Spark Streaming client* to consume logs.
- Use *Storm spout* to consume logs.
- Use *Flink Connector* to consume logs: including Flink Consumer and Producer.
- Use *LogHub Consumer Library* to consume logs: LogHub Consumer Library is an advanced mode provided to LogHub consumers. It provides a lightweight computing framework to implement automatic shard allocation and sequence guarantee when multiple LogHub consumers consume LogStores simultaneously.
- Use SDKs to consume logs: Log Service provides SDKs in several languages (Java and Python) with support for log consumption APIs. For more information about the SDKs, see *Cite LeftLog Service SDKsCite Right*.
- Use cloud products to consume logs:
  - *Use StreamCompute to consume logs*: Custom monitoring scenarios.
  - Use E-MapReduce to consume logs: See *Storm*.

LogSearch/Analytics

See *Index and query*:

- Query logs in the Log Service console: See *Index and query*.
- Query logs using the SDKs or APIs of Log Service: Log Service provides HTTP-enabled RESTful APIs. The APIs support full featured log query. For more information, see *Cite LeftLog Service API ReferenceCite Right*.

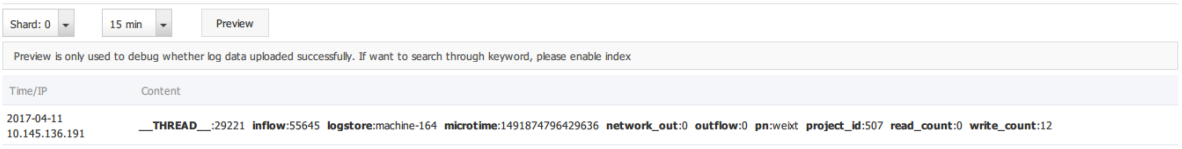
22.10.1 Regular consumption

Log preview is a regular type of log consumption. The Log Service console provides a dedicated preview page to help you preview a portion of the logs in the LogStore directly in your browser.

Procedure

1. *Log on to the Log Service Console*.
2. Click the name of a project to go to the **Logstores** page.
3. On the **Logstores** page, select a Logstore and click **Preview**.
4. On the **logstore** page that appears, select a shard, specify the time range, and click **Preview**.

   The **logstore** page displays the log data of the first 10 packets in the specified time range.



22.10.2 Consumption by consumer groups

The consumer library is an advanced method to consume logs in Log Service. It provides the consumer group feature to abstract and manage consumers. Different from SDKs for data reading , the consumer group allows you to focus on the service logic without paying attention to Log Service implementation details and the load balancing and failover between consumers.

Basic concepts

Before using the consumer library, get to know the consumer group and consumer.

- Consumer group

  A consumer group consists of multiple consumers. The consumers in the same consumer group consume data in the same Logstore. The data consumed by consumers is not repeated.
- Consumer

  Consumers form a consumer group and consume data. Consumers in the same consumer group must have different names.

In Log Service, a Logstore has several shards. The consumer library allocates shards to the consumers in a consumer group based on the following principles:

- Each shard is allocated to only one consumer.
- A consumer may have multiple shards.

After a consumer joins a consumer group, the shard subordination in this group is adjusted for load balancing, but the preceding principles remain unchanged. The allocation process is transparent.

The consumer library also saves checkpoints to enable consumption to resume from the breakpoint after programs recover. This avoids repeated data consumption.

Instructions for use

Maven dependency

```
<dependency> <groupId>com.google.protobuf</groupId> <artifactId>
protobuf-java</artifactId> <version>2.5.0</version> </dependency> <
dependency> <groupId>com.aliyun.openservices</groupId> <artifactId
>aliyun-log</artifactId> <version>0.6.11</version> </dependency> <
```

```
dependency> <groupId>com.aliyun.openservices</groupId> <artifactId>
loghub-client-lib</artifactId> <version>0.6.15</version> </dependency>
```

main .java file

```
public class Main { // Domain name of Log Service. Set it according to
 actual information. private static String sEndpoint = "cn-hangzhou.
log.aliyuncs.com"; // Name of a Log Service project. Set it according
 to actual information. private static String sProject = "ali-cn
-hangzhou-sls-admin"; // Name of a Logstore. Set it according to
actual information. private static String sLogstore = "sls_operat
ion_log"; // Name of a consumer group. Set it according to actual
information. private static String sConsumerGroup = "consumerGroupX
"; // AccessKey for data consumption. Set it according to actual
information. private static String sAccessKeyId = ""; private static
String sAccessKey = ""; public static void main(String []args) throws
 LogHubClientWorkerException, InterruptedException { // The consumer
 names (the second parameter) in the same consumer group must be
different. The names of consumer groups may be the same. Different
 consumer names are used to start multiple processes on several
machines to consume data in the same Logstore in the load balancing
manner. The consumer group names can be differentiated by the machine
 IP addresses. maxFetchLogGroupSize (the ninth parameter) indicates
the number of log groups retrieved from the server at a time. You can
keep the default value, or adjust it in the range (0,1000] as needed.
LogHubConfig config = new LogHubConfig(sConsumerGroup, "consumer_1",
sEndpoint, sProject, sLogstore, sAccessKeyId, sAccessKey, LogHubConf
ig.ConsumePosition.BEGIN_CURSOR); ClientWorker worker = new ClientWork
er(new SampleLogHubProcessorFactory(), config); Thread thread = new
Thread(worker); //The ClientWorker instance runs automatically after
 the thread is executed and extends the Runnable interface. thread
.start(); Thread.sleep(60 * 60 * 1000); //The shutdown function of
the ClientWorker instance is called to exit the consumption instance
. The associated thread is stopped automatically. worker.shutdown
(); //Multiple asynchronous tasks are generated when the ClientWorker
 instance is running. You are advised to wait 30s until all tasks are
exited after shutdown. Thread.sleep(30 * 1000); } }
```

SampleLogHubProcessor.java file

```
public class SampleLogHubProcessor implements ILogHubProcessor {
private int mShardId; // Records the last persistent checkpoint
time. private long mLastCheckTime = 0; public void initialize(int
shardId) { mShardId = shardId; } // Master logic of data consumptio
n. All the exceptions must be captured and cannot be thrown. public
 String process(List<LogGroupData> logGroups, ILogHubCheckPointTra
cker checkPointTracker) { // Print the retrieved data simply. for
(LogGroupData logGroup: logGroups){ FastLogGroup flg = logGroup.
GetFastLogGroup(); System.out.println(String.format("\tcategory\t:
\t%s\n\tsource\t:\t%s\n\ttopic\t:\t%s\n\tmachineUUID\t:\t%s", flg.
getCategory(), flg.getSource(), flg.getTopic(), flg.getMachineUUID
())); System.out.println("Tags"); for (int tagIdx = 0; tagIdx < flg
.getLogTagsCount(); ++tagIdx) { FastLogTag logtag = flg.getLogTags
(tagIdx); System.out.println(String.format("\t%s\t:\t%s", logtag
.getKey(), logtag.getValue())); } for (int lIdx = 0; lIdx < flg.
getLogsCount(); ++lIdx) { FastLog log = flg.getLogs(lIdx); System.
out.println("--------\nLog: " + lIdx + ", time: " + log.getTime() +
 ", GetContentCount: " + log.getContentsCount()); for (int cIdx = 0
; cIdx < log.getContentsCount(); ++cIdx) { FastLogContent content =
log.getContents(cIdx); System.out.println(content.getKey() + "\t:\t"
```

```
 + content.getValue()); } } } long curTime = System.currentTimeMillis
(); // Writes checkpoints to the server every 30s. If a ClientWork
er instance crashes during the 30s period, // the new ClientWorker
instance consumes data starting from the last checkpoint. Duplicate
data may exist. if (curTime - mLastCheckTime > 30 * 1000) { try { //
When the parameter is set to TRUE, checkpoints are updated to the
server immediately; when the parameter is set to FALSE, checkpoints
 are cached locally. The default update interval is 60s. checkPoint
Tracker.saveCheckPoint(true); } catch (LogHubCheckPointException e
) { e.printStackTrace(); } mLastCheckTime = curTime; } return null
; } // The ClientWorker instance calls this function upon exit,
during which you can perform cleanup. public void shutdown(ILogHubChe
ckPointTracker checkPointTracker) { //Save the consumption breakpoint
 to Log Service. try { checkPointTracker.saveCheckPoint(true); }
catch (LogHubCheckPointException e) { e.printStackTrace(); } } }
class SampleLogHubProcessorFactory implements ILogHubProcessorFact
ory { public ILogHubProcessor generatorProcessor() { // Generate a
consumption instance. return new SampleLogHubProcessor(); } }
```

Run the preceding code to print all the data in a Logstore. If you allow multiple consumers to consume the same Logstore, you can modify the program based on the comment to add different consumer names by using the same consumer group name, and start another consumption process.

Constraints and exception diagnostic

Up to 10 consumer groups can be created for each Logstore. The error `ConsumerGroupQuotaExceed` is returned when this limit is exceeded.

We recommend that you configure Log4j for the consumer program to throw error messages in consumer groups for troubleshooting. If you save the log4j.properties file to the resources directory and execute the program, the following error message is displayed:

```
[WARN ] 2018-03-14 12:01:52,747 method:com.aliyun.openservices.loghub
.client.LogHubConsumer.sampleLogError(LogHubConsumer.java:159) com.
aliyun.openservices.log.exception.LogException: Invalid loggroup count
, (0,1000]
```

You can refer to a simple example of log4j.properties configuration:

```
log4j.rootLogger = info,stdout log4j.appender.stdout = org.apache.
log4j.ConsoleAppender log4j.appender.stdout.Target = System.out log4j.
appender.stdout.layout = org.apache.log4j.PatternLayout log4j.appender
.stdout.layout.ConversionPattern = [%-5p] %d{yyyy-MM-dd HH:mm:ss,SSS}
method:%l%n%m%n
```

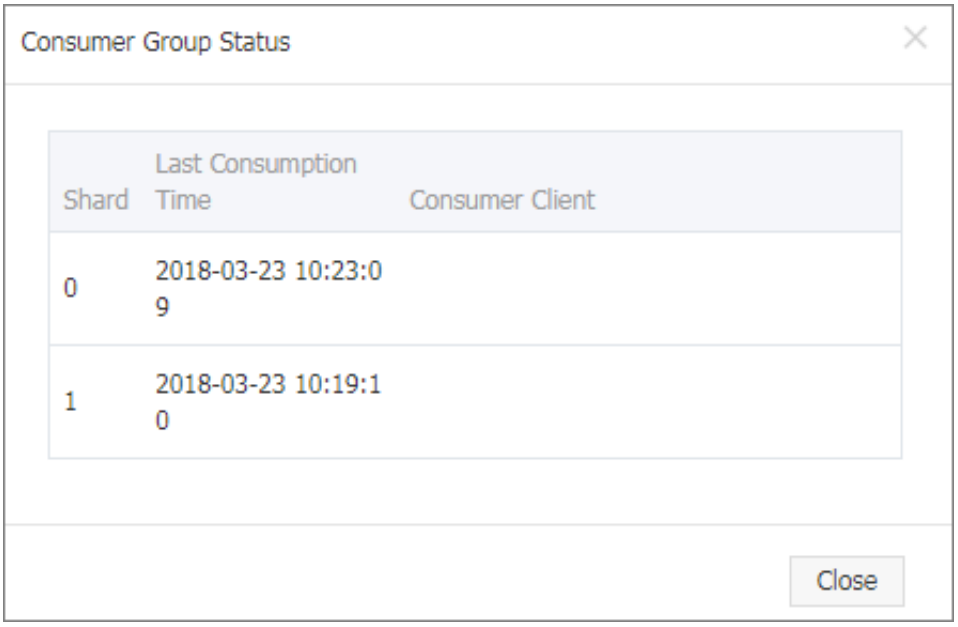
22.10.3 Consumer group status

Consumption by consumer groups is an advanced real-time data consumption mode. It provides automatic Logstore consumption load balancing for multiple consumer instances. Both Spark Streaming and Storm use ConsumerGroup as their basic mode.

View consumption progress in the Log Service console

1. *Log on to the Log Service Console*.
2. Click the name of a project.
3. In the left-side navigation pane, click **LogHub - Consume** > **Consumer Group**.
4. On the **Consumer Group** page, select a Logstore to check whether the collaborative consumption function is enabled.

5. Select a consumer group, and click **Consumption Status** to view the progress of data consumption for each shard.



.

As shown in the figure above, the page displays four shards of the Logstore, corresponding to four consumers. The most recent data consumption time is shown for each consumer in the second column. With data consumption time, you can determine if the current data processing can keep up with data production. If processing seriously lags behind (i.e. data consumption is slower than data production), you should consider increasing the number of consumers.

Use APIs and SDKs to view the consumption progress

Here, we will use the Java SDK as an example to show how to get consumption status using API.

```
package test; import java.util.ArrayList; import com.aliyun.openservic
es.log.Client; import com.aliyun.openservices.log.common.Consts.
CursorMode; import com.aliyun.openservices.log.common.ConsumerGroup;
import com.aliyun.openservices.log.common.ConsumerGroupShardCheckPoint
; import com.aliyun.openservices.log.exception.LogException; public
class ConsumerGroupTest { static String endpoint = ""; static String
 project = ""; static String logstore = ""; static String accesskeyI
d = ""; static String accesskey = ""; public static void main(String
[] args) throws LogException { Client client = new Client(endpoint,
accesskeyId, accesskey); //Retrieve all consumer groups under this
Logstore. If no consumer groups exist, the consumerGroups length is
 0. ArrayList<ConsumerGroup> consumerGroups; try{ consumerGroups =
client.ListConsumerGroup(project, logstore).GetConsumerGroups(); }
catch(LogException e){ if(e.GetErrorCode() == "LogStoreNotExist")
System.out.println("this logstore does not have any consumer group
"); else{ //internal server error branch } return; } for(ConsumerGr
oup c: consumerGroups){ ///Print consumer group attributes, including
 names, heartbeat timeout, and consumption order. System.out.println
("Name: " + c.getConsumerGroupName()); System.out.println("Heartbeat
 timeout: " + c.getTimeout()); System.out.println("Consumption order
: " + c.isInOrder()); for(ConsumerGroupShardCheckPoint cp: client.
GetCheckPoint(project, logstore, c.getConsumerGroupName()).GetCheckPo
ints()){ System.out.println("shard: " + cp.getShard()); //Reformat
 the returned time to be precise to milliseconds in long integer.
System.out.println("Last data consumption time: " + cp.getUpdateTime
()); System.out.println("Consumer name: " + cp.getConsumer()); String
 consumerPrg = ""; if(cp.getCheckPoint().isEmpty()) consumerPrg = "
Consumption not started"; else{ //Unix timestamp, in seconds; note
 the format during output. try{ int prg = client.GetPrevCursorTime(
project, logstore, cp.getShard(), cp.getCheckPoint()).GetCursorTime
(); consumerPrg = "" + prg; } catch(LogException e){ if(e.GetErrorCode
() == "InvalidCursor") consumerPrg = "Invalid, the previous consumptio
n time has exceeded the data lifecycle in the Logstore."; else{ //
internal server error throw e; } } } System.out.println("Consumptio
n progress: " + consumerPrg); String endCursor = client.GetCursor(
project, logstore, cp.getShard(), CursorMode.END).GetCursor(); int
endPrg = 0; try{ endPrg = client.GetPrevCursorTime(project, logstore
, cp.getShard(), endCursor).GetCursorTime(); } catch(LogException e)
{ //do nothing } //Unix timestamp, in seconds; note the format during
 output. System.out.println("Arrival time of the last piece of data: "
 + endPrg); } } } }
```

22.10.4 Use Flink to consume LogHub logs

The Flink log connector is a tool provided by Alibaba Cloud Log Service and used to connect to Flink. It consists of the consumer and producer.

The consumer reads data from Log Service. It supports the exactly-once syntax and shard-based load balancing.

The producer writes data into Log Service. When using the connector, you must add the Maven dependency to the project:

```
<dependency> <groupId>org.apache.flink</groupId> <artifactId>flink-
streaming-java_2.11</artifactId> <version>1.3.2</version> </dependency
> <dependency> <groupId>com.aliyun.openservices</groupId> <artifactId>
flink-log-connector</artifactId> <version>0.1.7</version> </dependency
> <dependency> <groupId>com.google.protobuf</groupId> <artifactId>
protobuf-java</artifactId> <version>2.5.0</version> </dependency> <
dependency> <groupId>com.aliyun.openservices</groupId> <artifactId
>aliyun-log</artifactId> <version>0.6.10</version> </dependency> <
dependency> <groupId>com.aliyun.openservices</groupId> <artifactId>log
-loghub-producer</artifactId> <version>0.1.8</version> </dependency>
```

Prerequisite

1. *Log on to the Log Service Console*.
2. There is a pair of AccessKeys, and you have created a project and a Logstore. For more information about how to create a project and a Logstore, see *Preparation*.

Log consumer

In the connector, the Flink log consumer provides the capability of subscribing to a specific Logstore in Log Service to achieve the exactly-once syntax. During use, you do not need to concern about the change of the number of shards in the Logstore. The consumer automatically senses the change.

Each sub-task in Flink consumes some shards in the LogStore. If shards in the LogStore are split or merged, shards consumed by the sub-task change accordingly.

Associated APIs

The Flink log consumer uses the following Alibaba Cloud Log Service APIs:

- GetCursorOrData

  This API is used to pull data from a shard. If this API is frequently called, data volume may exceed the shard quota of Log Service. You can use ConfigConstants.LOG_FETCH_DATA_INTERVAL_MILLIS and ConfigConstants.LOG_MAX_NUMBER_PER_FETCH to control the interval of API calls and number of logs pulled by each call. For details about the shard quota, see *Operate on shards*.

  ```
  configProps.put(ConfigConstants.LOG_FETCH_DATA_INTERVAL_MILLIS, "100
  "); configProps.put(ConfigConstants.LOG_MAX_NUMBER_PER_FETCH, "100
  ");
  ```

- ListShards

This API is used to obtain all shard lists and shard status in a LogStore. If your shards are always split and merged, you can adjust the API call cycle to locate shard changes in time.

```
// Call ListShards every 30s configProps.put(ConfigConstants.
LOG_SHARDS_DISCOVERY_INTERVAL_MILLIS, "30000");
```

- CreateConsumerGroup

  This API is called only when consumption progress monitoring is enabled. It is used to create a consumer group to synchronize checkpoints.

- ConsumerGroupUpdateCheckPoint

  This API is used to synchronize snapshots of Flink to a consumer group of Log Service.

Procedure

1. Configure the startup parameters.

```
Properties configProps = new Properties(); // Set the domain name used
 to access Log Service. configProps.put(ConfigConstants.LOG_ENDPOI
NT, "cn-hangzhou.log.aliyuncs.com"); // Set the AccessKey used
to access Log Service. configProps.put(ConfigConstants.LOG_ACCESS
SKEYID, ""); configProps.put(ConfigConstants.LOG_ACCESSKEY, ""); //
Configure Log Service project. configProps.put(ConfigConstants.
LOG_PROJECT, "ali-cn-hangzhou-sls-admin"); // Configure the Log
Service LogStore. configProps.put(ConfigConstants.LOG_LOGSTORE,  "
sls_consumergroup_log"); // Set the start position to consume Log
Service. configProps.put(ConfigConstants.LOG_CONSUMER_BEGIN_POSITION,
Consts.LOG_END_CURSOR); // Set the message deserialization method for
Log Service. RawLogGroupListDeserializer deserializer = new RawLogGrou
pListDeserializer(); final StreamExecutionEnvironment env = StreamExec
utionEnvironment.getExecutionEnvironment(); DataStream<RawLogGroupList
> logTestStream = env.addSource( new FlinkLogConsumer<RawLogGroupList
>(deserializer, configProps));
```

The preceding is a simple consumption example. As java.util.Properties is used as the configuration tool, configurations of all consumers can be located in ConfigConstants.

Note:

The number of sub-tasks in the Flink stream is independent of that of shards in the Log Service LogStore. If the number of shards is greater than that of sub-tasks, each sub-task consumes multiple shards only once. If the number of shards is smaller than that of sub-tasks, some sub-tasks are idle until new shards are generated.

2. Set the consumption start position.

You can set the start position for consuming a shard on the Flink log consumer. By setting ConfigConstants.LOG_CONSUMER_BEGIN_POSITION, you can set whether to consume a

shard from its header or tail or at a specific time. The connector also supports consumption restoration from a specific consumer group. The values are as follows:

- Consts.LOG_BEGIN_CURSOR: Indicates that the shard is consumed from its header, that is, from the earliest data of the shard.
- Consts.LOG_END_CURSOR: Indicates that the shard is consumed from its tail, that is, from the latest data of the shard.
- Consts.LOG_FROM_CHECKPOINT: Indicates that the shard is consumed from the saved checkpoint in a specific consumer group. The consumer group is specified by ConfigConstants .LOG_CONSUMERGROUP.
- UnixTimestamp: A string of an integer value, which is expressed in seconds from1970-01-01. It indicates that the shard is consumed from this time point.

Examples of the preceding four values are as follows:

```
configProps.put(ConfigConstants.LOG_CONSUMER_BEGIN_POSITION, Consts
.LOG_BEGIN_CURSOR); configProps.put(ConfigConstants.LOG_CONSUM
ER_BEGIN_POSITION, Consts.LOG_END_CURSOR); configProps.put(ConfigCons
tants.LOG_CONSUMER_BEGIN_POSITION, "1512439000"); configProps.put
(ConfigConstants.LOG_CONSUMER_BEGIN_POSITION, Consts.LOG_FROM_C
HECKPOINT);
```

Note:

If you have configured consumption restoration from the StateBackend of Flink when you start the Flink task, the connector ignores the preceding configurations and uses the checkpoint saved in StateBackend.

3. Configure consumption progress monitoring (optional).

The Flink log consumer supports consumption progress monitoring. The consumption progress indicates the real-time consumption position of each shard, which is expressed using the timestamp.

```
configProps.put(ConfigConstants.LOG_CONSUMERGROUP, "your consumer
group name");
```

Note:

The preceding code is optional. If it is set, the consumer creates a consumer group first. If the consumer group already exists, no further operation is required. Snapshots in the consumer are automatically synchronized to the consumer group of Log Service. You can check the consumption progress of the consumer on the Log Service console.

4. Configure support for disaster tolerance and exactly-once syntax.

If the checkpoint function of Flink is enabled, the Flink log consumer periodically stores the consumption progress of each shard. When a job fails, Flink restores the log consumer and starts consumption from the latest checkpoint that is stored.

The period of writing checkpoint defines the maximum amount of data to be rolled back (that is, re-consumed) if a failure occurs. The code is as follows:

```
final StreamExecutionEnvironment env = StreamExecutionEnvironment.
getExecutionEnvironment(); // Enable the exactly once syntax of Flink
. env.getCheckpointConfig().setCheckpointingMode(CheckpointingMode
.EXACTLY_ONCE); // Store the checkpoint every five seconds. env.
enableCheckpointing(5000);
```

For details about the Flink checkpoints, see the Flink documentation *Checkpoints*.

Log producer

The Flink log producer writes data to Alibaba Cloud Log Service.

Note:
The producer supports only the Flink at-least-once syntax. It means that when a job failure occurs, data written into Log Service may be duplicated but never lost.

Procedure

1. Initialize the producer.
   a. Initialize properties for the producer.

      The initialization process for the producer is similar to that for the consumer. The producer contains the following parameters. Set these parameters to the default values in general conditions or to custom values as required.

      ```
      // Number of I/O threads used for sending data. The default value
       is 8. ConfigConstants.LOG_SENDER_IO_THREAD_COUNT // Time when
      the log data is cached before being sent. The default value is
      3000. ConfigConstants.LOG_PACKAGE_TIMEOUT_MILLIS // Number logs
      in the cached package. The default value is 4096. ConfigConstants
      .LOG_LOGS_COUNT_PER_PACKAGE // Size of the cached package. The
      default value is 3 MB. ConfigConstants.LOG_LOGS_BYTES_PER_PACKAGE
       // Total memory size that the job can use. The default value is
      100 MB. ConfigConstants.LOG_MEM_POOL_BYTES
      ```

      The preceding parameters are not mandatory. You can retain the default values.

   b. Reload LogSerializationSchema to define the method for serializing data to RawLogGroup.

RawLogGroup is a collection of logs. For the meaning of each field, see **Data model** in *Cite LeftLog Service Development GuideCite Right*.

To use the shardHashKey function of Log Service, specify the shard to which data is written.

You can use **LogPartitioner** to generate the HashKey of data.

Example:

```
FlinkLogProducer<String> logProducer = new FlinkLogProducer<String
>(new SimpleLogSerializer(), configProps); logProducer.setCustomP
artitioner(new LogPartitioner<String>() { // Generate a 32-bit
 hash value. public String getHashKey(String element) { try {
MessageDigest md = MessageDigest.getInstance("MD5"); md.update(
element.getBytes()); String hash = new BigInteger(1, md.digest()).
toString(16); while(hash.length() < 32) hash = "0" + hash; return
 hash; } catch (NoSuchAlgorithmException e) { } return "0000000000
00000000000000000000000000000000000000000000000000000000"; } });
```

Note:

LogPartitioner is optional. If this parameter is not set, data is randomly written to a shard.

2. In the following example, a simulated string is written to Log Service:

```
// Serialize data to the data format of Log Service. class
SimpleLogSerializer implements LogSerializationSchema<String> {
public RawLogGroup serialize(String element) { RawLogGroup rlg =
new RawLogGroup(); RawLog rl = new RawLog(); rl.setTime((int)(System
.currentTimeMillis() / 1000)); rl.addContent("message", element);
rlg.addLog(rl); return rlg; } } public class ProducerSample { public
 static String sEndpoint = "cn-hangzhou.log.aliyuncs.com"; public
 static String sAccessKeyId = ""; public static String sAccessKey
 = ""; public static String sProject = "ali-cn-hangzhou-sls-admin
"; public static String sLogstore = "test-flink-producer"; private
 static final Logger LOG = LoggerFactory.getLogger(ConsumerSample
.class); public static void main(String[] args) throws Exception
 { final ParameterTool params = ParameterTool.fromArgs(args);
final StreamExecutionEnvironment env = StreamExecutionEnvironment
.getExecutionEnvironment(); env.getConfig().setGlobalJobParamete
rs(params); env.setParallelism(3); DataStream<String> simpleStri
ngStream = env.addSource(new EventsGenerator()); Properties
configProps = new Properties(); // Set the domain name used to
 access Log Service. configProps.put(ConfigConstants.LOG_ENDPOI
NT, sEndpoint); // Set the AccessKey used to access Log Service
. configProps.put(ConfigConstants.LOG_ACCESSSKEYID, sAccessKey
Id); configProps.put(ConfigConstants.LOG_ACCESSKEY, sAccessKey
); // Configure the Log Service project to which logs are written
. project configProps.put(ConfigConstants.LOG_PROJECT, sProject
); // Configure the Log Service Logstore to which logs are written.
Logstore configProps.put(ConfigConstants.LOG_LOGSTORE, sLogstore);
FlinkLogProducer<String> logProducer = new FlinkLogProducer<String>(
new SimpleLogSerializer(), configProps); simpleStringStream.addSink
(logProducer); env.execute("flink log producer"); } // Simulate
 log generation. public static class EventsGenerator implements
SourceFunction<String> { private boolean running = true; @Override
 public void run(SourceContext<String> ctx) throws Exception { long
 seq = 0; while (running) { Thread.sleep(10); ctx.collect((seq++) +
```

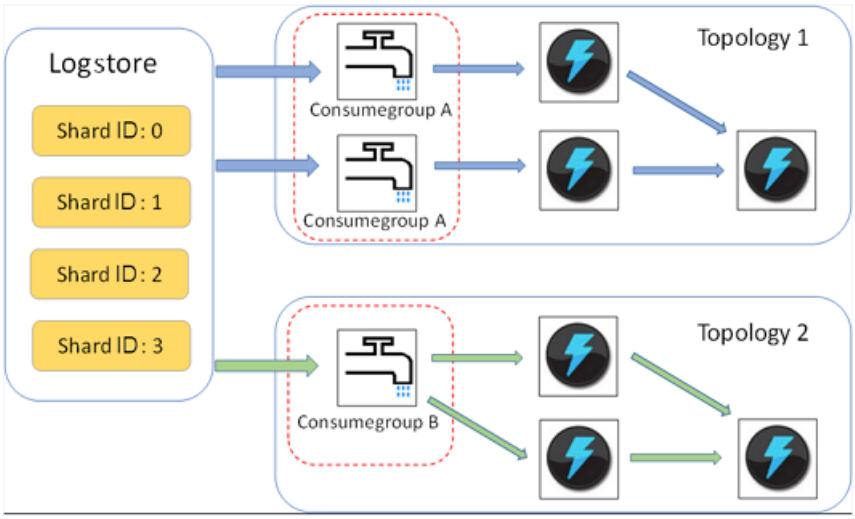
```
"-" + RandomStringUtils.randomAlphabetic(12)); } } @Override public
void cancel() { running = false; } } }
```

22.10.5 Storm consumption

LogHub of Log Service provides an efficient and reliable log channel for collecting log data through Logtail and SDKs. You can access real-time systems such as Spark Streaming and Storm to consume the data written to LogHub.

The LogHub Storm spout feature is provided to read data from LogHub in real time, reducing Storm users cost for LogHub consumption.

Basic architecture and process



- Enclosed in the red dotted boxes in the preceding figure are LogHub Storm spouts. Each Storm topology has a group of spouts that read all data from a Logstore. The spouts in different topologies are independent of each other.
- Each topology is identified by a unique LogHub consumer group name. The LogHub client library is used for load balancing and automatic failover among the spouts in the same topology .
- Spouts read data from LogHub in real time, send the data to the bolt nodes of the topology, and save consumption endpoints as checkpoints to the LogHub server periodically.

Restrictions

- To prevent misuse, each Logstore supports up to five consumer groups. You can use the DeleteConsumerGroup API of the Java SDK to delete unused consumer groups.
- We recommend that the number of spouts be equal to the number of shards. Otherwise, a single spout may be unable to process a large amount of data.

- If a shard contains a large amount of data which exceeds the processing capability of a single spout, you can use the shard split API to lower down the per-shard data volume.
- The dependency on the Storm ACK mechanism is mandatory in LogHub spouts, to confirm that spouts send messages correctly to bolts. Therefore, bolts must call ACK for such confirmation.

Example

- **Spout (used for topology creation)**

```
 public static void main( String[] args ) { String mode = "Local
"; // Use the local test mode. String conumser_group_name = ""; //
Each topology must be assigned a unique consumer group name,
which contains 3 to 63 characters including letters a–z, numbers
 0–9, underlines (_), and hyphens (-). The consumer group must
 be specified and must start and end with lowercase letters or
numbers. String project = ""; // Project of Log Service. String
logstore = ""; // Logstore of Log Service. String endpoint = ""; //
Domain name of Log Service. String access_id = ""; // AccessKey
 of the user. String access_key = ""; // Configurations required
 for creating a LogHub Storm spout. LogHubSpoutConfig config =
new LogHubSpoutConfig(conumser_group_name, endpoint, project,
logstore, access_id, access_key, LogHubCursorPosition.END_CURSOR
); TopologyBuilder builder = new TopologyBuilder(); // Create a
 LogHub Storm spout. LogHubSpout spout = new LogHubSpout(config
); // In the actual situation, the number of spouts may be equal
to the number of Logstore shards. builder.setSpout("spout", spout
, 1); builder.setBolt("exclaim", new SampleBolt()).shuffleGrouping
("spout"); Config conf = new Config(); conf.setDebug(false); conf
.setMaxSpoutPending(1); // The serialization method LogGroupDa
taSerializSerializer of LogGroupData must be configured explicitly
 when Kryo is used for data serialization and deserialization.
LogGroupDataSerializSerializer Config.registerSerialization(conf
, LogGroupData.class, LogGroupDataSerializSerializer.class); if (
mode.equals("Local")) { logger.info("Local mode..."); LocalClust
er cluster = new LocalCluster(); cluster.submitTopology("test-
jstorm-spout", conf, builder.createTopology()); try { Thread.sleep
(6000 * 1000); //waiting for several minutes } catch (Interrupte
dException e) { // TODO Auto-generated catch block e.printStackTrace
(); } cluster.killTopology("test-jstorm-spout"); cluster.shutdown
(); } else if (mode.equals("Remote")) { logger.info("Remote mode
..."); conf.setNumWorkers(2); try { StormSubmitter.submitTopology
("stt-jstorm-spout-4", conf, builder.createTopology()); } catch (
AlreadyAliveException e) { // TODO Auto-generated catch block e.
printStackTrace(); } catch (InvalidTopologyException e) { // TODO
Auto-generated catch block e.printStackTrace(); } } else { logger.
error("invalid mode: " + mode); } } }
```

- **Sample code of the bolts that consume data. (Only the content of each log is printed.)**

```
public class SampleBolt extends BaseRichBolt { private static
final long serialVersionUID = 4752656887774402264L; private static
 final Logger logger = Logger.getLogger(BaseBasicBolt.class);
private OutputCollector mCollector; @Override public void prepare(@
SuppressWarnings("rawtypes") Map stormConf, TopologyContext context
, OutputCollector collector) { mCollector = collector; } @Override
 public void execute(Tuple tuple) { String shardId = (String) tuple
 .getValueByField(LogHubSpout.FIELD_SHARD_ID); @SuppressWarnings("
unchecked") List<LogGroupData> logGroupDatas = (ArrayList<LogGroupDa
```

```
ta>) tuple.getValueByField(LogHubSpout.FIELD_LOGGROUPS); for (
LogGroupData groupData : logGroupDatas) { // Each log group consists
 of one or more logs. LogGroup logGroup = groupData.GetLogGroup();
for (Log log : logGroup.getLogsList()) { StringBuilder sb = new
StringBuilder(); // Each log has a time field and multiple key-value
 pairs. int log_time = log.getTime(); sb.append("LogTime:").append(
log_time); for (Content content : log.getContentsList()) { sb.append
("\t").append(content.getKey()).append(":") .append(content.getValue
()); } logger.info(sb.toString()); } } // The dependency on the
Storm ACK mechanism is mandatory in LogHub spouts, to confirm that
 spouts send messages correctly to bolts. //Therefore, bolts must
call ACK for such confirmation. mCollector.ack(tuple); } @Override
public void declareOutputFields(OutputFieldsDeclarer declarer) { //
do nothing } }
```

Maven

Use the following code for versions earlier than Storm 1.0 (for example, 0.9.6):

```
<dependency> <groupId>com.aliyun.openservices</groupId> <artifactId>
loghub-storm-spout</artifactId> <version>0.6.5</version> </dependency>
```

Use the following code for Storm 1.0 and later versions:

```
<dependency> <groupId>com.aliyun.openservices</groupId> <artifactId
>loghub-storm-1.0-spout</artifactId> <version>0.1.2</version> </
dependency>
```

22.10.6 Spark Streaming consumption

E-MapReduce has implemented a set of universal APIs for Spark Streaming to calculate real-time LogHub data consumption. To download SDKs, go to *GitHub*.

22.10.7 Consumption by StreamCompute

StreamCompute can be used to consume data directly in LogHub after data sources of the LogHub type are created.

```
CREATE STREAM TABLE source_test_galaxy ( $schema ) WITH ( type=
loghub, endpoint=$endpoint, accessId=$loghub_access_id, accessKey=$
loghub_access_key, projectName=$project, logstore=$logstore );
```

Table 22-15: Parameter list

Parameter	Description
`$schema`	The keys in logs that are mapped to the columns in the StreamCompute table, for example, `name STRING, age STRING, id STRING`.
`$endpoint`	Your endpoint.

Parameter	Description
`$loghub_access_id`	AccessID of the account (or subaccount) with the read permission.
`$loghub_access_key`	Accesskey of the account (or subaccount) with the read permission.
`$project`	Project where data is located.
`$logstore`	Logstore where data is located.

Example:

```
CREATE STREAM TABLE source_test_galaxy ( name STRING, age STRING, id
STRING ) WITH ( type=loghub, endpoint=http://cn-hangzhou-intranet.log
.aliyuncs.com, accessId=mock_access_id, accessKey=mock_access_key,
projectName=ali-cloud-streamtest, logstore=stream-test );
```

23 Key Management Service (KMS)

23.1 What is KMS

Key Management Service (KMS) is a secure and easy-to-use key hosting service provided by Alibaba Cloud Apsara Stack. With KMS, you can easily create and manage your keys and use them to encrypt your data.

KMS is integrated with multiple Alibaba Cloud products and services to protect your cloud data.

KMS can solve the problems shown in *Table 23-1: Problems solved by KMS*.

Table 23-1: Problems solved by KMS

Roles	Problems and requirements	KMS solution
Application/Website developers	• My program needs keys or certificates for encryption or signature, and I want secure and independent key management. • I want to securely access keys wherever my application is deployed. I do not accept plaintext keys deployed everywhere. That is too risky.	Using envelop encryption technology, you can store the customer master key (CMK) in a KMS instance and deploy only the encrypted data keys . You need to call the KMS instance to decrypt data keys only when necessary.
Service developers	• I do not accept responsibi lity for the security of users' keys and data. • I want users to manage their own keys. I want to use specified keys to encrypt their data with their authorization. This way, I can focus on developing service features.	Based on envelop encryption technology and KMS APIs, service developers can use specified CMKs to encrypt and decrypt data keys, so **`plaintext is not directly stored on a storage device`**. This removes service developers' worries about how to manage users' keys.
Chief Security Officer (CSO)	• I expect our key management activities to meet compliance requirements.	KMS can connect to RAM for unified authorization management.

Roles	Problems and requirements	KMS solution
	• I need to ensure that keys are reasonably authorized and any use of keys is audited.	

23.2 Log on to the KMS console

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

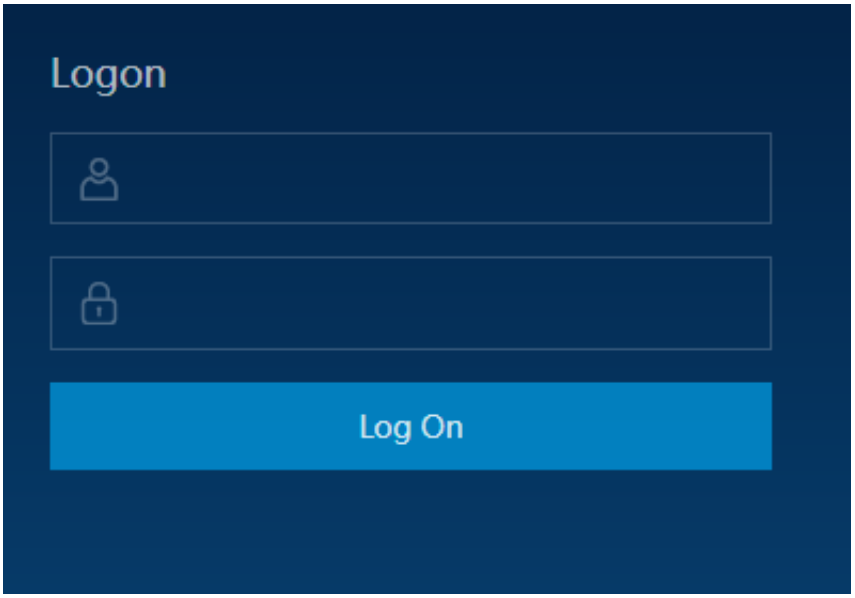
  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 23-1: Log on to the Apsara Stack console*.

Figure 23-1: Log on to the Apsara Stack console



3. Enter the correct username and password.
   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **Log On** to go to the **Home** page.
5. On the top navigation bar, choose **Console** > **Compute, Storage & Networking** > **Key Management Service**.

23.3 Create a CMK

Create a CMK on the cloud console for subsequent encryption and decryption operations.

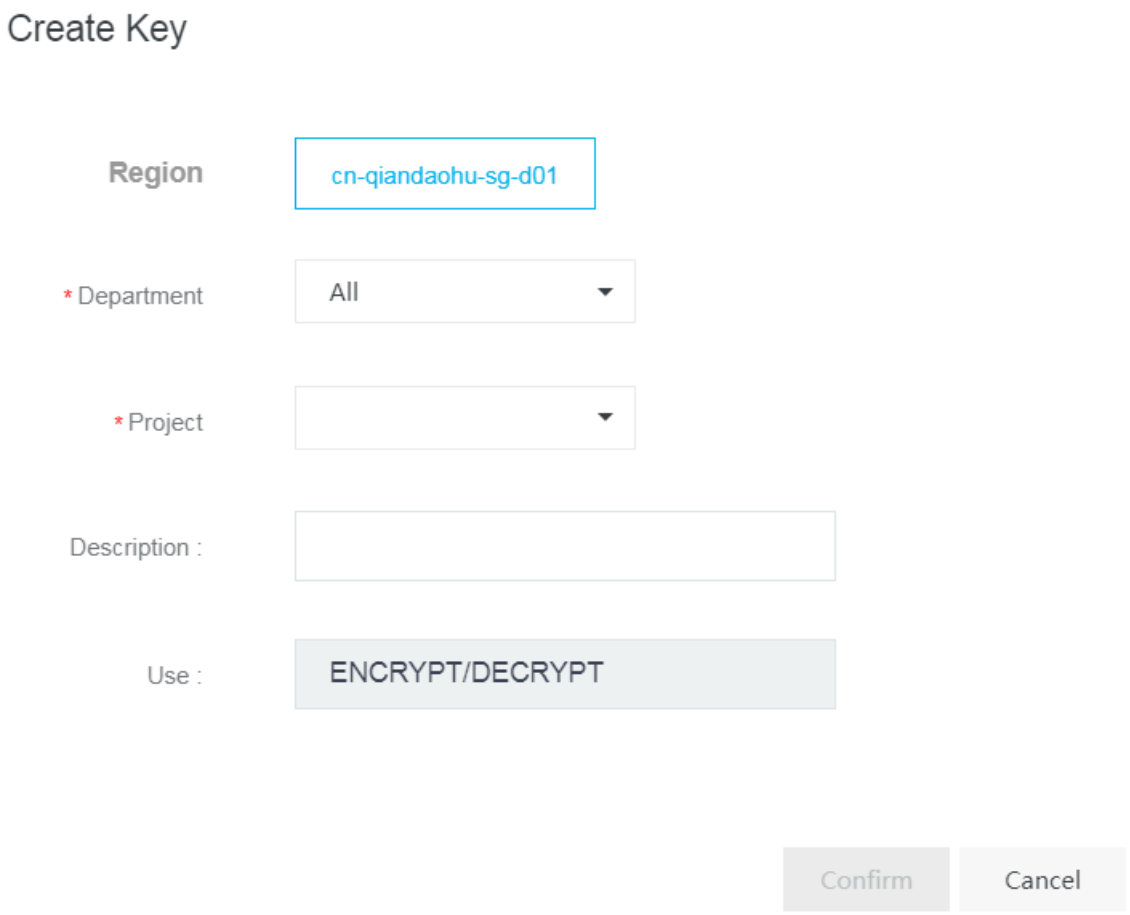
Procedure

1. *Log on to the Apsara Stack console*.
2. Choose**Cloud Resource Center** > **Cloud Product Management** > **Key Management Service** to go to the **Key Management Service** page.

3. Click **Create Key**.

   The **Create Key** dialog box appears, as shown in *Figure 23-2: Create a key*.

   Figure 23-2: Create a key



4. Select your desired region, department, and project, and enter descriptive information. Then click **OK**.

   After the CMK is successfully created, call the KMS API for programming in accordance with the application scenario and the *Cite LeftKMS Development GuideCite Right*.

23.4 View key details

After a key is successfully created, you can view the key ID, key status, key usage, and creator information.

Procedure

1. *Log on to the Apsara Stack console*.

2. Choose**Cloud Resource Center** > **Cloud Product Management** > **Key Management Service** to go to the **Key Management Service** page.
3. In the key list, select the key you want to check and click the key ID link, or click and select **Details**.

   The **Key Details** page appears.
4. The **Basic Information** area shows the key ID, key status, key usage, and creator information.

23.5 Enable a key

You can flag a key as Enabled, to assign the permission for a user to use this key.

Procedure

1. *Log on to the Apsara Stack console*.
2. Choose**Cloud Resource Center** > **Cloud Product Management** > **Key Management Service** to go to the **Key Management Service** page.
3. Select a key in **`Disabling`** state, click , and select **Enable Key**.

   After the key is successfully enabled, the key status changes from **`Disabling`** to **`Enabling`**.

23.6 Disable a key

You can disable a key so that it cannot be used for encryption or decryption. The ciphertext encrypted using the key cannot be decrypted until the key is enabled again.

Context

When a key is created, it is in **`Enabled`** state by default.

Procedure

1. *Log on to the Apsara Stack console*.
2. Choose**Cloud Resource Center** > **Cloud Product Management** > **Key Management Service** to go to the **Key Management Service** page.
3. Select a key in **`Enabling`** state, click , and select **Disable Key**.

   The **Disable Key** confirmation window appears.
4. Click **Confirm** to disable the key.

   After the key is successfully disabled, the key status changes from **`Enabling`** to **`Disabling`**.

23.7 Delete a key on schedule

You can apply to delete CMK within a specified pre-deletion period (7 to 10 days).

Context

You must specify a pre-deletion period when you apply for CMK deletion. The period ranges from 7 to 30 days.

You can use **CancelKeyDeletion** to cancel the CMK deletion application after the application is submitted, but before the pre-deletion period ends.

Note:

- During the key pre-deletion period, the key is in the PendingDeletion state and cannot be used for encryption, decryption, or data key generation.
- Deleting CMK has a very serious impact. Normally, we recommend that you select *Disable a key*.
- A deleted CMK cannot be recovered, and the content encrypted along with the data key generated using the CMK cannot be decrypted. Therefore, we do not allow you to directly delete CMK, but you can apply to do so.
- We agree to delete the key within 24 hours after the pre-deletion period.

  For example, if you apply for CMK deletion at 14:00, September 10, 2017, and the pre-deletion period is 7 days, then KMS will delete the CMK within 24 hours after 14:00, September 17, 2017.

Procedure

1. *Log on to the Apsara Stack console*.
2. Choose**Cloud Resource Center** > **Cloud Product Management** > **Key Management Service** to go to the **Key Management Service** page.
3. Select a key, click , and select **Plan to Delete Key**.

   The **Plan to Delete Key** dialog box appears, as shown in *Figure 23-3: Plan to delete a key*.

Figure 23-3: Plan to delete a key

4. Enter the pre-deletion period (in days) in the text box, and click **OK**.

   Then the key is in the **`PendingDeletion`** state.

   If you want to cancel the deletion application before the pre-deletion period ends, click and select **Cancel Key Deletion**.

24 StreamCompute

24.1 What is streaming computing?

The growing demand for high availability and operability of information requires software systems to process more data in less time. Traditional large data processing models separate online transaction processing from offline analysis, but it is clear that the architecture cannot satisfy the need for real-time processing of large amounts of data.

StreamCompute is designed to meet the strict requirements for quick data processing: The business value of data rapidly decreases with the passage of time, so it must be calculated and processed as soon as possible after it is generated. However, traditional big data processing modes follow the traditional daily processing mode. This means that current data is accumulate d and processed in computing cycles of hours or even days. Clearly, such processing modes cannot meet the needs of real-time data computing. In terms of real-time data analysis, risk control and alarm, real-time forecasting, financial trading, and many other business scenarios, batch (or offline) processing cannot meet the business needs for applications with strict data processing latency demand. However, as a real-time computing model for streaming data, StreamCompute can effectively shorten the full-link data stream latency, enable real-time computing logic, reduce computing costs, and finally meet business needs for real-time processing of big data.

What is streaming data?

Generally speaking, the generation of data can be seen as a series of discrete events. These discrete events form event streams/data streams along the timeline. Unlike traditional offline data , streaming data is continuously generated from thousands of data sources. Streaming data is usually sent in the form of data records, but it is generally smaller in size compared with the offline data. Streaming data is generated from endless event streams, such as log files generated by your mobile or Web applications, online shopping data, in-game player activities, social network information, financial transactions or geographic positioning services, as well as telemetry data from the connected devices or instruments in the data center.

In general, StreamCompute has three features:

- Real-time and unbound data streams. StreamCompute is used to compute the data streamed in real time from the data source. Streaming data is subscribed to and consumed by StreamCompute in time sequence. Due to the continuity of data generation, the data stream will be continuously integrated into the StreamCompute system over a long period. For

example, for the visit log stream of a website, as long as the website does not close the log stream, the data stream will continue to be generated and integrated into the StreamCompute system. Therefore, for a streaming system, the data is real-time and will not be terminated (unbound).

- Continuous and efficient computing. StreamCompute is an **event triggered** computing mode. The trigger source is the above-mentioned unbound streaming data. Once new streaming data enters StreamCompute, StreamCompute will immediately initiate and perform a computing task, so the entire StreamCompute is constantly computing.
- Streaming and real-time data integration. The streaming data triggers the computing result of StreamCompute. The result can be directly written into the destination data storage. For example, the computed report data is directly written into the RDS for report display. Therefore , the computing result of the streaming data can be written into the destination data storage like a streaming data source.

24.2 Quick start

24.2.1 Host word statistics

Statistical analysis on hot words is widely applied in different scenarios, including hot word searching, forum hot word, and label hot word. For example, real-time Weibo hot word statistics enables you to know the latest hot words on Weibo. The statistical analysis on hot words is a simple WordCount task. For streaming real-time statistical analysis on hot words, the WordCount processing logic is completely converted into streaming real-time processing. This service enables real-time statistical analysis on hot words and displays real-time statistical results. The following describes how to compile the first streaming computing task by using a WordCount streaming task as an example:

Q: What is WordCount?

A: A WordCount task of big data, like the "Hello World" in programming instructions, is often an essential task for newcomers. The following describes how to develop a streaming version of WordCount by using WordCount of Alibaba Cloud StreamCompute as an example. The newcomers can learn the basic syntax format of BlinkSQL and the basic task compiling/publishing operations by performing a WordCount task.

24.2.1.1 Code development

1. Enter the home page of Alibaba Cloud StreamCompute. On the navigation bar at the top, click the **`Development`** tab to display the IDE page.

2. Enter the **Newbie Task** folder that is built in Alibaba Cloud StreamCompute. Right-click the folder to create a task named **WordCount**.

The **WordCount** task for streaming computing follows the principle of a batch task. The only difference is that the streaming **WordCount** data source is continuous and unbound; therefore, theoretically, the **WordCount** task for streaming computing should not stop unless it is killed by the user.

The StreamSQL codes are as follows:

```
1    create stream table stream_source(word string);
2    create result table stream_result(word string, cnt bigint);
3
4    insert into stream_result select
5       t.word
6       ,count(1)
7     from stream_source t
8     group by t.word;
```

In Line 1, we have declared a streaming data table named stream_source, which contains only one column. The column is of the String type and named word.

Note:
As mentioned above, streaming data is the source that drives streaming computing. Therefore, the stream_source here is the data-driven source and each entry (batch) of stream_source data triggers a downstream streaming computing task.

In Line 2, we have declared a result table to store the computing result of the WordCount task. The table is named stream_result, which contains two columns: one column is of the String type and named word, and the other is of the bigint type and named cnt.

Note:
As mentioned above, StreamCompute itself does not have any data storage system. Theoretically, all the result data storage systems are common storage systems, such as RDS and Table Store. We have declared a result table here to store the computing result.

From Line 4, we start the formal WordCount computing logic. If you are familiar with SQL, you should be clear about the meaning of the SQL codes, that is, reading data from the stream_source table and counting the number/frequency of each word for each entry of incoming data.

Note:

To minimize your costs and difficulties for learning streaming computing, the StreamSQL provided by Alibaba Cloud StreamCompute is basically the same to the standard SQL format.

24.2.1.2 Code debugging

To facilitate BlinkSQL debugging, Alibaba Cloud StreamCompute provides the online debugging function to help you construct debugging data and perform regression tests. The powerful debugging function of Alibaba Cloud StreamCompute can perform simulation debugging of streaming storage, static storage, and result storage, to help you construct all kinds of data for SQL verification.

 **Note:**

- To prevent impact on the read and write operations of the online storage system, the StreamCompute debugging process requires all the input tables to provide test data. It is not allowed to read data from the online storage system.
- All the write (insert) operations are printed on the screen without affecting the online system.

On the **Data Development** IDE page, click **Debug** to start the **Debugging Task**. When the debugging task is started for the first time, the system displays a dialog box to remind you of uploading a test data file.

On the left side, there is a list of reference data tables declared by the BlinkSQL user, including streaming input tables and static dimension tables. You need to provide debugging data for each input table before debugging. To facilitate data debugging, Alibaba Cloud StreamCompute automatically builds a data template for each table and provides a title in the format of **`field name`** (field type) in the first line of the template. You can modify the test data based on the information in the title. On the debugging page, click **Download Debugging Template** and enter debugging data according to your test policies. StreamCompute applies strict definitions of uploaded debugging data:

- A debugging data file can contain a maximum of 1 MB or 1 K records.
- Only debugging files in UTF-8 format are supported.
- Multiple .csv files must be separated by commas, but the content should not contain a comma.
- Only common numerical types are supported and scientific notation is not supported.

StreamCompute uses .csv debugging files, so we recommend that you use Excel software on the Windows platform or VIM/Sublime software on the Mac platform to open the template and

modify the data. **We do not recommend the Number tool on the Mac platform because it will generate a large amount of field information**.

The newbie task helps you get familiar with streaming computing quickly. We provide a *test data* sample, and you can download it and upload it again on the test interface.

Note:
In a .pdf file, *Cite Lefttest data based on hot word statisticsCite Right* cannot be downloaded through the link. StreamCompute provides test data as an attachment. Please contact the system administrator for the test data.

Click **Debug**. StreamCompute immediately starts a test stream computing task for debugging. The test task runs with the test data you provided. The final test result is directly output on the screen.

For streaming computing, the computing operations are triggered by the streaming data. In the test status, each data entry of stream_source directly triggers a streaming computing task and outputs the computing result, so we can find three data entries in the test file and three data entries on the result page. The computing traces are respectively as follows:

The first line of source data (aliyun) reaches StreamCompute. At this moment, StreamCompute finds that the word aliyun does not exist, so it outputs the computing result `<aliyun, 1>` on the screen.

The second line of source data (still aliyun) reaches StreamCompute. At this moment, StreamCompute finds that a record of `<aliyun, 1>` already exists, so it adds 1 to the record value and outputs the result `<aliyun, 2>` on the screen.

The third line of source data (still aliyun) reaches StreamCompute. At this moment, StreamCompute finds that a record of `<aliyun, 2>` already exists, so it adds 1 to the record value and outputs the result `<aliyun, 3>` on the screen.

The final result is based on the last output, that is, `<aliyun, 3>` represents the final result of the data debugging. We also provide one additional [*test data*] sample for you to observe the output on the debugging interface when different entries of test data (words) are used.

24.2.1.3 Data O&M

Once the codes are tested and verified for accuracy, they can be published to the **Data O&M** module. A task can be submitted to the streaming computing cluster for production.

Procedure

1. On the **Development** page, click **Launch**. The **Launch the Latest Version** window is displayed.
2. Enter launching comments and click **Launch**. The latest version is launched.
3. Click the **O&M** tab. You can find the newly launched WordCount task in the job list.
4. Click **Start** in the action column corresponding to the WordCount task. The **Start Job** window is displayed.
5. Select **Specify Data Read Time** and click **Start with Above Configurations**. The streaming computing task can be scheduled by the production cluster.

Result

Once the task starts successfully, the task status turns green.

You may have this question: The computing task starts to run in the distributed streaming computing cluster, but it has no streaming data input and data output. Why does this happen and how to run the task successfully? This happens because we defined tables `my_source` and `my_result` without specifying the type of external reference data source. When no data source type is specified, StreamCompute considers the input Stream table as an internal random table of randomly generated strings/numbers and directly discards the output result table.

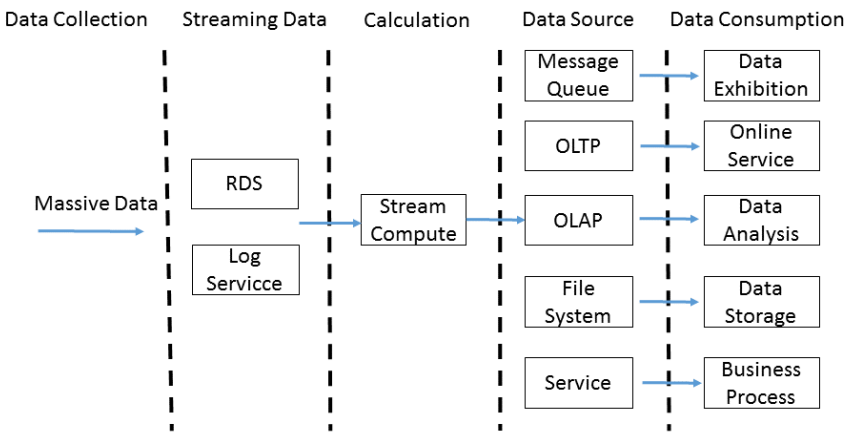
24.3 Operation guide

This chapter describes how to use the Alibaba Cloud StreamCompute console for real-time data analysis on the cloud. This chapter provides guidance on operations such as data collection, data storage, and data development.

24.3.1 Data collection

One cannot make bricks without straw. Similarly, a big data system cannot work without data collection. To enable full use of your streaming storage systems, Alibaba Cloud StreamCompute is interoperable with multiple types of streaming storage. Therefore, you can use existing streaming storage systems without the need for data collection or data integration.

Figure 24-1: System Architecture



Alibaba Cloud StreamCompute supports two data storage systems. You need to complete data collection using different data integration tools. StreamCompute can interoperate with the following systems:

- **Log Service**

  LogHub is an all-in-one service for log data. It has been honed by countless big data applicatio ns at Alibaba Group. This log service provides multiple functions for logs, including log collection, consumption, shipping, query, and analysis.
- **RDS**

24.3.2 Data storage

24.3.2.1 Storage overview

By registering your data storage, you can experience more convenience provided by the all-in-one StreamCompute platform to manage the data storage easily. Alibaba Cloud StreamCompute provides a management UI for various data storage systems, including RDS and Log Service. This platform allows you to manage your data storage on the cloud without switching between management pages of different products.

Note:

Authorization is required before data storage registration. For details, see *Cite LeftRole AuthorizationCite Right*.

24.3.2.1.1 Storage Types

Streaming Storage

Streaming storage provides data drive for the downstream StreamCompute platform and data output for StreamCompute jobs.

Table 24-1: Streaming Storage

Support	Input	Output
Log Service (LogHub)	Supported	Supported

Static storage

Static storage provides data association query for StreamCompute and can also be used as data output of StreamCompute jobs.

Table 24-2: Static Storage

Support	Dimension Table	Output
ApsaraDB (RDS)	Supported	Supported

24.3.2.1.2 Storage usage

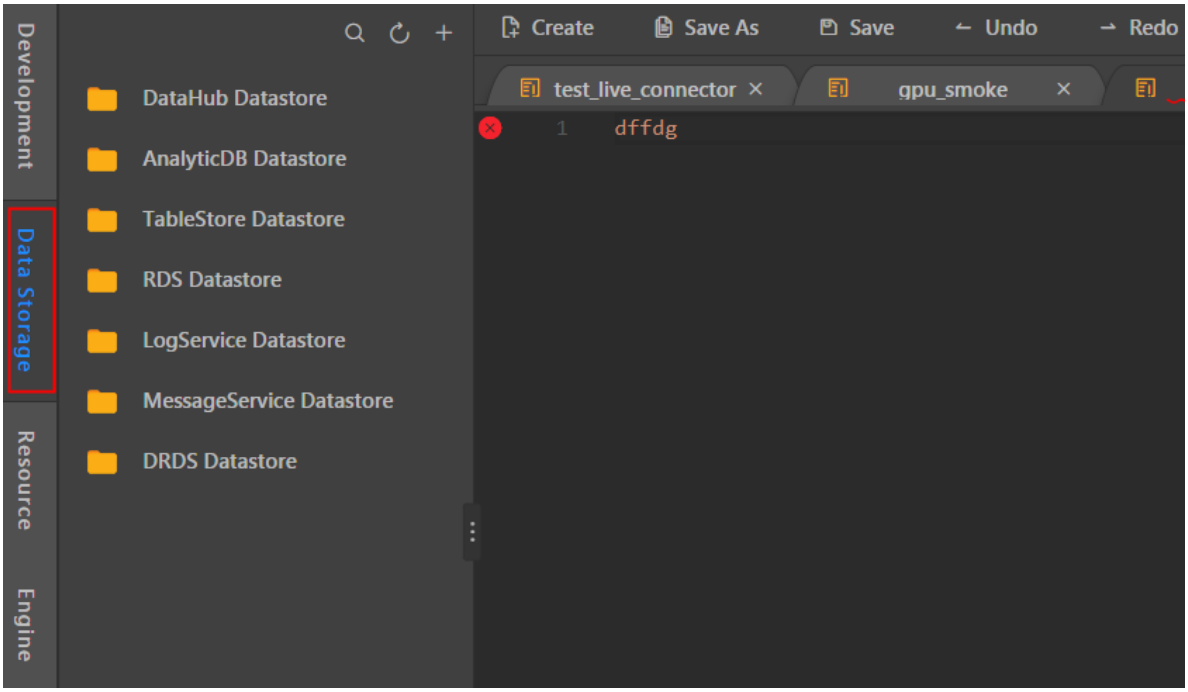
Registered data sources can be used in the following scenarios, enabling you to experience more convenience provided by the all-in-one StreamCompute platform.

Note:
To use the data resources of another user, you can directly reference the data source in a DDL statement by using the AccessId and AccessKey. In this case, you cannot handle the data source on the UI, but the job can run directly.

- **Data Registration**

  You cannot use any data storage function before registering a data storage. Therefore, you must **register required data storage information with StreamCompute** first. Open the **Development** UI, choose**Data Storage** > **+** from the left-side navigation menu to access *Figure 24-2: Data Storage Registration page*.

Figure 24-2: Data Storage Registration page



Note:
The data storage function of StreamCompute only supports registration of the storage resources of the same account and does not allow cross-account authorization.

- **Data Preview**

  StreamCompute allows you to preview each data storage you have registered. Click **Data Storage** and select a data storage type to preview the data.

- **Automatic DDL Creation**

  Before referencing an external storage in StreamCompute, declare the external storage.

  StreamCompute provides an auxiliary DDL creation function. This function allows you to generate table creation DDL statements with one click.

  Click the task you want to edit on the data development page, click **`Data Storage`** and then **Reference as Input Table**. The StreamCompute system then creates the DDL statement in the position of the cursor.

Annex: Reference data resources of another account

You cannot register or use data storage resources of other accounts on the StreamCompute UI. The data storage function of StreamCompute supports only registration of the storage resources

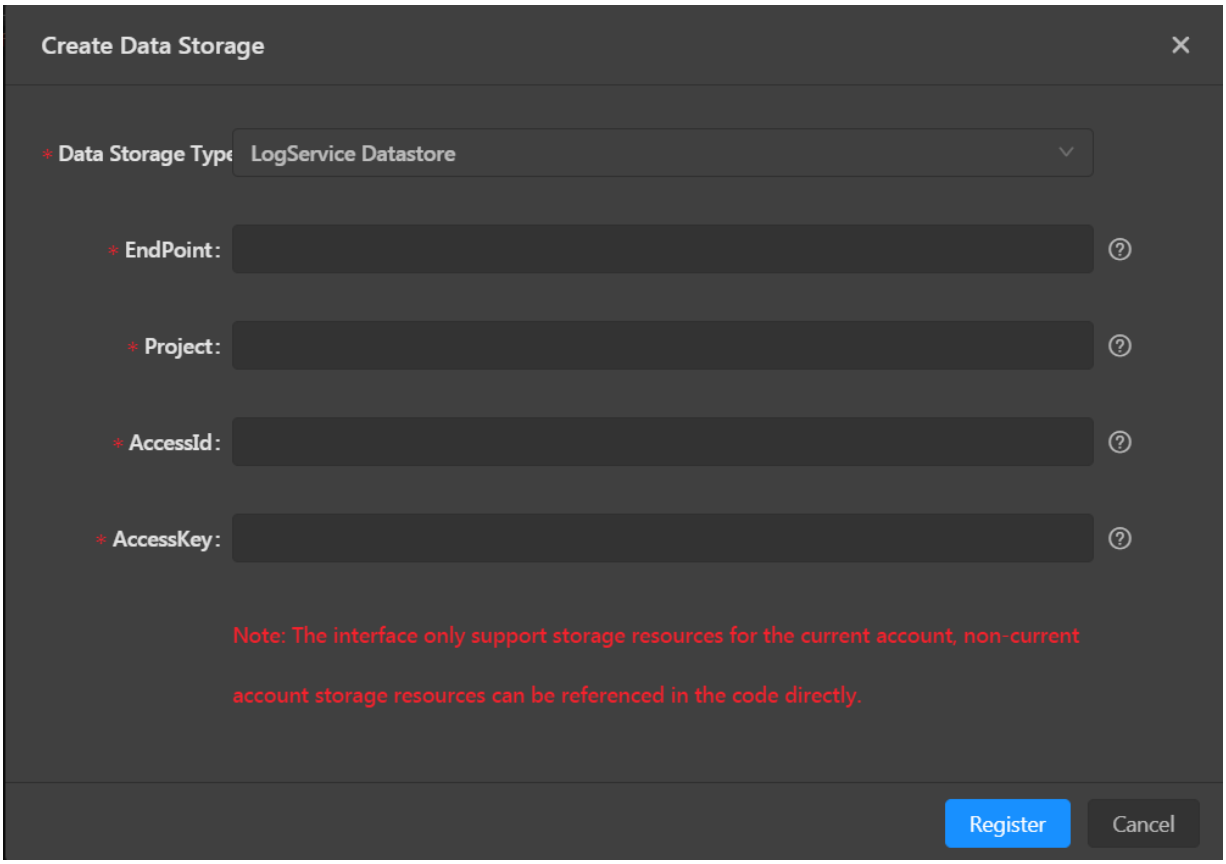
under the same account and does not allow cross-account authorization. To use the data storage of another account, you can reference it as external data source in a DDL statement.

24.3.2.2 Log Service

Registration

Log Service (LOG/previous SLS) is an end-to-end solution designed for the log scenario. This service supports the collection/subscription, dump, and query of large amounts of log data. Log Service is a complete log management platform of Alibaba Cloud. When you use LOG for ECS log management, the streaming computing can directly store LogHub to avoid the work for data transferring.

Figure 24-3: LogService Data Storage



- Endpoint

  Enter the endpoint of LOG. The projects of LOG vary in different regions.

  Note:

You can consult your Apsara Stack system administrator about how to enter the LOG endpoint address.

- Project

  Enter a LOG project.

  Note:
  You cannot register data storages of other accounts. For example, if Project A of DataHub is owned by user A, user B cannot use Project A in StreamCompute.

- **AccessId**

  Enter the AccessKey ID of the current account.

- **AccessKey**

  Enter the AccessKey Secret of the current account, so that StreamCompute can access the project of Log Service.

Scenario

LOG is a streaming data storage; therefore, StreamCompute can only use it as **streaming data input and output, and cannot reference it as a dimension table**.

FAQ

- **Q: Why do I fail to register a data storage and see the failure cause XXX?**

  A: The data storage page of StreamCompute only helps you with data management. It uses the corresponding storage SDK to access various storage systems for you. Therefore, most registration failures are caused by mistakes in the registration process. Check and ensure the following:

  - You have activated and owned the LOG project. Log on to the LOG console and check whether you have the permission to access the project.
  - You are the owner of the LOG project. Note that you cannot register the data storage of another user. For example, if Project A of LOG is owned by user A, user B cannot use Project A in StreamCompute.
  - The LOG endpoint and project you entered are correct.

    Note:

The LOG endpoint must start with http and cannot end with /. For example, `http://cn-hangzhou.log.aliyuncs.com` is correct but `http://cn-hangzhou.log.aliyuncs.com/` is incorrect.

  - Do not register a data storage repeatedly. StreamCompute provides registration check to prevent repeated registration.
- Q: Why does the data sampling function support only time-based sampling?

  A: LOG is a streaming data storage and can only provide the time parameter as the external interface. Therefore, StreamCompute only provides time-based sampling.

  Note:
  Log on to the LOG console to use the retrieval function if required.

24.3.2.3 ApsaraDB (RDS)

Overview

ApsaraDB for RDS (RDS) is a stable, reliable, and scalable online database service. RDS is based on the Apsara distributed file system and high-performance full-SSD disk storage. This service supports MySQL, SQL Server, PostgreSQL, and PPAS (high Oracle compatibility) engine. Currently, StreamCompute supports MySQL.

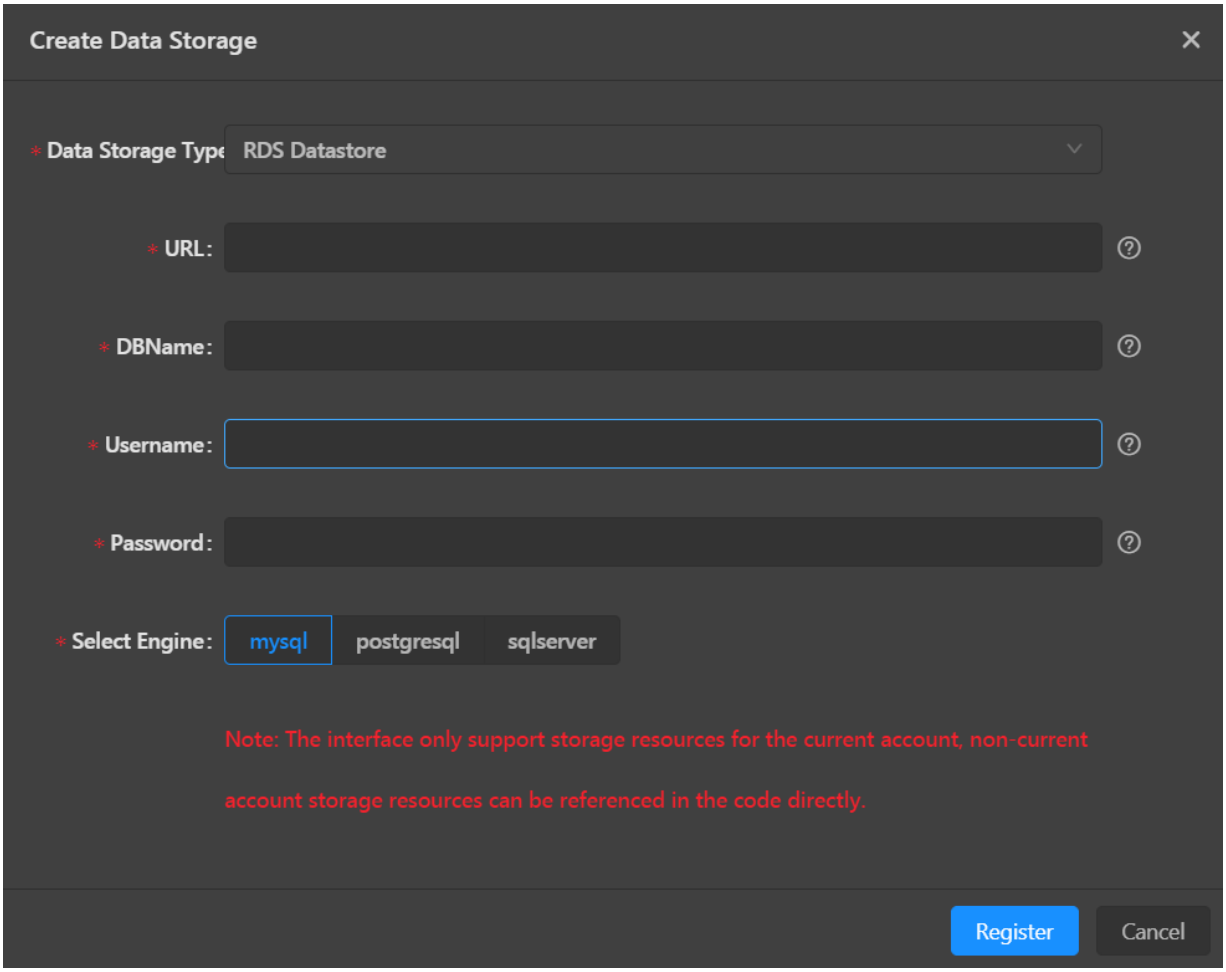
Due to limitations of the relational model, RDS is not as powerful as Table Store in terms of processing large amounts of concurrent requests, so StreamCompute mainly uses RDS as a result table in such cases. For small-volume low-concurrency requests, StreamCompute can also use RDS as a dimension table.

Use of RDS

Data storage method

Note:
We recommended that you use the data storage registration function for RDS, to avoid unnecessary loss that may be caused by connection failure after upgrade or expansion. We are not responsible for any consequences from your nonuse of the data storage registration function. The StreamCompute team reserves the right of final interpretation.

Figure 24-4: Register data storage



Configuration item information:

- **URL**

  Enter the connection URL of RDS instance.

- **DBName**

  Enter the name of the RDS database accessed (not the RDS instance name).

  Currently, RDS uses whitelists for security management. StreamCompute needs to automatically add a whitelist for RDS; otherwise, it may fail to access RDS. For details, see *Cite LeftHow to add an RDS whitelistCite Right*.

- **Username**

  Database logon name. To be compatible with database communication protocols, the user name/password for RDS does not support Alibaba Cloud AccessId and AccessKey. We recommend that you provide a separate user name/password for StreamCompute to read from

and write into RDS. Account-based management can easily ensure security of the RDS system .

- **Password**

  Database logon password. To be compatible with database communication protocols, the user name/password for RDS does not support Alibaba Cloud AccessId and AccessKey. We recommend that you provide a separate user name/password for StreamCompute to read from and write into RDS. Account-based management can easily ensure security of the RDS system .

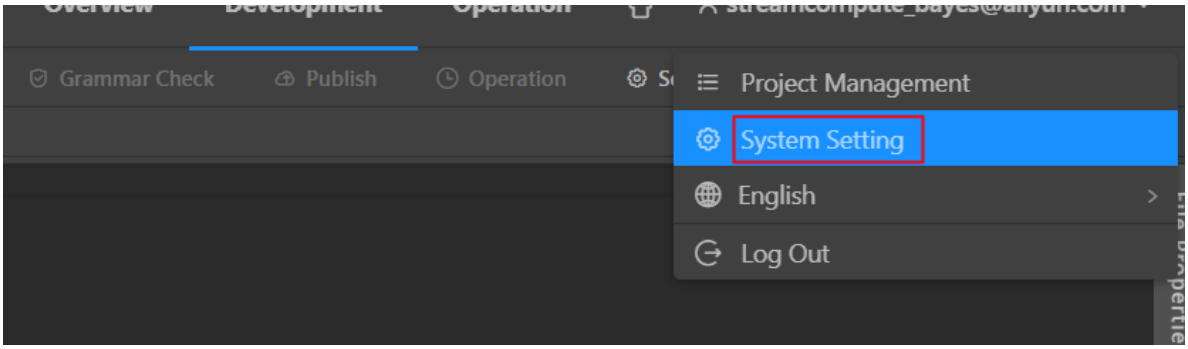
- **Select Engine**

  The type of RDS.

After the data storage is registered, select a result table

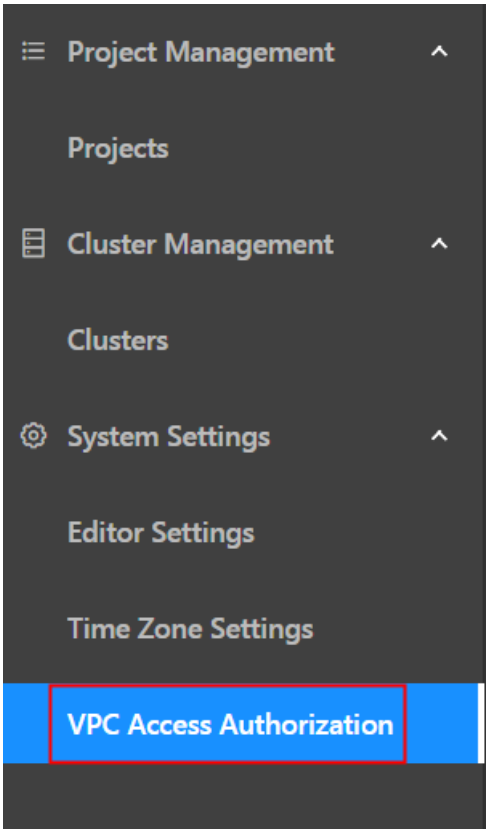
1. Click the account name in the top-right corner and select **System Setting**.

   Figure 24-5: Management configurations



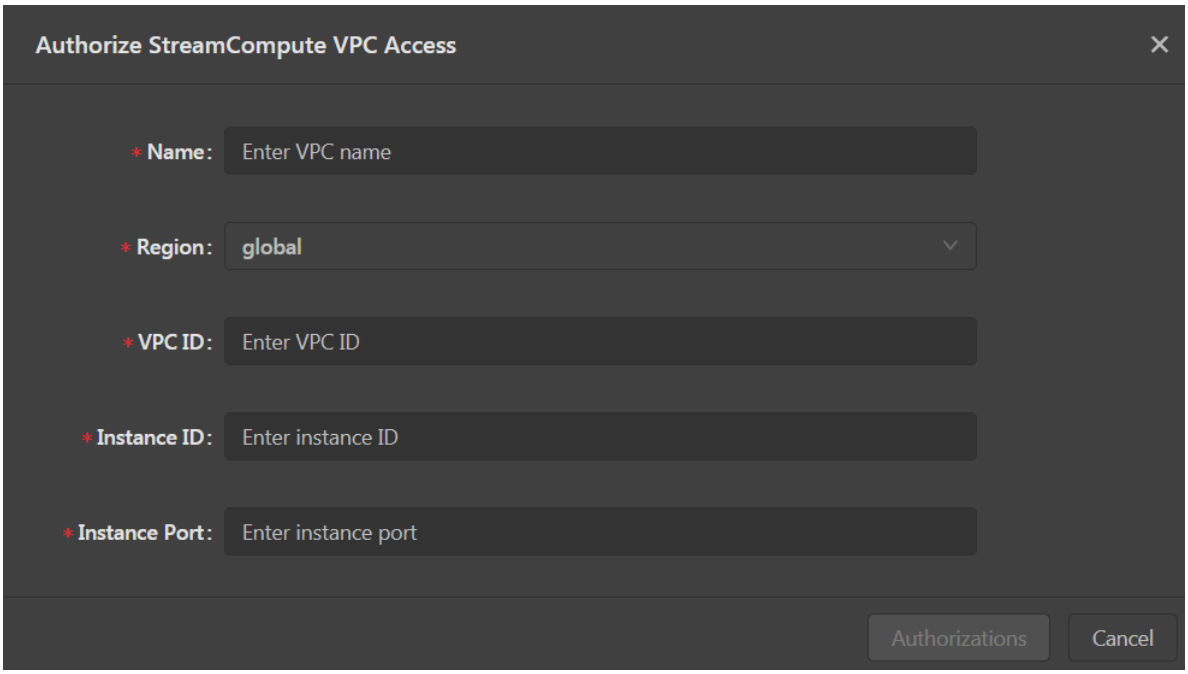
2. Seclect **VPC Access Authorization** in the left navigation pane.

Figure 24-6: VPC authorization



3. Authorize StreamCompute to access VPC.

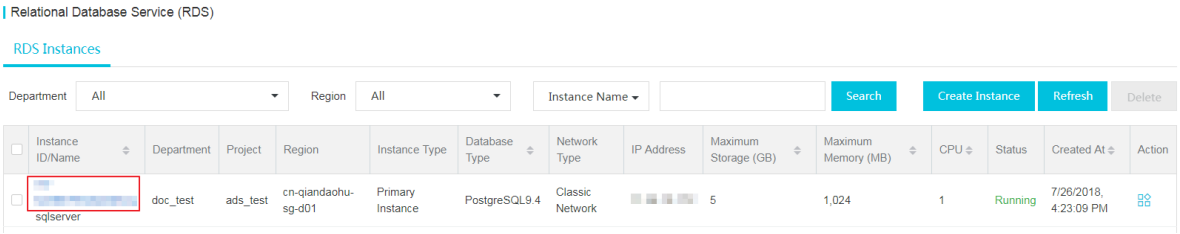
Figure 24-7: Authorization



Region refers to an RDS region.

4. View the RDS instance ID.

Figure 24-8: Instance information



5. Select the data storage and enter registration information.

Plaintext mode:

If you are using the storage resources under the primary account, the plaintext mode is not recommended for RDS.

If you are not using the storage resources under the primary account, you cannot register the data storage of another user. For example, if instance A of RDS is owned by user A, user B cannot use instance A in StreamCompute. You need to inject the database into the task in plaintext.

User B needs to enter URL, username, password, and tablename under the With parameter according to instance A, as shown in the following figure:

Figure 24-9: Configuration information

In the plaintext mode, you need to set a whitelist. Follow these steps:

How to add an RDS whitelist?

Part of data storage aims to meet security requirements. If the whitelist mechanism is enabled, only the whitelisted IP addresses can access the links, but this will prevent other Alibaba Cloud products from writing into data storage. For RDS, the created RDS database completely rejects all external accesses. You need to add a whitelist so that the whitelisted IP addresses can access

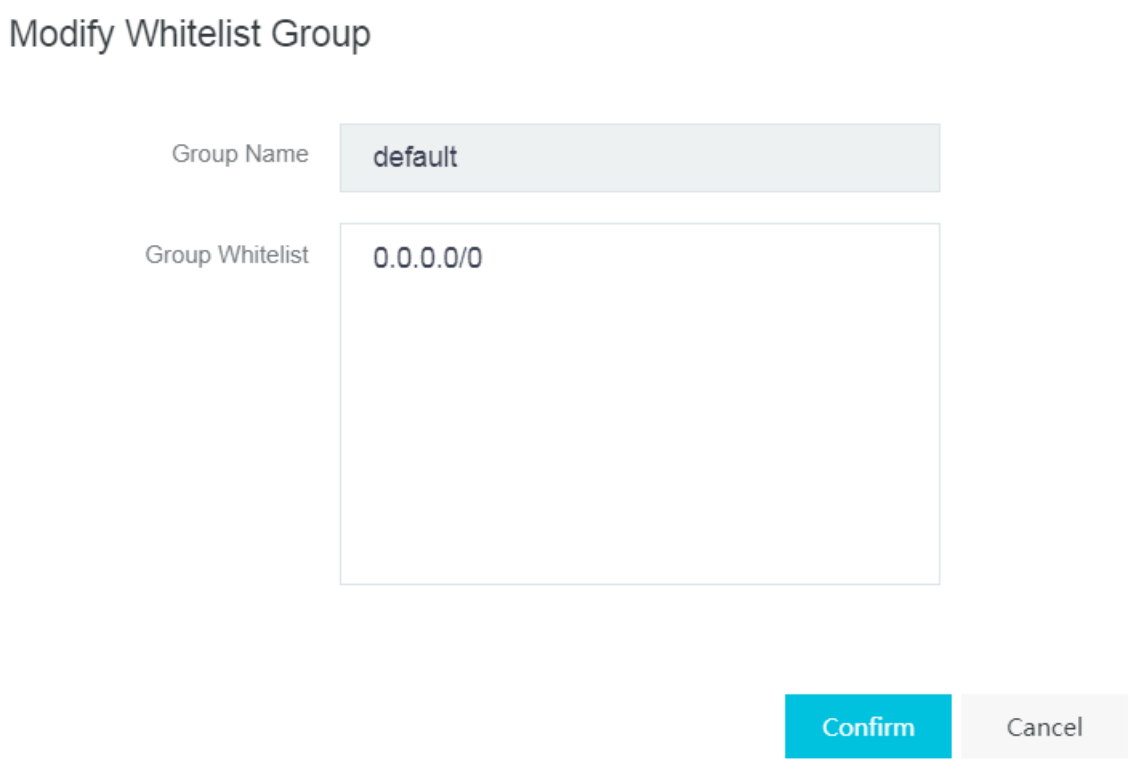
RDS. Similarly, StreamCompute needs to read from and write into the RDS database multiple times if it uses RDS as a dimension table or result table. If the whitelist mechanism is enabled, you need to whitelist WebConsole and Worker of StreamCompute before they access RDS.

RDS supports access from internal network IP addresses and Internet IP addresses (StreamCompute currently does not support VPC addresses). You only need to add an address segment whitelist for StreamCompute to access RDS. The specific IP address range is as follows:

Follow the following steps:

1. Log on to the RDS console and click the target instance name.
2. Select **Security Control** > **Whitelist Settings** in the left navigation pane.
3. Select the default group and click the edit icon.
4. On the **Modify Whitelist Group** page, delete the default whitelist 0.0.0.0/0, enter a custom whitelist, if you want to set more than one IP, separate them with a comma(,). Click **Confirm**.

Figure 24-10: Modify a whitelist



You can also click clear icon after the default group to remove the whitelist from the default group. Then, click **Add Whitelist Group** to create a custom group.

1. Click **Add Whitelist Group**.

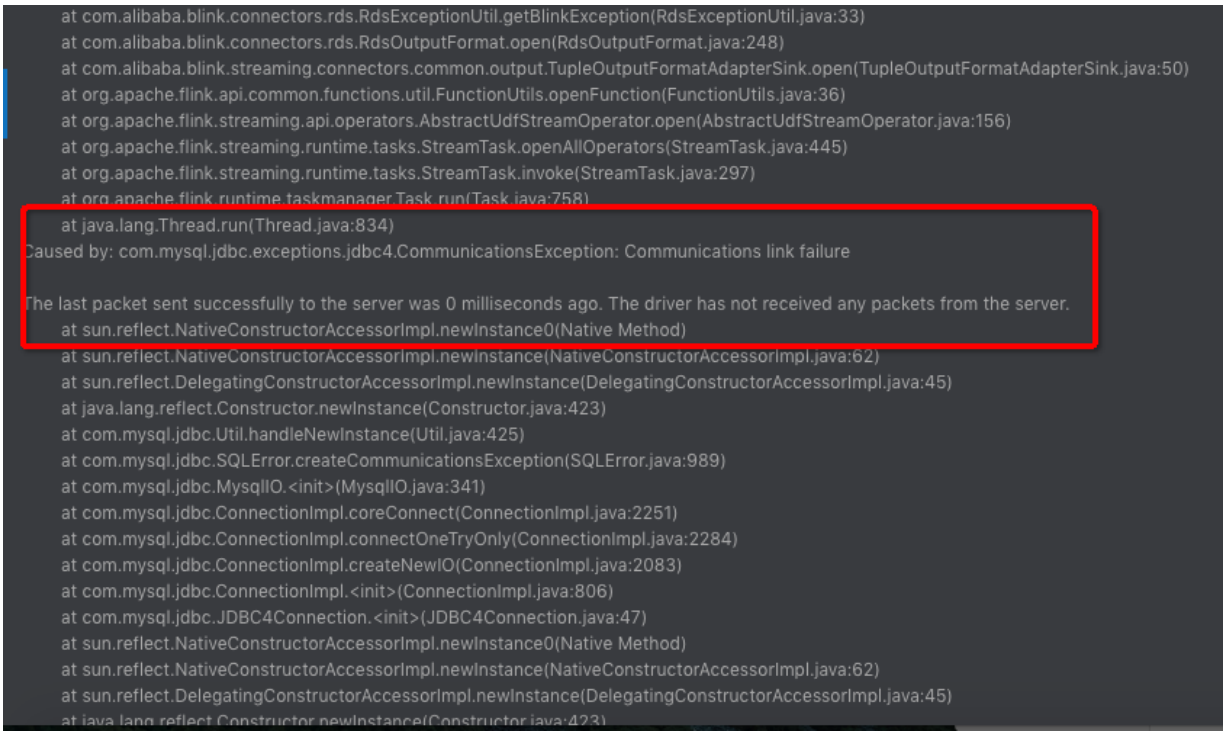
> **Note:**
> If the whitelist only contains 0.0.0.0/0, this instance can be accessed from any IP address.

2. Set the **`Group Name`** and **`Group Whitelist`**, and then click **Confirm**.
   - Group name: The group name contains 2 to 32 characters which consist of lowercase letters, digits, or underscores. The group name must start with a lowercase letter and end with a letter or digit. The default group cannot be modified or deleted.
   - Group whitelist: Enter the IP addresses or IP address segments allowed to access the database. IP addresses or IP address segments are separated by commas (,).

Note

Problem Description

Exception stacks are reported during running, as shown in the following figure:



Solution

Add the IP address of your region to the whitelist. For detailed steps, see How to add an RDS whitelist in this document.

24.3.3 Data development

24.3.3.1 Development stage

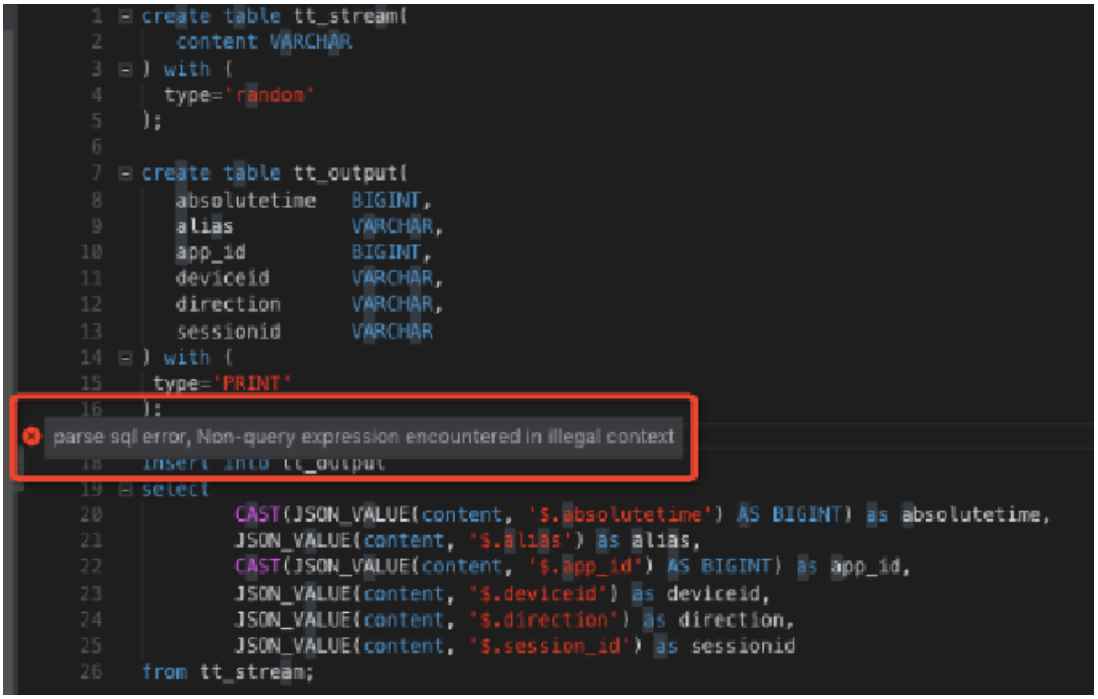
24.3.3.1.1 SQL assistance

The data development module provides a complete set of online SQL IDE tools with the following functions to assist you in BlinkSQL development:

- **BlinkSQL syntax check**

  The IDE text is saved automatically after modification, which can trigger the SQL syntax check function. If an error is found, the error line, column, and cause are prompted on the IDE interface.

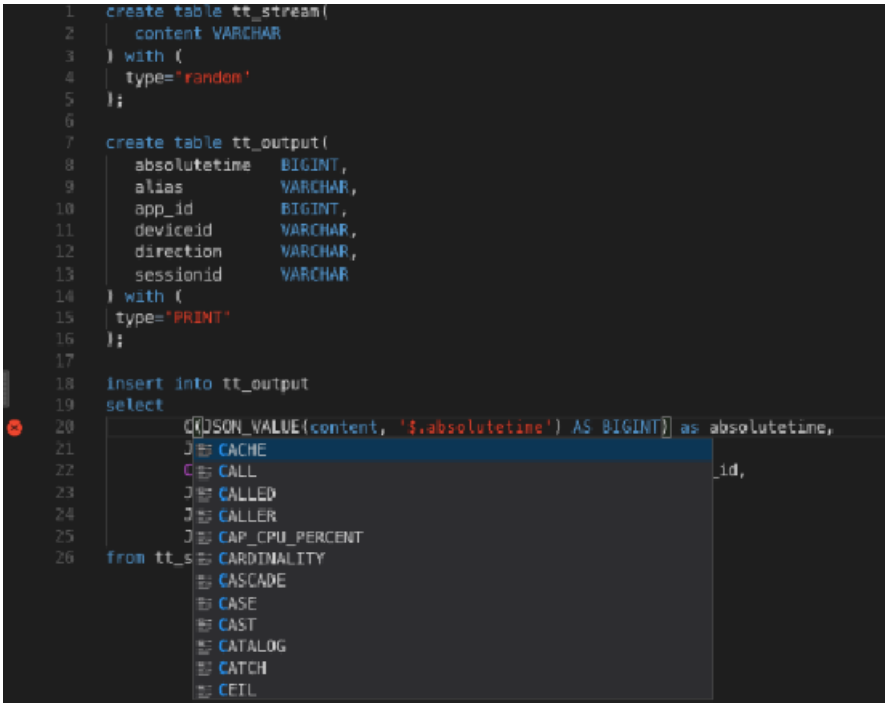
  Figure 24-11: Exception information chart



- BlinkSQL intelligent prompt

  When you enter BlinkSQL codes, IDE provides intelligent prompts for keywords, built-in functions, tables/fields, and so on.

Figure 24-12: Exception information chart



- BlinkSQL syntax highlight

  For BlinkSQL keywords, IDE provides different syntax highlighting colors to distinguish different BlinkSQL structures.

24.3.3.1.2 SQL version management

Our data development covers key areas of routine development, including code assistance and code version. The data development module provides a code version management function. Each submission can generate a code version for tracking and rollback in the future.

- Version management

  Each time you submit or launch a task, StreamCompute generates a code snapshot for code tracking in the future. Click **Data Development** > **Task Attribute**. All version information of the task is listed under **`Task Attribute`**.

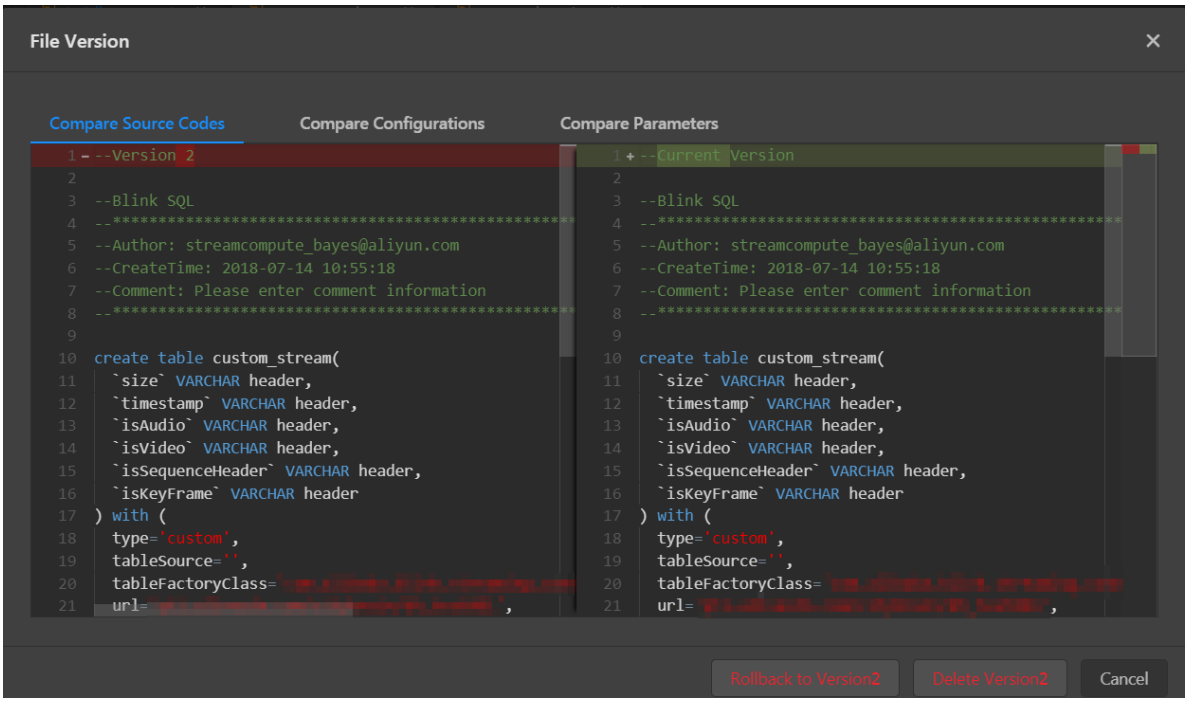
- Version deletion

  Each time you submit or launch a task, StreamCompute generates a code snapshot for code tracking in the future. StreamCompute sets the maximum number of versions for each user . By default, VPC can have 20 versions at most. For details in other environments, consult the StreamCompute system administrator. If the number of generated versions exceeds the

limit, the system rejects the submission and reports an error to instruct you to delete some old versions.

Navigate to **Development** > **Task Attribute** and click **Details** under **Version List** to delete the invalid versions. Then, the task can be launched.

Figure 24-13: Code comparison



24.3.3.1.3 Data storage management

The **Development** page provides a complete set of tools for data storage management. You can register a data source on the**Development** page to enjoy multiple traversal data storage services, as shown in the following figure:

- **Data preview**

  The Data Development page provides the data preview function for various types of data storage. The data preview function can effectively assist you in learning upstream and downstream data features. This function enables you to identify key service logic, and quickly complete service development.

- **Assistant DDL generation**

  Most DDL generation tasks of StreamCompute are monotonous translation work. The DDL statements for data storage that need to be mapped are manually translated into the DDL statements of StreamCompute. StreamCompute provides the assistant DDL generation

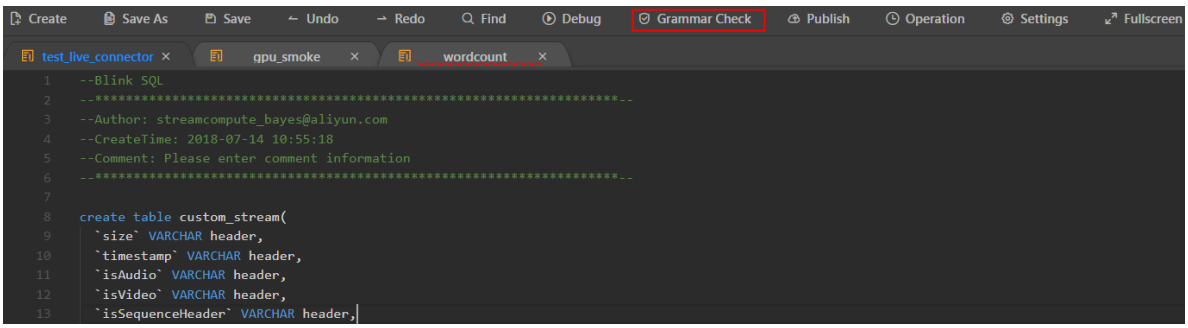
function to further simplify your streaming task coding work. This service effectively reduces the error rate of SQL codes and improves the output efficiency of streaming computing services.

24.3.3.2 Debugging stage

The data development module provides a simulated running environment for you to upload data, simulate running, and check output results. When you complete all service logic, follow the steps shown in the following figure:

1. Check the syntax. Check whether the SQL csde has a syntax error. An error message will be directly displayed if there is an error.

   Figure 24-14: Syntax check



2. Debug the task. StreamCompute supports the following two modes for task debugging:
   - Construct test data independently.
   - Extract data from a data source table either randomly or sequentially. **Note that this mode is available only when you use the data storage registration function**. **It is recommended to use plaintext instead of Excel files to extract data; otherwise, the data test results may be inaccurate**.
3. View debugging results.

The following functions can be realized by running BlinkSQL in this environment:

- **Completely isolated production**

  In the debugging environment, BlinkSQL runs in a separate debugging container and all outputs are directly rewritten to the debugging result screen. This has no impact on the online streaming computing tasks and data storage systems, allowing you to run tasks freely.

  The data debugging output is not really written into the external data source. It is intercepte d by StreamCompute and output to the debugging result screen. Failures may occur during online running due to the format written into the target data source. Such failures cannot be

completely avoided at the debugging stage because they can only be found during online running. If the result data is output to the RDS system and the length of string data from some fields is larger than the maximum value allowed for the RDS table, we cannot detect such a problem in the debugging environment. However, an exception will occur during online production and running. In the future, StreamCompute will support local debugging and write debugging results to real data sources. This function helps users shorten the gap between debugging and production, and solves problems at the debugging stage.

- **Supports test data construction**

  In the debugging environment, BlinkSQL does not read data from the source data storage system, including DataHub streaming input and RDS dimension table input. You need to construct a test data set and upload it to the **`data development`** module.

  StreamCompute provides different test data templates for different tasks. You can download a data template to construct data for testing.

  > **Note:**
  > To avoid errors, we recommend that you download a data template to construct data.

- **Debugging data separator**

  By default, the data in a debugging file is separated by commas (,). For example, in the following test file sample:

  ```
  id,name,age 1,alicloud,13 2,stream,1
  ```

  If you have not specified a separator, the data is separated by commas (,). However, if you use the JSON format for a field, the field content contains commas (,). In this case, you need to specify a separator other than comma (,).

  > **Note:**
  > For separators, StreamCompute only supports a single English character instead of a string (such as aaa).

  ```
  id|name|age 1|alicloud|13 2|stream|1
  ```

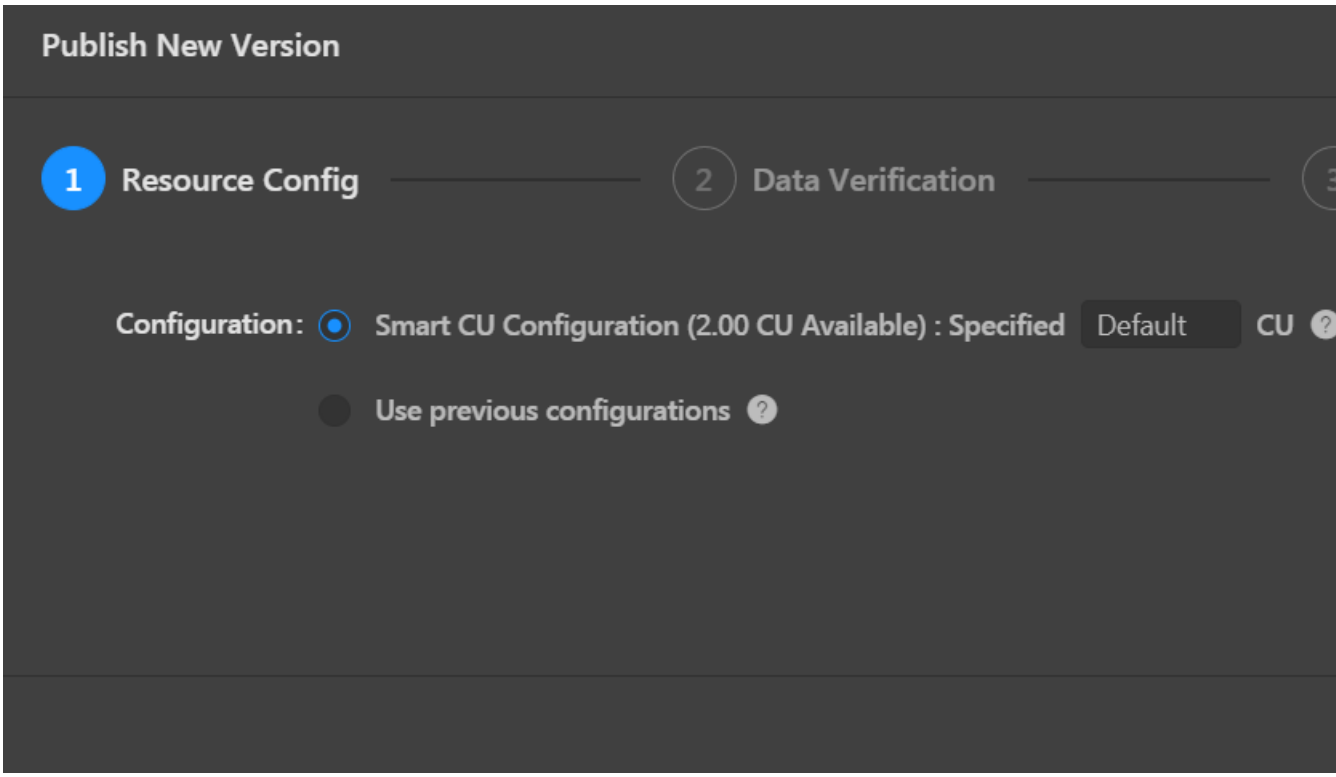
  You need to set debug.input.delimiter=| for the task parameter in the data storage system.

24.3.3.3 Publishing stage

Procedure

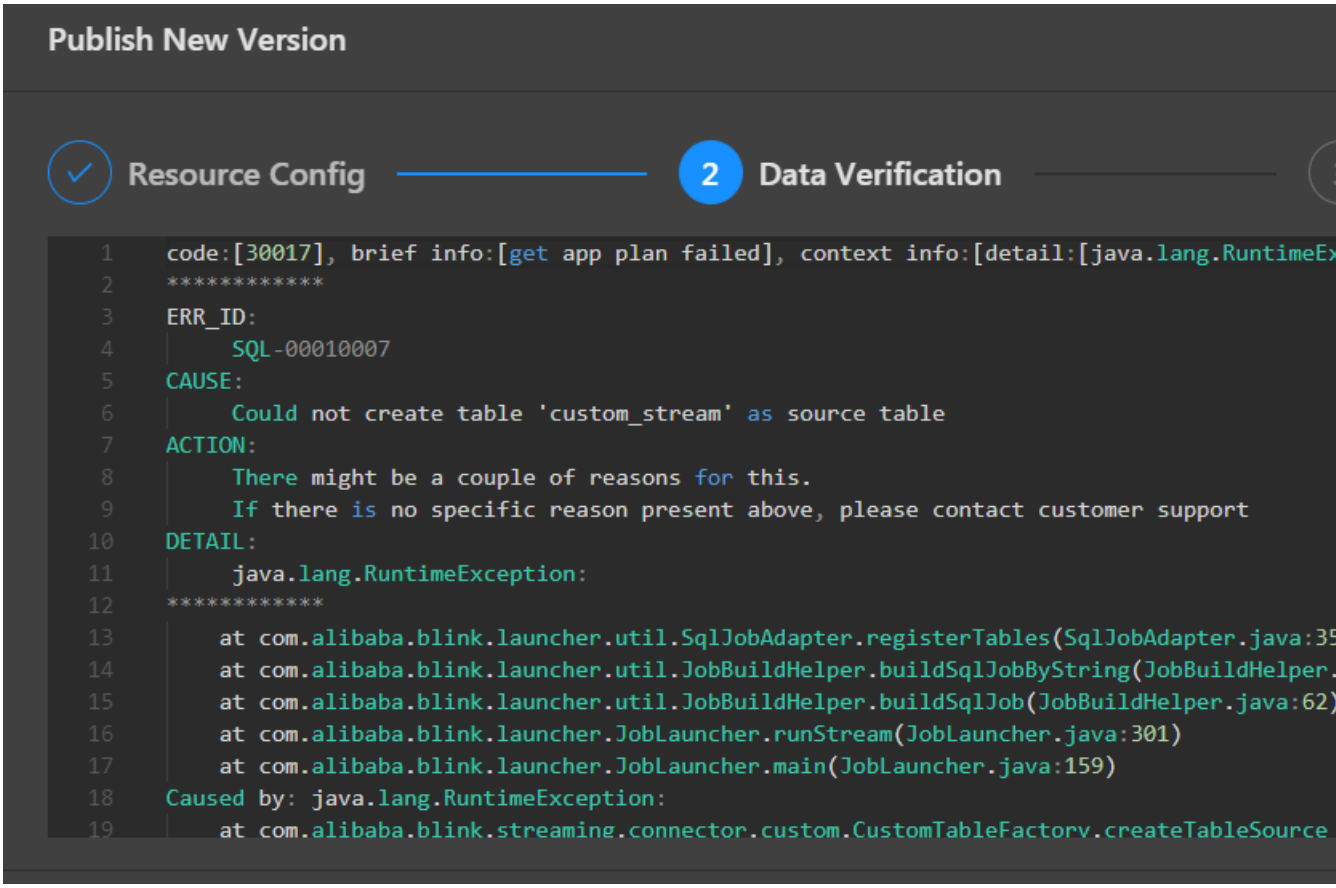
1. After the development and debugging stages are completed and no error of BlinkSQL is found, click **Publish**.
2. Click **Smart CU Configuration** and select the default system configuration for the first time (no need to specify the number of CUs). Click **Next**.

Figure 24-15: Resource configuration



3. Check data. Then, click **Next**.

Figure 24-16: Data check



4. Click **Publish** to publish the job.
5. On the **O&M** page, click a task to start it.

   All your modifications and debugging operations on the **Development** page are completely isolated from the production debugging on the **O&M** page. This avoids interference between development and production.

25 E-MapReduce

25.1 Product introduction

25.1.1 What is EMR

E-MapReduce (EMR) is a one-stop big data processing and analysis service that uses resources of the open-source big data ecosystem, including Hadoop, Spark, Kafka, and Storm, to provide users with the cluster, job, and data management functions.

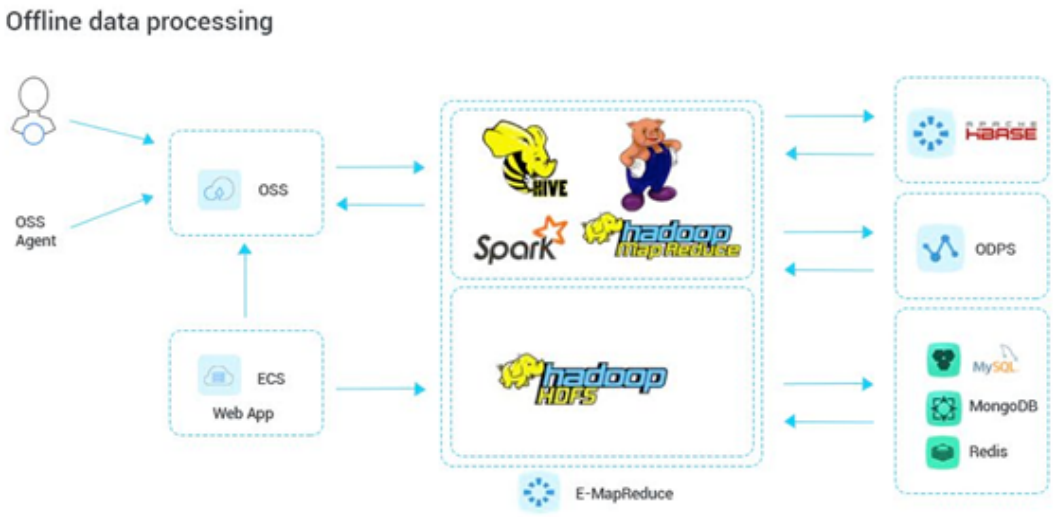
EMR is built on Alibaba Cloud Elastic Compute Service (ECS) and is based on open-source Apache Hadoop and Apache Spark. It allows you to use other peripheral systems, such as Apache Hive, Apache Pig, and HBase, in the Hadoop and Spark ecosystems to analyze and process your own data. Moreover, you can use EMR to easily import and export the data to other cloud data storage and database systems of Alibaba Cloud, such as Alibaba Cloud Object Storage Service (OSS) and ApsaraDB for RDS.

25.1.2 Scenarios

- **Offline data analysis**

  After synchronizing massive logs of games, Web applications, mobile apps, and other services from servers to EMR data nodes, you can quickly obtain data insights using tools such as Hue and mainstream computing frameworks such as Hive, Spark, and Presto. You can also load data distributed among ApsaraDB for RDS instances or other storage engines using tools, such as Sqoop, and synchronize the analyzed data to ApsaraDB for RDS instances, to provide data support for data visualization products.

Figure 25-1: Offline data analysis

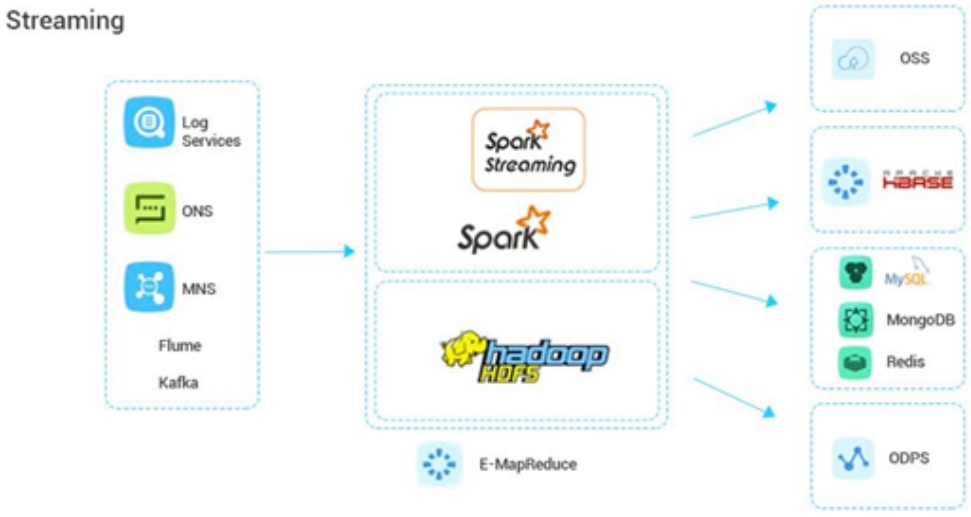


- **Streaming data analysis**

  With Spark Streaming and Storm, you can use and process real-time data of Alibaba Cloud Log Service (Log), Message Queue (MQ), Message Service (MNS), Apache Kafka, or other streaming data of data streams.

  You can also analyze the streaming data in a fault-tolerant way and write the corresponding results to Alibaba Cloud Object Storage Service (OSS) or Hadoop Distributed File System (HDFS).

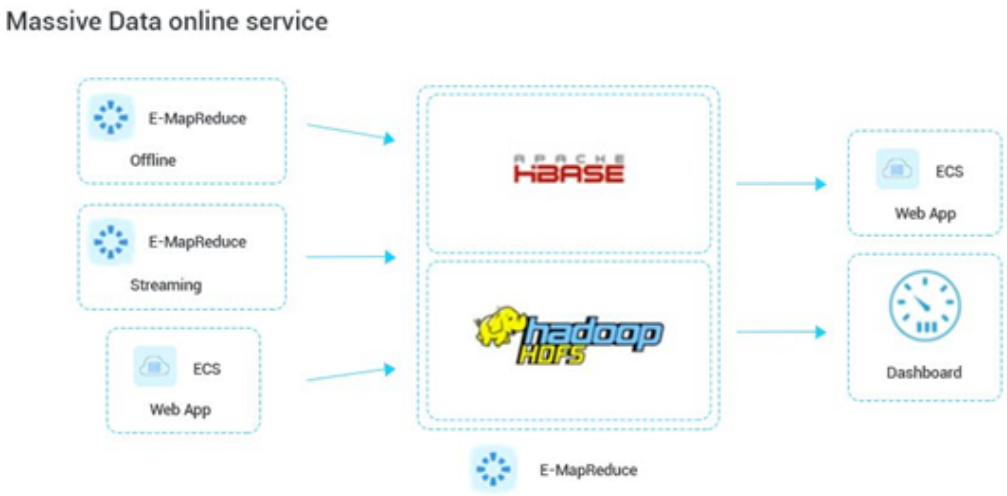
Figure 25-2: Streaming data analysis



- **Online massive data analysis**

EMR analyzes petabytes of structured, semi-structured, and unstructured data generated by your Web and mobile applications online, facilitating the Web applications or data visualization products to obtain the analysis results and display them in real time.

Figure 25-3: Online massive data analysis



25.2 Software configuration

25.2.1 Software environment

Table 25-1: Software environment describes the software environment.

Table 25-1: Software environment

Software	Description
Operating system	CentOS 7 64-bit kernel-3.10.0-693.2.2.el7.x86_64
JDK	OpenJDK 1.8.0

25.2.2 Software list

Table 25-2: Software list lists the software and versions.

Table 25-2: Software list

Software	Version
Hadoop	2.7.2

Software	Version
Hive	2.0.1
Tez	0.8.4
Spark	2.1.1
Oozie	4.2.0
Hue	3.12.0
Zeppelin	0.7.1
Sqoop	1.4.6
Knox	0.13.0
ZooKeeper	3.4.6
Ganglia	3.7.2
Pig	0.14.0
Kafka	2.11_0.10.1.0
HBase	1.1.1
Phoenix	4.10.0
Presto	0.188

25.2.3 Software description

- **Hadoop**
  - YARN

    Schedules tasks and manages cluster resources.
  - HDFS

    Provides a distributed file storage system.
- **Hive**

  Hadoop-based offline data processing system that provides SQL-like query syntax for data analysis and processing, and stores data in tables with table management capabilities.
- **Spark**

  Memory-based new-generation distributed computing framework that supports offline and real-time computing, SQL syntax, and machine learning.
- **Oozie**

Job scheduling engine that supports complex DAG orchestration of jobs of various types.

- **Hue**

  Visualized platform that manages open source components such as Hadoop, Hive, Oozie, and HBase.

- **Sqoop**

  Data migration tool that supports migration of data between ApsaraDB for RDS and HDFS.

- **ZooKeeper**

  Distributed open source application coordination service as an open source implementation of Google Chubby and an important component of Hadoop and HBase. As the software offering the consistency service for distributed applications, it provides features such as configuration maintenance, domain name services, distributed synchronization, and group services.

- **Kafka**

  Kafka is a high-throughput distributed message system featuring scalability, high reliability, and high performance. It is widely used for real-time computing, log processing, aggregation, and other scenarios.

- **HBase**

  As a distributed and column-oriented open source database, HBase is a subitem of Apache Hadoop. Different from common relational databases, HBase is applicable to unstructured data storage and works using the column oriented storage, instead of the row oriented storage.

- **Phoenix**

  Provides SQL-like syntax for analysis of HBase data.

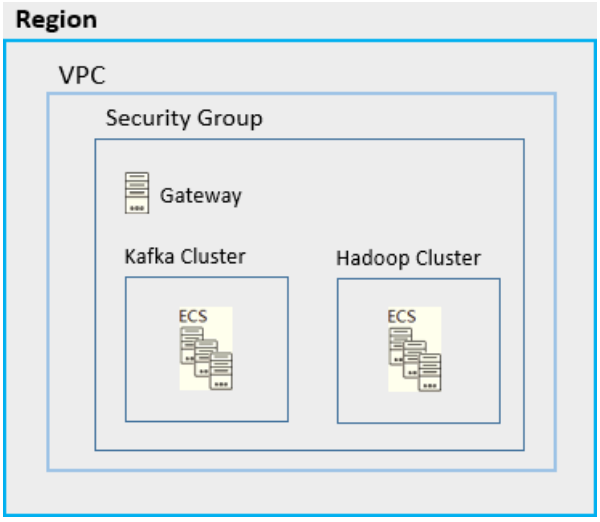
- **Presto**

  Presto is a distributed SQL query engine that queries big data sets distributed among one or more data sources.

25.3 Hardware description

Figure 25-4: Hardware architecture shows the hardware architecture.

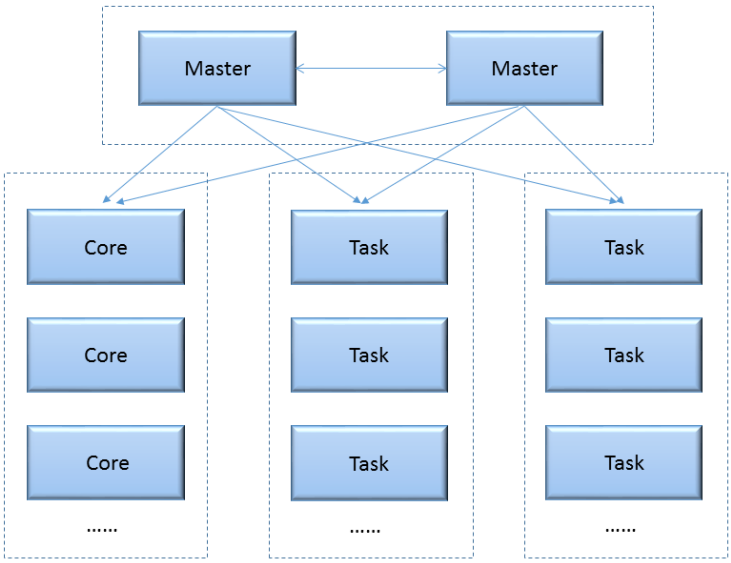
Figure 25-4: Hardware architecture



25.3.1 Node composition

Figure 25-5: Node composition shows the node composition.

Figure 25-5: Node composition



25.3.2 Hardware selection

Currently, EMR provides services based on the ECS platform. You can use all ECS-supported models for cluster nodes.

Table 25-3: Hardware list

Hardware	Configuration
CPU	Minimum configuration: 4-core Recommended configuration: 32-core The standalone mode is supported.
Memory	Minimum configuration: 32 GB Recommended configuration: ≥ 64 GB
Disk	Management node (master): • System disk ≥ 200 GB cloud disk; 1 disk • Data disk ≥ 500 GB cloud disk; 1 disk Data node (core): • System disk ≥ 100 GB cloud disk; 1 disk • Data disk: — Ephemeral disk, configured based on the model; at least 4 disks and at most 12 disks — Cloud disk, configured as required; 4 disks Computing node (task): • System disk ≥ 100 GB cloud disk; 1 disk • Data disk ≥ 500 GB cloud disk; 4 disks; configured based on the actual capacity
Network	The classic network and VPC are supported. VPC is recommended.

25.4 Deployment description

25.4.1 Deployment modes

EMR supports the following deployment modes:

- Hybrid deployment

  EMR supports the full-cluster hybrid deployment mode in which all the components are deployed in one cluster. Multiple services are deployed on one node and run synchronously.
- Independent deployment

  Only one service is deployed on an EMR cluster and runs independently.

25.4.2 Service list

Table 25-4: Service list lists EMR-supported services.

Table 25-4: Service list

Service	Role	Deployment method
Hadoop HDFS	NameNode	It is deployed on the master node. In HA mode, it is deployed on two master nodes.
	DataNode	It is deployed on the core node.
	ZKFC	It is deployed on the master node. In HA mode, it is deployed on two master nodes.
	JournalNode	• In non-HA mode, it is deployed on the master node and the first and second core nodes. • In HA mode, it is deployed on two master nodes and the first core node.
	KMS	It is deployed on the master node and supports only single-node deployment.
	HttpFS	It is deployed on the master node and supports only single-node deployment.
Hadoop YARN	ResourceManager	It is deployed on the master node. In HA mode, it is deployed on two master nodes.
	NodeManager	It is deployed on the core node.
	JobHistory	It is deployed on the master node and supports only single-node deployment.

Service	Role	Deployment method
	TimeLineServer	It is deployed on the master node and supports only single-node deployment.
	WebAppProxyServer	It is deployed on the master node and supports only single-node deployment.
Hive	HiveServer	It is deployed on the master node. In HA mode, it is deployed on two master nodes .
	HiveMetaStore	It is deployed on the master node. In HA mode, it is deployed on two master nodes .
Spark	JobHistory	It is deployed on the master node and supports only single-node deployment.
Ganglia	GMond	It is deployed on all nodes to collect information.
	GMetad	It is deployed on the master node and supports only single-node deployment.
HBase	HMaster	It is deployed on the master node. In HA mode, it is deployed on two master nodes .
	HRegionServer	It is deployed on the core node .
	ThriftServer	It is deployed on the master node and supports only single-node deployment.
ZooKeeper	ZooKeeper	• In non-HA mode, it is deployed on the master node and the first and second core nodes.

Service	Role	Deployment method
		• In HA mode, it is deployed on two master nodes and the first core node.
Hue	Hue	It is deployed on the master node. In HA mode, it is deployed on two master nodes .
Oozie	Oozie	It is deployed on the master node. In HA mode, it is deployed on two master nodes .
HAS	HASServer	It is deployed on the master node. In HA mode, it is deployed on two master nodes .
Knox	Knox	It is deployed on the master node. In HA mode, it is deployed on two master nodes .

25.5 O&M

25.5.1 Complete GUI-based O&M

Service component monitoring overview

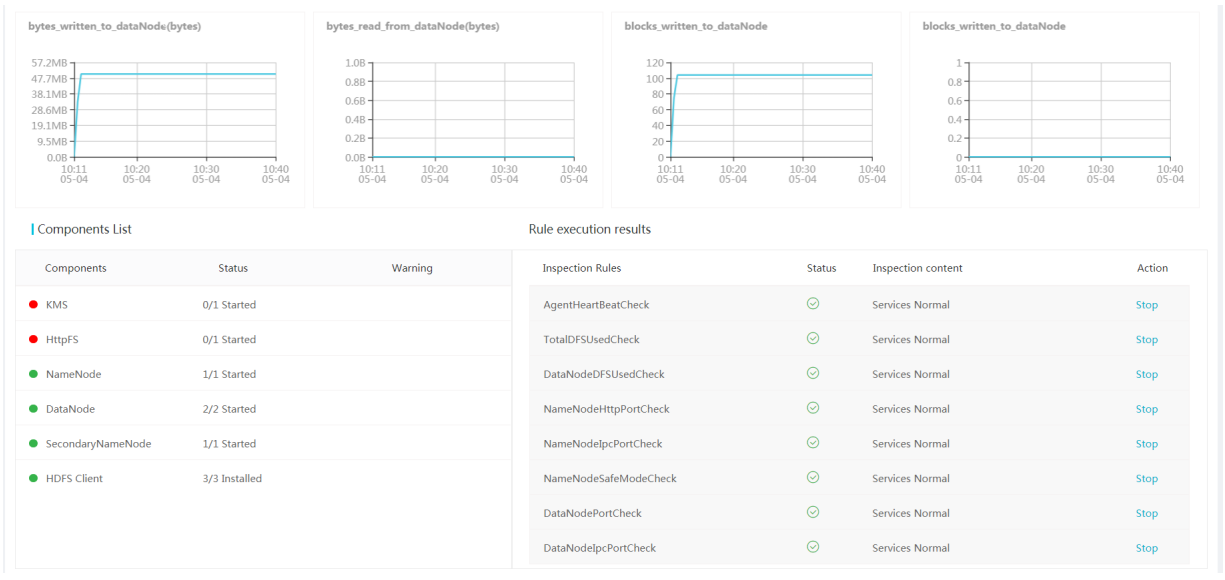
You can monitor all components at the background and view historical records at any time. Data is updated in real time.

Figure 25-6: Service component monitoring overview

Detailed monitoring of a single service component and periodic automatic inspection

If a service contains multiple sub-services, you can view details of all the sub-services. Health check is enabled for all services by default, which is updated with the system upgrade.

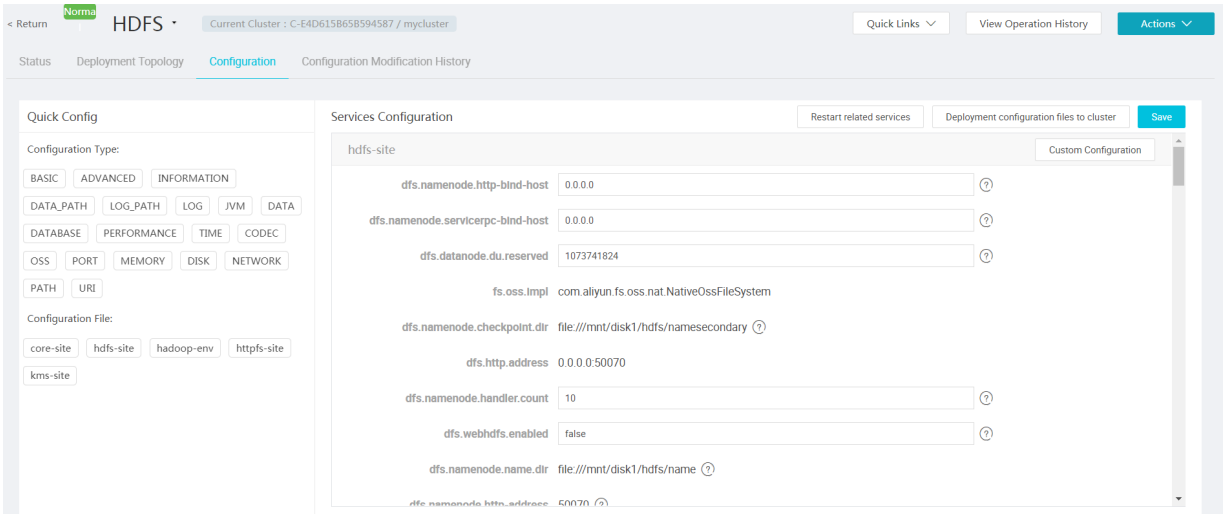
Figure 25-7: Monitoring of a single service component



Modification of service component configuration

You can configure service parameters on the interface and add custom parameters as required. The system automatically prompts the affected services based on the modified parameters. Ripple restart is supported, without affecting external services.

Figure 25-8: Modification of service component configuration



25.5.2 O&M methods

The EMR smart control system performs O&M of the basic cluster environment. In the case of hardware damage or shutdown, services are migrated in the background without user intervention.

All service processes in the cluster software environment are monitored in real time. A collapsed process is restarted in real time. This function applies to most of the scenarios. If problems cannot be automatically solved by the system, you can contact the EMR attendant to manually solve the problems. In practice, you can configure the services, such as modifying the memory, to solve most of the similar problems.

We do not recommend that you modify the cluster without permission. Inform us if any problem occurs due to the modification. We will assist you in troubleshooting and solving the problem.

Problems in your business scenarios are not technical issues in the cluster environment. In this case, we provide consultation rather than solving the problems.

25.6 Disaster tolerance

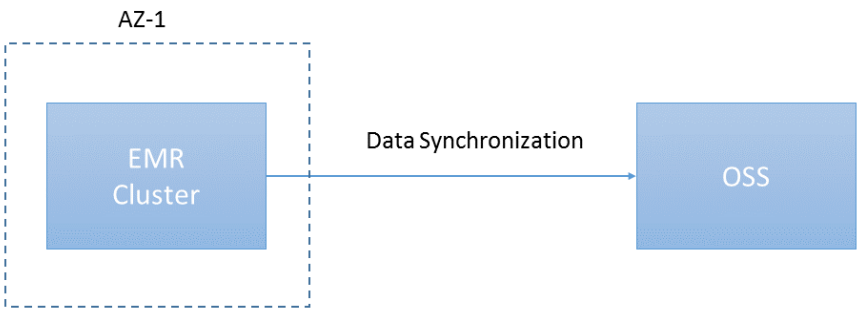
25.6.1 Data disaster tolerance

Hadoop Distributed File System (HDFS) stores data of each file by block and stores multiple copies of each data block (three copies of each data block by default) to ensure that the data block copies are distributed to different racks. In most cases, HDFS stores three copies in two nodes of the local rack and a node of another rack, respectively.

HDFS regularly scans the data copies. If HDFS finds that a data copy is lost, it quickly duplicates data to ensure sufficient copies. If finding that a node is lost, HDFS quickly duplicates all data on the node to restore services. On Alibaba Cloud, if cloud disks are used, each cloud disk on the background has three data copies. If any copy is faulty, the copy data is automatically switched and recovered to ensure data reliability.

Hadoop HDFS is a high-reliability data storage system that has passed through long-term tests. It has the capability to reliably store massive data. Based on the cloud features, HDFS can perform additional data backup on services such as OSS to achieve higher data reliability.

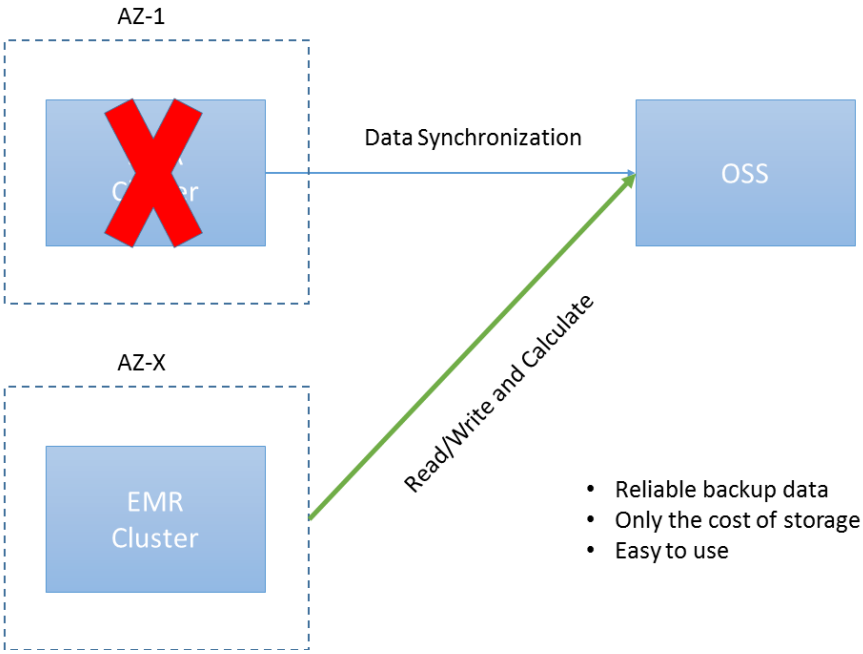
Figure 25-9: Data synchronization



You can use EMR scheduled tasks or other scheduled tasks to periodically synchronize data to OSS.

The synchronization interval determines the tolerable time range of data loss. For example, if data is synchronized each hour, data within an hour may be lost. If data is synchronized every 30 minutes, data within 30 minutes may be lost.

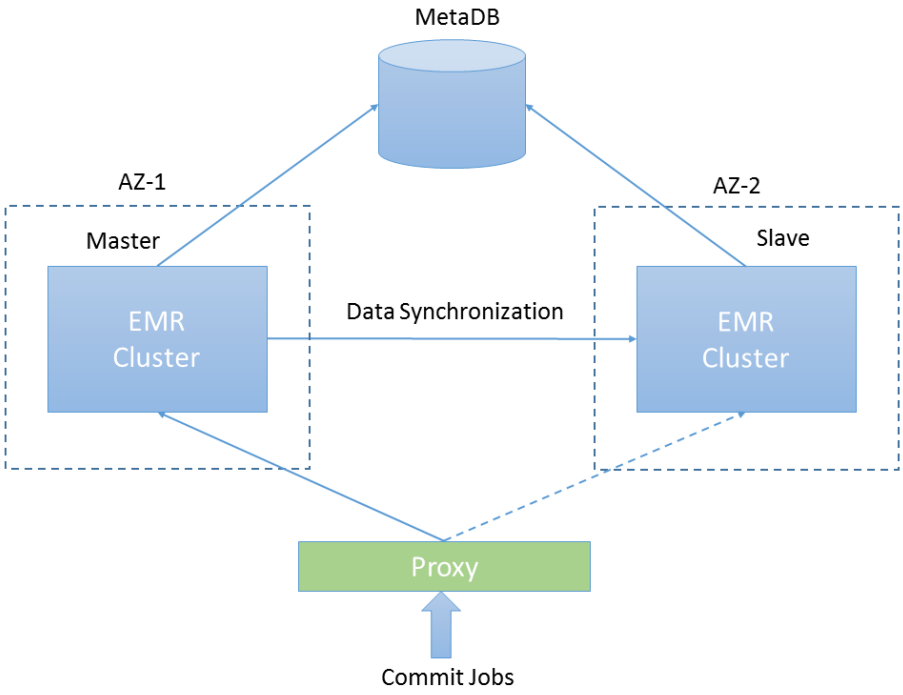
Figure 25-10: Data backup



When an exception occurs in a cluster, another cluster is created to read and write OSS data. If the original cluster contains metadata, the metadata must be rebuilt to ensure the service. When the original cluster is restored, changed data is directly synchronized from OSS to the original cluster.

25.6.2 Service disaster tolerance

Figure 25-11: Service disaster tolerance



In two-node cluster mode, two identical clusters with the same computing and storage capabiliti es are created on two AZs, respectively. A disaster-tolerant database similar to the ApsaraDB for RDS three-node database is used as the meta database for both clusters. Data is synchronized between the clusters to ensure that data in the master cluster is synchronized to the slave cluster in near real time. A method for data synchronization is to start the scheduled task DistCp, which provides low data disaster-tolerant timeliness. Another method is to monitor data in the master cluster. The incremental data is synchronized to the slave cluster in real time, to ensure data synchronization in seconds.

You can use the proxy provided on the front end to submit jobs. You do not need to know the target cluster to which jobs are submitted. By default, the jobs are submitted to the master cluster. If the master cluster cannot be accessed, jobs are automatically submitted to the slave cluster.

Note the scenarios in which external data is written to the cluster. When the master cluster fails, all external data is switched to the standby cluster. Therefore, the external data sources must be disaster-tolerant.

25.7 User operation description

You can access services of a cluster in either of the following ways:

- Access the cluster services through Gateway and submit computing jobs. You can also deploy standalone apps on the cluster to control the jobs.
- Access the cluster services directly from other positions, for example, an independent app.

25.7.1 Log on to E-MapReduce Console

Procedure

1. In the menu bar at the top of the Apsara Stack Console page, select **Console** > **Big Data** > **E-MapReduce**.
2. In the E-MapReduce product page, select your region and department in the drop-down menus respectively, and then click **EMR** to enter the E-MapReduce console.

25.7.2 Gateway

Gateway provides the following clients:

- Hadoop
- Spark
- Hive
- Oozie
- HBase

Gateway also provides the Kerberos authentication environment. You must be authenticated using the corresponding account to access a cluster.

Gateway updates information about all IP addresses of the cluster in /etc/hosts. To use the information in other locations, copy all the Hadoop cluster hosts from this directory to the target directory.

25.7.3 Log on to Gateway

Developers have their own Gateway with qualified configuration.

Obtain the logon address and password from the administrator.

Logon method: `sshroot@gateway_ip`

Gateway provides a root account for node operations and a Hadoop cluster authentication account for cluster operations such as job submission. You can run the su command to switch between the accounts.

Switch to the authentication account: `su user_account`

Note:

user_account indicates the distributed authentication account.

25.7.4 Software environment description

The basic environment has been configured on Gateway.

- MR

  Software location: /usr/lib/hadoop-current

  Configuration location: /etc/ecm/hadoop-conf
- Spark

  Software location: /usr/lib/spark-current

  Configuration location: /etc/ecm/spark-conf
- Hive

  Software location: /usr/lib/hive-current

  Configuration location: /etc/ecm/hive-conf
- Oozie

  Software location: /usr/lib/oozie-current

  Configuration location: /etc/ecm/oozie-conf
- HBase

  Software location: /usr/lib/hbase-current

  Configuration location: /etc/ecm/hbase-conf
- Phoenix

  Software location: /usr/lib/phoenix-current

  Configuration location: /etc/ecm/phoenix-conf

All paths have been added and can be used directly.

Hosts have been configured on Gateway. You do not need to modify any configuration.

When the nodes of the cluster change, the hosts must be updated. The administrator periodically updates the host information.

25.7.5 Security authentication description

Kerberos authentication is enabled for clusters by default. You can obtain the Kerberos account and password from the cluster administrator.

The following uses user ali as an example to describe how to perform security authentication.

- User name: ali
- Password: ali123
- principal: aliyun@EMR.xxxxx.COM

You can run klist on Gateway to view the current user authentication information.

Example:

```
Ticket cache: FILE:/tmp/krb5cc_1000
Default principal: xxx@EMR.xxxx.COM
Valid starting Expires Service principal
10/19/2017 10:49:16 07/03/2023 18:49:16 krbtgt/EMR.xxxxx.COM@EMR.xxxxx
.COM
renew until 10/21/2017 10:49:16
```

The preceding response indicates that the local Gateway is valid until 2023.

You can run kinit for authentication. To always validate Gateway, run

```
kinit -l 1000h
```

.

Obtain the keytab file.

Gateway provides the keytab file that is permanently valid. You can find this file in the user root directory `/home/${user}/${user}.keytab`.

For example, if you access user ali, the file is located in /home/ali/ali.keytab.

25.7.6 HDFS environment description

The default HDFS access domain is `hdfs://emr-cluster`, which must be used as the prefix whenever an HDFS domain is used.

The default user HDFS path is `/user/`*`user name`*. For example, the directory of user ali is `/user/ali`.

Hive data storage path: `/user/hive/warehouse/`

25.8 Job submission description

25.8.1 MR

You can run the following Hadoop command to submit common Hadoop MR jobs:

```
hadoop jar /usr/lib/hadoop-current/share/hadoop/mapreduce/hadoop-
mapreduce-examples-2.7.2.jar pi 10 10
```

25.8.2 Spark

- Spark-Core

  ```
  spark-submit --class org.apache.spark.examples.SparkPi --master yarn
  -client --driver-memory 512m --num-executors 1 --executor-memory 1g
   --executor-cores 2 /usr/lib/spark-current/lib/spark-examples-1.6.3-
  hadoop2.7.2.jar 10
  ```

- Spark-SQL

  ```
  spark-sql -e "select * from demo"
  ```

- Spark Streaming

  Same as Spark-Core

  ```
  spark-submit --class org.apache.spark.examples.SparkPi --master yarn
  -client --driver-memory 512m --num-executors 1 --executor-memory 1g
   --executor-cores 2 /usr/lib/spark-current/lib/spark-examples-1.6.3-
  hadoop2.7.2.jar 10
  ```

- Spark-shell

  ```
  spark-shell
  ```

25.8.3 Hive

- Hive console

  After entering Hive, go to the interactive console and enter the commands.

- Hive commands

  ```
  hive –e "select * from ali.xxx"
  ```

```
hive –f example.hql
```

25.8.4 Oozie

Oozie can schedule the MR, Spark, and Hive jobs.

See the following example to configure the HDFS domain and jobTracker used in the job:

```
nameNode=hdfs://emr-cluster
```

```
jobTracker=emr-header-1.cluster-xxxxx:8032
```

In practice, set xxxxx based on the actual conditions of the cluster.

25.8.4.1 Schedule an MR job

job.properties

```
nameNode=hdfs://emr-cluster
jobTracker=emr-header-1.cluster-xxxxx:8032
```

Configure other parameters as required.

workflow.xml

Use the map-reduce action as the job node.

Demo path: `/home/${user}/examples/apps/map-reduce/`

25.8.4.2 Schedule a Spark job

Problems may occur when Kerberos authentication is implemented for Spark jobs that use Spark actions. Therefore, you must run the Spark jobs using the shell.

job.properties

```
nameNode=hdfs://emr-cluster
jobTracker=emr-header-1.cluster-xxxxx:8032
oozie.use.system.libpath=true
```

Configure other parameters as required.

workflow.xml

Run the following command for Kerberos authentication:

```
<credentials>
    <credential name='hcat_auth' type='hcat'>
    <property>
        <name>hcat.metastore.uri</name>
        <value>thrift://emr-header-1.cluster-xxxxx:9083</value>
    </property>
    <property>
```

```
        <name>hcat.metastore.principal</name>
        <value>hive/_HOST@EMR.xxxxx.COM</value>
    </property>
    </credential>
    </credentials>
```

Demo path: `/home/${user}/examples/apps/spark/`

25.8.4.3 Schedule a Hive job

job.properties

```
nameNode=hdfs://emr-cluster
jobTracker=emr-header-1.cluster-xxxxx:8032
oozie.use.system.libpath=true
jdbcURL=jdbc:hive2://emr-header-1.cluster-xxxxx:10000/default
jdbcPrincipal=hive/emr-header-1.cluster-xxxxx@EMR.xxxxx.COM
```

Configure other parameters as required.

workflow.xml

Run the following command to configure the authentication:

```
<credentials>
    <credential name="hs2-creds" type="hive2">
      <property>
        <name>hive2.server.principal</name>
          <value>${jdbcPrincipal}</value>
      </property>
      <property>
       <name>hive2.jdbc.url</name>
         <value>${jdbcURL}</value>
      </property>
    </credential>
  </credentials>
```

Demo path: `/home/${user}/examples/apps/hive2/`

25.9 Software access page

You can quickly view the Web UIs of most core software through the unified cluster portal.

- HDFS UI

  https://{cluster_eip}:8443/gateway/cluster-topo/hdfs/

- Yarn UI

  https://{cluster_eip}:8443/gateway/cluster-topo/yarn/

- SparkHistory UI

  https://{cluster_eip}:8443/gateway/cluster-topo/sparkhistory/

- Ganglia UI

  https://{cluster_eip}:8443/gateway/cluster-topo/ganglia/

- Oozie UIs

  https://{cluster_eip}:8443/gateway/cluster-topo/oozie/

- Hue UI

  http://{cluster_eip}:8888/

26 Quick BI

26.1 Product overview

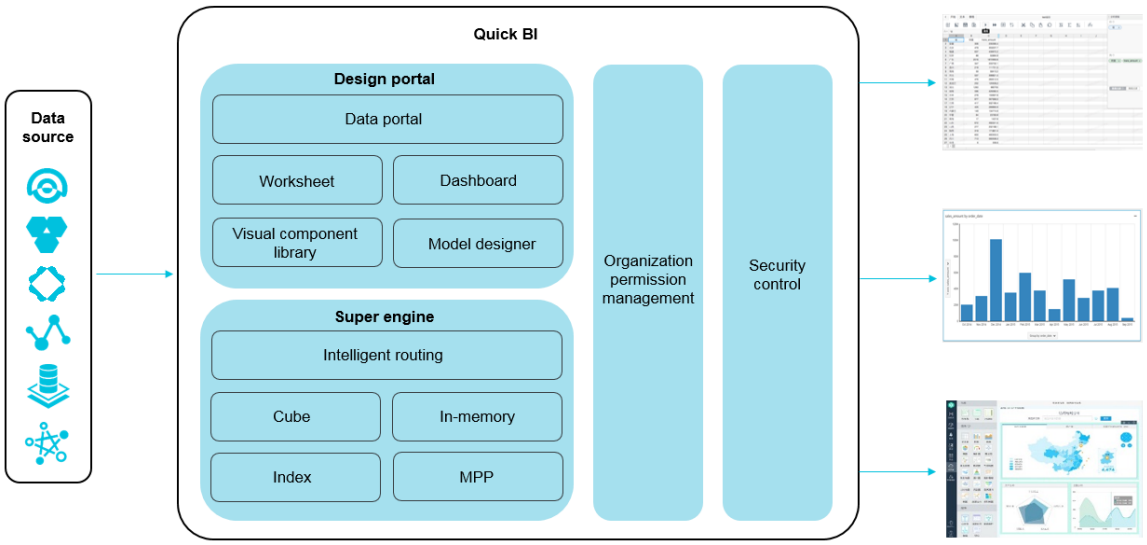
This section describes the features of Quick BI.

Quick BI is a flexible and lightweight self-service platform that provides BI tools based on cloud computing. Quick BI can connect to multiple data sources, including cloud data sources such as MaxCompute (ODPS), ApsaraDB for RDS, AnalyticDB, and HybridDB (Greenplum), as well as the user-created MySQL database and SQL Server database in your ECS instance.

Quick BI supports real-time online analysis of massive data and can return responses in seconds without large amount of data pre-processing. Quick BI also supports terabytes of incremental data on a daily basis.

With an intelligent data modeling tool and a variety of visual chart tools, Quick BI reduces data acquisition costs greatly and makes it much easier to use, allowing you to easily complete data perspective analysis, self-service data acquisition, business data query, and report making.

Figure 26-1: Architecture of Quick BI



26.2 Log on to the QuickBI console

Take the Chrome browser as an example to describe how to log on to the Quick BI console through the Apsara Stack console as cloud product users.

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

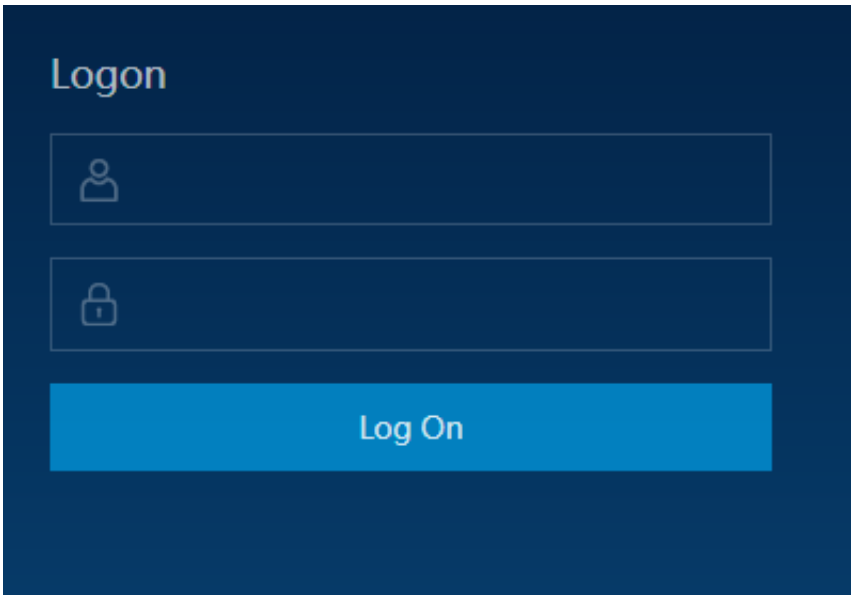
  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 26-2: Log on to the Apsara Stack console*.

   Figure 26-2: Log on to the Apsara Stack console

3. Enter the correct username and password.

- The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
- You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).

4. Click **Log On** to go to the **Home** page.
5. In the menu bar, choose **Console** > **Big Data** > **QuickBI** .
6. Select a department,click**QuickBI**or click**Quick BI Management Console**
   - Click**QuickBI**, you can enter the Quick BI product page directly.
   - Click**Quick BI Management Console**, you can manage the organizational unit and workspaces

26.3 Data modeling

This section describes how to use Quick BI for data modeling.

The data modeling procedure has the following steps.

- Create a data source
- Select a table from the data source to create a dataset
- (Optional) Use custom SQL to create a dataset

26.3.1 Manage data sources

This section describes the data source types that Quick BI supports.

Quick BI supports the following types of data sources.

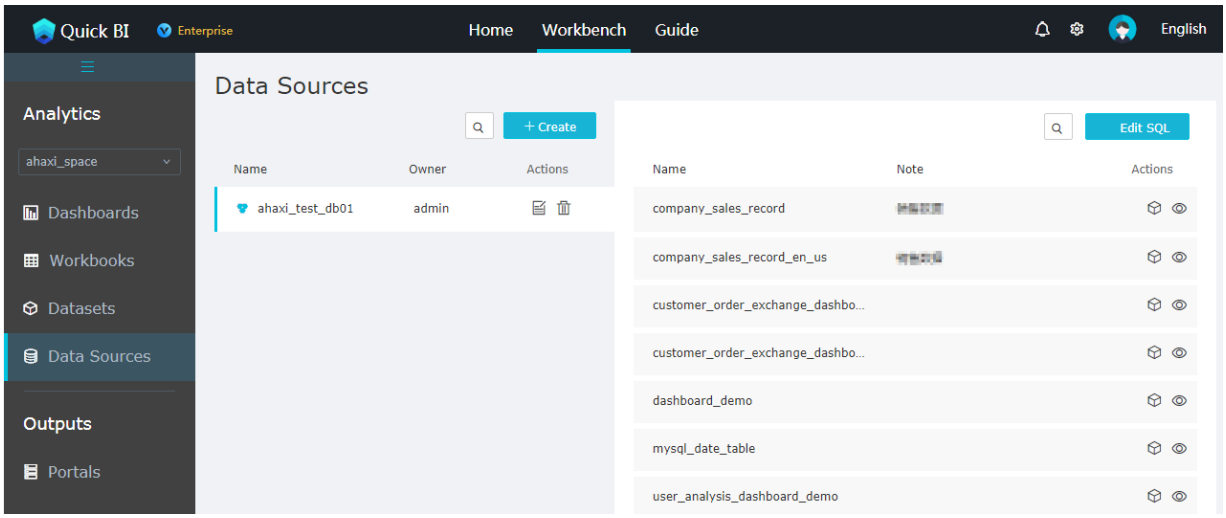
- Data sources from cloud databases such as MaxCompute, MySQL, SQL Server, AnalyticDB, HybirdDB for MySQL, HybirdDB for PostgreSQL, PostgreSQL, and PPAS
- Data sources from external database such as MySQL, SQL Server, and Oracle

Note:
Currently, views cannot be used on SQL server.

26.3.1.1 Data source list

You can manage all data sources in the data source list in a centralized manner and perform operations such as **create a data source**, **edit a data source**, and **delete a data source**, as shown in *Figure 26-3: Data source list*.

Figure 26-3: Data source list



26.3.1.2 Create a data source

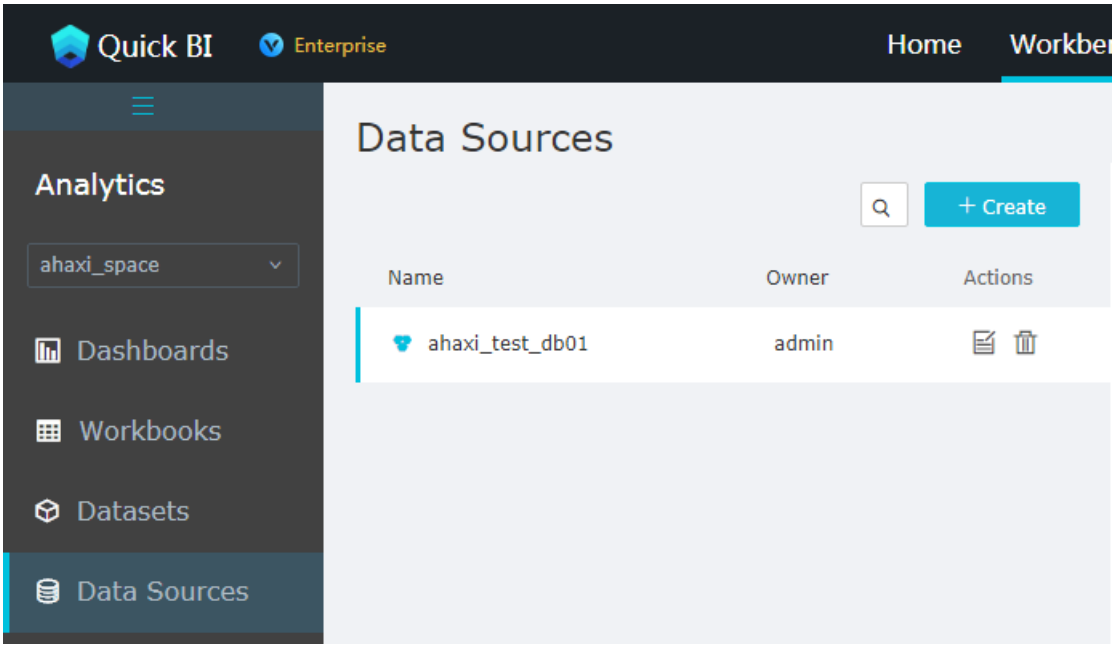
Context

Data sources must be used as the basis to operate datasets, worksheets, dashboards, data portals, and others. This section describes how to create a data source.

Procedure

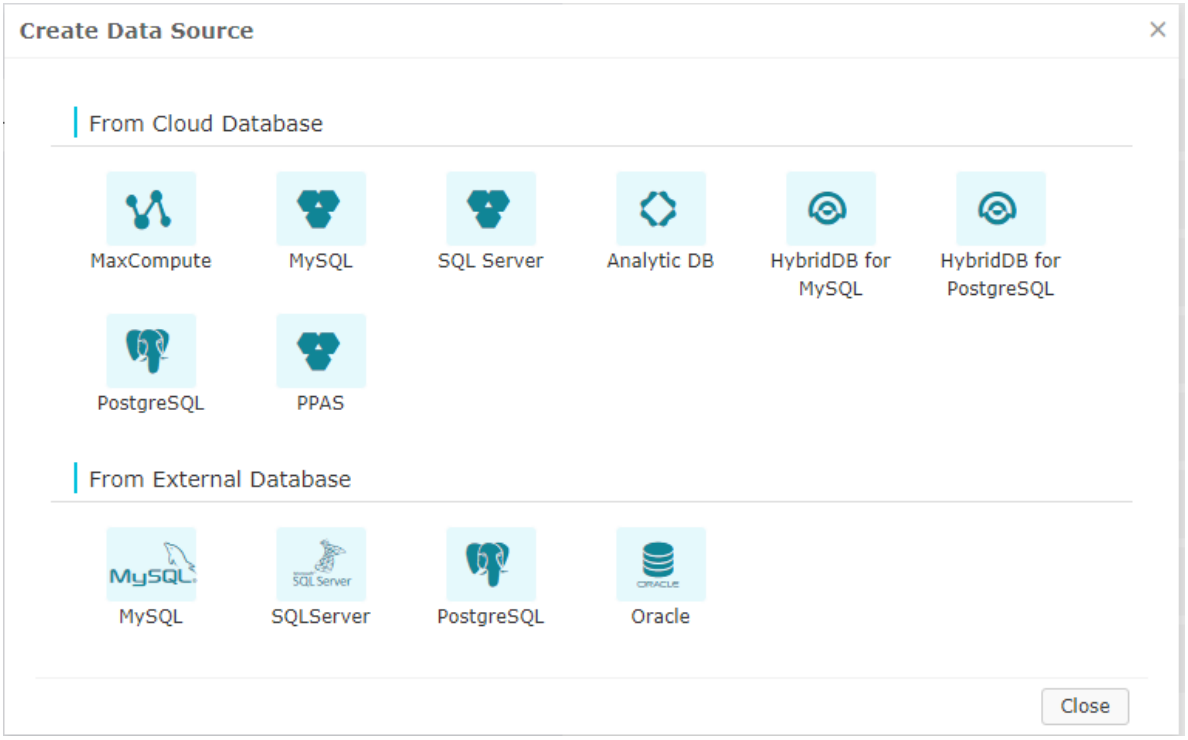
1. *Log on to the Quick BI console*, click **Workbench** to go to the workbench management page.
2. Click the **Data Sources**. The data source management page appears.
3. Click **Create**, as shown in Figure *Figure 26-4: Create a data source*.

Figure 26-4: Create a data source



4. In the dialog box that appears, select a source for the new data source, as shown in *Figure 26-5: Select a source for the new data source*.

Figure 26-5: Select a source for the new data source



5. Enter information about the new data source as instructed and then click **Add**.

26.3.1.2.1 Data sources from cloud databases

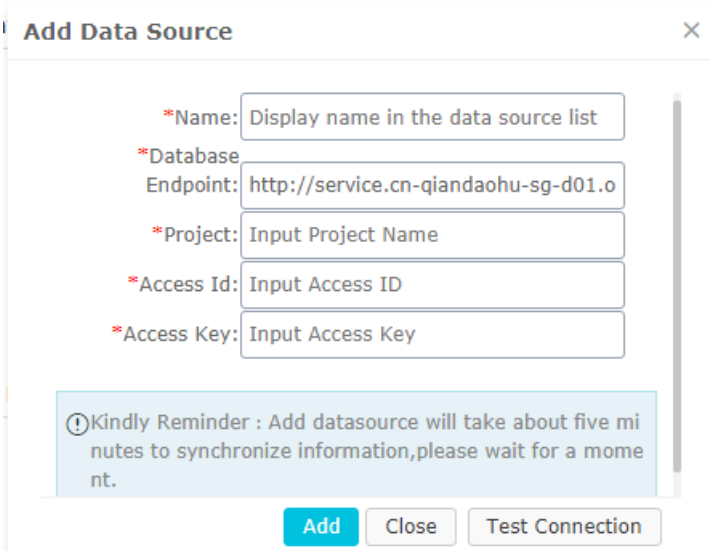
This section describes how to create a data source from a cloud database.

26.3.1.2.1.1 MaxCompute

Procedure

1. Click the **MaxCompute** icon. A new dialog box appears, as shown in *Figure 26-6: Add a data source*.

   Figure 26-6: Add a data source



2. Enter the required information for connecting to the new data source.
   - Name: Name of the data source.
   - Database Endpoint: The default address is used, which does not need to be modified.
   - Project: Name of the project.
   - Access Id: Access Key ID of the Alibaba Cloud console.
   - Access Key: Access Key Secret of the Alibaba Cloud console.
3. Click **Test Connection** to test data source connectivity .
4. After the connection test is successful, click **Add**.

26.3.1.2.1.2 MySQL

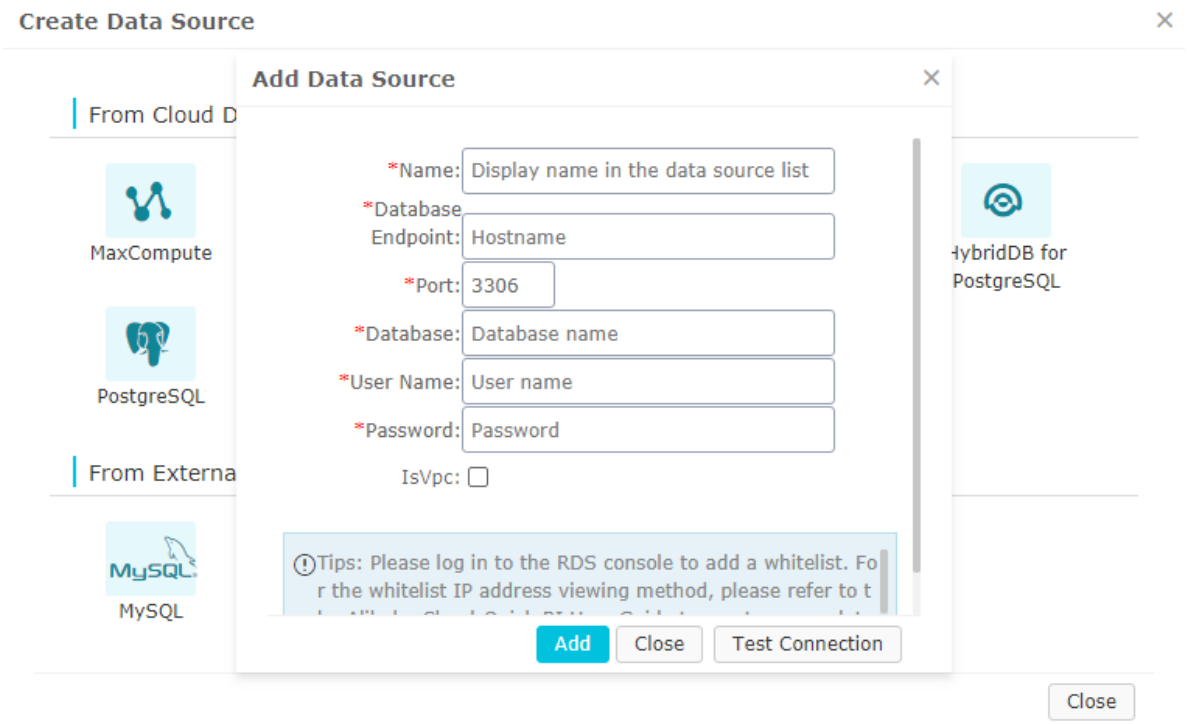
Context

Given the limitations imposed by the whitelist policy of ApsaraDB for RDS, before creating a MySQL data source, you need to query available IP addresses and add them to the RDS console manually.

Procedure

1. *Log on to the Quick BI console*
2. Choose **Workbench** > **Data Sources** > **Create**.
3. Click **MySQL** and enter the required information for connecting to the new data source, as shown in *Figure 26-7: Add a MySQL data source*.

Figure 26-7: Add a MySQL data source



- Name: Name of the data source list.
- Database Endpoint: Enter the host name or IP address.
- Port: Enter the correct port number.
- Database: The name of the database to be connected to.
- Username: The user name of the database.
- Password: The password of the database.

Note:

If you do not know the username and password, contact your data warehouse administrator.

4. Click **Test Connection** to test data source connectivity.
5. After the connection test is successful, click **Add**.

   If a data source with the same configuration already exists, a message indicating a conflict is displayed.

26.3.1.2.1.3 SQL Server

Context

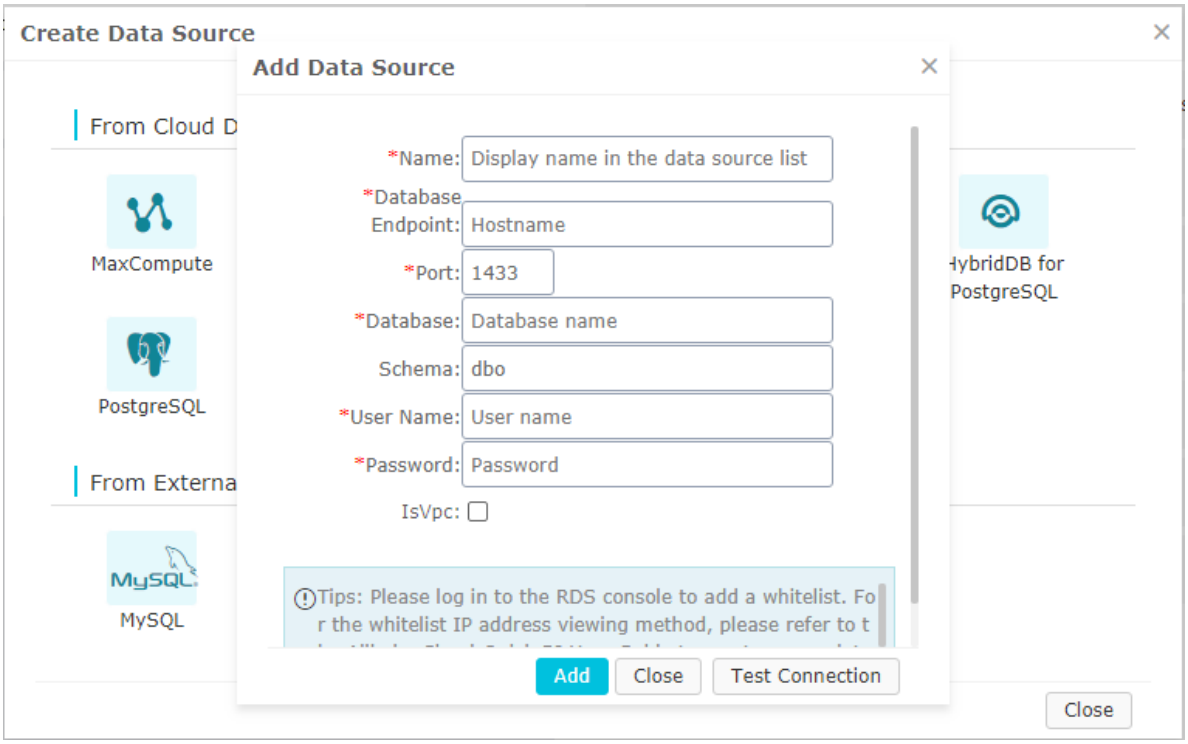
Given the limitations imposed by the whitelist policy of ApsaraDB for RDS, before creating an SQL Server data source, you need to query available IP addresses and add them to the RDS console manually.

The method for configuring SQL Server data sources is similar to that for configuring MySQL data sources. The difference being that SQL Server data sources have a unique configuration item called **schema**.

Procedure

1. Click the SQL Server icon and enter the required information for connecting to the new data source, as shown in *Figure 26-8: Add an SQL Server data source*.

Figure 26-8: Add an SQL Server data source



- Name: Name of the data source list.
- Database Endpoint: Enter the host name or IP address.
- Port: Enter the correct port number.
- Database: The name of the database to be connected to.
- Schema: dbo
- Username: The user name of the database.
- Password: The password of the database.

2. Click **Test Connection** to test data source connectivity.
3. After the connection test is successful, click **Add**.

26.3.1.2.1.4 Analytic DB

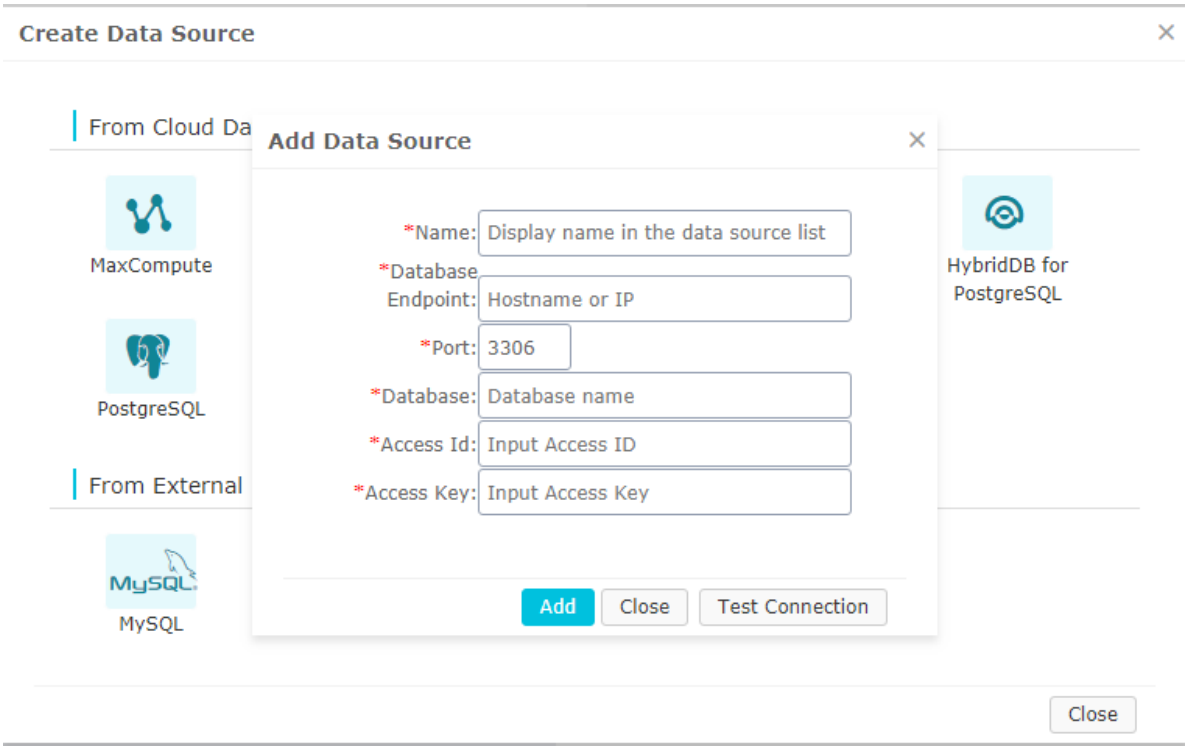
Context

AnalyticDB is formerly called ADS.

Procedure

1. Click the **AnalyticDB** icon and enter the required information for connecting to the new data source, as shown in *Figure 26-9: Add an AnalyticDB data source*.

Figure 26-9: Add an AnalyticDB data source



- Name: Name of the data source list.
- Database Endpoint: Enter the host name or IP address.
- Port: Enter the correct port number.
- Database: The name of the database to be connected to.
- Access Id: The Access Key ID of the Alibaba Cloud console.
- Access Key: The Access Key Secret of the Alibaba Cloud console.

2. Click **Test Connection** to test data source connectivity.
3. After the connection test is successful, click **Add**.

26.3.1.2.1.5 HybirdDB for MySQL

Context

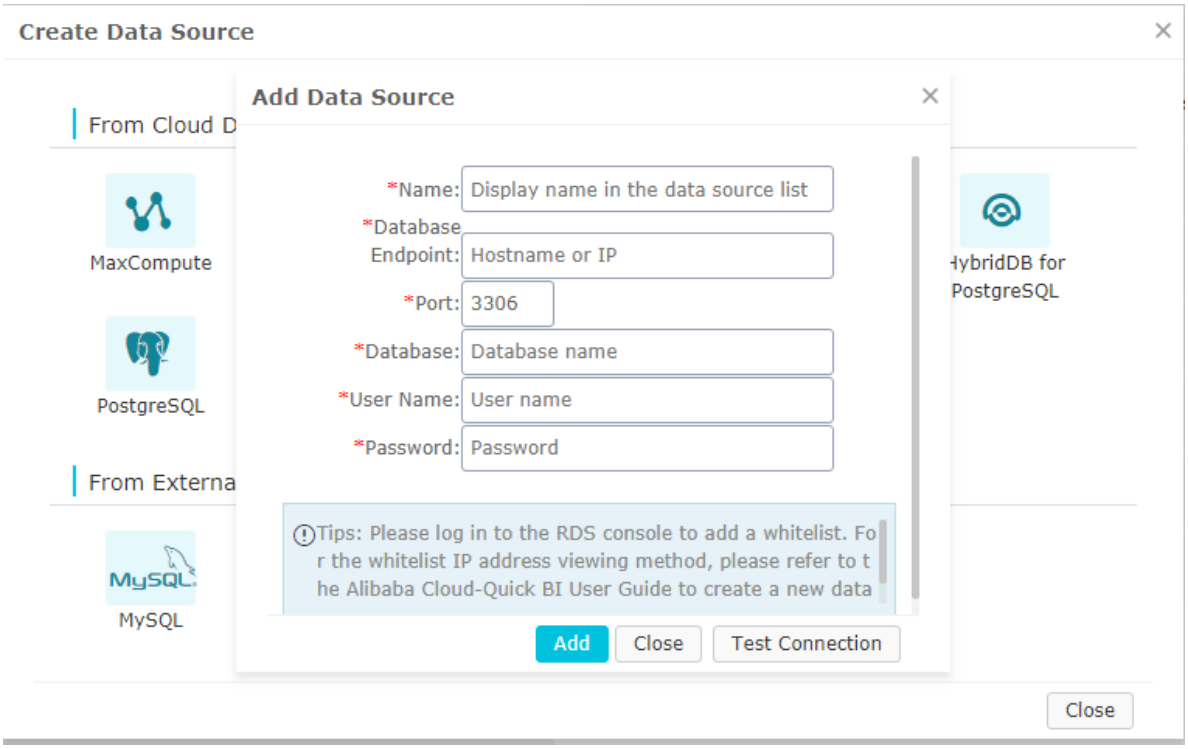
Given the limitations imposed by the whitelist policy of ApsaraDB for RDS, before creating a HybirdDB for MySQL data source, you need to query available IP addresses and add them to the RDS console manually.

The method for configuring HybirdDB for MySQL data sources is similar to that for configuring SQL Server data sources. The default port is the port specific to HybirdDB for MySQL.

Procedure

1. Click **HybirdDB for MySQL** and enter the required information for connecting to the new data source, as shown in *Figure 26-10: Add a HybirdDB for MySQL data source*.

   Figure 26-10: Add a HybirdDB for MySQL data source



   - Name: Name of the data source list.
   - Database Endpoint: Enter the host name or IP address.
   - Port: Enter the correct port number.
   - Database: The name of the database to be connected to.
   - Username: The user name of the database.
   - Password: The password of the database.

2. Click **Test Connection** to test data source connectivity.
3. After the connection test is successful, click **Add**.

26.3.1.2.1.6 HybirdDB for PostgreSQL

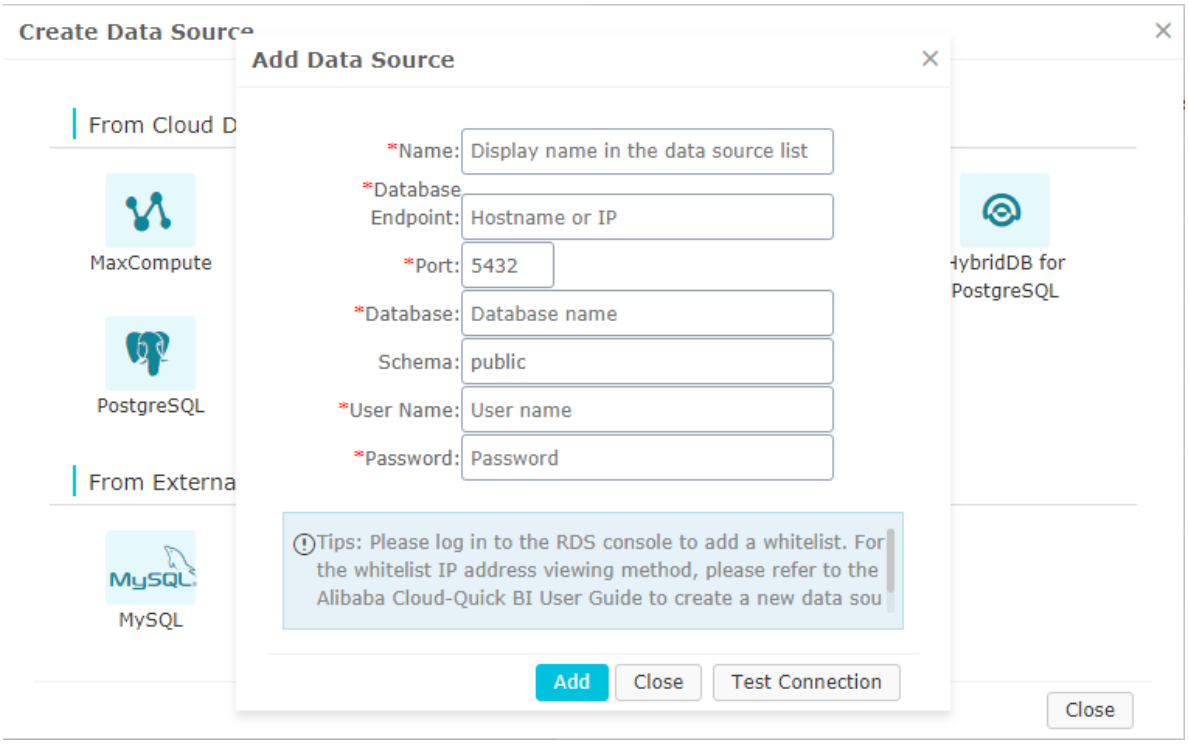
Context

The method for adding a data source from HybridDB for PostgreSQL is similar to that from SQL Server. The default port is the port specific to HybridDB for PostgreSQL.

Procedure

1. Click **HybridDB for PostgreSQL** and enter the required information for connecting to the new data source, as shown in *Figure 26-11: Add a data source from HybirdDB for PostgreSQL* .

   Figure 26-11: Add a data source from HybirdDB for PostgreSQL



   - Name: Name of the data source list.
   - Database Endpoint: Enter the host name or IP address.
   - Port: Enter the correct port number.
   - Database: The name of the database to be connected to.
   - Schema: public
   - Username: The user name of the database.
   - Password: The password of the database.
2. Click **Test Connection** to test data source connectivity.
3. After the connection test is successful, click **Add**.

26.3.1.2.1.7 PostgreSQL

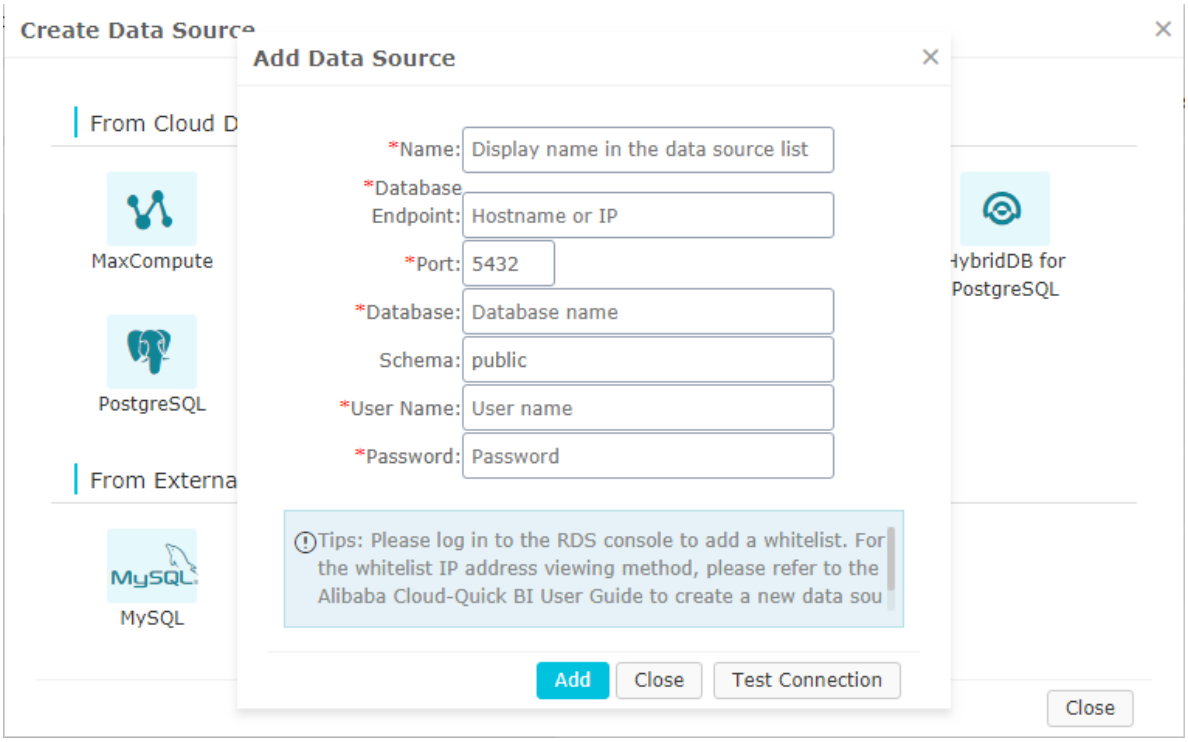
Context

The method for adding a data source from ApsaraDB for PostgreSQL is similar to that from HybridDB for PostgreSQL.

Procedure

1. Click **PostgreSQL** and enter the required information for connecting to the new data source, as shown in *Figure 26-12: Add a data source from ApsaraDB for PostgreSQL*.

   Figure 26-12: Add a data source from ApsaraDB for PostgreSQL



   - Name: Name of the data source list.
   - Database Endpoint: Enter the host name or IP address.
   - Port: Enter the correct port number.
   - Database: The name of the database to be connected to.
   - Schema: public
   - Username: The user name of the database.
   - Password: The password of the database.
2. Click **Test Connection** to test data source connectivity.
3. After the connection test is successful, click **Add**.

26.3.1.2.1.8 PPAS

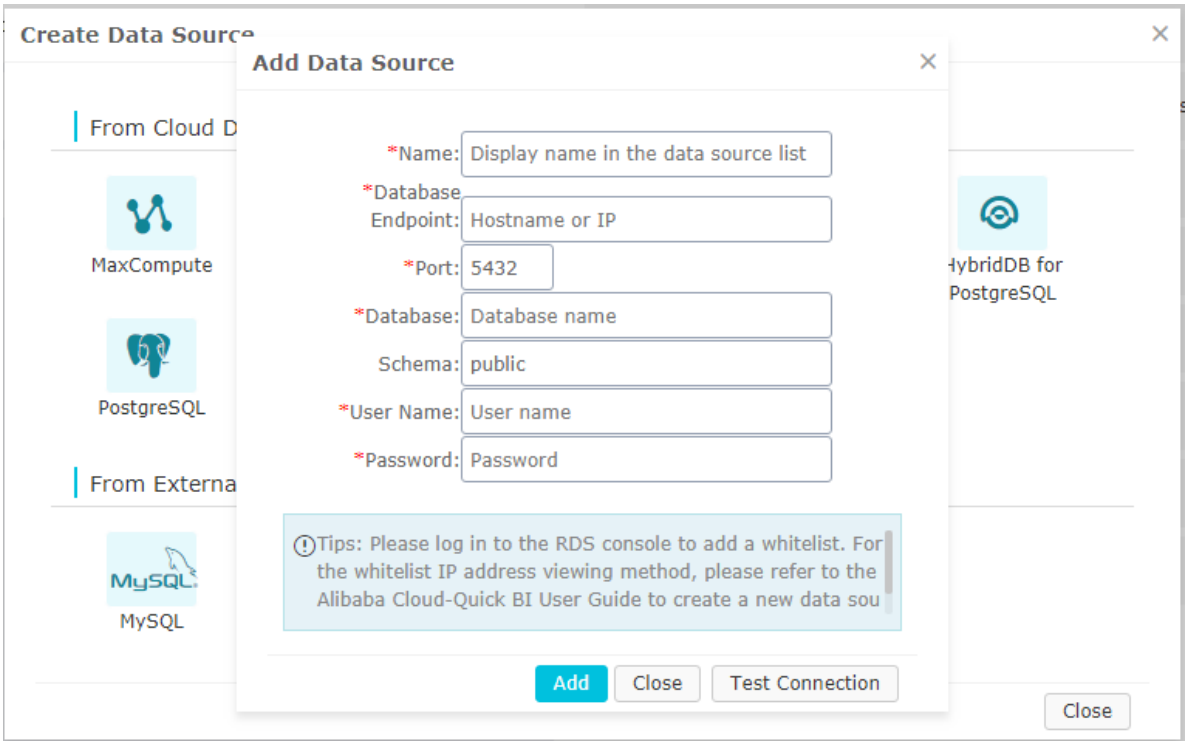
Context

The method for adding a data source from ApsaraDB for PPAS is similar to that from PostgreSQL.

Procedure

1. Click **PPAS** and enter the required information for connecting to the new data source, as shown in the following figure.

Figure 26-13: Add a data source from ApsaraDB for PPAS



- Name: Name of the data source list.
- Database Endpoint: Enter the host name or IP address.
- Port: Enter the correct port number.
- Database: The name of the database to be connected to.
- Schema: public
- Username: The user name of the database.
- Password: The password of the database.

2. Click **Test Connection** to test data source connectivity.
3. After the connection test is successful, click **Add**.

26.3.1.2.2 Data sources from external database

This section describes how to use Quick BI to add a data source from external database.

26.3.1.2.2.1 MySQL

Context

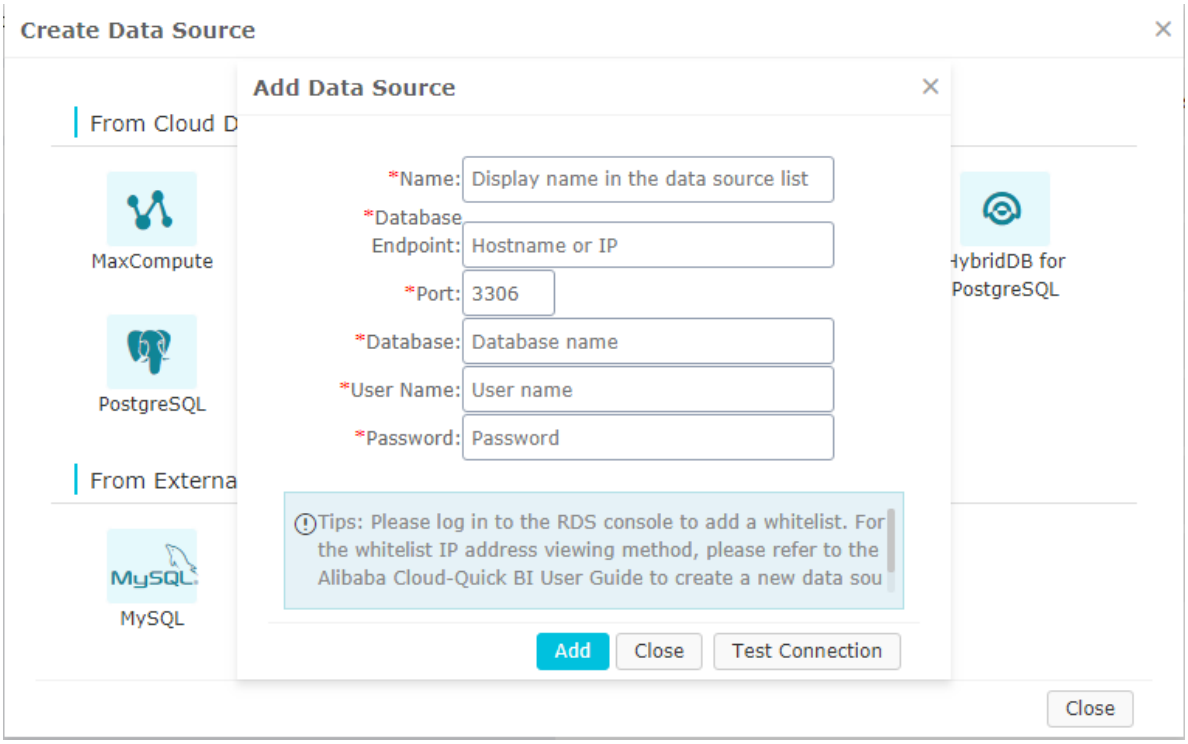
Given the limitations imposed by the whitelist policy of ApsaraDB for RDS, before creating a MySQL data source, you need to query available IP addresses and add them to the RDS console manually.

The method for configuring MySQL data sources from external database is similar to that for configuring data sources from MySQL in Cloud database.

Procedure

1. Click **MySQL** and enter the required information for connecting to the new data source, as shown in *Figure 26-14: Add a MySQL data source*.

   Figure 26-14: Add a MySQL data source



   - Name: Name of the data source list.
   - Database Endpoint: Enter the host name or IP address.
   - Port: The default value is 3306.
   - Database: The name of the database to be connected to.
   - Username: The user name of the database.
   - Password: The password of the database.

2. Click **Test Connection** to test data source connectivity.
3. After the connection test is successful, click **Add**.

4. You must enable the firewall under the data source in external database to allow external access to MySQL. Run the following command to access the firewall configuration file.

   ```
   vi /etc/sysconfig/iptables
   ```

5. Add the following command to the firewall configuration file.

   ```
   -A RH-Firewall-1-INPUT -m state –state NEW -m tcp -p tcp –dport 3306 -j ACCEPT
   ```

6. After the configuration is successful, run the following command to restart **iptable**.

   ```
   service iptables restart
   ```

26.3.1.2.2.2 SQL Server

Context

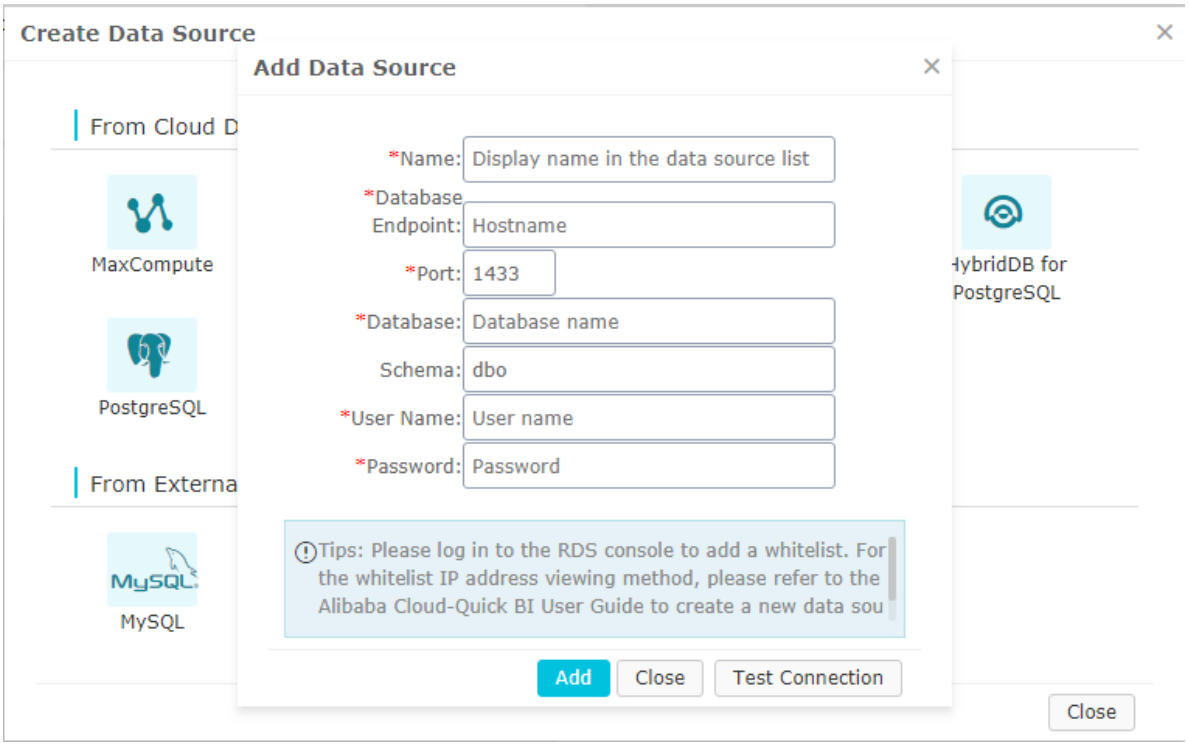
Given the limitations imposed by the whitelist policy of ApsaraDB for RDS, before creating an SQL Server data source, you need to query available IP addresses and add them to the RDS console manually.

The method for configuring SQL Server data sources from external database is similar to that for configuring data sources from SQL Server in cloud database.

Procedure

1. Click **SQL Server** and enter the required information for connecting to the new data source, as shown in *Figure 26-15: Add a SQL Server data source from external database*.

Figure 26-15: Add a SQL Server data source from external database



- Name: Name of the data source list.
- Database Endpoint: Enter the host name or IP address.
- Port: The default value is 1433.
- Database: The name of the database to be connected to.
- Schema: dbo
- Username: The user name of the database.
- Password: The password of the database.

2. Click **Test Connection** to test data source connectivity.
3. After the connection test is successful, click **Add**.

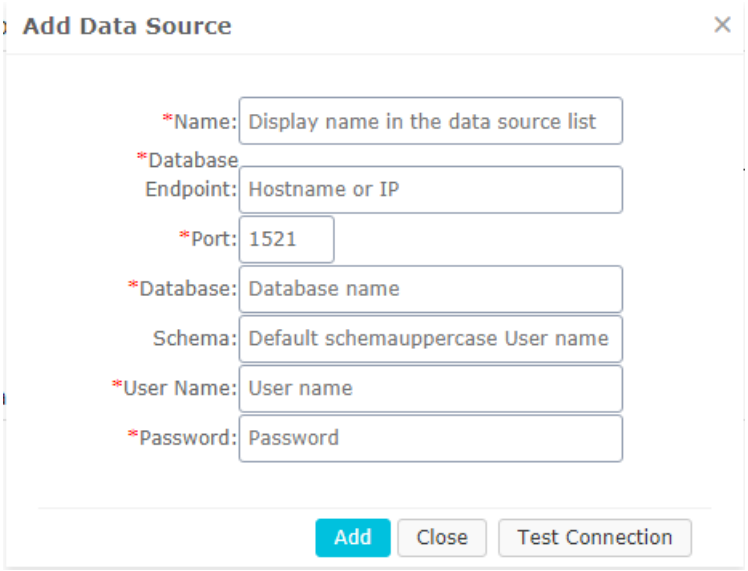
Note:
Duplicate data sources cannot be added; otherwise, an error is returned.

26.3.1.2.2.3 Oracle

Procedure

1. Click **Oracle** and enter the required information for connecting to the new data source, as shown in *Figure 26-16: Add a data source from Oracle*.

Figure 26-16: Add a data source from Oracle



- Name: Name of the data source list.
- Database Endpoint: Enter the host name or IP address.
- Port: The default value is 1521.
- Database: Name of the database to be connected to.
- Schema: public
- Username: Username for database logon.
- Password: Password for database logon.

2. Click **Test Connection** to test data source connectivity.
3. After the connection test is successful, click **Add**.

26.3.1.3 Edit a data source

Procedure

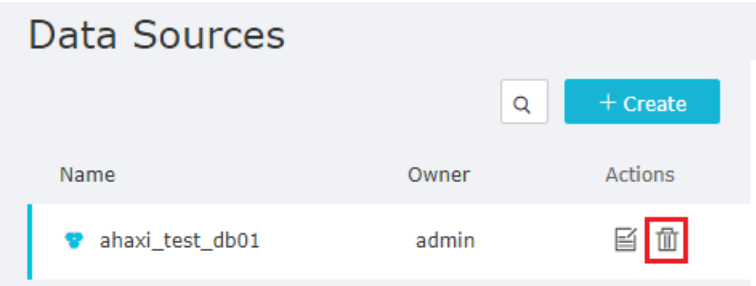
1. Select a data source from the data source list.
2. Click **Edit** to edit the selected data source.

26.3.1.4 Delete a data source

Context

If the data source you want to delete has been used for creating a dataset, the data source cannot be deleted, and an error message as shown in *Figure 26-17: Delete a data source* is displayed.

Figure 26-17: Delete a data source



Procedure

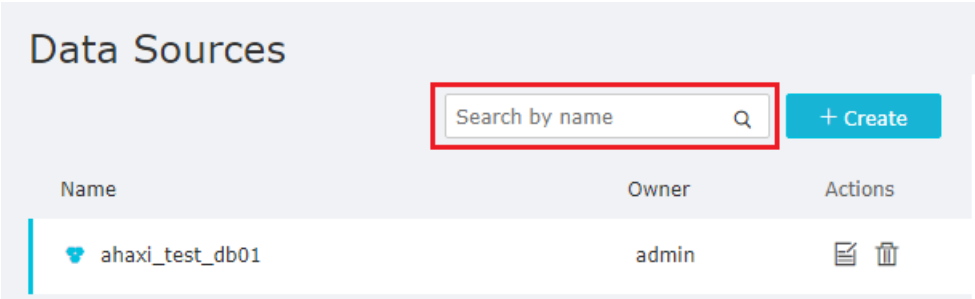
1. Select a data source from the data source list.
2. Click **Delete** to delete the selected data source.

26.3.1.5 Query a data source

Procedure

1. *Log on to the Quick BI console*, click **Data Sources** to go to the data source management page.
2. Enter a keyword in the search box to search for the expected data source, as shown in *Figure 26-18: Query a data source*.

Figure 26-18: Query a data source



26.3.1.6 Query the tables of a data source

Procedure

1. *Log on to the Quick BI console*, click **Data Sources** to go to the data source management page.
2. Select a data source from the list. All tables of the selected data source are listed on the right of the page.

3. Enter a keyword in the search box on the right to search for the expected table, as shown in *Figure 26-19: Query the tables of a data source*.

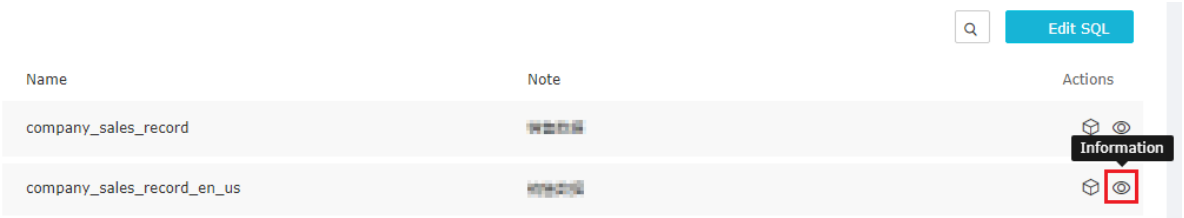
Figure 26-19: Query the tables of a data source

26.3.1.7 Query the details of a table of a data source

Procedure

1. *Log on to the Quick BI console*, click **Data Sources** to go to the data source management page.
2. Select a data source from the list. All tables of the selected data source are listed on the right of the page.
3. Select a table and click **Information** next to it to view the table details and related fields, as shown in *Figure 26-20: Query the details of a table of a data source*.

Figure 26-20: Query the details of a table of a data source



26.3.2 Manage datasets

This section describes how to use Quick BI to manage datasets.

You can create different tables of a data source as datasets. The following sections describe how to create, edit, and query datasets.

26.3.2.1 Create a dataset

You can create a dataset in either of the following ways.

- Create a dataset from a data source

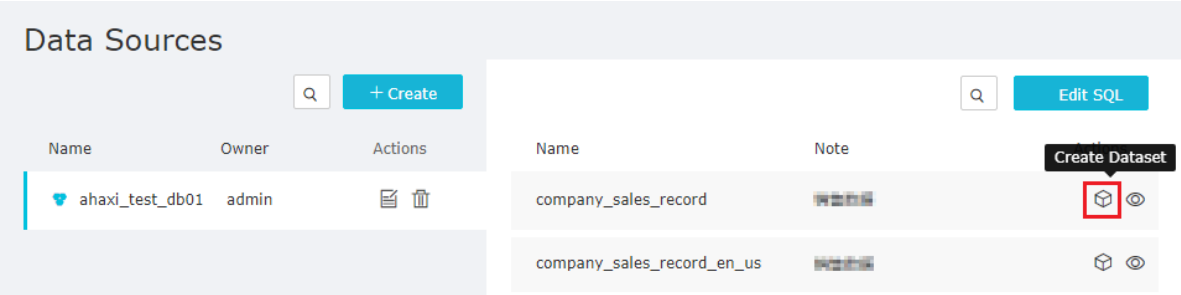
- Create a dataset through custom SQL

26.3.2.1.1 Create a dataset from a data source

Procedure

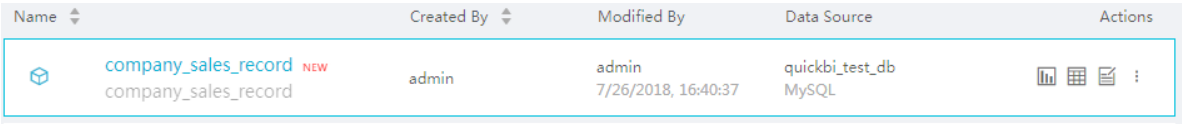
1. *Log on to the Quick BI console*, click **Data Sources** to go to the data source management page.
2. Select a data source from the list. All tables of the selected data source are listed on the right of the page.
3. Select a table and click **Create Dataset** next to it, as shown in *Figure 26-21: Create a dataset from a data source*.

Figure 26-21: Create a dataset from a data source



After the dataset is created, the **Dataset** tab is automatically displayed. **New** is displayed for the newly created dataset, allowing you to fast locate the dataset., as shown in *Figure 26-22: New dataset*.

Figure 26-22: New dataset



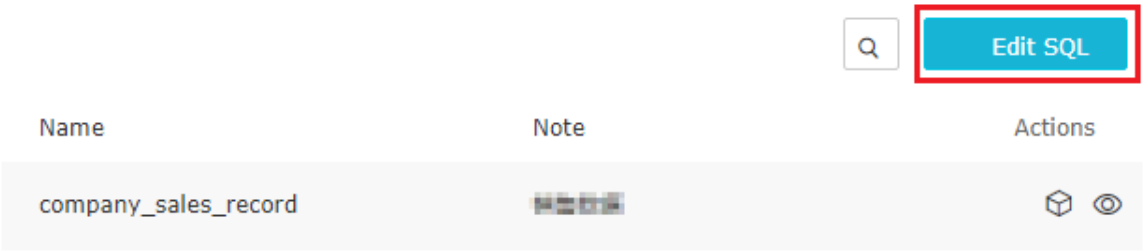
26.3.2.1.2 Custom SQL under a MaxCompute data source

Currently, only MaxCompute (ODPS) data sources support custom SQL.

Procedure

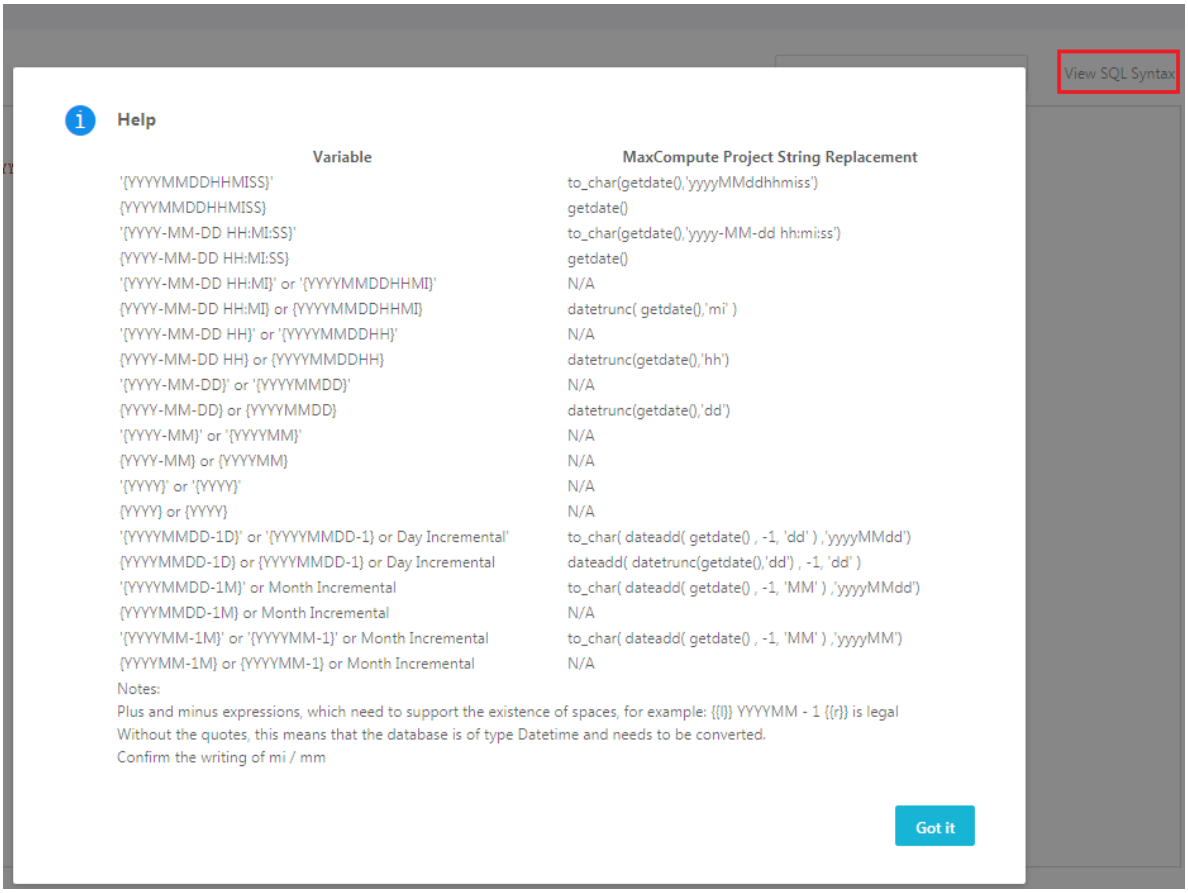
1. *Log on to the Quick BI console*, click the **Data Sources**.
2. Select a MaxCompute data source from the list.
3. On the right of the page, click **Edit SQL** to create a dataset, as shown in *Figure 26-23: Custom SQL*.

Figure 26-23: Custom SQL



4. Edit the data source. If needed, you can click **View SQL Syntax** to open the SQL help page, as shown in *Figure 26-24: SQL rules*.

Figure 26-24: SQL rules



5. Click **Save** to save the data source as a dataset directly.

26.3.2.2 Set the default name of a dataset

Context

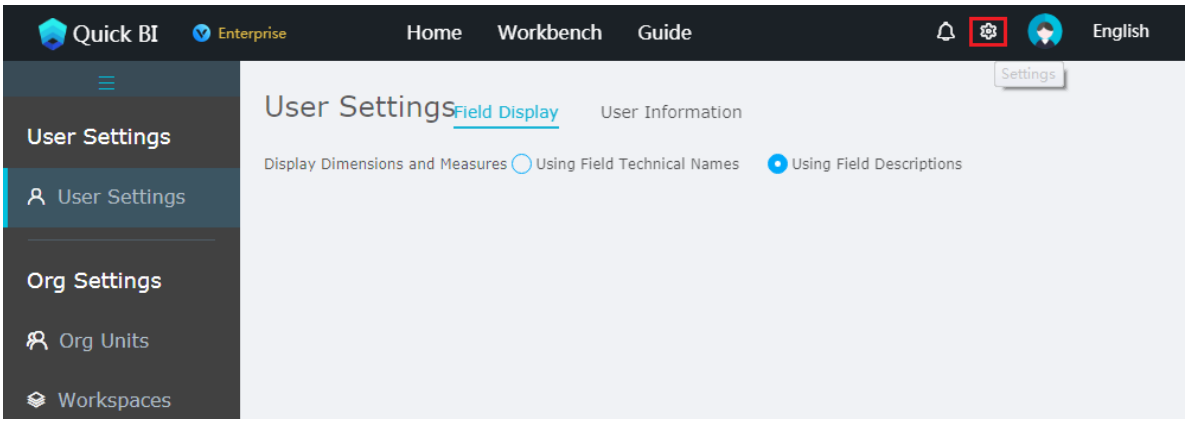
Quick BI automatically creates datasets based on the metadata of physical tables and converts the fields of physical tables to the dimension fields or measurement fields of datasets. Dimension

fields and measurement fields are automatically named after the names or comments of physical fields. In **Preference Setting** of the management console, you can select preferences as needed, and the dimension fields and measurement fields of subsequently created datasets are automatically named according to the selected preferences.

Procedure

1. *Log on to the Quick BI console*, click **Settings** to go to the setting page.
2. Choose **User Settings** > **Field Display**, as shown in *Figure 26-25: Set the default name of a dataset*.

Figure 26-25: Set the default name of a dataset



26.3.2.3 Edit a dataset

Procedure

1. *Log on to the Quick BI console*, click **Datasets** to go to the dataset management page.
2. Select a dataset and click the dataset name. The dataset editing page appears.

26.3.2.3.1 Edit a dimension field

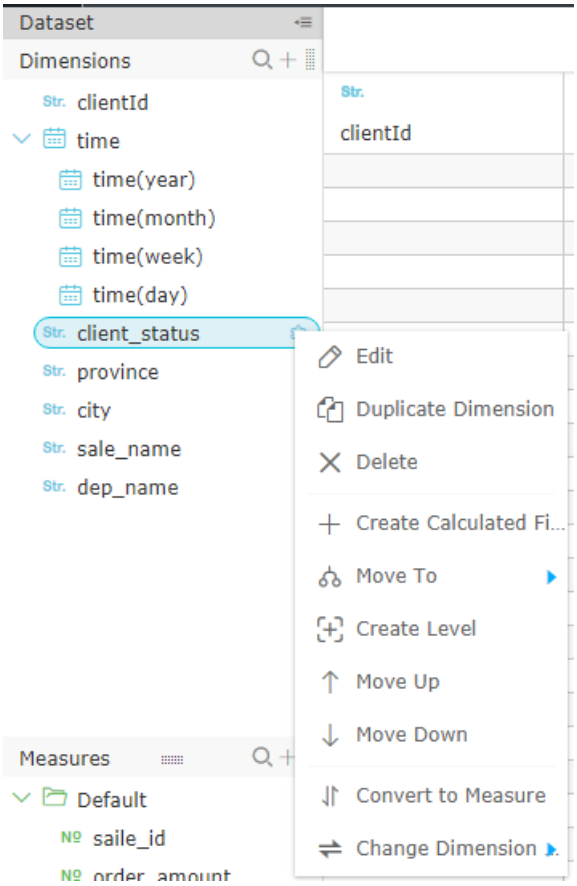
Context

Table fields of the character type and other types are considered as dimension fields by default. You can edit fields in the dimension or measurement list.

Procedure

1. Select a field from the dimension list.
2. Right-click the field, and a shortcut menu showing editing options appears.

Figure 26-26: Shortcut menu for dimension field editing



- Edit: To modify the display name and remarks of a dimension field.
- Duplicate Dimension: To quickly copy a dimension. **Copy** is automatically displayed for the generated dimension.
- Delete: To delete a field.
- Create Calculated Field (Dimension): To create a dimension field and customize the calculation mode.
- Move To: To quickly include a dimension field in an existing level for drilling.
- Create Level: To quickly include a dimension field in a created level.
- Move Up/Move Down: To move a field. You can drag the field or right-click the field to move it.
- Convert to Measure: To convert the current dimension field to a measurement field.
- Change Dimension Type: To switch a dimension field to the default, date, or geographical type.

26.3.2.3.2 Edit a measurement field

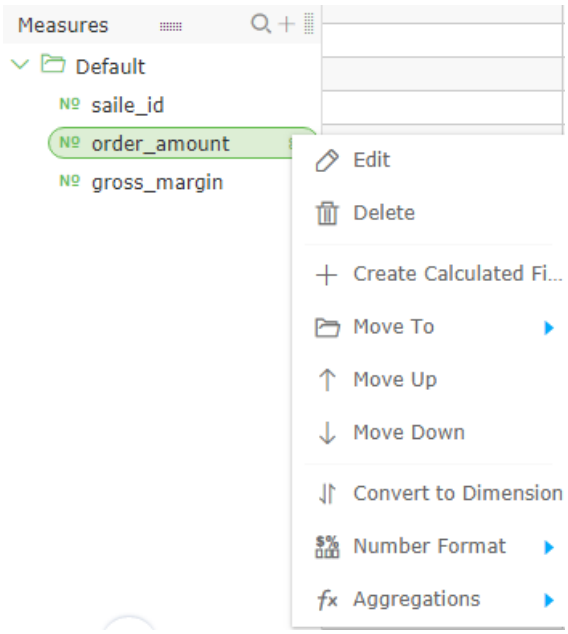
Context

Table fields of the numeric value type are considered as measurement fields by default.You can edit fields in the dimension or measurement list.

Procedure

1. Select a field from the measurement list.
2. Right-click the field, and a shortcut menu showing editing options appears, as shown in *Figure 26-27: Shortcut menu for measurement field editing*.

Figure 26-27: Shortcut menu for measurement field editing



- Edit: To modify the display name and remarks of a dimension field.
- Delete: To delete a field.
- Create Calculated Field (Measurement): To create a measurement field and customize the calculation mode.
- Move To: To quickly include a measurement field in an existing folder.
- Move Up/Move Down: To move a field. You can drag the field or right-click the field to move it.
- Convert to Dimension: To convert the current measurement field to a dimension field.
- Number Format: To set the display format of a number.

- Aggregations: You can select an aggregation mode, such as sum, max, or min, on the menu .

26.3.2.3.3 Tool buttons

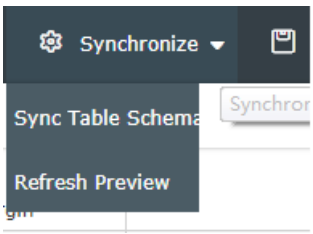
Quick BI provides tool buttons for editing datasets, as shown in *Figure 26-28: Tool button area*.

Figure 26-28: Tool button area

Synchronize

Click the **Synchronize** button to synchronize table structures and refresh datasets, as shown in *Synchronize*.

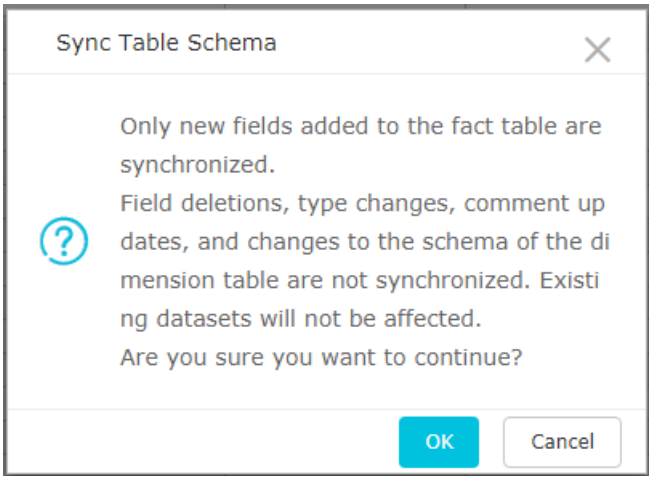
Figure 26-29: Synchronize



- Synchronize Table Schema: To synchronize changes (such as new fields) made to an online physical table.

  The system does not delete datasets if the fields of an online table are deleted, the field name or comment is modified, or the structure of a dimension table is changed. See *Figure 26-30: Synchronize the table structure*.

Figure 26-30: Synchronize the table structure

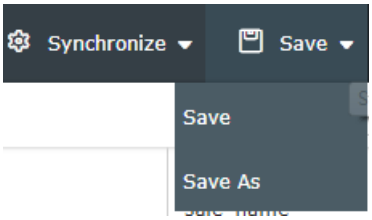


- Refresh Preview: To refresh the preview data of datasets. If you want to view the latest data in real time, refresh datasets after saving them.

Save

Click the **Save** button to save a dataset or save it with a different name. See *Figure 26-31: Save*.

Figure 26-31: Save



- Save: To save a dataset.

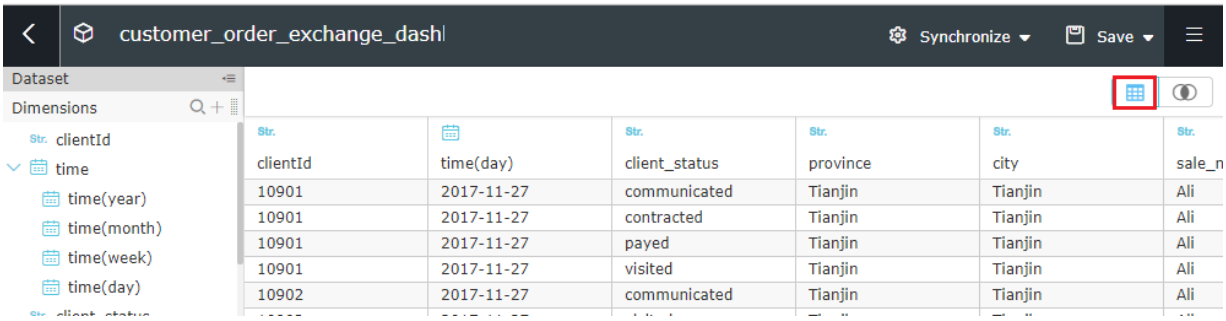
  Click this button to save datasets and then refresh the datasets to view the latest data after you create fields (measurement), delete fields, or convert between measurement and dimension fields on the dataset editing page.
- Save as: To save the current dataset as a new one. This operation can be used to quickly copy a new dataset or back up a dataset.

26.3.2.3.4 Preview data

Context

Click **Preview** to enter data preview mode, as shown in *Figure 26-32: Data preview*.

Figure 26-32: Data preview

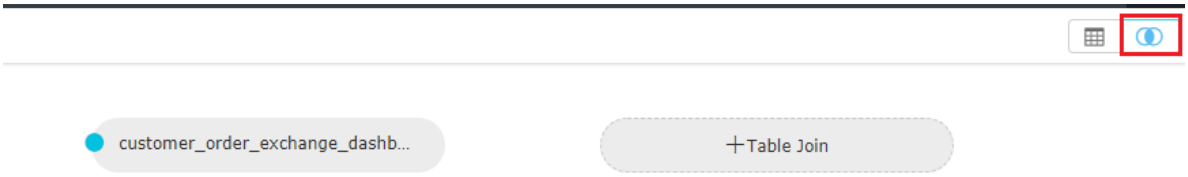


26.3.2.3.5 Join a worksheet

Procedure

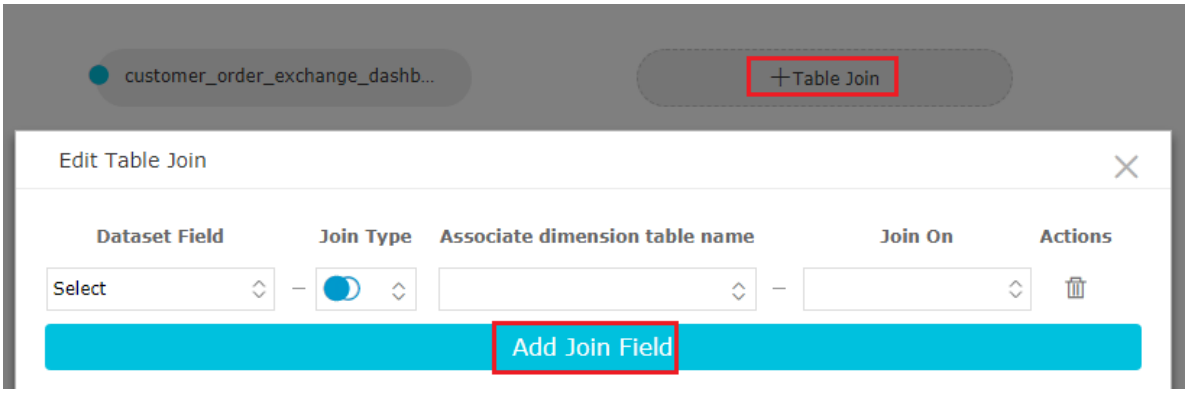
1. Click **Join** to change to join mode, as shown in *Figure 26-33: Data joining*.

   Figure 26-33: Data joining

2. Click **+ Table Join**. The **Create Join Model** dialog box appears.
3. Select the fields to be joined and a joining mode, as shown in *Figure 26-34: Create a join model*.

   Figure 26-34: Create a join model

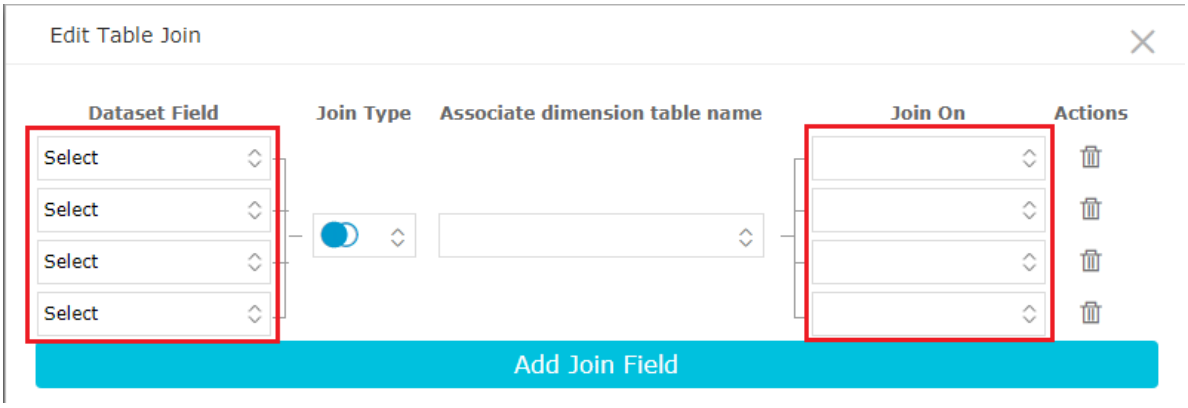
   The following two joining modes are supported.

   - Inner join
   - Left outer join

> **Note:**
> Currently, the mode in which similar Table A joins Table B which then joins Table C is not supported, and non-same-source table joining is not supported either.

4. Click **Add Join Field** to join multiple table fields, as shown in *Figure 26-35: Add join Field*.

Figure 26-35: Add join Field



5. Click **OK** to create the join model.
6. Click **Preview** to enter data preview mode.
7. In preview mode, click **Save**.
8. After the join model is saved, click **Refresh Preview** to check the result of table joining.

26.3.2.3.6 Example of joining a worksheet

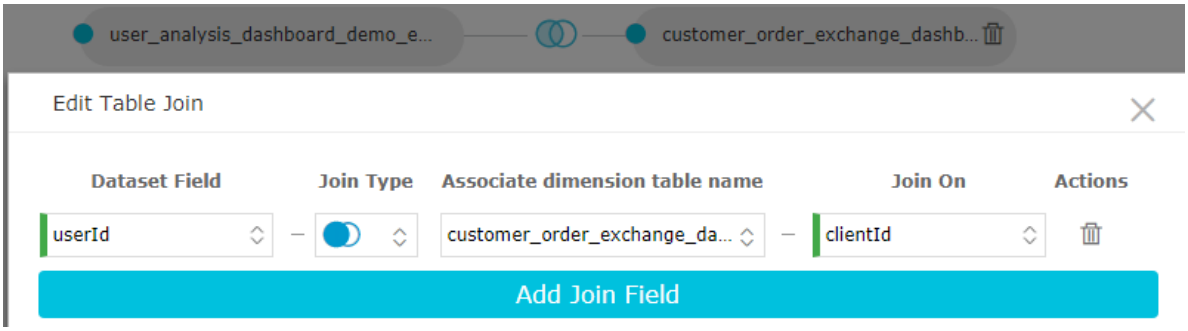
Context

Perform the **left outer join** operation on the tables user_analysis_dashboard_demo_en_us and customer_order_exchange_dashboard_demo_en_us by using the ID field.

Procedure

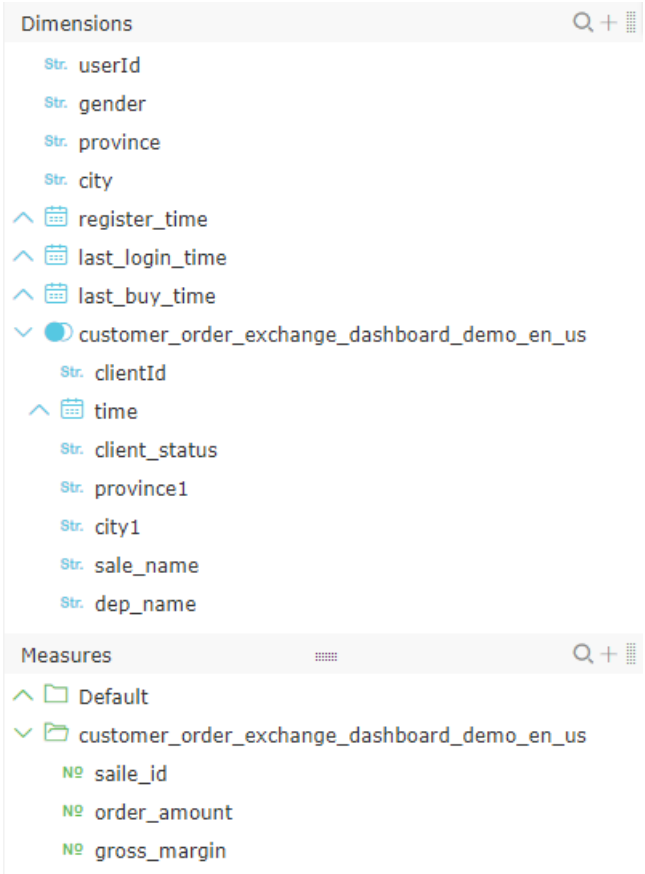
1. Select user_analysis_dashboard_demo_en_us from the dataset list.
2. Click **Edit**. The dataset editing page appears.
3. Click **Join Table**. The data table joining page appears.
4. Click **+ Table Join**. The **Create Join Model** dialog box appears.
5. Click the drop-down arrow to select the fields to be joined and a joining mode, as shown in *Figure 26-36: Left outer join*.

Figure 26-36: Left outer join



6. Click **OK** to create the join model.
7. Click **Preview**. The data preview page appears.
8. Click **Save** to save the new joining model.
9. Click **Refresh Preview** and view the joined data, as shown in *Figure 26-37: Preview data after the left outer join operation*.

Figure 26-37: Preview data after the left outer join operation



26.3.2.3.7 Drilling

Context

The following procedure describes how to perform drilling based on the company_sales_record dataset to query the company's sales order amount and profit in each area of China. Click an area to drill data at the provincial level, and click a province to drill data at the city level.

Procedure

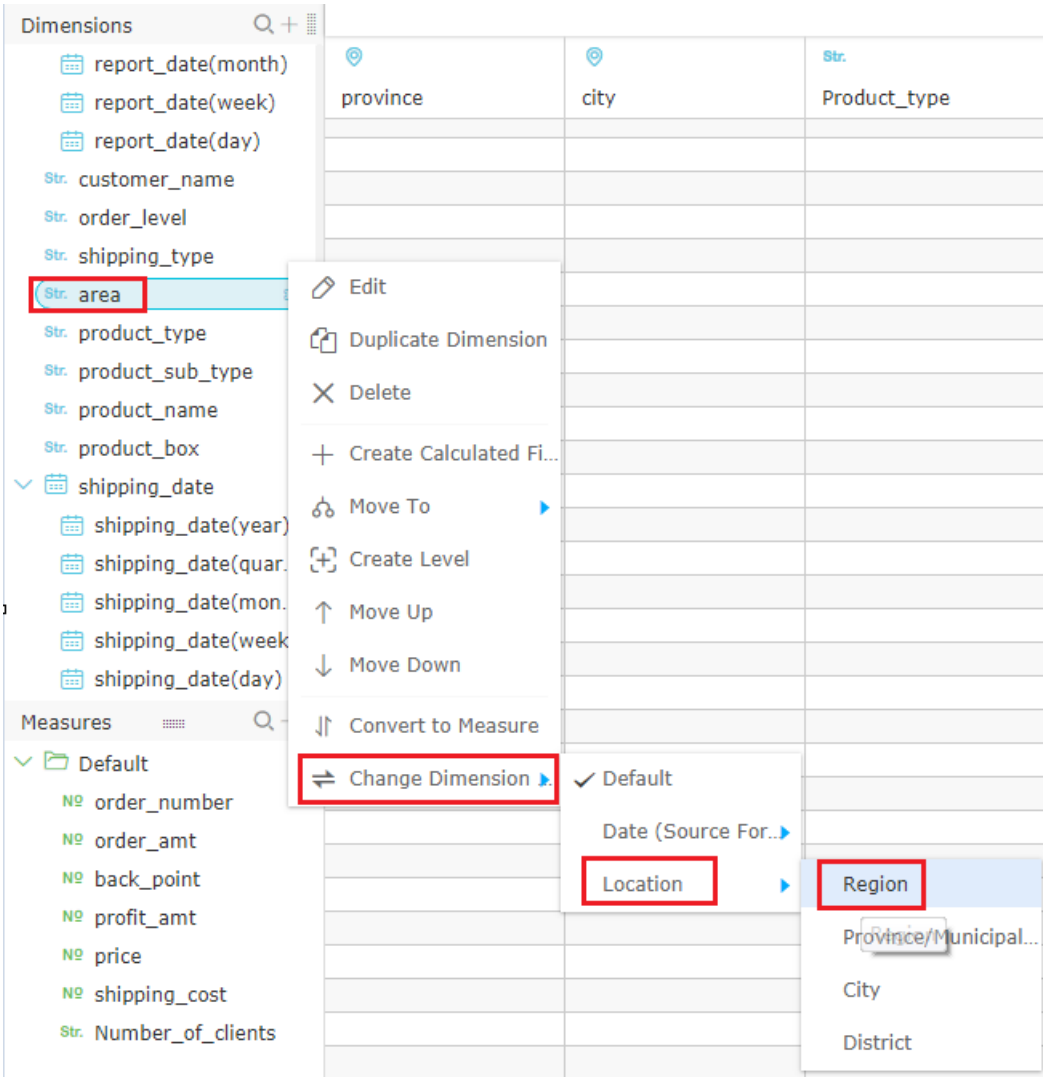
1. In the **Data Source** list, find company_sales_record.
2. Click **Create Dataset** next to it.

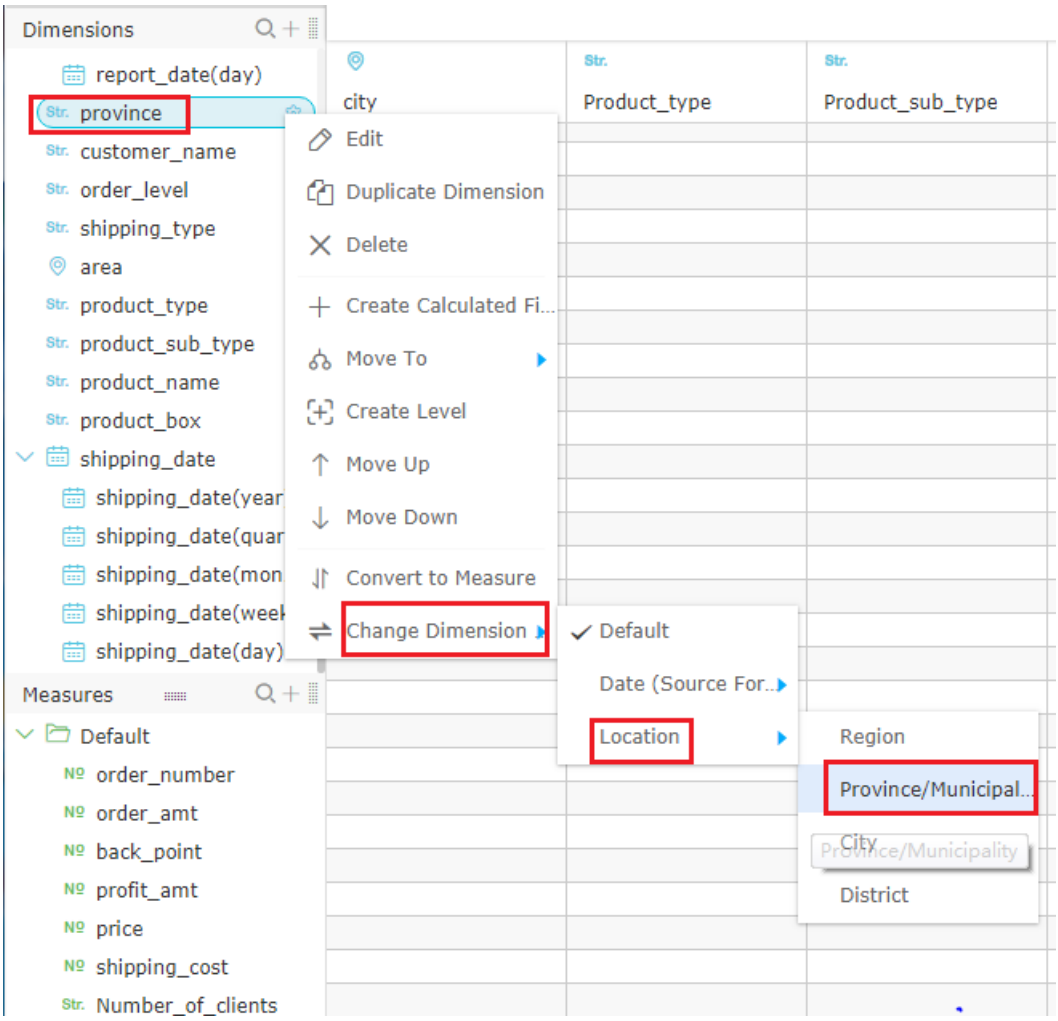
Change the field type

3. Convert text fields to the geographical type.

   The sales record has three fields that contain geographical information to indicate a specific city, province, and area, respectively. Specify the three fields as the geographical type and make sure that the contained information is real with one-to-one mappings, as shown in *Figure 26-38: Change the field type*.

Figure 26-38: Change the field type

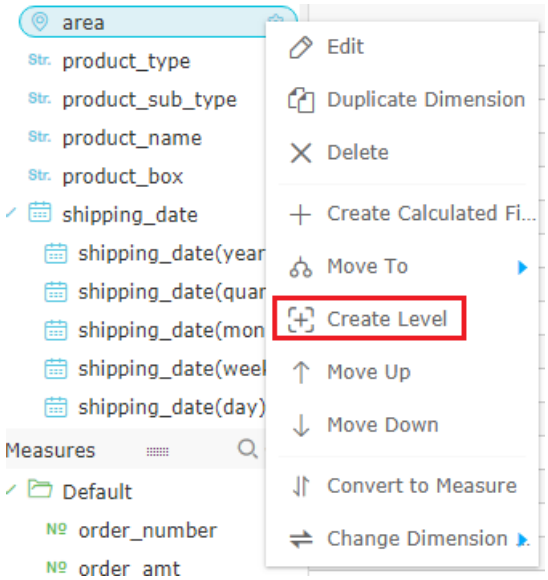




You can also create a level that contains geographical information.

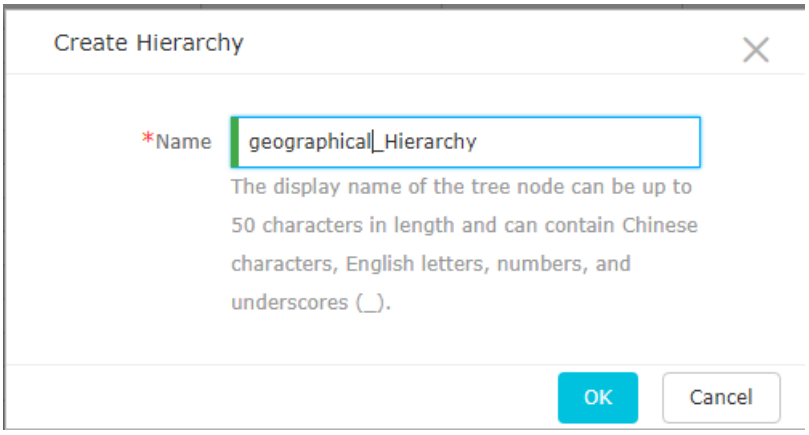
4. Select a field from the dimension list.
5. Right-click the field, and a shortcut menu showing editing options appears.
6. Choose **Create Level** from the shortcut menu, as shown in *Figure 26-39: Create a level*.

Figure 26-39: Create a level



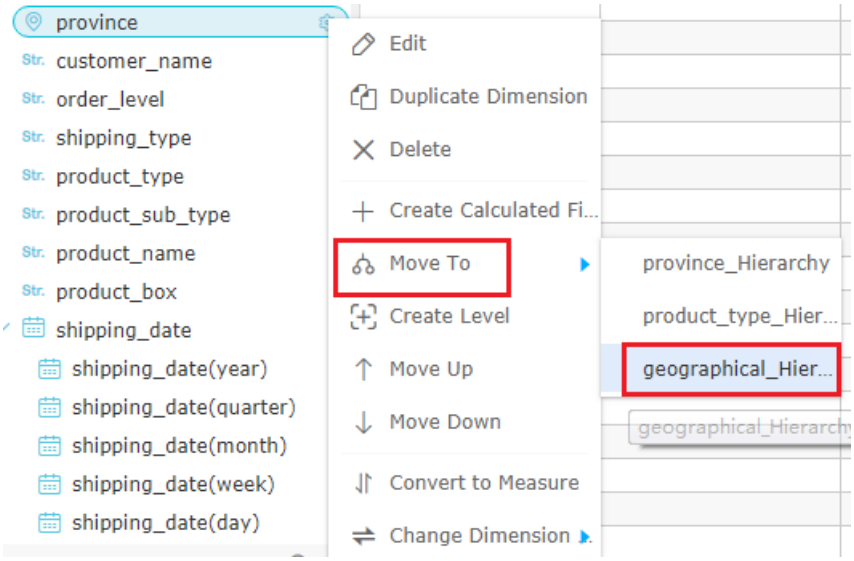
7. Enter a name for the level and click **OK**, as shown in *Figure 26-40: Level name*.

Figure 26-40: Level name



8. Move the dimension fields with geographical information to the new level, as shown in *Figure 26-41: Add dimension fields*.

Figure 26-41: Add dimension fields



9. Click **Save** to save the dataset.

10. Add the "province" and "order_amt" fields to the analysis panel.

Figure 26-42: Add fields to the analysis panel

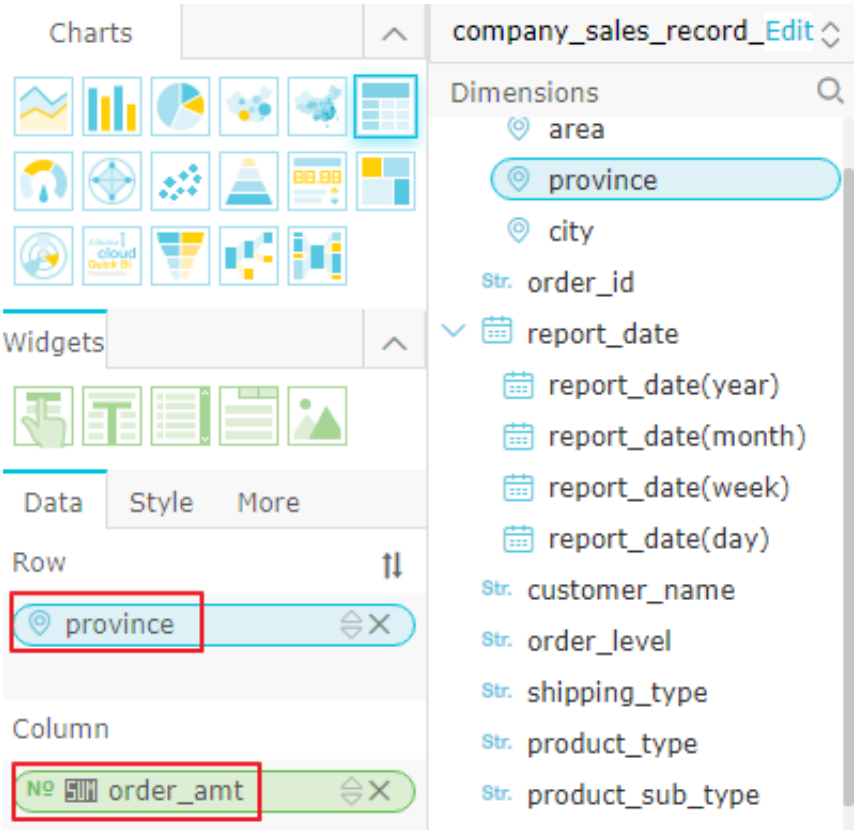
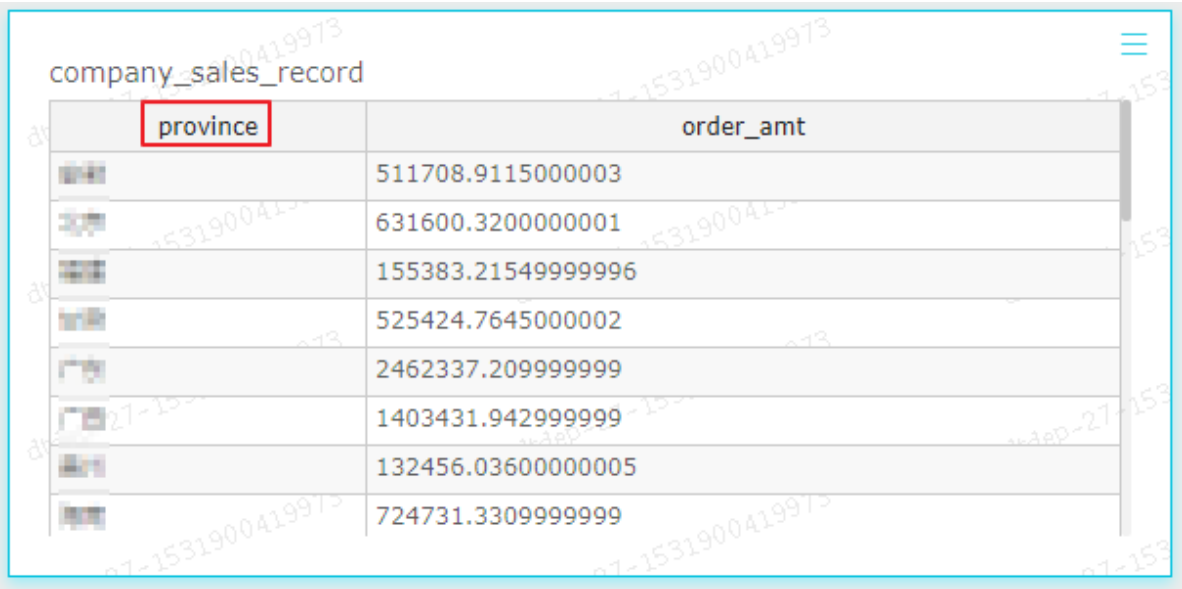
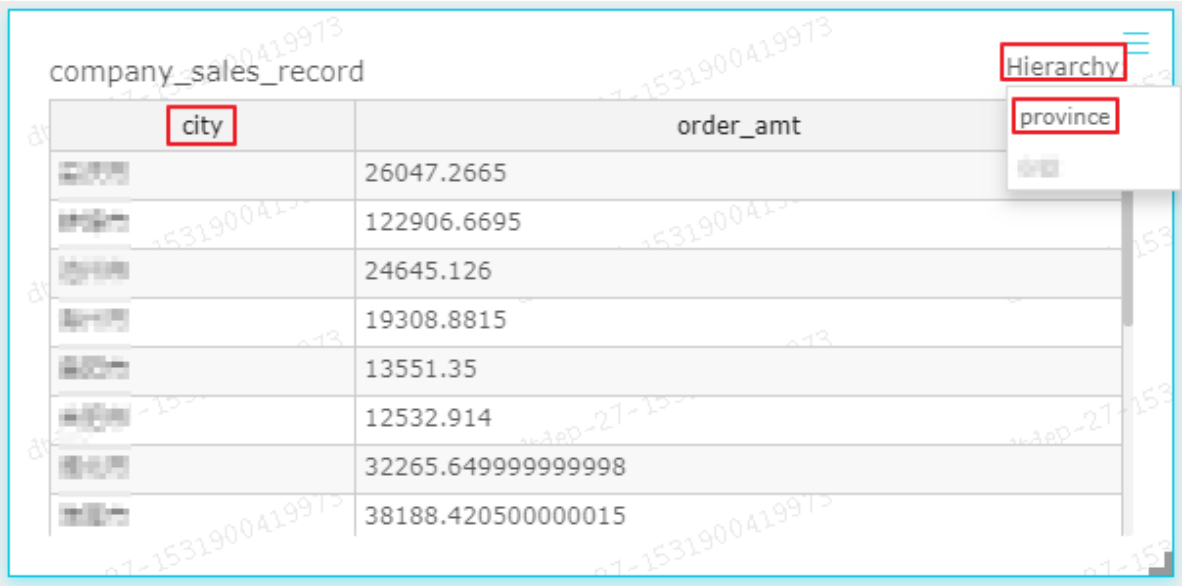


Figure 26-43: Province level data



Place the cursor in a province. Click the province to drill data at the city level.

Figure 26-44: Drill down from a province



After you drill to the city of the selected province, when you click **Hierarchy**, the details about drilling are displayed, you can click **province**to back to the province level.

11. Click **Save** to save the edited worksheet.

26.3.2.3.8 Calculated field

This section introduces calculated fields and provides instructions on using these fields.

A calculated field is a new column that is constructed using the existing fields and functions supported by SQL and meets the definition syntax rules of the SQL column of the current data source.

If you want to perform calculation based on existing data in the data source, you can add calculated fields. When constructing calculated fields, you can use semantic dimensions or measurements that are easily understood by business personnel as expression parameters. When semantic logic expressions of calculated fields generate actually executed SQL expression s in the Quick BI engine, the Quick BI engine translates the expressions to column expressions consisting of bottom-layer physical field names.

On the dataset editing page, click "+" next to "Dimension" and "Measurement". In the calculated field editor dialog box that appears, combine supported functions and existing fields.

Calculated fields created in "Dimension" and "Measurement" are automatically used as the calculation dimensions and calculation measurements, respectively.

In the calculated field expression editing box, you can use syntax of all functions and column expressions supported by the current data source.

You need to enter function names manually. You can manually enter field names in the format of [field name]. Alternatively, you can enter "[" in QWERTY mode to select the fields in the field name list or double-click the nodes in the left-side dimension and measurement tree to insert dimension or measurement field names to the expression editing box. After you enter correct SQL expression s in the editing box, the syntax is automatically colored.

When the calculated field expressions are compiled, the most common mistake is to mix Chinese and English punctuation marks such as quotation marks, comma, and parentheses, which leads to syntax resolution errors. In fact, only English punctuation marks can be used as syntax symbols in the SQL column expressions. If an error is returned for a calculated field, first check whether Chinese punctuation marks are entered.

After you complete the settings on the dataset editing page, save the dataset before refreshing data.

Currently, the calculated fields that are already added as the expressions cannot be used for other calculated fields. If the physical layer of the original basic field in a calculated field is deleted, the calculated field is invalid.

26.3.2.3.8.1 How to use a calculated field

Non-aggregated calculated fields can be used as dimensions, or as measurements after the aggregation mode is set. Aggregated calculated fields can only be used as dimensions. They cannot be used as measurements.

You can set "Data Type" for a calculated field. Currently, "Data Type" can be set to "Number", " Text", or "Datetime".

For a calculated field, if you set "Data Type" to "Text", actual content to text, and aggregatio n mode to either of "sum" or "avg", the query result is not displayed due to the report type conversion error.

Similar to dimensions and measurements generated by the original fields in data sources, the calculation dimensions or measurements can be used in rows and columns, the attribute panel, and filter. You can also perform conversion between dimensions and measurements of calculated fields.

26.3.2.3.8.2 Calculated field examples

- Sum aggregation: sum([order amount])
- Average value aggregation: avg([order amount])
- Maximum value aggregation: max([order amount])
- Minimum value aggregation: min([order amount])
- Count aggregation: count([customer name]
- Count (deduplicated) aggregation: count(distinct[customer name])

Arithmetic

- Order cost ([Order amount] – [Profit amount])/100

String truncation

- substring([customer name], 1, 1)

Case measurement interval group

- Order amount interval

  Case when [order amount] < 500 then 'small order' when [order amount] >= 500 and [order amount] < 2,000 then 'medium order' when [order amount] >= 2,000 and [order amount] < 5, 000 then 'large order' else 'ultra-large order' end

- Case dimension member group (region after combination of specific provinces)

case when [province] in ('Heilongjiang', 'Liaoning', 'Jilin') then 'Northeast region' else [province] end

- Composite aggregate measurement —— Average order amount per person

  sum([order amount])/count(distinct[customer name])

- Unix timestamp preparation

  from_unixtime([order No.] + 1234567890)

- Extract different days in a month

  day([order date])

  Return a number ranging from 1 to 31

- Extract different hours in a day

  hour([order date])

  Return a number ranging from 0 to 23

- Ad effectiveness conversion rate

  case when sum([access times]) > 0 then sum([conversion times])/sum([access times]) else 0 end

  sum(case when [access times] > 0 then [conversion times]/[access times] else 0 end) is incorrect. The division operation cannot be performed before the sum operation for rate metrics. Instead, you must perform the sum operation first and then the division operation.

You can use various functions supported by the current database for calculated field expressions:

- MySQL function list applicable to AnalyticDB
- Greenplum function list

  http://www.postgres.cn/docs/9.5/functions.html?spm=5176.7730345.2.4.K8Pak6

- SQL Server function list

  https://msdn.microsoft.com/zh-cn/library/ms174318.aspx?spm=5176.7730345.2.5.b6orHl

26.3.2.3.8.3 Calculation measurement types

Calculation measurements can be classified into common measurements and aggregate measurements.

Measurements that are composed of expressions without aggregate functions are common measurements. Measurements that are composed of expressions with aggregate functions are aggregate measurements.

You can use the count() or count(distinct) function to form a deduplicated aggregate measurement using dimension fields as function parameters.

Examples of aggregate measurements: Average purchase amount per user sum(purchase amount)/countd(user ID), order cost proportion sum(order cost)/sum(order amount). However, avg (order cost/order amount) is an incorrect example.

Common measurements cannot be used together with aggregate measurements. Therefore, sum(order cost)/order amount is incorrect.

The aggregate modes of common measurements that do not contain aggregate functions can be used to change the aggregate functions. The menu options for changing the aggregate functions are not provided for aggregate measurements, and the aggregate measurements cannot be converted to dimensions.

The aggregate measurements support the following aggregate functions: SUM, AVG, MIN, MAX, COUNT, and COUNT distinct.

26.3.2.3.8.4 Create a calculated field

Context

For the usage instructions on calculated fields and related examples, see *How to use a calculated field* and *Calculated field examples*.

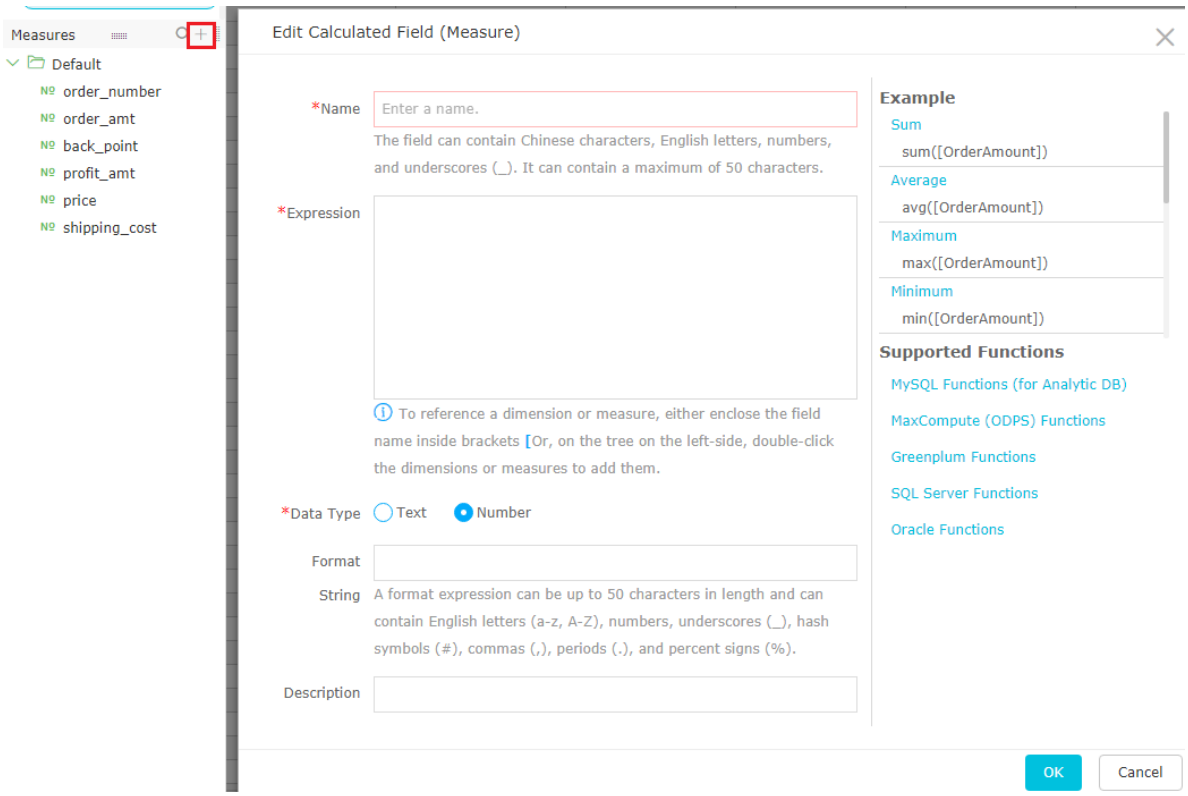
Calculated fields can be used as calculation dimensions and calculation measurements. For more information about calculation measurements, see *Calculation measurement types*.

The following uses company_sales_record as an example to calculate the average profit of orders .

Procedure

1. *Log on to the Quick BI console*, click **Data** to go to the data management page.
2. Select the **Dataset** tab and find the company_sales_record dataset.
3. Click **Edit** next to the dataset. The dataset editing page appears.
4. Click the plus sign in the measurement list. The "Create Calculated Field" dialog box appears, as shown in *Figure 26-45: Create a calculated field*.

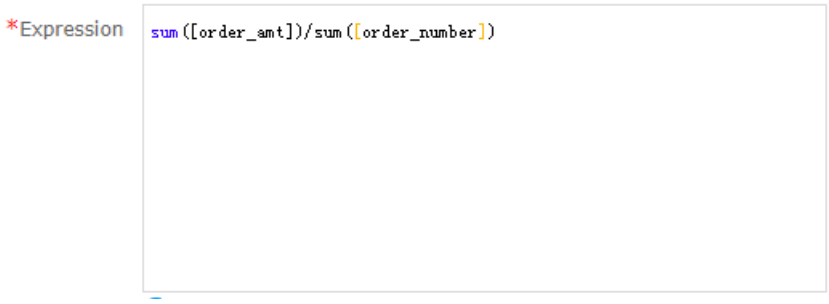
Figure 26-45: Create a calculated field



5. Enter a measurement name and expression, as shown in *Figure 26-46: Create a calculated field expression*.

   For example, if you want to calculate the average profit of orders, enter the expression Total profit of orders/Number of orders.

Figure 26-46: Create a calculated field expression



Note:

You must enter the expression in QWERTY mode.

6. Select a data type.

For example, if the average is a numeric value, select **Number** as the data type.

7. Click **OK** to create the field.
8. Click **Save** to save the dataset.
9. Click **Refresh Dataset** to view the new calculated field .

26.3.2.4 Delete a dataset

Context

If you delete a dataset with worksheets, an error occurs, as shown in *Figure 26-47: Error message for failed dataset deletion (1)*.

If you delete a dataset with dashboards, an error occurs when the dashboards are opened, as shown in *Figure 26-48: Error message for failed dataset deletion (2)*.

Figure 26-47: Error message for failed dataset deletion (1)

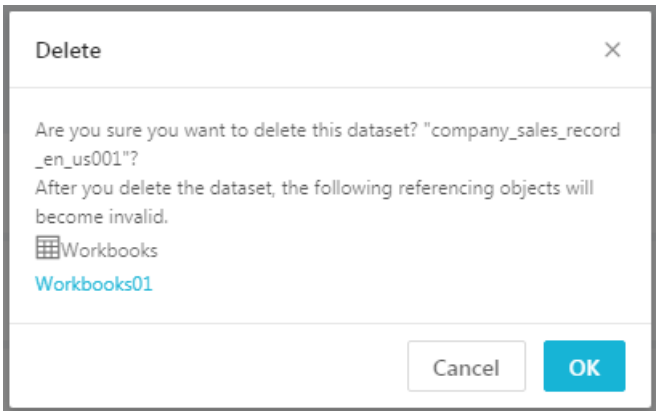
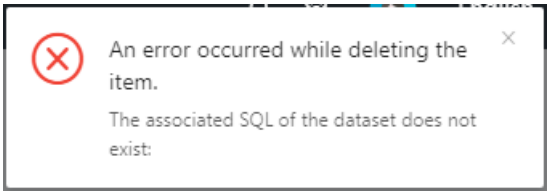


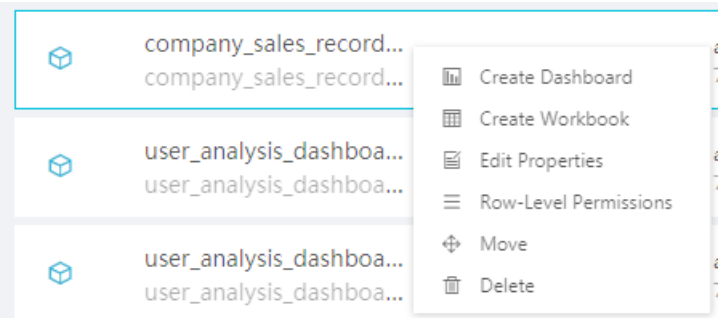
Figure 26-48: Error message for failed dataset deletion (2)



Procedure

1. On the dataset management page, select a dataset that you want to delete.
2. Right-click the dataset and choose **Delete** from the shortcut menu to delete the selected dataset, as shown in *Figure 26-49: Delete a dataset*.

Figure 26-49: Delete a dataset



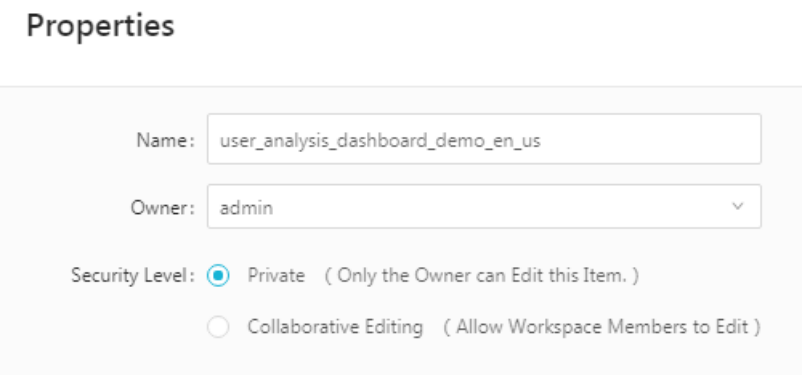
26.3.2.5 Rename a dataset

Context

Procedure

1. On the dataset management page, select a dataset that you want to rename.
2. Right-click the dataset and choose **Edit Properties** from the shortcut menu to rename the selected dataset, as shown in *Figure 26-50: Rename a dataset*.

Figure 26-50: Rename a dataset



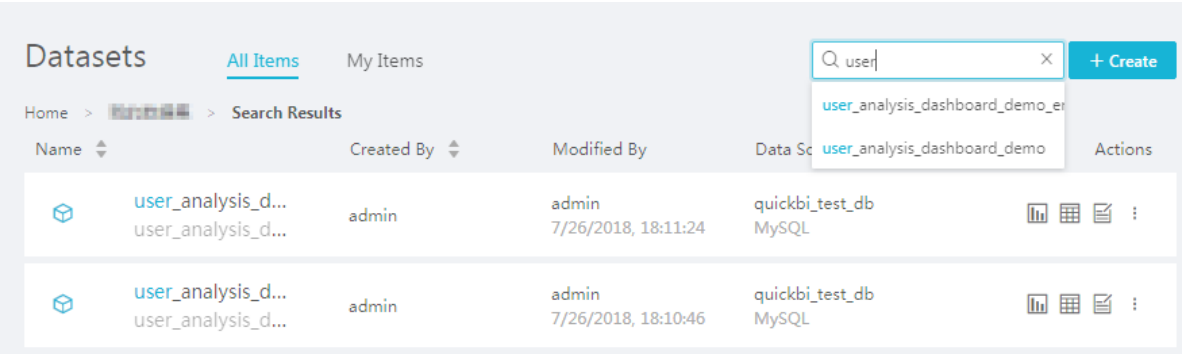
26.3.2.6 Query a dataset

Context

Procedure

1. On the dataset management page, find the search box.
2. Enter a keyword and click **Search** to search for the target dataset, as shown in *Figure 26-51: Query a dataset*.

Figure 26-51: Query a dataset



26.3.2.7 Create a dataset folder

Context

On the dataset management page, you can create multiple folders to facilitate dataset management.

Procedure

1. Select **Workbench** > **Datasets** to enter the dataset management page.
2. Select **Create** > **Folder Create Folder** from the shortcut menu.
3. Enter a folder name and click **OK** to complete the folder creation.

26.3.2.8 Rename a dataset folder

Procedure

1. On the dataset management page, select a folder that need to be renamed.
2. Click **Rename** next to the selected folder.
3. Enter a new folder name and click **OK** to complete the folder renaming.

26.3.2.9 Set row-level permissions on datasets

Context

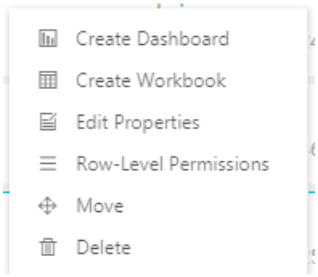
You can set row-level permissions on datasets to greatly reduce the workload of the permission administrator to maintain special organizational unit members who have relatively higher permissions.

You can control row-level permissions on datasets.

Procedure

1. On the **Dataset** List page, select a dataset.
2. Click the ellipsis (...) next to the dataset. The edit menu appears.
3. Click **Row-level Permissions**. The dialog box for dataset row-level permission control appears, as shown in *Figure 26-52: Row-level permission control*.

   Figure 26-52: Row-level permission control



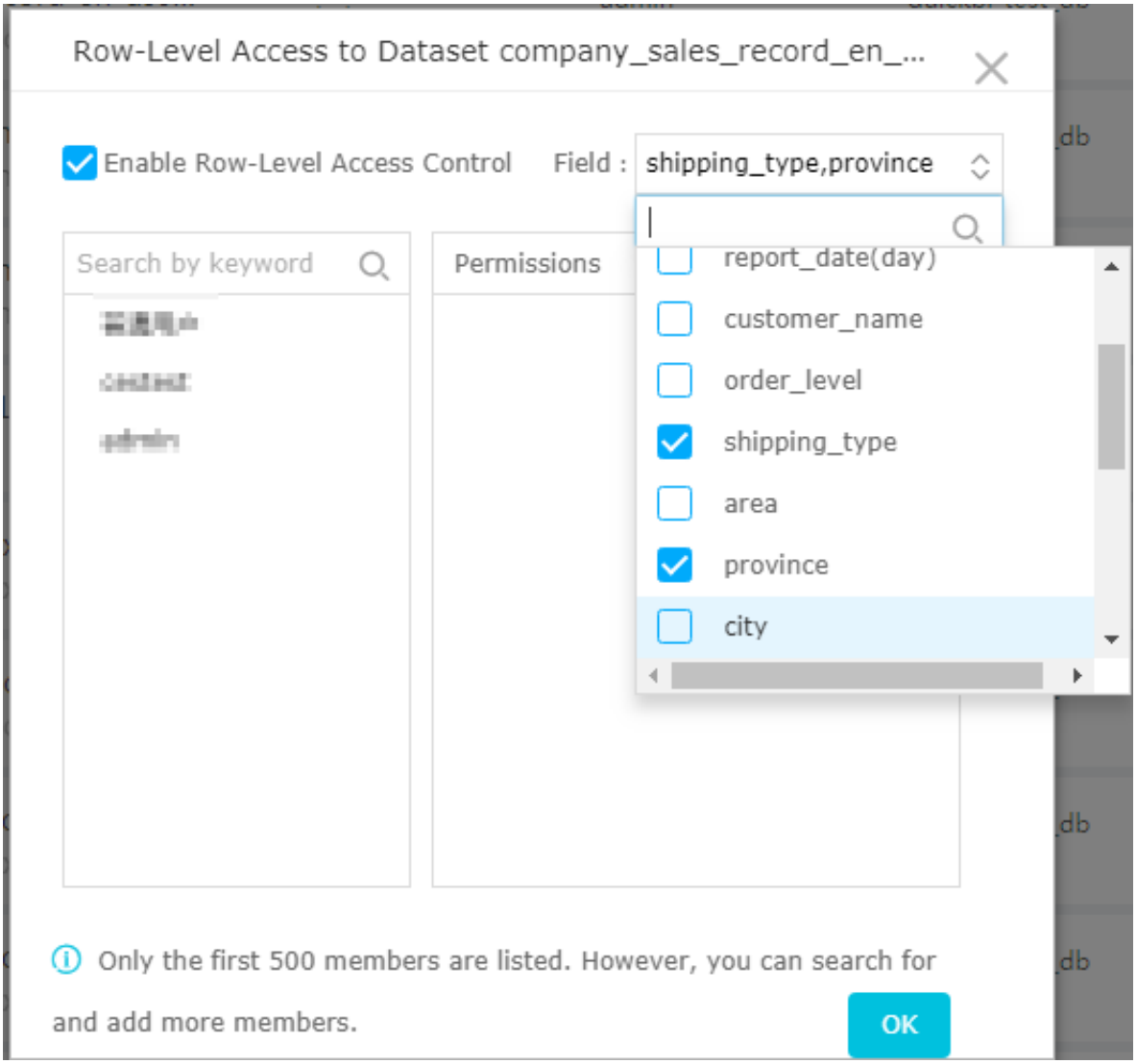
4. Select the fields you want to control on the dataset, as shown in *Figure 26-53: Select fields to be controlled*.

   Not all fields in the dataset require row-level permission control. Select the fields for row-level permission control as needed. For example, select the transportation method and province fields for row-level permission control.

   A dataset consists of dimensions and measurements. The list of all measurements forms a special field which is called measurement value.

   Members of the measurement value are all measurements in the dataset. You can control the measurement value field to display different measurements to different users.

Figure 26-53: Select fields to be controlled



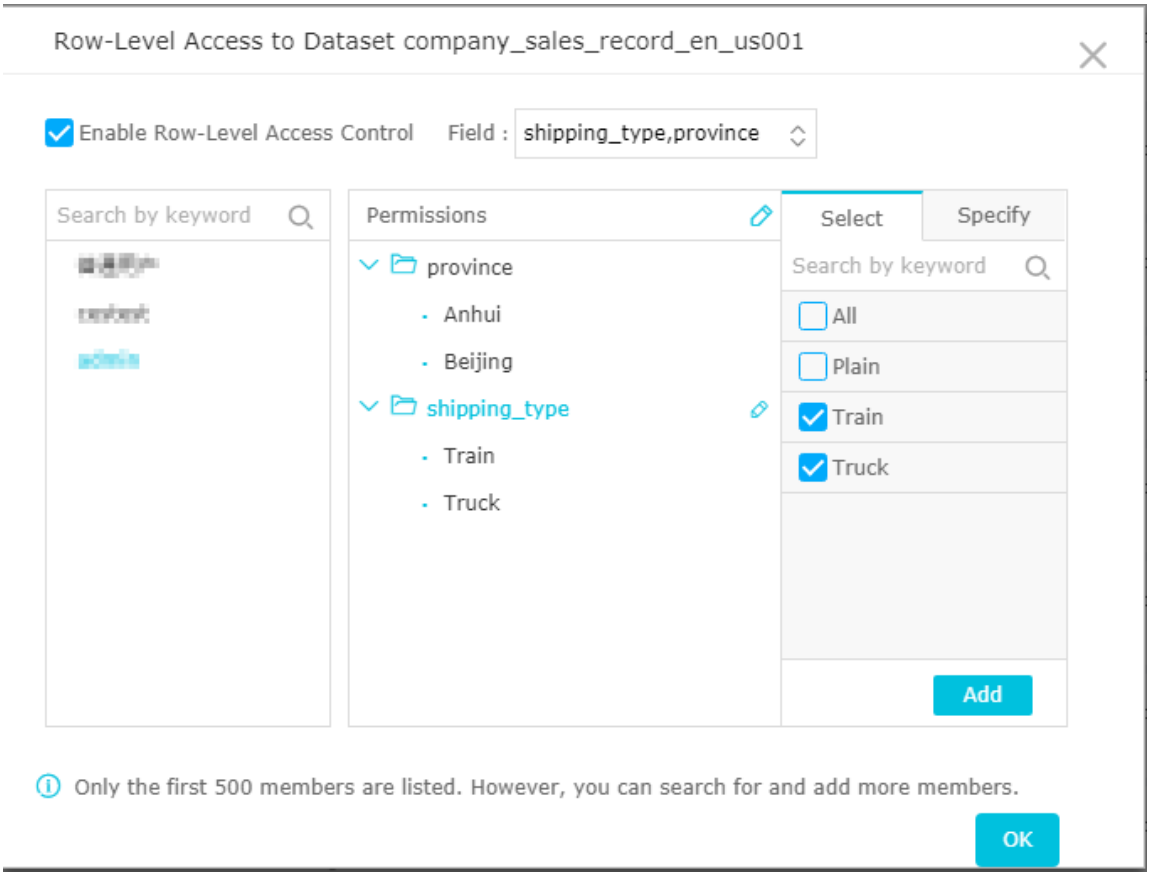
5. Set the list of field members who have access to the controlled fields for different users.

   In a dataset, if one field requires row-level permission control, you must specify the list of field members who have access to the controlled field in the dataset for all members in the organizational unit. Otherwise, no data is displayed due to lack of permission when the members access any data reports generated by the dataset.

   The following shows the procedure of setting row-level permissions for two different users, while in practice, you must set row-level permissions for all users.

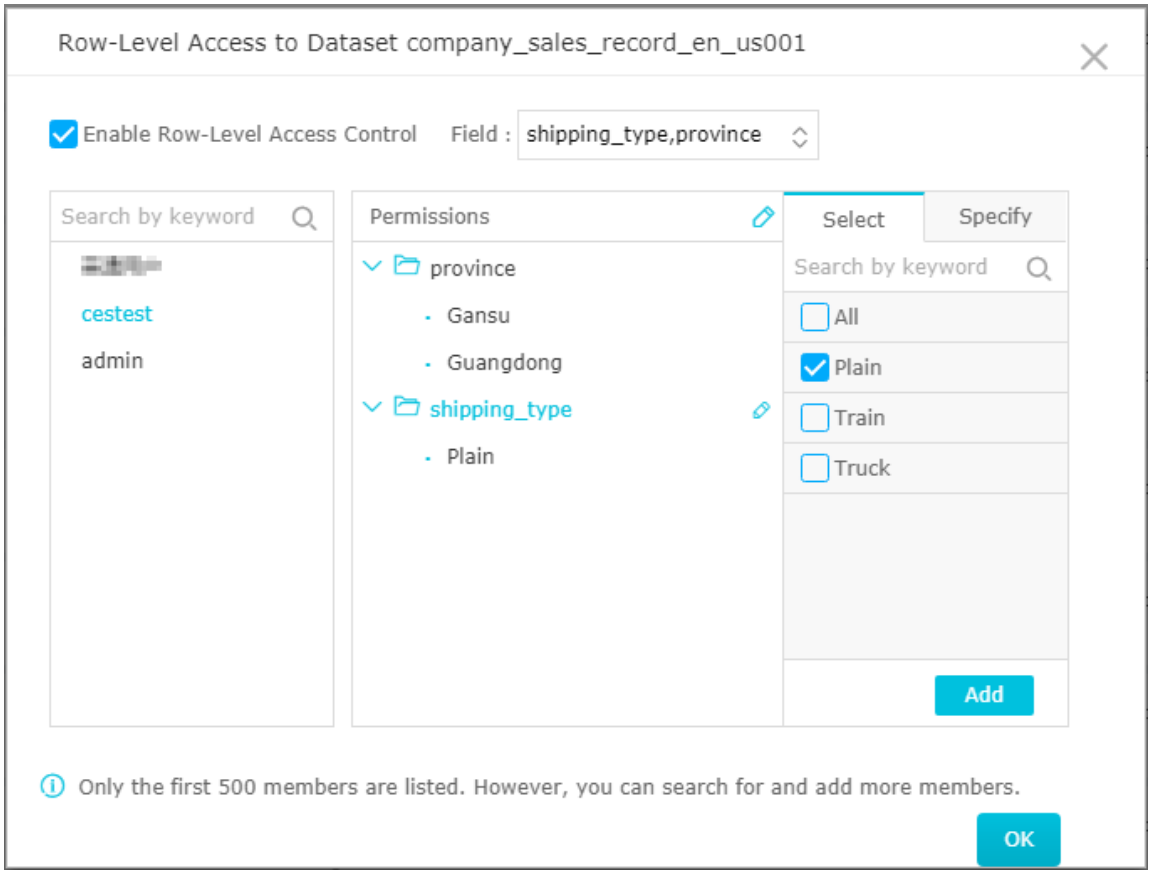
   - Set row-level permissions for User 1.

Figure 26-54: Set row-level permissions for User 1



- Set row-level permissions for User 2.

Figure 26-55: Set row-level permissions for user 2



The list only displays 500 member values under the field. If the number of member values under the field exceeds 500, and some members cannot be searched but actually exist, these members can be manually added to the list.

In the dialog box for selecting members from the member list, a special member called **`All`** exists. If you assign the member to an organizational unit member, this member is not restricted by row-level permissions on the field, no matter the number of field members increases or decreases in the future. After you select the **`All`** member option for this field, selection of other member options of this field has no restrictive effect.

6. Verify row-level permission control.

   If an organizational unit member has not been assigned the view permission on the controlled field, the report cannot be executed, and a message indicating that the member has no permission on a controlled field is displayed.

26.4 Manage dashboards

This section describes how to use dashboards to filter and query datasets and then use different data charts to visually display data and query dynamic data.

26.4.1 Dashboard

This section introduces the basic concepts of a dashboard, including the types, use cases, and data elements of charts in the dashboard.

26.4.1.1 Features of a dashboard

Quick BI provides powerful controls, allowing you to build pages for various products by dragging controls.

Quick BI also provides a variety of dashboard components, allowing you to prepare all types of reports easily by setting chart elements.

Dashboards use a more flexible tile layout to show interactions between report data. A dashboard not only visualizes data but also supports data filtering and query and multiple data display modes to highlight the key fields of data.

In terms of data display, dashboards display data in a more intuitive and clearer way through the wizard and drag, drop, and click operations on fields. In terms of data analysis, dashboards improve your interaction experience by prompts.

The data display performance is also greatly improved. You can query dynamic data on the dashboard editing page.

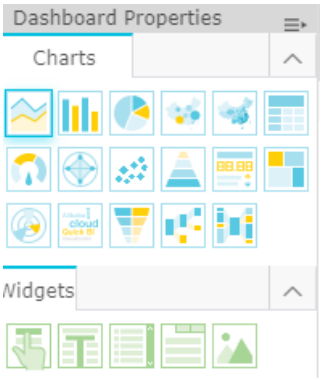
26.4.1.2 Dashboard optimizations and new features

Compared with the previous version, the current dashboard version has the following optimizati ons and new features:

Optimizations

- Optimizes the page layout, drag and drop interaction, and tile layout functions.
- Optimizes chart functions. For example, a scatter chart can support up to 1,000 dimension values.
- Merges the date control into the query control.
- Optimizes the color schemes of charts.
- Provides 17 chart styles and 5 controls.

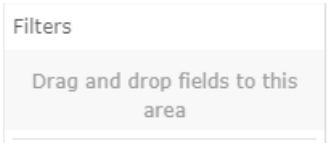
Figure 26-56: Chart styles and controls



New features

- Adds the field filtering function.

Figure 26-57: Filter



- Supports non-same-data-source association for the query condition control.
- Adds the function of canceling association to the advanced association functions of charts.

26.4.1.3 Types and use cases of data charts

Different chart types are required for displaying different data. Currently, Quick BI supports 17 types of data charts, including line charts, bar charts, bubble maps, and funnel charts.

For instructions on preparing a specific chart, see *Create a dashboard*.

The following table lists the analysis type and common use cases of each type of chart.

Table 26-1: Types and use cases of data charts

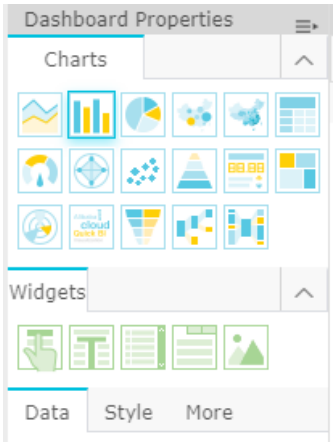
Analysis type	Description	Use case	Available chart
Comparison	Compares the differences between values or displays a simple comparison of measurements	Compare the sales/ income differences of different countries or regions.	Bar chart, radar chart , funnel chart, cross tabulation, polar chart , tornado funnel chart, and word cloud

Analysis type	Description	Use case	Available chart
	between different categories.		
Percentage	Displays the percentage of a part or the ratio of a value to the whole.	Identify the salesperson who contributes the most to the total sales.	Pie chart, funnel chart, dashboard, and matrix tree
Relation	Displays the relationship between values or compares multiple measurement values.	View the relativity between two values and analyze the influence of the first value on the second value.	Scatter chart, matrix tree, indicator panel, tree chart, and source direction chart
Trend	Displays the trend of a value (especially the trend changes with time, for example, by year, month, or day) or the progress or possible modes of an indicator.	View the sales or income trend of a product in a period.	Line chart
Geographic chart	Intuitively shows the size and distribution of related data indicators in a country or region on a map. The datasets used must contain geographical data.	View the income of each region in a country.	Bubble map and color map

26.4.1.4 Data elements of data charts

Each chart has the **Data**, **Style**, and **More** tabs, as shown in *Figure 26-58: Chart tabs*.

Figure 26-58: Chart tabs



- The "Data" tab determines the data displayed on the chart.
- The "Style" tab determines the chart appearance and displayed details.
- The "Advanced" tab determines whether data is associated with multiple charts and whether to dynamically display interaction and comparison between data as needed.

Each chart is differentiated from other charts by its core data elements. For example, the core element of a map is its geographic latitude. Without the core element, data cannot be displayed on the map.

The following table lists the core data elements of each type of chart.

Table 26-2: Data elements of data charts

Chart name	Data element	Composition of data elements
Line chart	Category axis and value axis	The category axis has at least one dimension, and the value axis has at least one measurement.
Bar chart	Category axis and value axis	The category axis has at least one dimension, and the value axis has at least one measurement.
Pie chart	Slice label and slice angle	Slice labels have only one dimension, and the dimension value is smaller than or equal

Chart name	Data element	Composition of data elements
		to 12. The slice angle has only one measurement.
Bubble map	Geographic region and bubble size	Geographic regions have only one dimension, which is the geographic latitude. The bubble size has one to five measurements.
Color map	Geographic region and color saturation	Geographic regions have only one dimension, which is the geographic latitude. The color saturation has one to five measurements.
Cross tabulation	Row and column	Rows have unlimited dimensions, and columns have unlimited measurements.
Dashboard	Pointer angle and tooltip	A dashboard has only one measurement.
Radar chart	Branch label and branch length	Branch labels have one to two dimensions, and the branch length has at least one measurement.
Scatter chart	Color legend, X axis, and Y axis	Color legends have only one dimension, and the maximum number of dimension members is 1,000. The X axis has one to three measurements, and the Y axis has only one measurement.
Funnel chart	Funnel layer label and funnel layer width	Funnel layer labels have only one dimension, and the funnel layer width has only one measurement.
Indicator panel	Panel label and panel indicator	Panel labels have at most one dimension. Panel indicators have at least one and at most 10 measurements.

Chart name	Data element	Composition of data elements
Matrix tree	Color block label and color block size	Color block labels have only one dimension, and the dimension value is smaller than or equal to 12. The color block size has only one measurement.
Polar chart	Slice label and slice length	Slice labels have only one dimension, and the dimension value ranges from 3 to 12. The slice length has only one measurement.
Word cloud	Word size and word label	The word size has only one dimension, and word labels have only one measurement.
Tornado funnel chart	Comparison subject and comparison indicator	The comparison subject has only one dimension, and comparison indicators have at least one measurement.
Tree chart	Tree parent and child node label and tree parent and child node indicator	Tree parent and child node labels have at least two dimensions. Tree parent and child node indicators have at least one measurement.
Source direction chart	Previous page, current page , and next page; PV of the previous, current, and next pages, UV of the previous, current, and next pages, path conversion rate, and page bounce rate	Each data element has only one dimension or measuremen t.

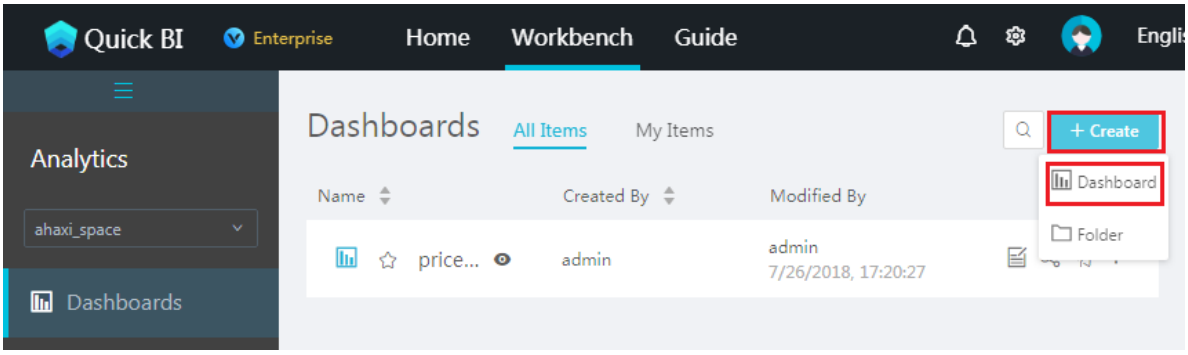
26.4.2 Access a dashboard

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.

3. Choose **Create** > **Dashboard** to enter the dashboard editing page, as shown in *Figure 26-59: Create a dashboard*.

Figure 26-59: Create a dashboard

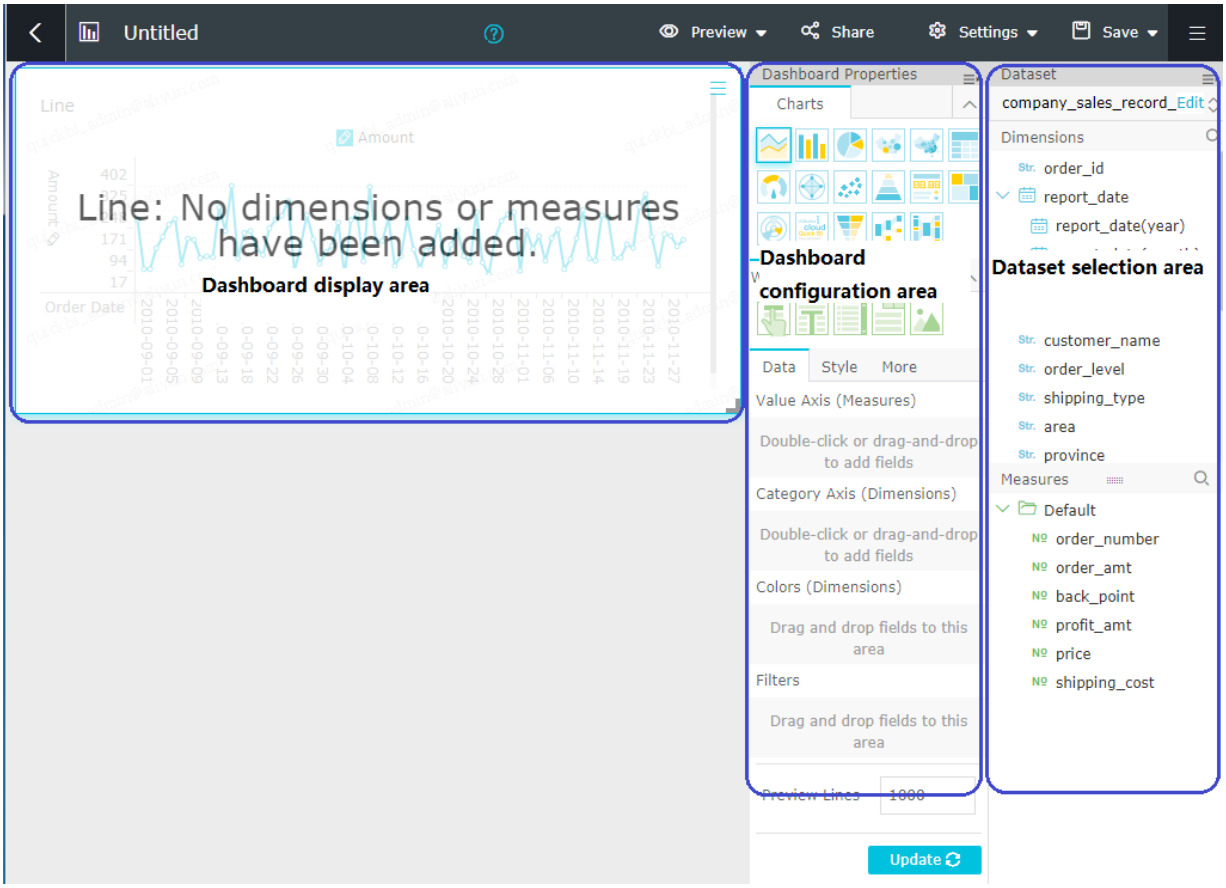


26.4.3 Areas of a dashboard

The dashboard editing page has three areas, as shown in *Figure 26-60: Dashboard*.

- Dataset selection area
- Dashboard configuration area (drawing board configuration)
- Dashboard display area (canvas)

Figure 26-60: Dashboard



- Dataset selection area: In this area, you can select a dataset. The fields of the selected dataset are displayed in the dimension and measurement lists based on the data types preset in the system. You can select the required dimension and measurement fields from the lists according to the data elements provided by data charts.
- Dashboard configuration area (drawing board configuration): In this area, you can select the expected chart type, and edit the title, layout, and legends of a chart as needed. By using the advanced feature, you can associate multiple charts to display data analysis results from multiple perspectives. You can also filter data by using the filter feature, and insert a query control to query key data in charts.
- Dashboard display area (canvas): In this area, you can drag charts to change their positions and select different chart types. For example, you can change from a bar chart to a bubble map. Based on the components of different charts, the system displays the missing or error element information if any. The function icons in the upper part of the dashboard display area enable you to save, preview, and create a dashboard. The dashboard editing page provides the guiding feature, which helps you learn how to create a dashboard.

26.4.3.1 Dataset selection area

In the dataset selection area, you can select a dataset or change from the current dataset to another one, and search for dimension and measurement fields.

26.4.3.1.1 Change from the current dataset to another one

Context

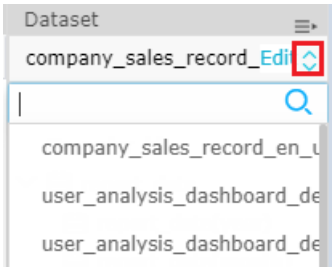
If you cannot find the expected dataset in the drop-down list, go back to the dataset management page to check whether the dataset has been successfully created.

Create a dataset. For more information, see *Create a dataset*.

Procedure

1. In the dataset selection area, click the **switchover** icon.
2. Select an expected dataset from the drop-down list, as shown in *Figure 26-61: Change from the current dataset to another one*.

Figure 26-61: Change from the current dataset to another one

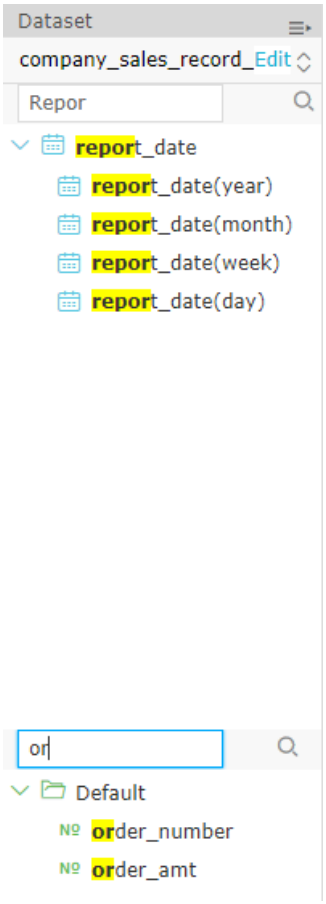


26.4.3.1.2 Search for dimension and measurement fields

Procedure

1. Enter a keyword, such as product in the field search box.
2. Click **Search**, as shown in *Figure 26-62: Search for dimension and measurement fields*.

Figure 26-62: Search for dimension and measurement fields



For information about how to edit dimension and measurement fields, see *Edit a dimension field* and *Edit a measurement field*.

26.4.3.2 Dashboard configuration area (drawing board configuration)

In the dashboard configuration area, you can select a chart and edit the chart.

26.4.3.2.1 Select a field

Context

Before creating a chart, make sure that you have selected a dataset in the dataset selection area and finished editing of the dataset.

For information about how to edit a dataset, see *Edit a dataset*.

Procedure

1. In the dashboard configuration area, select a chart type.

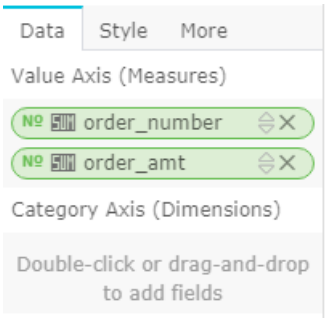
2. Double-click the chart type icon. A chart of this type is displayed in the dashboard display area.

   To change from the current chart type to another one, click the target chart type icon.

3. On the **Data** tab, select the required field, as shown in *Figure 26-63: Select a field*.

   Double-click the field, and the field appears in the dimension and measurement display area. Alternatively, you can also drag the field to the display area.

   Figure 26-63: Select a field



   - To delete a field from the display area, click **Delete** next to the field.
   - To sort the values of a field in ascending or descending order in the chart, click the up or down arrow next to the field.

4. Click **Update**. The system automatically draws the chart.

26.4.3.2.2 Enable color legend

Context

The color legend feature displays the values of the selected field in different colors in a chart.

You can only add dimension fields to the color legend area.

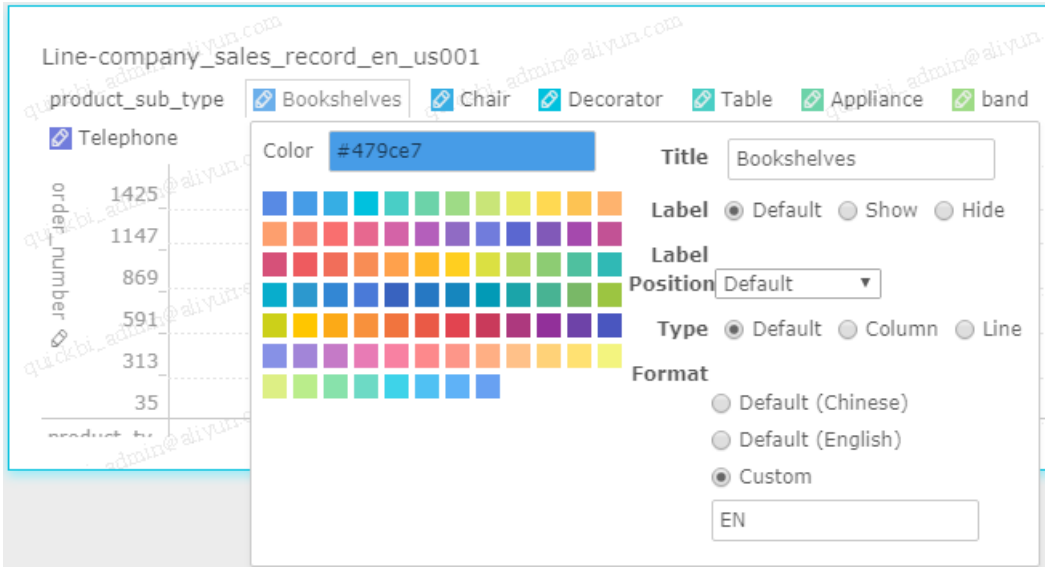
Procedure

1. Drag a dimension field, such as product_sub_type to the **Colors** area, as shown in *Figure 26-64: Color legend*.

Figure 26-64: Color legend

2. Click **Update**. The values of the selected field are marked by different colors in the chart.
3. By clicking the color icon before a product type, you can change the color scheme. See *Figure 26-65: Legend color scheme*.

Figure 26-65: Legend color scheme



26.4.3.2.3 Sort

Context

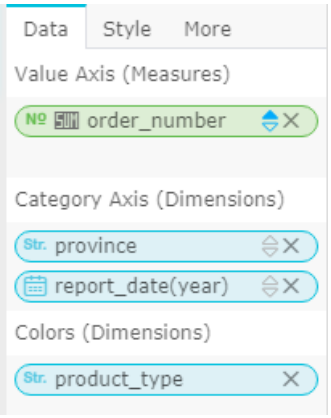
On the **Data** tab, you can sort data based on the specified dimension and measurement fields.

Procedure

1. Select a field, for example, order_amt.
2. Click the up arrow next to the field, as shown in *Figure 26-66: Set sorting*.

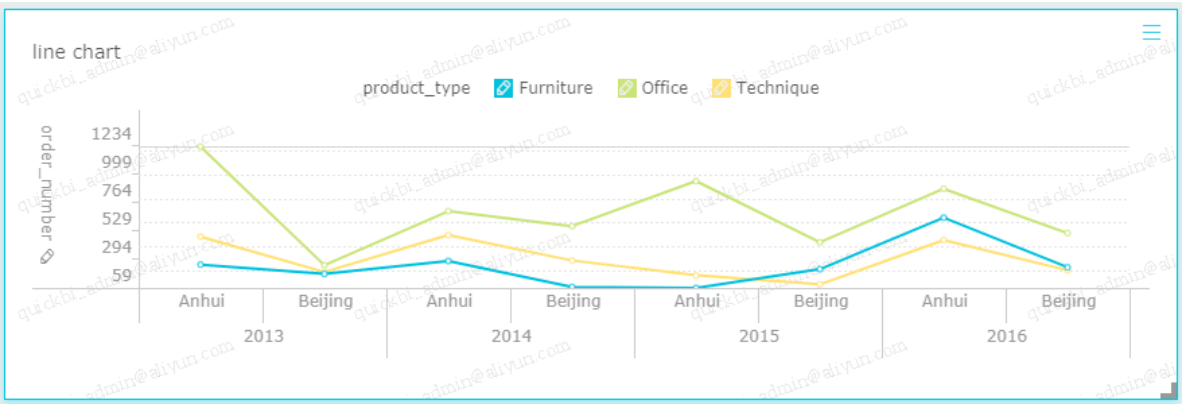
The up arrow indicates the ascending order, and the down arrow indicates the descending order.

Figure 26-66: Set sorting



3. Click **Update**. *Figure 26-67: Sorting result* shows the updated chart.

Figure 26-67: Sorting result



26.4.3.2.4 Filter a field

Context

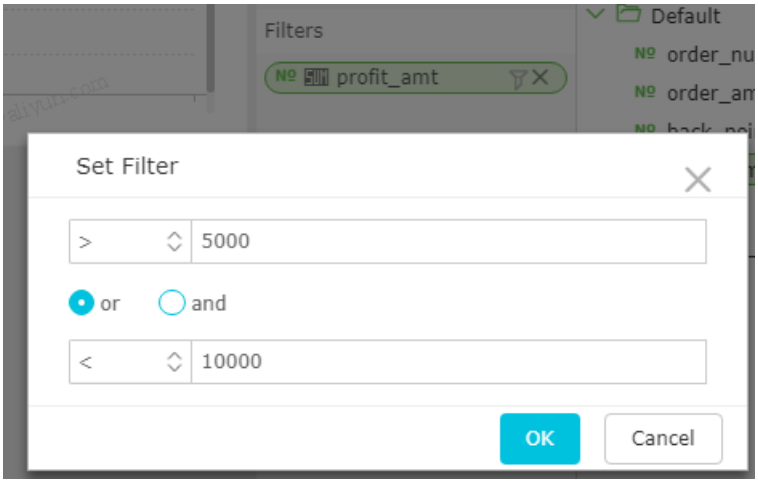
Drag a dimension or measurement field to the **Filter** area to filter the content of the field.

The following procedure uses the **`profit_amt`** field as an example to describe how to filter data.

Procedure

1. Drag the **`profit_amt`** field to the Filter area.
2. Click **Filter**. In the dialog box that appears, set a filter, as shown in *Figure 26-68: Set a filter*.

Figure 26-68: Set a filter



3. Select the expected filter condition, such as Greater Than, Less Than, and Equal To.
4. After the preceding settings, click **OK**.
5. Click **Update**. The system redraws the chart based on the filter settings.

26.4.3.2.5 Associate multiple charts

Context

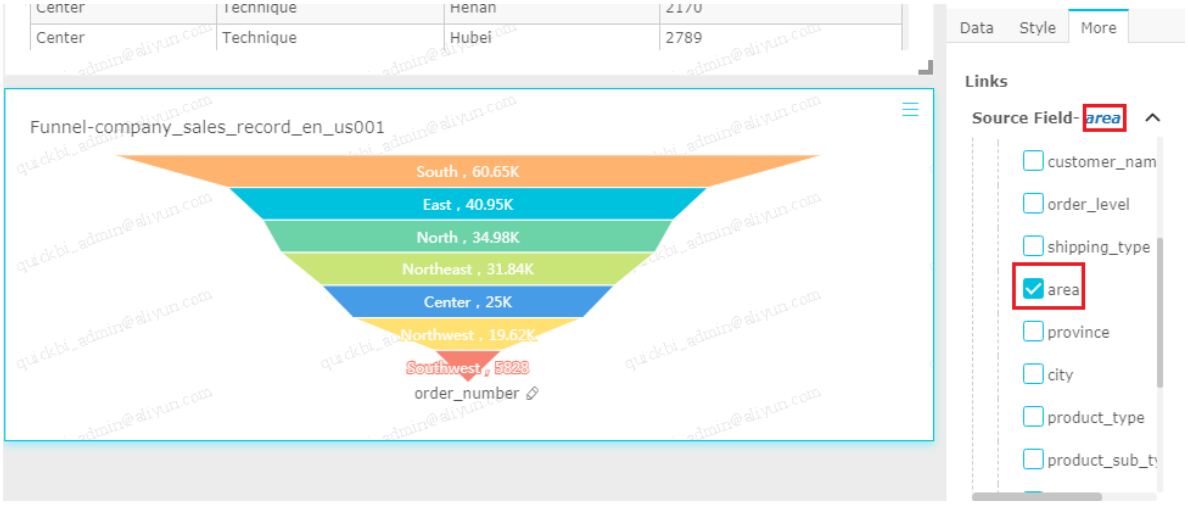
On the **More** tab in the dashboard configuration area, associate multiple charts by performing the following steps:

Before associating charts, make sure that at least two charts are available in the dashboard display area.

Procedure

1. Select a chart, for example, a funnel chart.
2. In the dashboard configuration area, select the **More** tab.
3. The **More** tab displays the available charts that can be associated with the selected funnel chart.
4. In an available chart, select the same field as the source field to associate this chart with the funnel chart. See *Figure 26-69: Set association*.

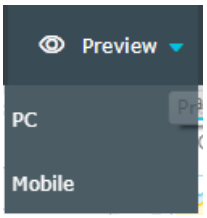
Figure 26-69: Set association



If the selected field is different from the source field, the system displays an error message.

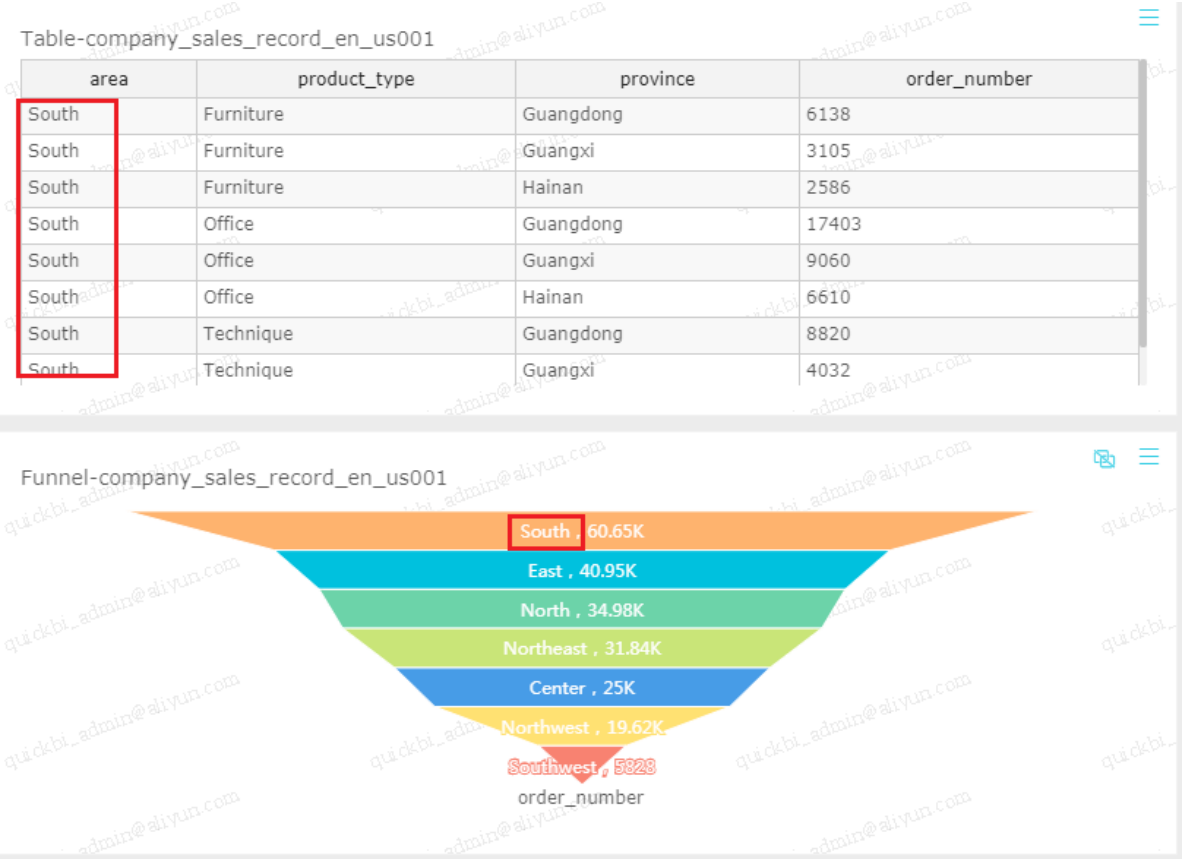
5. In the upper part of the dashboard display area, select **Preview** > **PC** . The preview page appears.

Figure 26-70: Dashboard preview icon



6. Click **China North** in the funnel chart, and the associated crosstab chart automatically displays other data of China North, as shown in *Figure 26-71: Association result*.

Figure 26-71: Association result



7. Click **Unlink** in the upper-right corner of the funnel chart can make the table goes back to the original status.

26.4.3.3 Dashboard display area (canvas)

You can perform the following operations on one or more charts in the dashboard display area.

- Manage dashboards
- Adjust the positions of charts.
- View chart data.
- Delete charts

26.4.3.3.1 Toolbar

In the dashboard display area, you can save, preview, or create a dashboard. See *Figure 26-72: Dashboard toolbar*.

Figure 26-72: Dashboard toolbar

26.4.3.3.2 Adjust the positions of charts

Context

The dashboard display area (canvas) can contain one or more charts. You can drag charts to adjust their positions.

Procedure

1. Select a chart or widget.
2. Click and hold on the chart and then drag it to the specified position.

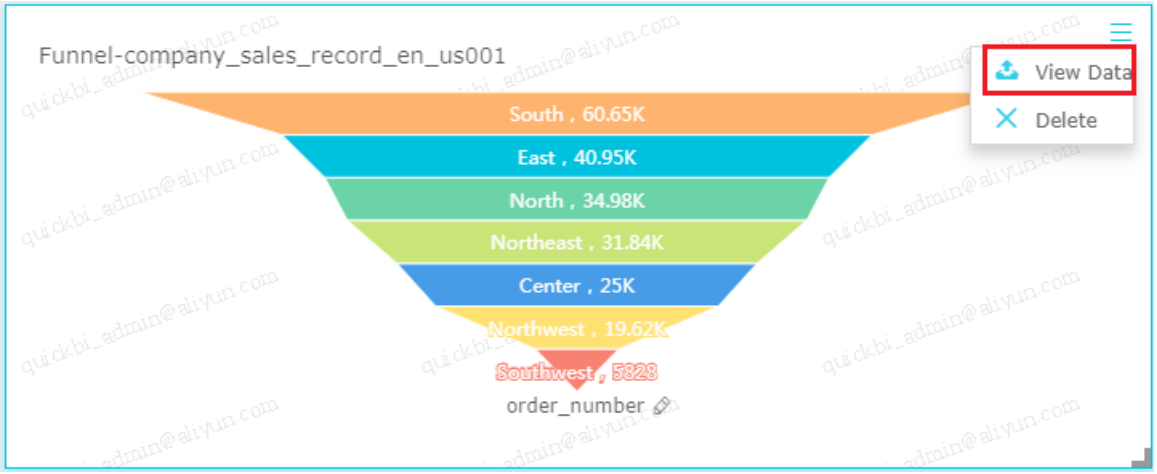
> **Note:**
> You can drag charts only to the dashboard display area (canvas).

26.4.3.3.3 View chart data.

Procedure

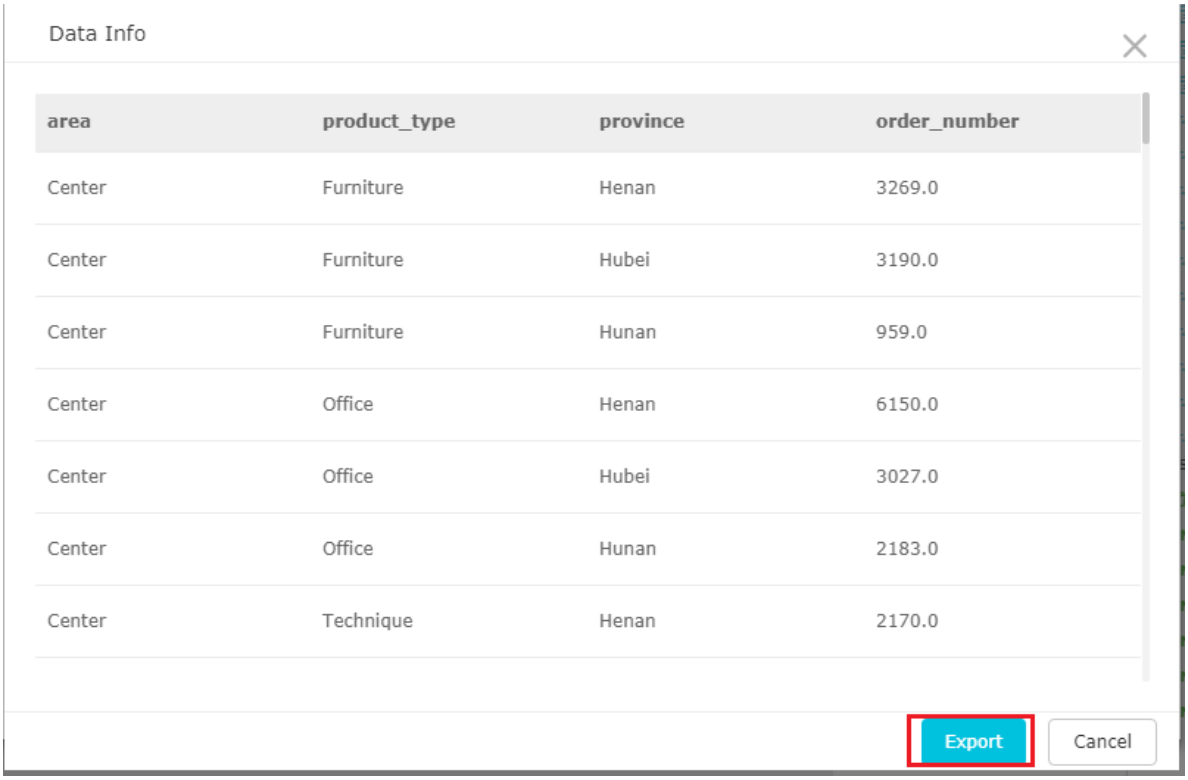
1. Select a chart, for example, a funnel chart.
2. Place the cursor in the upper-right corner of the chart.
3. Click **View Data**, as shown in Figure *Figure 26-73: View chart data.*.

Figure 26-73: View chart data.



4. Click **Export** to download the data to your local computer, as shown in *Figure 26-74: Export chart data*.

Figure 26-74: Export chart data

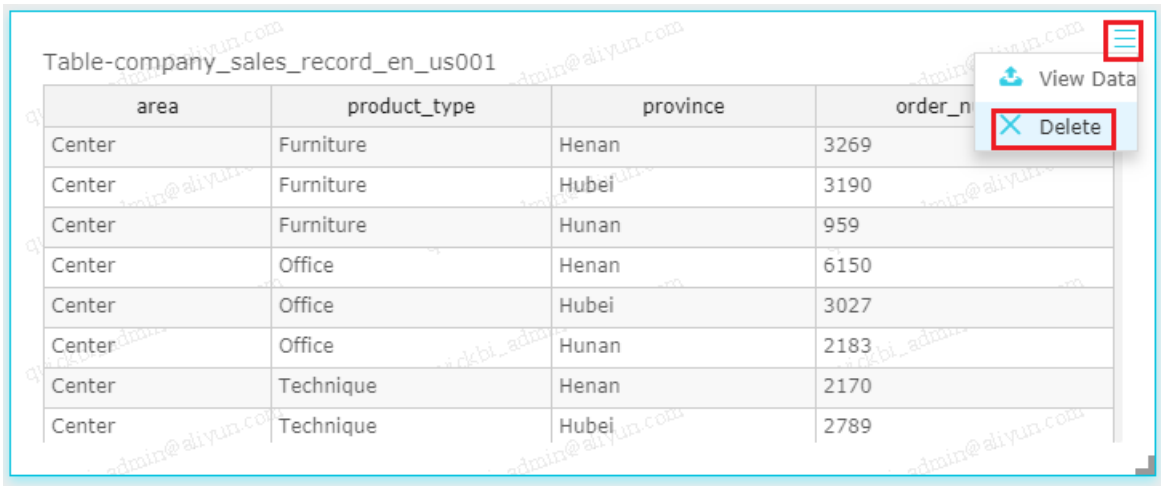


26.4.3.3.4 Delete charts

Procedure

1. Select a chart, for example, a radar chart.
2. Place the cursor in the upper-right corner of the chart.
3. Click **Delete**, as shown in *Figure 26-75: Delete charts*.

Figure 26-75: Delete charts



26.4.3.3.5 Select different chart types

Context

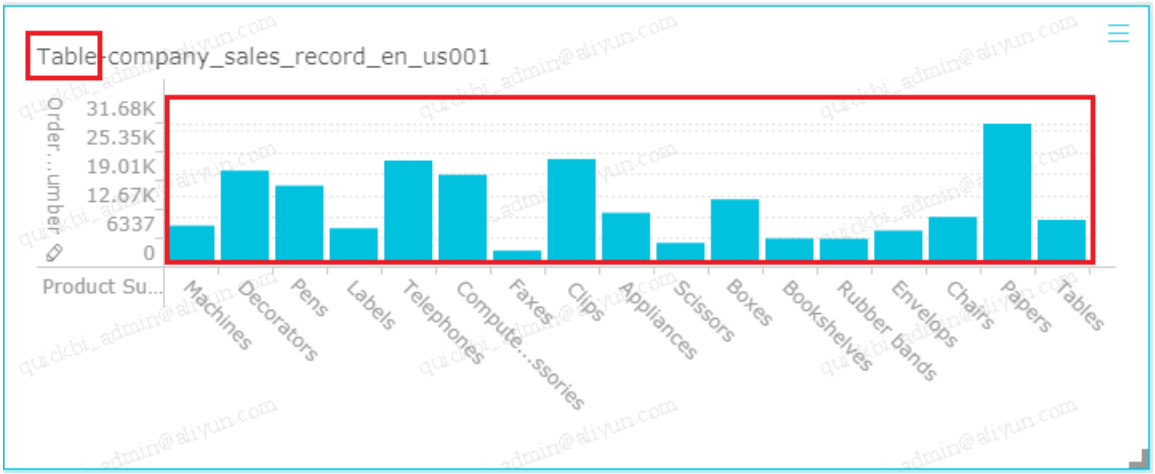
You can select different chart types in the dashboard display area (canvas).

Procedure

1. Select a chart type, for example, crosstab chart.
2. In the dashboard configuration area (canvas), select another chart type, for example, **bar chart** .
3. Click the **bar chart** icon to change the chart type.

   The system automatically changes from the crosstab chart to the bar chart.

   Figure 26-76: Select different chart types



If the system fails to change between chart types, the elements of the selected chart type do not match those of the current chart type. In this case, manually adjust chart elements.

The system displays a message indicating the names of elements to be adjusted, based on the chart type you want to change to. See *Figure 26-77: System prompt*.

Figure 26-77: System prompt

Manually adjust the dimension and measurement fields as prompted to complete chart type change.

26.4.3.3.6 Guiding feature

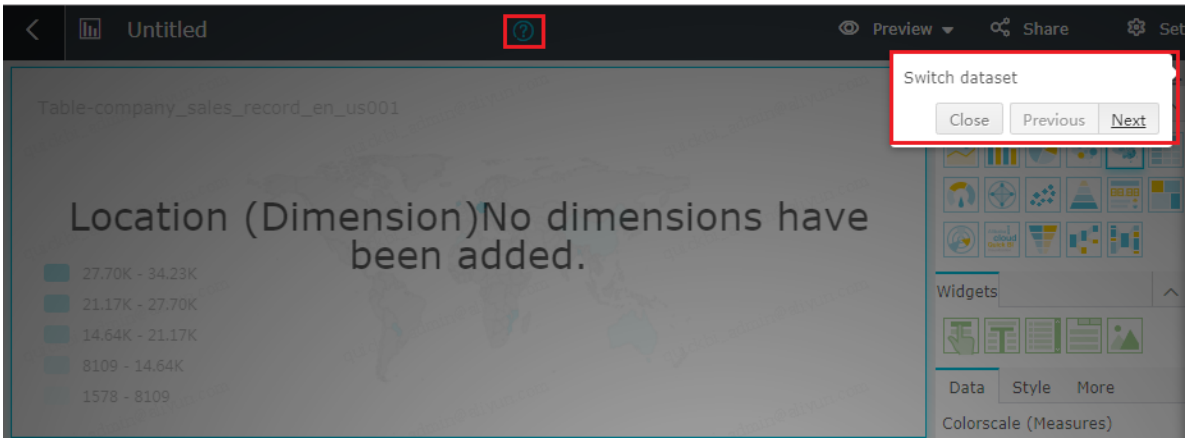
Context

You can enable the guiding feature in the dashboard display area (canvas). The system automatically demonstrates how to create a dashboard and you can create a dashboard accordingly.

Procedure

1. In the dashboard display area (canvas), click the **question mark (?)**.
2. You can click **Previous**, **Next**, or **Close** to learn how to create a dashboard, as shown in *Figure 26-78: Guiding feature*.

Figure 26-78: Guiding feature



26.4.3.3.7 Widgets

The dashboard display area supports the following widgets.

- Filter bar
- Text area
- Iframe
- TAB
- PIC

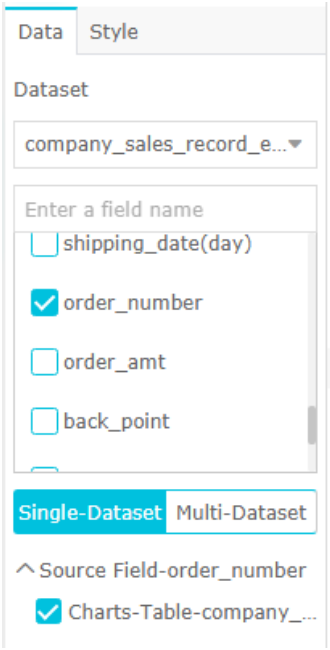
26.4.3.3.7.1 Filter bar

In the dashboard configuration area, you can select the filter bar to query data in one or more charts.

Context

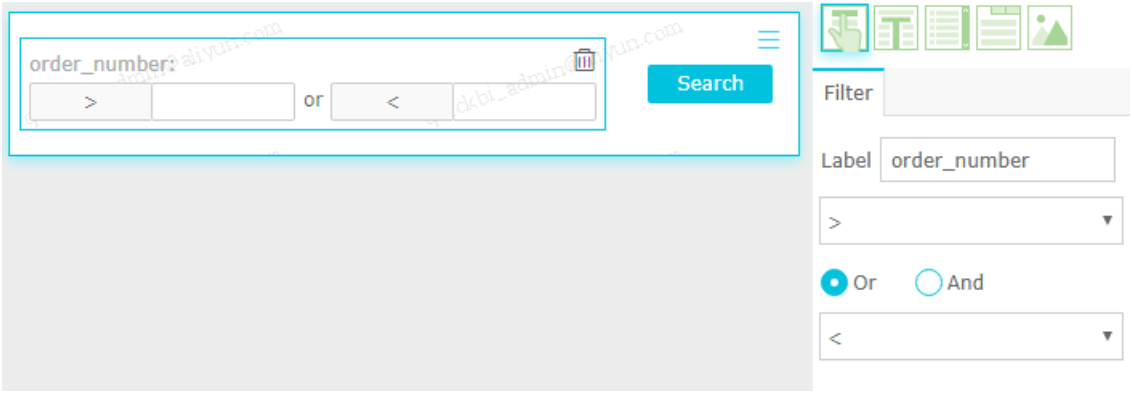
Double-click the **Filter Bar**. The editing menu of the control appears, as shown in *Figure 26-79: Filter conditions*.

Figure 26-79: Filter conditions



Click Filter Condition in the Query Condition control. The filter condition editing menu appears, as shown in *Figure 26-80: Filter condition editing menu*.

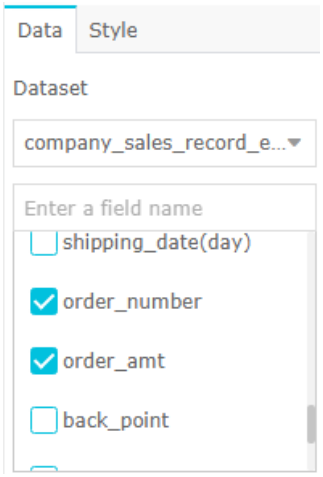
Figure 26-80: Filter condition editing menu



Procedure

1. Double-click **Filter Bar** .
2. On the Data tab, select a dataset and query condition fields, as shown in *Figure 26-81: Set query conditions*.

Figure 26-81: Set query conditions

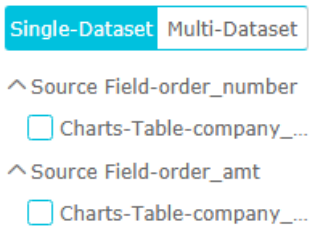


Currently, the Query Condition control supports single-dataset and multi-dataset.

Example of single-dataset association

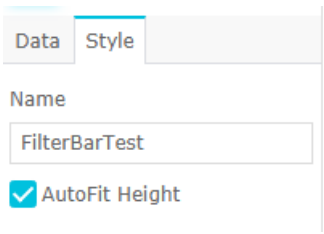
3. Select the **Single-Dataset** tab and select charts to which each query field applies. See *Figure 26-82: Single-dataset*.

Figure 26-82: Single-dataset



4. Select the **Style** tab, and edit the title of the widget and the position of the **Search** button, as shown in *Figure 26-83: Edit query conditions*.

Figure 26-83: Edit query conditions



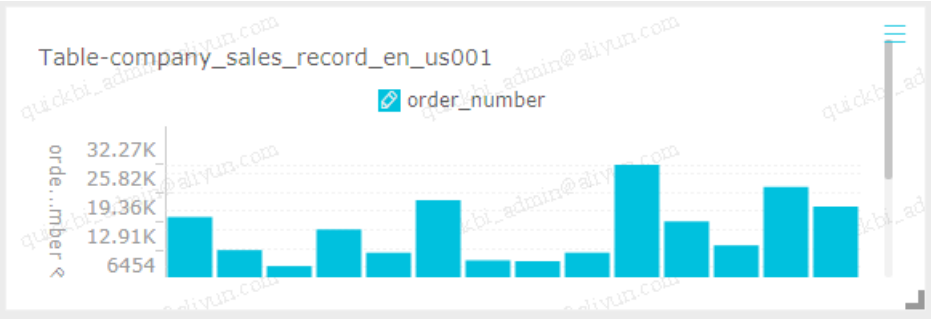
5. Select a field, for example, **order_number**.
6. Set the value is 5,000, as shown in *Figure 26-84: Set query conditions*.

Figure 26-84: Set query conditions

7. Click **Search**. The charts to which the query field applies are automatically updated, as shown in *Figure 26-85: Query results*.

   In the matrix tree, products with an order amount less than 5,000 are filtered out.

Figure 26-85: Query results



Example of multi-dataset

The Filter bar can also associate data from different datasets. Make sure that the values of data members of association items are consistent. Otherwise, the association is invalid.

The following procedure describes how to associate order level data from different datasets.

8. Select a dataset.
9. Edit dimension and measurement types.
10. Select a chart type, for example, table chart.
11. Select the expected fields add them to the corresponding rows and columns of the chart.
12. Click **Update**. The system automatically draws the chart.
13. In the **Style** tab, you can modify the title and layout of the table.
14. Click the dataset switchover icon to change from the current dataset to another one.
15. Edit dimension and measurement types.
16. Select a chart type, for example, table chart.
17. Select the expected fields add them to the corresponding rows and columns of the chart.

18.Click **Update**. The system automatically draws the chart.

19.In the **Style** tab, you can modify the title and layout of the table.

20.Double-click **Filter bar** and select a dataset and query condition fields.

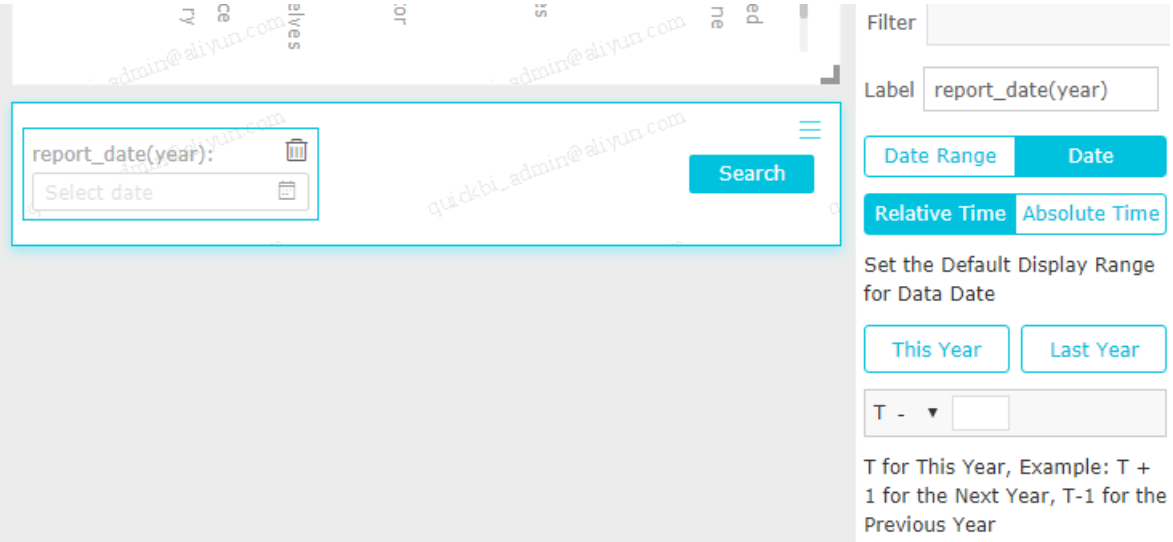
21.Select the **multi-dataset** tab and select association items according to field types.

26.4.3.3.7.1.1 Query data by date

Procedure

1. On the **Data** tab, select a dataset and a query condition field, for example, order_date.
2. Select charts to which the query field applies.
3. Click **report_date(year)**. The date selection page appears, as shown in *Figure 26-86: Query data by date*.

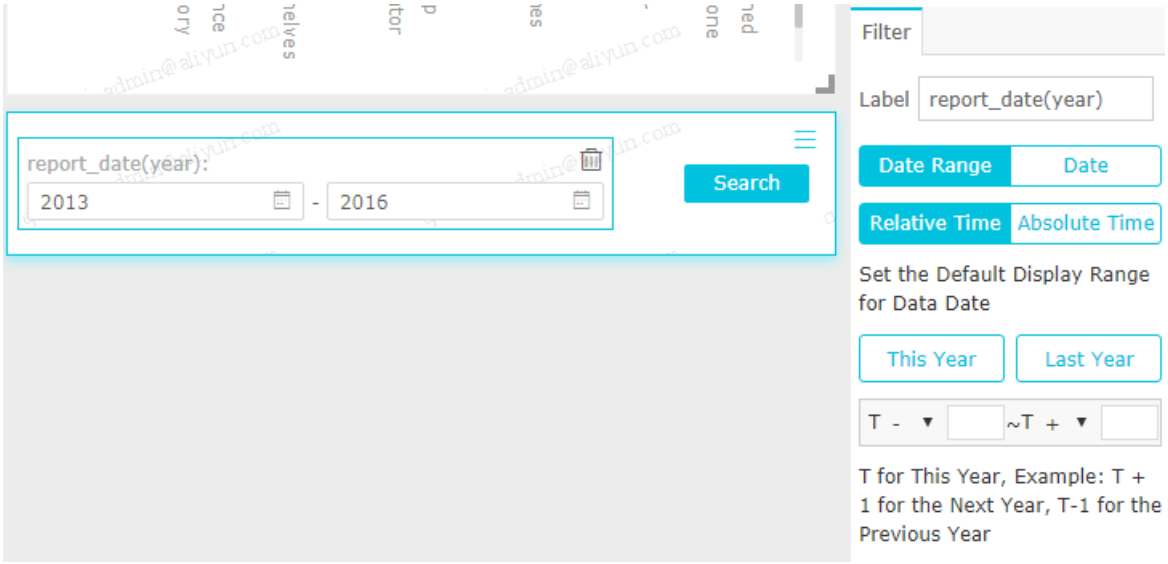
   Figure 26-86: Query data by date



4. Select a date range.

   The start date and end date of the last month are automatically displayed in the widget, as shown in *Figure 26-87: Edit the date range*.

Figure 26-87: Edit the date range



5. Click **Search**.

26.4.3.3.7.1.2 Query data based on text

Procedure

1. On the **Data** tab, select a dataset and a query condition field, for example, order_level.
2. Select charts to which the query field applies.
3. Click **order_level**. The text seection page appears, as shown in *Figure 26-88: Query data based on text*.

Figure 26-88: Query data based on text

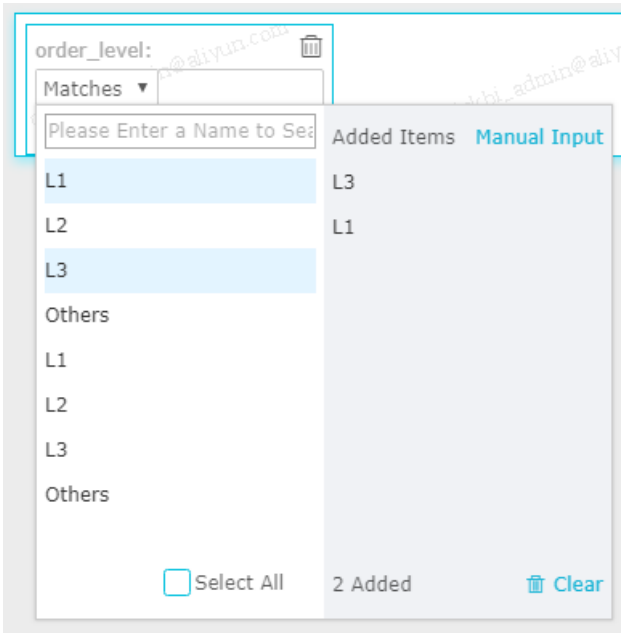
4. Select a text query condition, for example, **Enumeration**.

   The system automatically loads all options of order_level field to the Query Condition control.

   Select **Single-dataset** or **Multi-dataset**.

5. Click the drop-down arrow and select the option to be queried, as shown in *Figure 26-89: Enumeration*.

Figure 26-89: Enumeration



6. Click a blank area.

   The Query Condition control displays the content to be queried.

7. Click **Search**.

26.4.3.3.7.2 Text box

Context

You can enter fixed text in a text box and use the text as a report title.

Procedure

1. Double-click the text box icon.
2. Enter text in the text box, as shown in *Figure 26-90: Text box*.

Figure 26-90: Text box

26.4.3.3.7.3 IFrame

Context

iFrame enables you to insert a required webpage in a dashboard to query network data in real time or to browse the webpage or website related to the current data.

Procedure

1. Double-click the IFrame icon.
2. In the link area, enter a webpage address .

   Note:
   Webpage addresses must start with https.

   In the title area, you can change the title of the iFrame widget.

   If you want to delete the current widget, place the cursor in the upper-right corner of the widget and choose **Delete** from the shortcut menu.

26.4.3.3.7.4 TAB

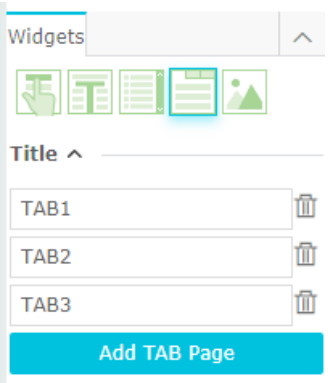
Context

The tab feature enables multiple charts to be displayed on different tabs.

Procedure

1. Double-click the **TAB** icon.
2. Click **Add TAB** to add a new tab, as shown in *Figure 26-91: Tab editing menu*.

   Figure 26-91: Tab editing menu



3. Click a tab and insert a chart in it.

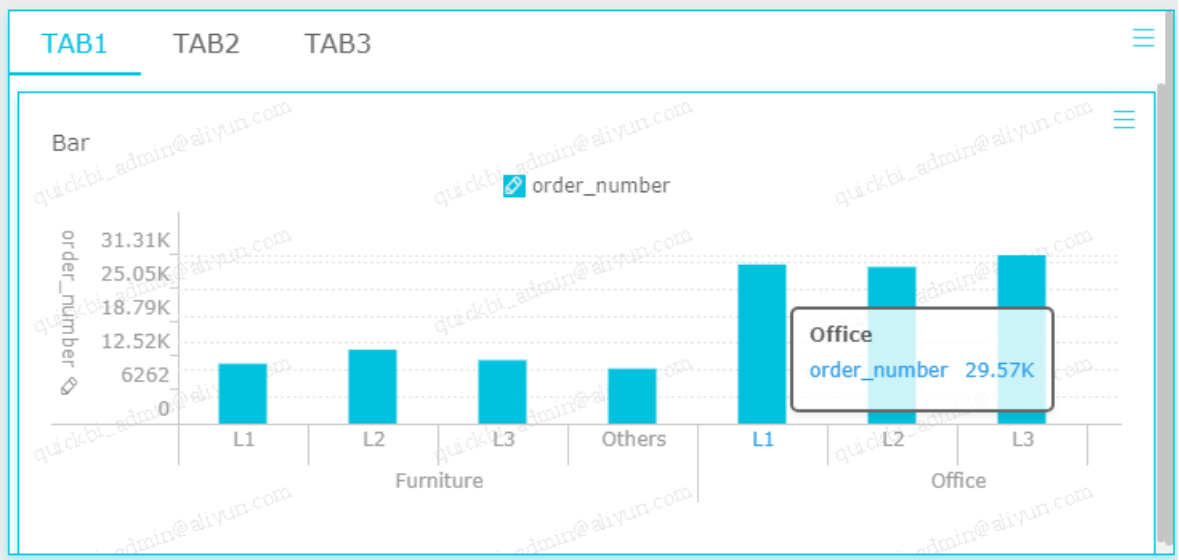
   Click TAB1. TAB1 then becomes blue.

4. Double-click the required chart icon. A chart is automatically inserted in TAB1, as shown in *Figure 26-92: Add a chart*.

   Figure 26-92: Add a chart

Create the chart by following the chart creation process. *Figure 26-93: "TAB" tab* shows the tab control after the chart is created.

Figure 26-93: "TAB" tab



If you want to delete the current widget, place the cursor in the upper-right corner of the widget and choose **Delete** from the shortcut menu.

26.4.3.3.7.5 PIC

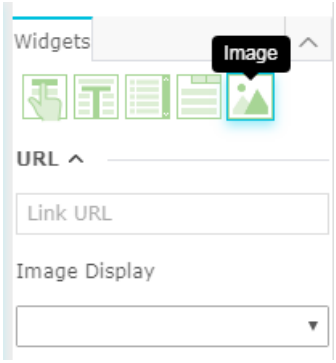
Context

The PIC function allows you to insert an image to a dashboard and adjust the image position and effect according to the display requirement.

Procedure

1. Double-click the **PIC** icon.
2. Enter the URL of an image.
3. Select display effect for the image, as shown in *Figure 26-94: PIC editing menu*.

Figure 26-94: PIC editing menu



26.4.4 Create a dashboard

This section describes how to prepare different types of charts.

26.4.4.1 Line chart

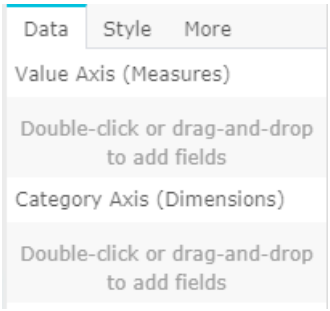
A line chart shows the trend of data changes using lines and displays the continuous time-varying data. It is well suited to analyze and display the data trend at equal time intervals. A line chart can also be used to analyze the changing mutual effect between multiple groups of data in a period. For example, use a line chart to show how the sales volume of one or more types of product changes with time, and further predict the future sales performance.

Context

A line chart consists of a category axis and a value axis. A category axis is horizontal and contains only dimension fields such as the date, province, and product type. The value axis is vertical and contains only measurement fields, such as the business indicator of the analysis object and order quantity.

In the dashboard, the system automatically matches the category axis, value axis, dimension fields, and measurement fields, as shown *Figure 26-95: Category axis and value axis*. Select the expected fields from the dimension and measurement lists as instructed by on-screen prompts.

Figure 26-95: Category axis and value axis



Select at least one dimension for the category axis and at least one measurement for the value axis. If you want to use the color legend function, select a maximum of one dimension for the color legend.

Note:
The color legend can be used only when one measurement field is set for the value axis. Otherwise, this feature is unavailable.

The following uses the company_sales_record dataset as an example to describe how to use a line chart to illustrate the order quantity of each type of products in each province per year.

Procedure

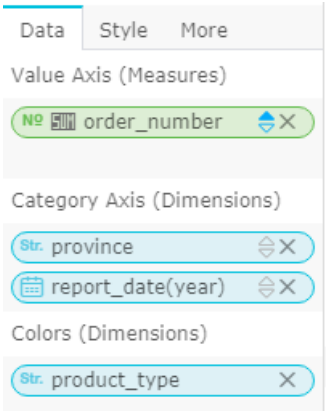
1. Log on to the Quick BI console.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **line chart** icon.
5. On the **Data** tab, select required dimension fields and measurement fields.

   In the dimension list, locate the order date (year), province, product type options, and add them to the category axis area in sequence. In the measurement list, locate the order quantity option, and add it to the value axis area, as shown in Figure *Figure 26-96: Select line chart fields*.

   Note:
   Make sure that the dimension type of the province field has been changed from string to geographical information.

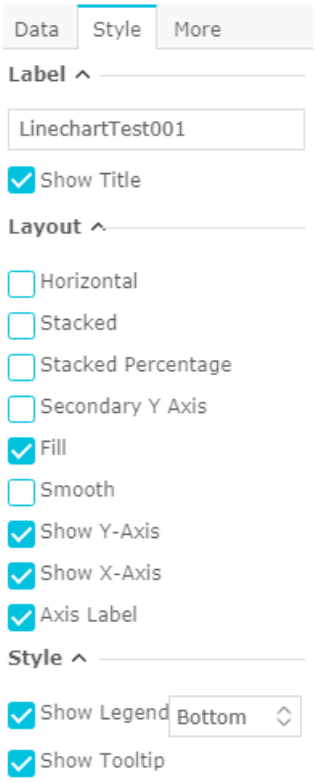
   For instructions on how to change the dimension field type, see *Edit a dimension field*.

Figure 26-96: Select line chart fields



6. Drag the **product type** field to the **Colors** area.
7. Click **Update**. The system automatically draws the chart.
8. On the **Style** tab, edit the title, layout, and legends of the chart, as shown in *Figure 26-97: Line chart after editing*.

Figure 26-97: Line chart after editing



9. Click **Save** and enter a name for the dashboard.
10. Click **OK** to save the dashboard.

By default, the dashboard is saved to **My Dashboard**.

26.4.4.2 Bar chart

For a clear demonstration of the differences among multiple groups of data, a bar chart can be used. It helps show data variations in a specific period or compares different items, for example, comparison of the traffic flows of a road crossing during different periods.

Context

A bar chart and a *line chart* have similar components. They both include a category axis and a value axis. A bar chart can be used to show data changes in a period or comparison between multiple items.

This section describes how to use a filter and how to create a double Y axis graph in the following two scenarios:

- Scenario 1: Compare the transportation costs of different products in provinces of China East.
- Scenario 2: Compare the order quantity and average profits of different products in the provinces.

Set at least one dimension for the category axis of a bar chart, such as a province or a product type. Set at least one measurement for the value axis, such as the order quantity or profit amount. Colors fuction only support dimension fields. A maximum of one dimension can be set for each color legend.

Note:
The colors fynction can be used only when one measurement field is set for the value axis. Otherwise, this feature is unavailable.

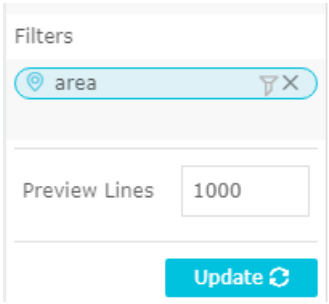
Scenario 1: The following uses the company_sales_record dataset as an example to describe how to use a bar chart to compare the transportation costs of different products in provinces of China East.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **bar chart** icon.
5. Select a region and add it to the filter, as shown in *Figure 26-98: Filter*.

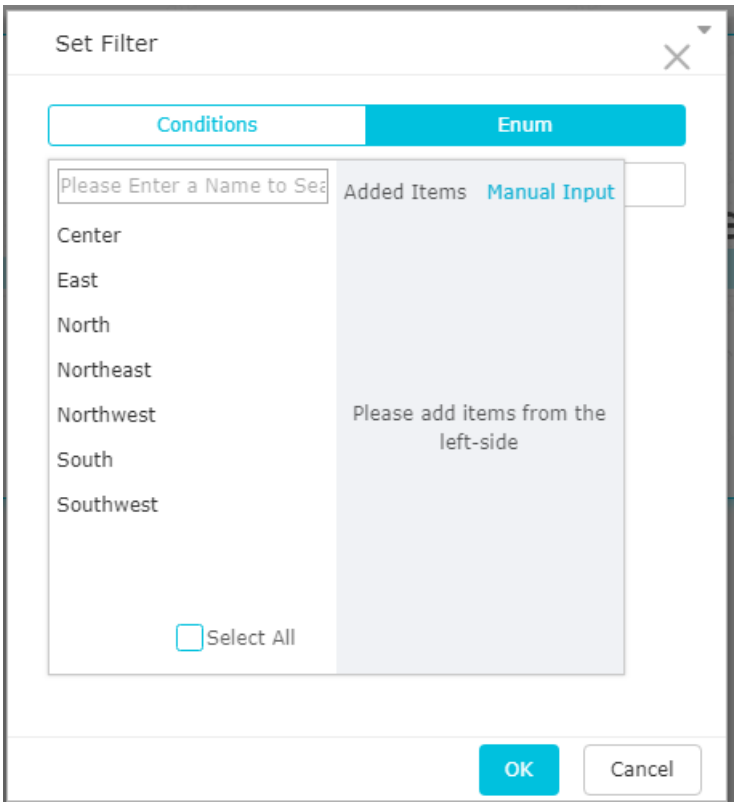
Use the filter to select **China East** from the region list.

Figure 26-98: Filter



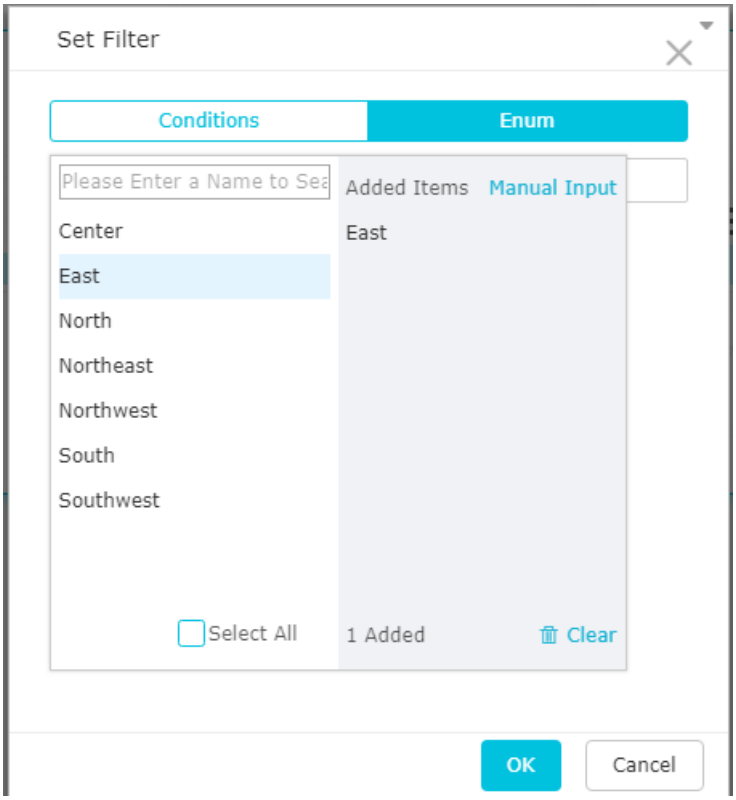
6. Click the **filter** icon and set a filter condition.
7. Select **Enumeration**. The system lists all region options, as shown in *Figure 26-99: Enumeration*.

Figure 26-99: Enumeration



8. Select **China East** and click **OK**, as shown in *Figure 26-100: Select "Enumeration"*.

Figure 26-100: Select "Enumeration"



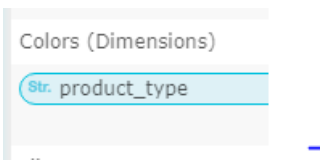
9. Locate the province and product type options, and add them to the category axis area in sequence.
10. Locate the transportation cost option, and add it to the value axis area.

> **Note:**
> Make sure that the dimension type of the province field has been changed from string to geographical information.

For instructions on how to change the dimension field type, see *Edit a dimension field*.

11. Drag the **product type** field to the **colors** area, as shown in *Figure 26-101: Enable color legend*.

Figure 26-101: Enable color legend



12. Click **Update** to update the chart.

13.On the **Style** tab, select **Stacked**.

Scenario 2: The following uses the company_sales_record dataset as an example to describe how to use a bar chart to compare the order quantity and average profits of different products in the provinces.

Data modeling may be required in this scenario. For instructions on how to perform data modeling, see *Create a calculated field*.

14.On the **Data** tab, select required dimension fields and measurement fields.

In the dimension list, locate the province and product type options and add them to the category axis area in sequence. In the measurement list, locate the order quantity and average profits options and add them to the value axis area in sequence.

15.Click **Update** to update the chart.

16.On the **Style** tab, select **Secondary Y Axis**.

17.Click **Save** and enter a name for the dashboard.

18.Click **OK** to save the dashboard.

By default, the dashboard is saved to **My Dashboard**.

26.4.4.3 Pie chart

A pie chart shows a data series and each data series has a unique color or pattern. A pie chart can be used to show the volume of each item its percentage in the total amount. For example, you can use a pie chart to show the ratio of the expense of five insurances (endowment, medical, unemployment, employment injury, and maternity insurances) and housing fund to the personal income, or the sales of a car brand to the total car sales.

Context

A pie chart consists of multiple slices. The label of each slice is determined by the data dimension , such as the region or product type. The angle (size) of each slice is determined by the data measurement, such as the order quantity or order amount.

The slice label area of a pie chart has at most one measurement, such as the region or product type, and the measurement value must be less than or equal to 12. The slice angle area has at most one measurement, such as the order quantity or profit amount.

The following uses the company_sales_record dataset as an example to describe how to use a pie chart to compare the transportation costs of different regions.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **pie chart** icon.
5. On the **Data** tab, select required dimension fields and measurement fields.

   In the dimension list, locate the region option and add it to the slice label area. In the measurement list, locate the transportation cost option, and add it to the slice size area, as shown in *Figure 26-102: Select pie chart fields*.

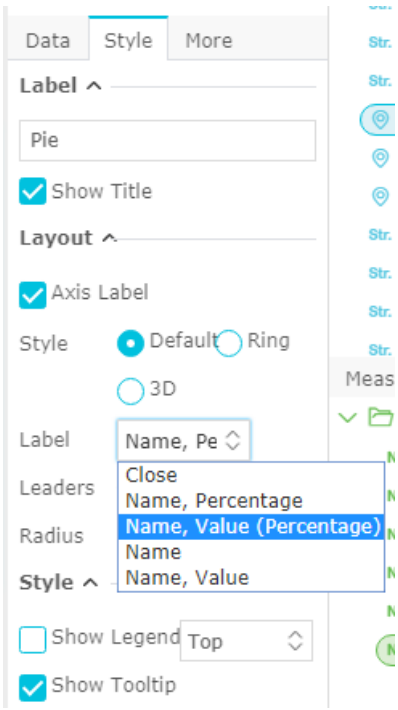
   > **Note:**
   > Make sure that the dimension type of the region field has been changed from string to geographical information. For instructions on how to change the dimension field type, see *Edit a dimension field*.

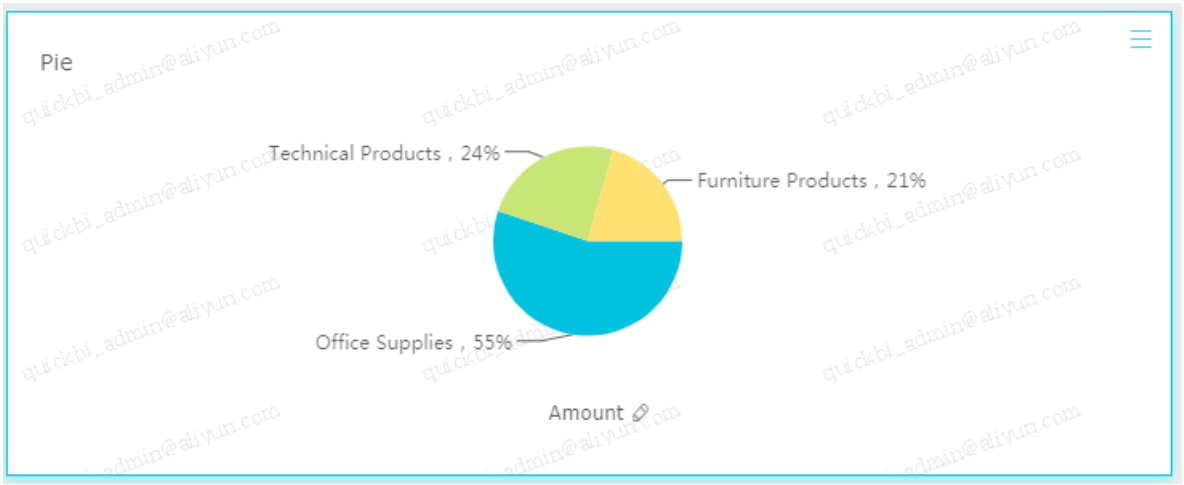
   Figure 26-102: Select pie chart fields

6. Click **Update** to update the chart.
7. On the **Style** tab, select **3D** and **Name, value (percentage)**, as shown in *Figure 26-103: Pie chart after editing*.

Figure 26-103: Pie chart after editing





8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

26.4.4.4 Geo bubble

A bubble map uses a map profile as its background and attaches bubbles to the map to indicate data values. It intuitively shows the related data indexes and data ranges of a country or region.

For example, a bubble map can be used to show the passenger flows of multiple tourist sites or the per capita incomes of multiple regions.

Context

A bubble map consists of a geographical region and a bubble size. The geographical region is determined by the data dimension, such as a province. The bubble size is determined by the data measurement, such as a transportation cost or an order quantity.

You can set only one dimension for the geographical region of a bubble map, and its type must be of geographical information, such as a region, a province, or a city. Set at least one bubble size and a maximum of five measurements.

The following uses the company_sales_record dataset as an example to describe how to use a bubble map to compare the order quantity and average profits of multiple provinces.

Data modeling may be required in this scenario.

For instructions on how to perform data modeling, see *Create a calculated field*.

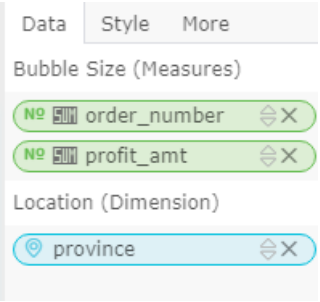
Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **Geo bubble** icon.
5. On the **Data** tab, select required dimension fields and measurement fields.

   In the dimension list, locate the province option and add it to the geographical area. In the measurement list, locate the order quantity and average profits options and add them to the bubble size area in sequence, as shown in *Figure 26-104: Select bubble map fields*.

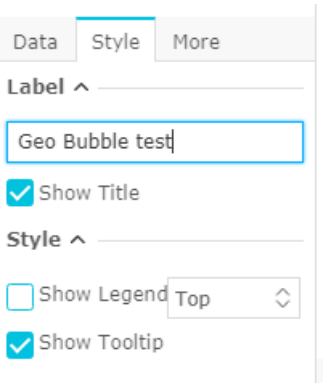
   Note:
   Make sure that the dimension type of the province field has been changed from string to geographical information. For instructions on how to change the dimension field type, see *Edit a dimension field*.

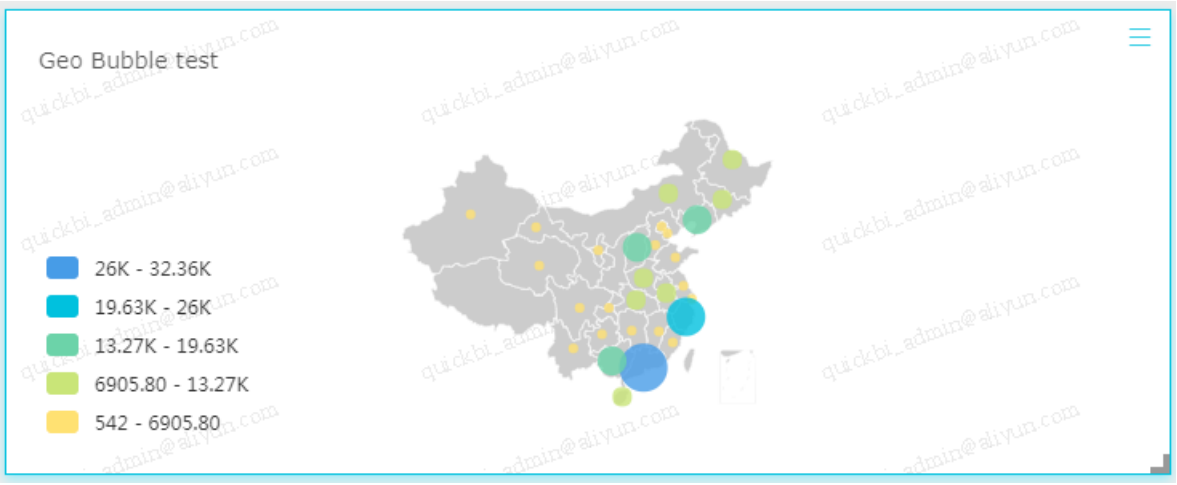
Figure 26-104: Select bubble map fields



6. Click **Update** to update the chart.
7. On the **Style** tab, edit the title and legends of the chart, as shown in *Figure 26-105: Bubble map after editing*.

Figure 26-105: Bubble map after editing





- (1) You can click the name of a legend to display the expected map information.
- (2) You can click a color icon to hide unnecessary data.

- (3) You can click a direction arrow to adjust the map position, and click + (plus sign) or - (minus sign) to zoom in or zoom out the map.

8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

26.4.4.5 Geo map

Like a *Geo bubble*, a geo map uses different colors to demonstrate the data values and ranges.

Context

A geo map consists of a geographical region and a color saturation. The geographical region is determined by the data dimension, such as a province. The color saturation is determined by the data measurement, such as an order amount or a profit amount.

You can set only one dimension for the geographical region of a geo map, and its type must be of geographical information. Set at least one colorscale and a maximum of five locations.

The following uses the company_sales_record dataset as an example to describe how to use a geo map to compare the transportation costs, order amounts and profit amounts of different regions.

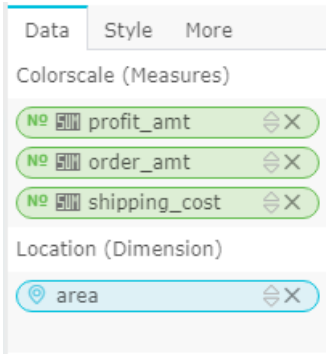
Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **geo map** icon.
5. On the **Data** tab, select required dimension fields and measurement fields.

   In the dimension list, locate the region option and add it to the location area. In the measurement list, locate the order amount, profit amount, and transport cost options and add them to the colorscale area in sequence, as shown in Figure *Figure 26-106: Select geo map fields*.

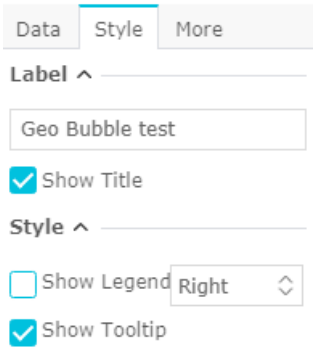
   Note:
   Make sure that the dimension type of the region field has been changed from string to geographical information. For instructions on how to change the dimension field type, see *Edit a dimension field*.

Figure 26-106: Select geo map fields



6. Click **Update** to update the chart.
7. On the **Style** tab, select **Right**, as shown in *Figure 26-107: Color map after editing*.

Figure 26-107: Color map after editing



In a geo map, you can select the legend you want to view, adjust the position and size of the map, and hide unnecessary data. For more information, see *Bubble map*.

8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

By default, the dashboard is saved to **My Dashboard**.

26.4.4.6 Table

A table is used to show the values of a table field and divide the values into two groups. One group is listed on the left of the data table, and the other is listed on the top of the data table. Multiple calculations can be performed in the intersection between a column and a row, for example, sum, average, count, maximum, or minimum.

Context

A table consists of rows and columns. The horizontal rows are determined by the data dimension , such as a province or a product type. The vertical columns are determined by the data measurement, such as an order quantity or a profit amount.

No limitations are imposed on the values of the dimension and measurement to define the rows and columns of a cross tabulation.

The following uses the company_sales_record dataset as an example to describe how to use a table to compare the packages, transportation costs, order quantities, and average profits of different types of products in multiple provinces.

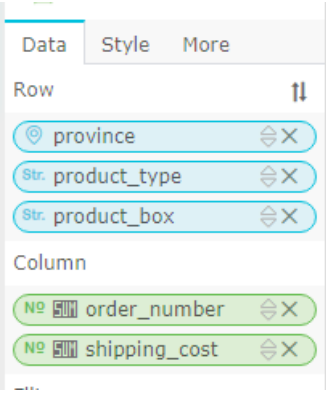
Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **Table** icon.
5. On the **Data** tab, select a required dimension field and measurement field for the cross tabulation.

   In the dimension list, locate the province, product type, and product package options and add them to the row area in sequence. In the measurement list, locate the order quantity, transportation cost, and average profit options and add them to the column area in sequence, as shown in Figure *Figure 26-108: Select cross tabulation fields*.

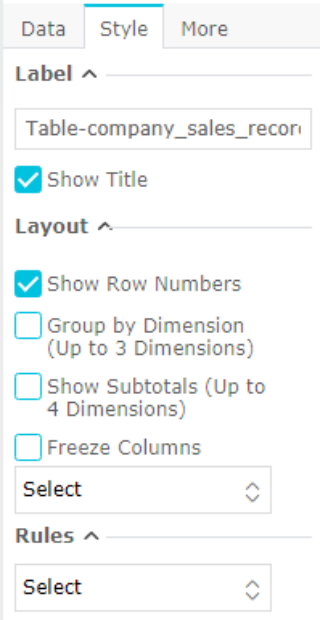
   Note:
   Make sure that the dimension type of the province field has been changed from string to geographical information. For instructions on how to change the dimension field type, see *Edit a dimension field*.

Figure 26-108: Select cross tabulation fields



6. Click **Update** to update the chart.
7. On the **Style** tab, select **Show Row Numbers**, as shown in *Figure 26-109: Cross tabulation after editing*.

Figure 26-109: Cross tabulation after editing



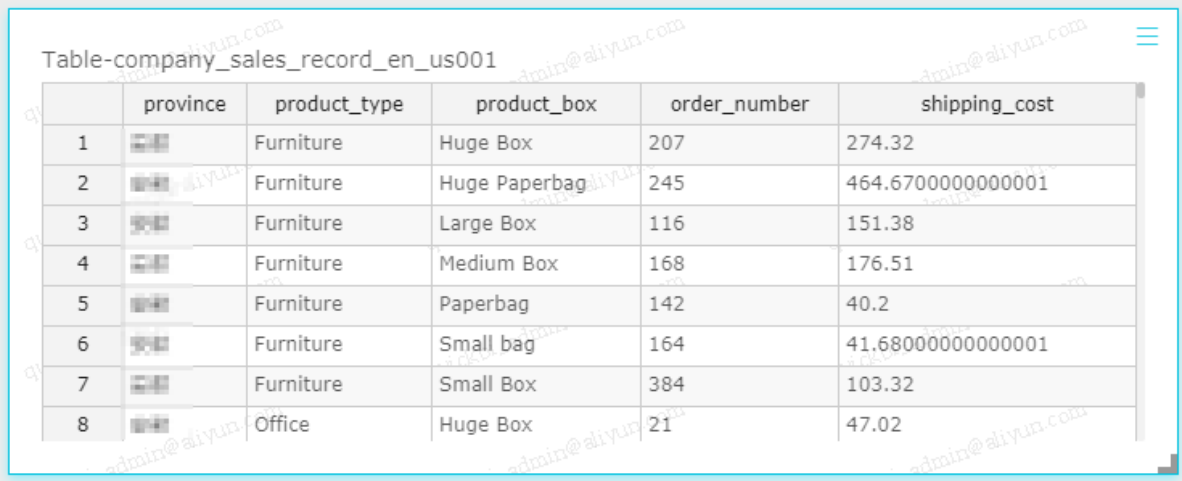


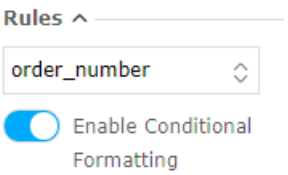
Table-company_sales_record_en_us001

	province	product_type	product_box	order_number	shipping_cost
1	[redacted]	Furniture	Huge Box	207	274.32
2	[redacted]	Furniture	Huge Paperbag	245	464.6700000000001
3	[redacted]	Furniture	Large Box	116	151.38
4	[redacted]	Furniture	Medium Box	168	176.51
5	[redacted]	Furniture	Paperbag	142	40.2
6	[redacted]	Furniture	Small bag	164	41.68000000000001
7	[redacted]	Furniture	Small Box	384	103.32
8	[redacted]	Office	Huge Box	21	47.02

In the rule menu, you can edit the data display effect for the table. For example, you can change the font color, add a data identifier, or highlight a data area, to help readers efficiently locate their expected data.

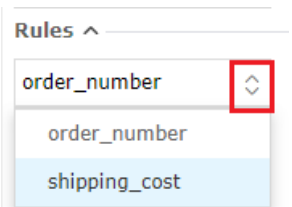
1. Click **Enable Conditional Formatting** to enable the rule function, as shown in *Figure 26-110: Enable conditional formatting*.

   Figure 26-110: Enable conditional formatting



2. Click the **switchover** icon to set rules for the chart data, as shown in *Figure 26-111: Set rules*.

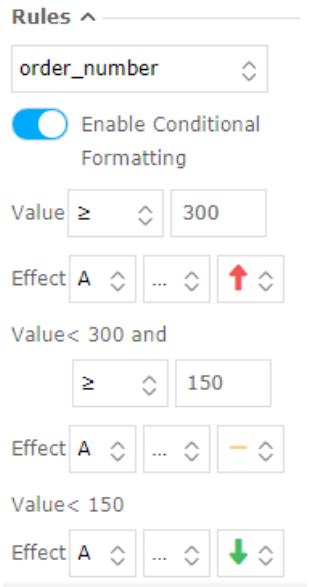
   Figure 26-111: Set rules



3. Click the drop-down arrow, and edit the values and set the display effect of the values.

   Using order number as an example, set specific rules for data values greater than 300, those ranging from 150 to 300, and those less than 150.

Figure 26-112: Style editing example



- When a data value is greater than 300, the font color of the value is red, its surrounding area is highlighted, and a red upward arrow is next to it as an identifier.
- When a data value ranges from 150 to 300, the font color of the value is orange, its surrounding area is not highlighted, and an orange dash is next to it as an identifier.
- When a data value is less than 150, the font color of the value is the default color, its surrounding area is not highlighted, and a green downward arrow is next to it as an identifier.

Figure 26-113: Cross tabulation after editing

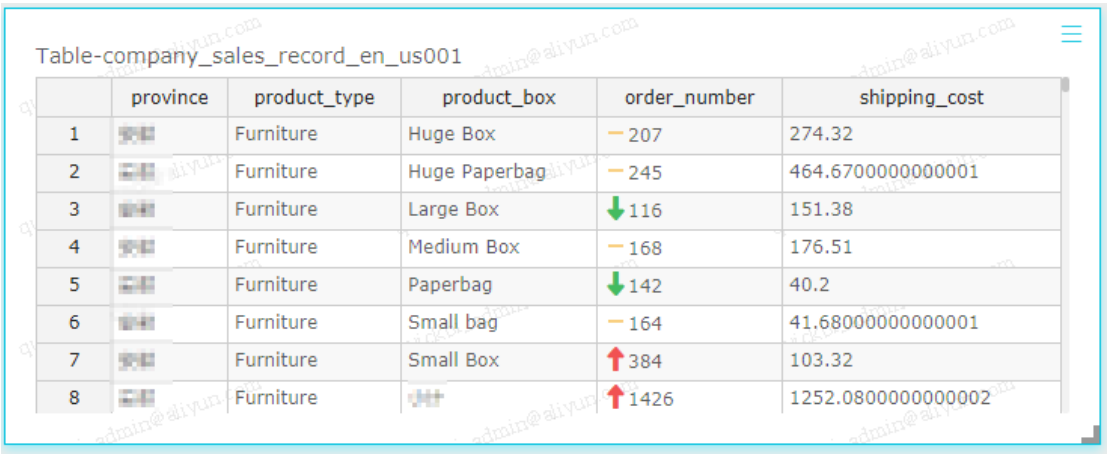


Table-company_sales_record_en_us001

	province	product_type	product_box	order_number	shipping_cost
1	[redacted]	Furniture	Huge Box	207	274.32
2	[redacted]	Furniture	Huge Paperbag	245	464.6700000000001
3	[redacted]	Furniture	Large Box	116	151.38
4	[redacted]	Furniture	Medium Box	168	176.51
5	[redacted]	Furniture	Paperbag	142	40.2
6	[redacted]	Furniture	Small bag	164	41.68000000000001
7	[redacted]	Furniture	Small Box	384	103.32
8	[redacted]	Furniture	[redacted]	1426	1252.0800000000002

8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

By default, the dashboard is saved to **My Dashboard**.

26.4.4.7 Gauge

Like a dashboard on a car, the gauge on Quick BI distinctly shows the range of a specific indicator. You can intuitively see the progress of your task and check whether data is under control or exceeds expectation. For example, you can use a gauge to show whether the stock of a kind of commodities is sufficient or requires replenishment.

Context

A gauge consists of pointer angles and tooltips. Pointer angles and tooltips are determined by the data measurements, such as the discount or profit amount.

Both the pointer angle and the tooltip can have only one measurement.

The following uses the company_sales_record dataset as an example to describe how to use a gauge to illustrate the order amount.

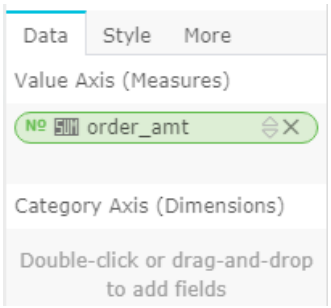
Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **Gauge** icon.
5. On the **Data** tab, select a required measurement field.

   Note:
   The system automatically adjusts the number of fields of indicator angles and tooltips according to the selection order of fields.

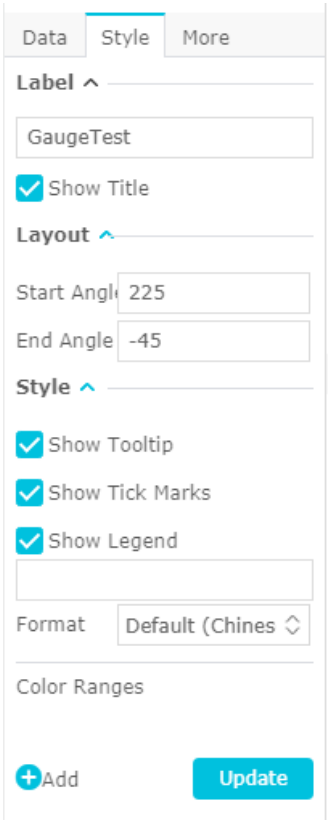
   In the measurement list, locate the order amount option and add it to the indicator angle area or tooltip area, as shown in *Figure 26-114: Select gauge fields*.

Figure 26-114: Select gauge fields



6. Click **Update** to update the chart.
7. On the **Style** tab, edit the title, layout, and legends of the gauge and set whether to display color ranges, as shown in *Figure 26-115: Edit the gauge style*.

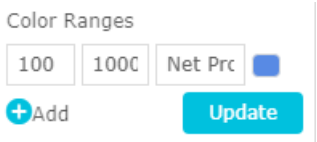
Figure 26-115: Edit the gauge style



8. In the display range setting area, click **Add** and set the start and end value.

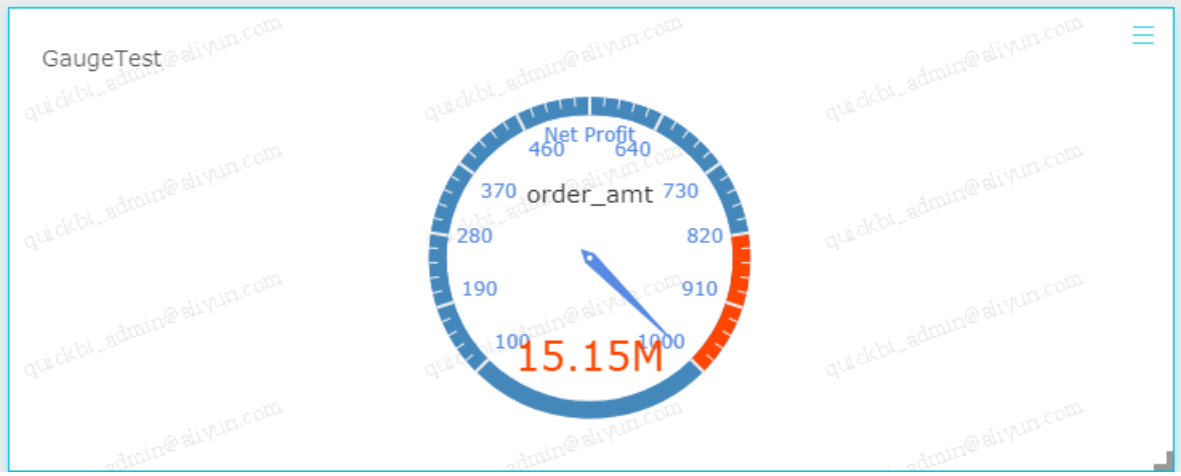
   For example, you can set the start value to 100, end value to 1,000, and range title to Net Profit, as shown in Figure *Figure 26-116: Set the display range*.

Figure 26-116: Set the display range



9. Click a color block to change the color scheme of the display range.
10. Click **Update** to update the chart, as shown in *Figure 26-117: Gauge after editing*.

Figure 26-117: Gauge after editing



11. Click **Save** and enter a name for the dashboard.
12. Click **OK** to save the dashboard.

By default, the dashboard is saved to **My Dashboard**.

26.4.4.8 Radar chart

A radar chart can be used to show analyzed numbers or ratios. You can intuitively know the changes and trend of each indicator. For example, you can use a radar chart to show the sales of each region.

Context

A radar chart consists of radius labels and the radius. radius labels are determined by the data dimension, such as the product type. The radius is determined by the data measurement, such as the transportation costs.

Radius labels of a radar chart have one to two dimensions. The dimension value must be greater than 3 and less than or equal to 12. The radius must have at least one measurement.

The following uses the company_sales_record dataset as an example to describe how to use a radar chart to compare the order quantity and order amount of different regions.

Procedure

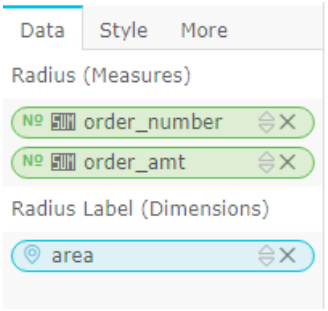
1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **radar** icon.
5. On the **Data** tab, select required dimension fields and measurement fields.

   In the dimension list, locate the region option and add it to the radius label area. In the measurement list, locate the order number and order amount options, and add them to the radius area in sequence, as shown in *Figure 26-118: Select radar chart fields*.

   > **Note:**
   > Make sure that the dimension type of the region field has been changed from string to geographical information.

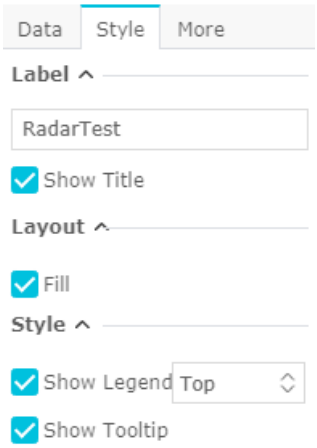
   For instructions on how to change the dimension field type, see *Edit a dimension field*.

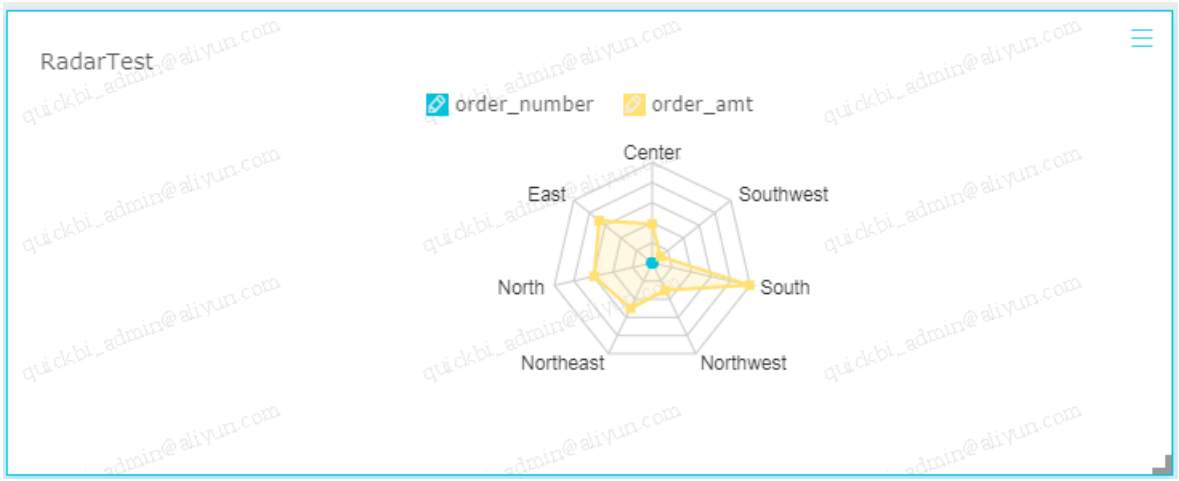
   Figure 26-118: Select radar chart fields

6. Click **Update** to update the chart.
7. On the **Style** tab, edit the title, layout, and legends of the radar chart, as shown in *Figure 26-119: Radar chart after editing*.

Figure 26-119: Radar chart after editing





8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

26.4.4.9 Scatter chart

A scatter chart shows distribution and convergence of data.

Context

A scatter chart consists of the X axis and Y axis. Color legends of a scatter chart are determined by the data dimension, such as the product type. The X axis and Y axis are determined by the data measurements.

Color legends of a scatter chart can have only one dimension, with the maximum number of dimension members as 1,000.

X axis: Has one to three measurements.

Y axis: Has only one measurement.

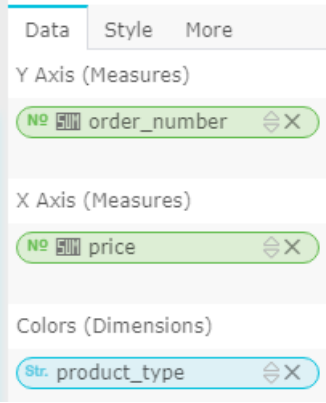
The following uses the company_sales_record dataset as an example to describe how to use a scatter chart to compare the unit price and order quantity of different types of products.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **scatter** icon.
5. On the **Data** tab, select a required dimension field and measurement field for the scatter chart.

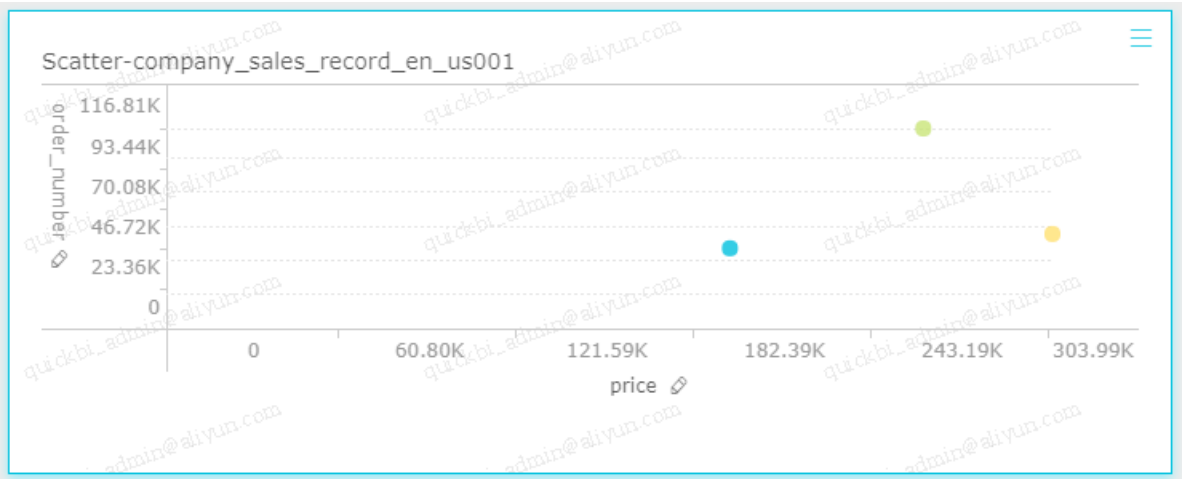
   In the dimension list, locate the product type option, and add it to the color legend area. In the measurement list, locate the unit price and order quantity options, and add them to the Y axis and X axis in sequence, as shown in *Figure 26-120: Select scatter chart fields*.

   Figure 26-120: Select scatter chart fields

6. Click **Update** to update the chart.
7. On the **Style** tab, edit the title, layout, and legends of the scatter chart, as shown in *Figure 26-121: Scatter chart after editing*.

Figure 26-121: Scatter chart after editing



8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

26.4.4.10 Funnel chart

The funnel chart can be used to analyze a business procedure that is relatively standard and has a long cycle with many phases. By comparing business data at each phase using a funnel chart, you can intuitively locate the problem and find the cause. The funnel chart can also be used to show the conversion rate between each phase. It can apply to analysis of complex business procedures. For example, a funnel chart can intuitively show the rate of visitors that finally buy any product after accessing the website.

Context

A funnel chart consists of tier labels and tier area. A tier label is determined by the data dimension , such as the region. A tier area is determined by the data measurement, such as the order amount.

Both the tier label and the funnel layer width can have only one measurement.

The following uses the company_sales_record dataset as an example to describe how to use a funnel chart to compare the order amount of different regions.

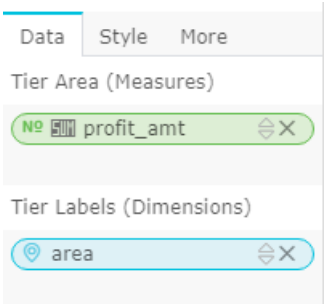
Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.

3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **funnel** icon.
5. On the **Data** tab, select required dimension fields and measurement fields.

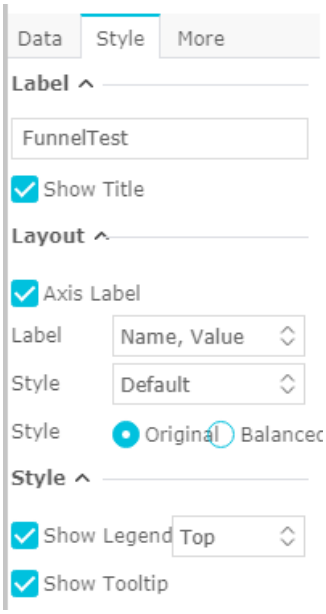
   In the dimension list, locate the region option and add it to the tier label area. In the measurement list, locate the order amount option and add it to the tier area, as shown in *Figure 26-122: Select funnel chart fields*.

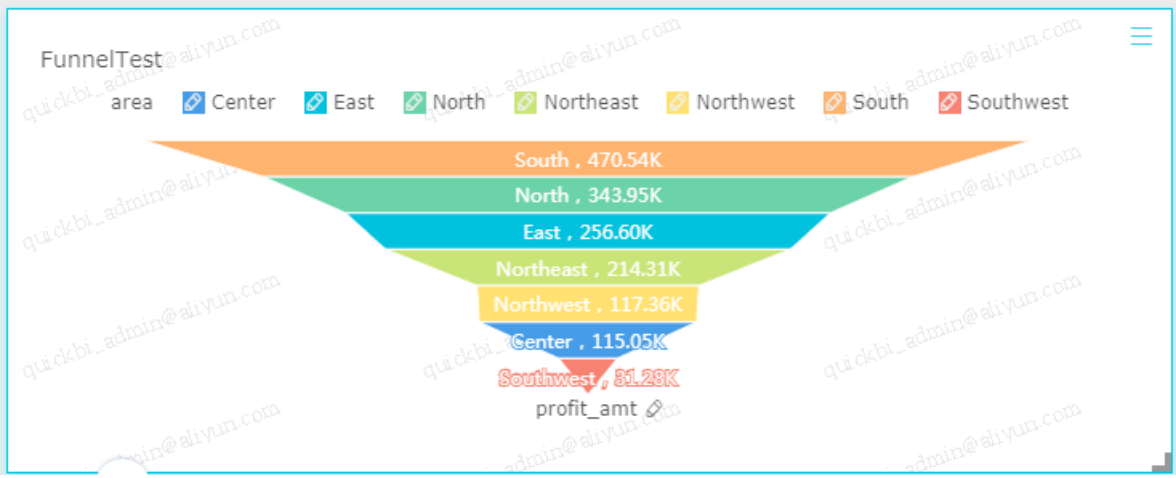
   Figure 26-122: Select funnel chart fields



6. Click **Update** to update the chart.
7. On the **Style** tab, edit the title, layout, and legends of the funnel chart, as shown in *Figure 26-123: Funnel chart after editing*.

   Figure 26-123: Funnel chart after editing





8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

26.4.4.11 Card

A card can intuitively show the data or sales performance. This helps the participants of a project to promptly know the sales information or management status and quickly take countermeasures. Therefore, they are one of the most efficient and intuitive methods of discovering and solving problems.

Context

A card consists of card labels and card metrics. Card labels are determined by the data dimension , such as the region. Card metrics are determined by the data measurement, such as the order quantity or order amount.

Card labels have up to one dimension. Card metrics have at least one, and up to 10 measuremen ts.

The following uses the company_sales_record dataset as an example to describe how to use a card to compare order quantity, order amount, transportation costs, and profit amount of different provinces.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **Card** icon.

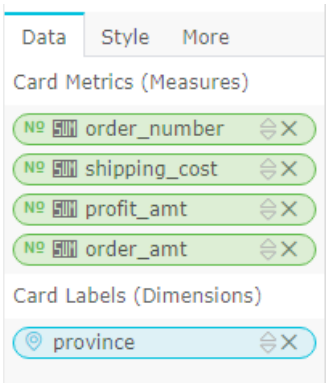
5. On the **Data** tab, select required dimension fields and measurement fields.

   In the dimension list, locate the province option and add it to the card label area. In the measurement list, locate the order number, order amt, shipping cost, and profit am options. Then add them to the card metrics area in sequence, as shown in *Figure 26-124: Select card fields*.

   > **Note:**
   > Make sure that the dimension type of the province field has been changed from string to geographical information.

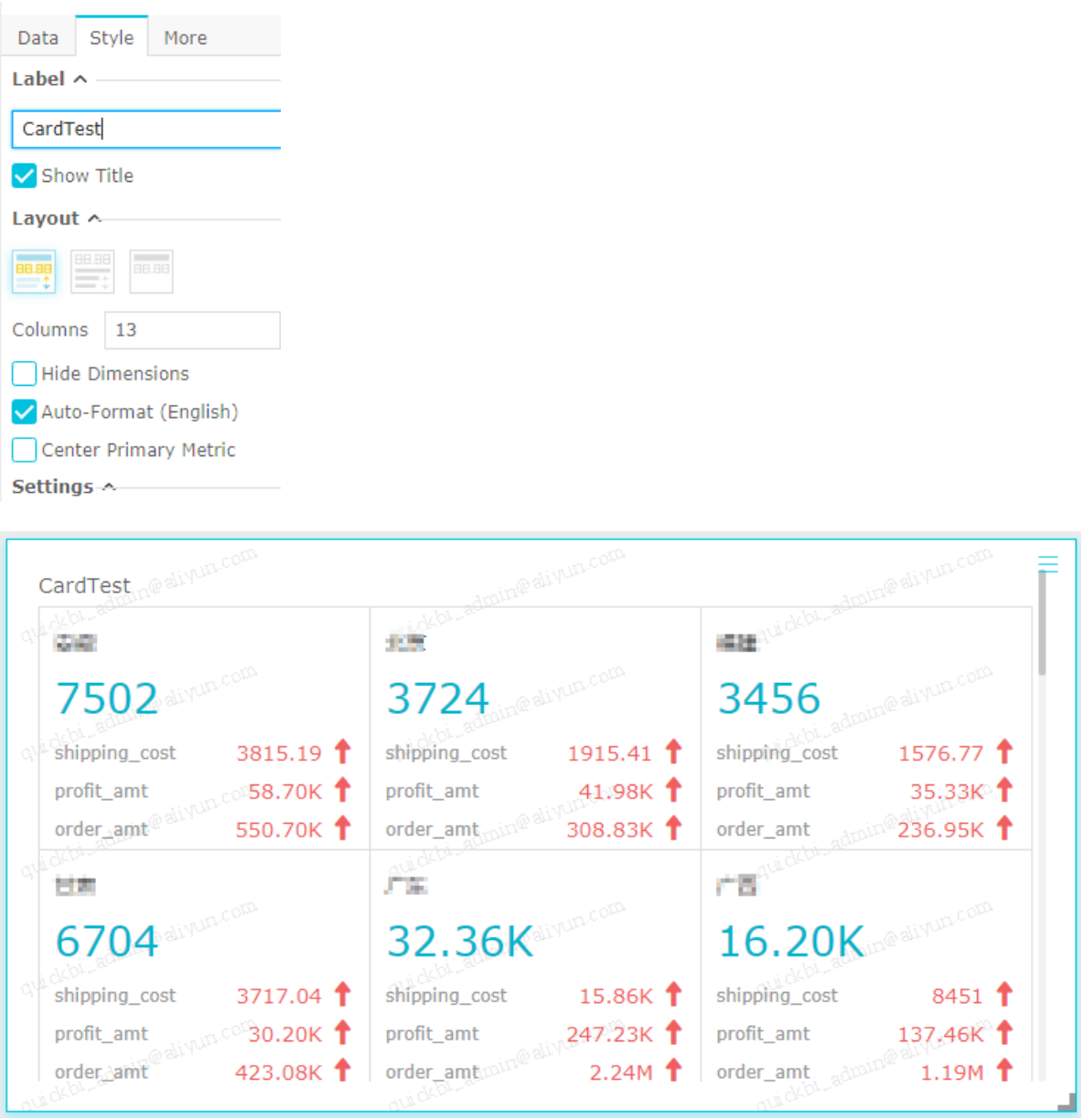
   For instructions on how to change the dimension field type, see *Edit a dimension field*.

   Figure 26-124: Select card fields

6. Click **Update** to update the chart.
7. On the **Style** tab, set the number of labels displayed on each line to three, as shown in *Figure 26-125: Card after editing*.

Figure 26-125: Card after editing



8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

26.4.4.12 TreeMap

A treemap describes the relative proportions of multiple data indicators of a specific object.

Context

A treemap consists of rectangle labels and the rectangle size. Rectangle labels are determined by the data dimension, such as the product packaging box. The rectangle size is determined by the data measurement, such as the transportation costs.

Rectangle labels of a treemap can have only one dimension. The dimension value must be less than or equal to 12. The rectangle size can have only one measurement.

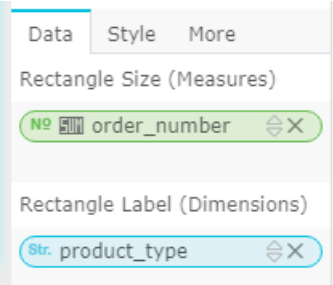
The following uses the company_sales_record dataset as an example to describe how to use a treemap to compare the order quantity of different types of products.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **TreeMap** icon.
5. On the **Data** tab, select required dimension fields and measurement fields.

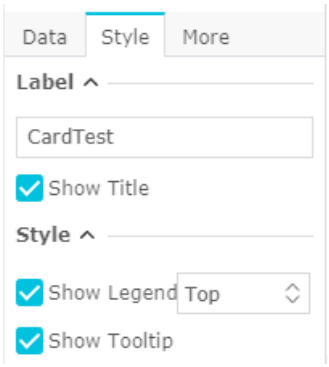
   In the dimension list, locate the product type option and add it to the rectangle label area. In the measurement list, locate the order number option and add it to the rectangle size area, as shown in *Figure 26-126: Select treemap fields*.

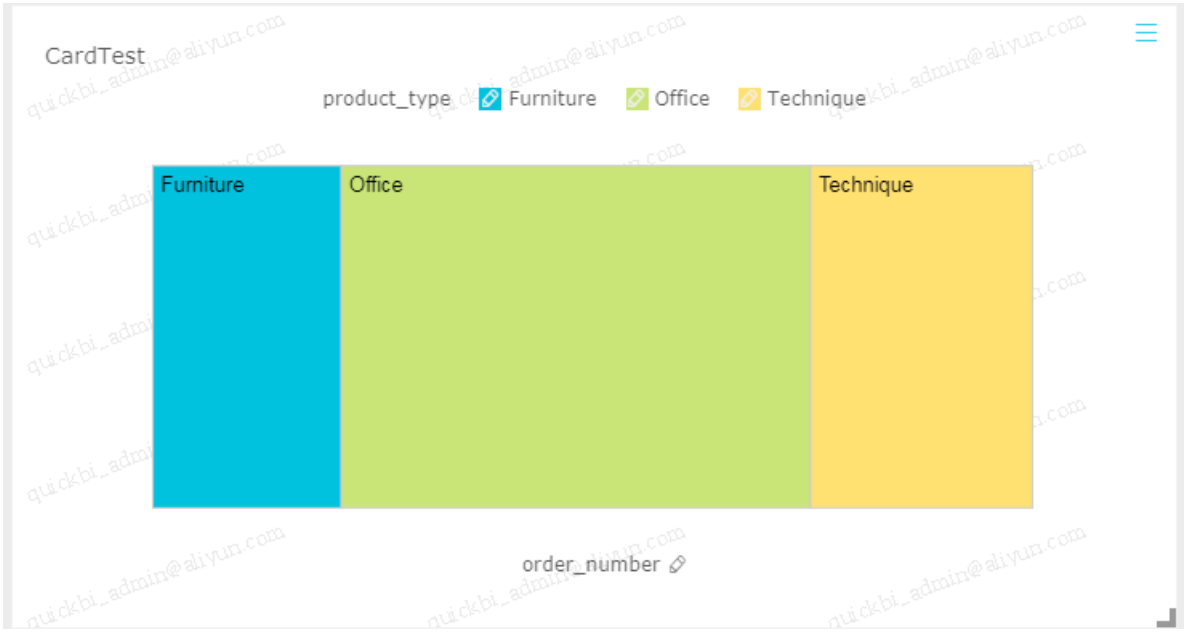
   Figure 26-126: Select treemap fields

6. Click **Update** to update the chart.
7. On the **Style** tab, edit the title and legends of the treemap, as shown in *Figure 26-127: Treemap after editing*.

Figure 26-127: Treemap after editing





8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

26.4.4.13 Polar chart

A polar chart can be used to show data changes in a period or comparison between multiple items. Polar charts apply to enumerated values, for example, comparison between data of different regions.

Context

Like a *pie chart*, a polar chart consists of multiple slices. The slice label is determined by the data dimension, such as the region or product type. The arc radius is determined by the data measurement, such as the order quantity or order amount.

Slice labels of a polar chart can have only one dimension. The number of dimension members must be greater than 3 and less than or equal to 12. The arc radius can have only one measurement.

The following uses the company_sales_record dataset as an example to describe how to use a polar chart to compare the order quantity of different regions. (The number of regions must be greater than 3 and less than or equal to 12.)

Procedure

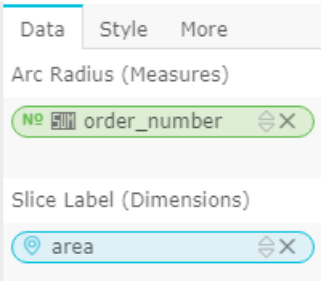
1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **polar chart** icon.
5. On the **Data** tab, select a required dimension field and measurement field for the polar chart.

   In the dimension list, locate the area option and add it to the slice label area. In the measurement list, locate the order number option, and add it to the arc radius area, as shown in *Figure 26-128: Select polar chart fields*.

   > **Note:**
   > Make sure that the dimension type of the region field has been changed from string to geographical information.

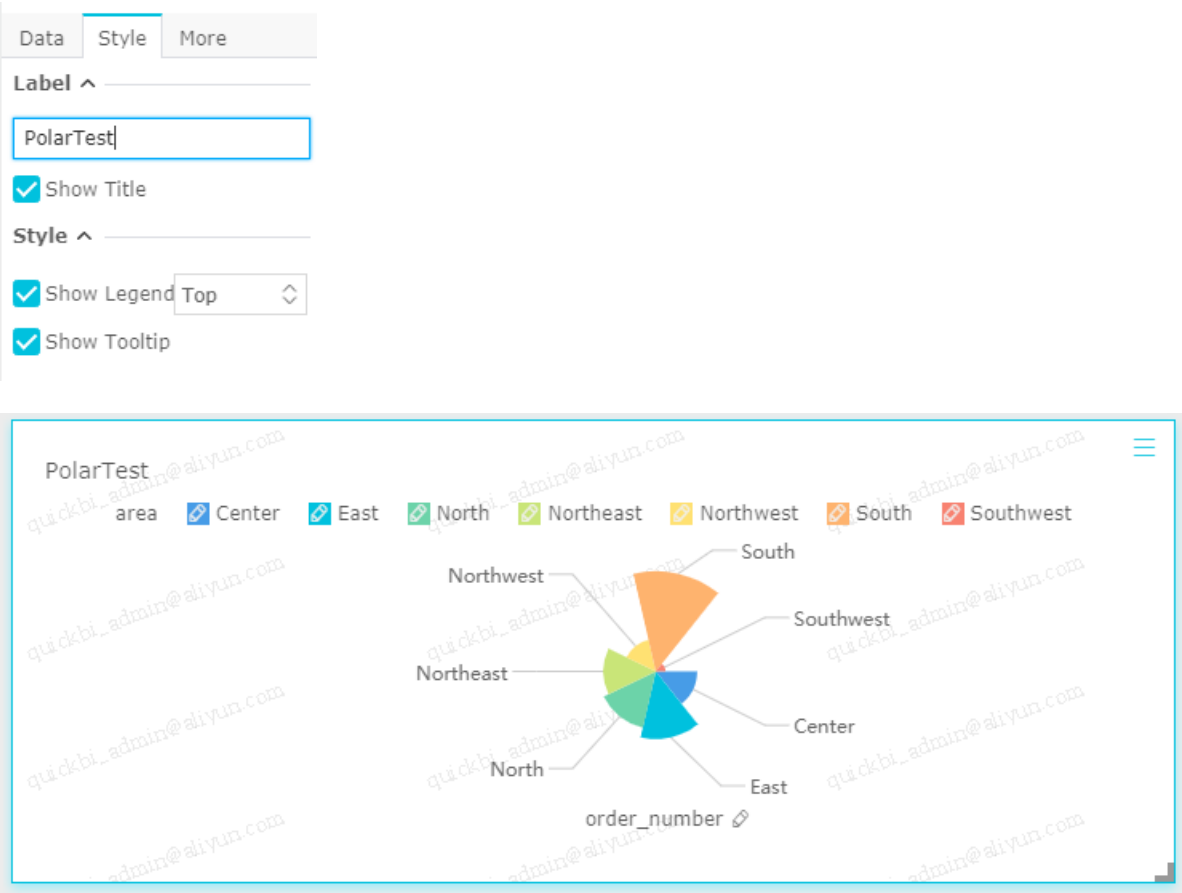
   For instructions on how to change the dimension field type, see *Edit a dimension field*.

   Figure 26-128: Select polar chart fields

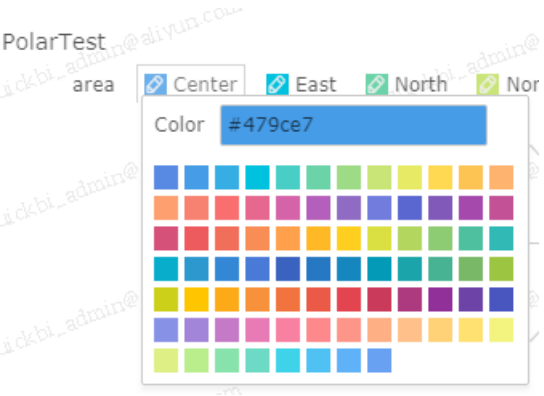
6. Click **Update** to update the chart.
7. On the **Style** tab, edit the title and legends of the chart, as shown in *Figure 26-129: Polar chart after editing*.

Figure 26-129: Polar chart after editing



8. Click the color block before a legend to change the color scheme of the legend.

Figure 26-130: Change the color scheme



9. Click **Save** and enter a name for the dashboard.
10. Click **OK** to save the dashboard.

    By default, the dashboard is saved to **My Dashboard**.

26.4.4.14 Word cloud

A word cloud intuitively shows the word frequency. It can be used to prepare user persona and user labels.

Context

A word cloud consists of word and the word size. Each word is determined by the data dimension , such as the customer name or product type. Each word size is determined by the data measurement, such as the profit or unit price.

Word of a word cloud can have only one dimension. The word size can have only one measurement.

The following uses the company_sales_record dataset as an example to describe how to use a word cloud to compare the order quantity of different provinces.

Procedure

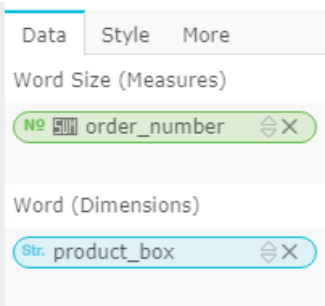
1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **word cloud** icon.
5. On the **Data** tab, select required dimension fields and measurement fields.

   In the dimension list, locate the product box option and add it to the word area. In the measurement list, locate the order number option and add it to the word size area, as shown in the *Figure 26-131: Select word cloud fields*.

   Note:
   Make sure that the dimension type of the province field has been changed from string to geographical information.

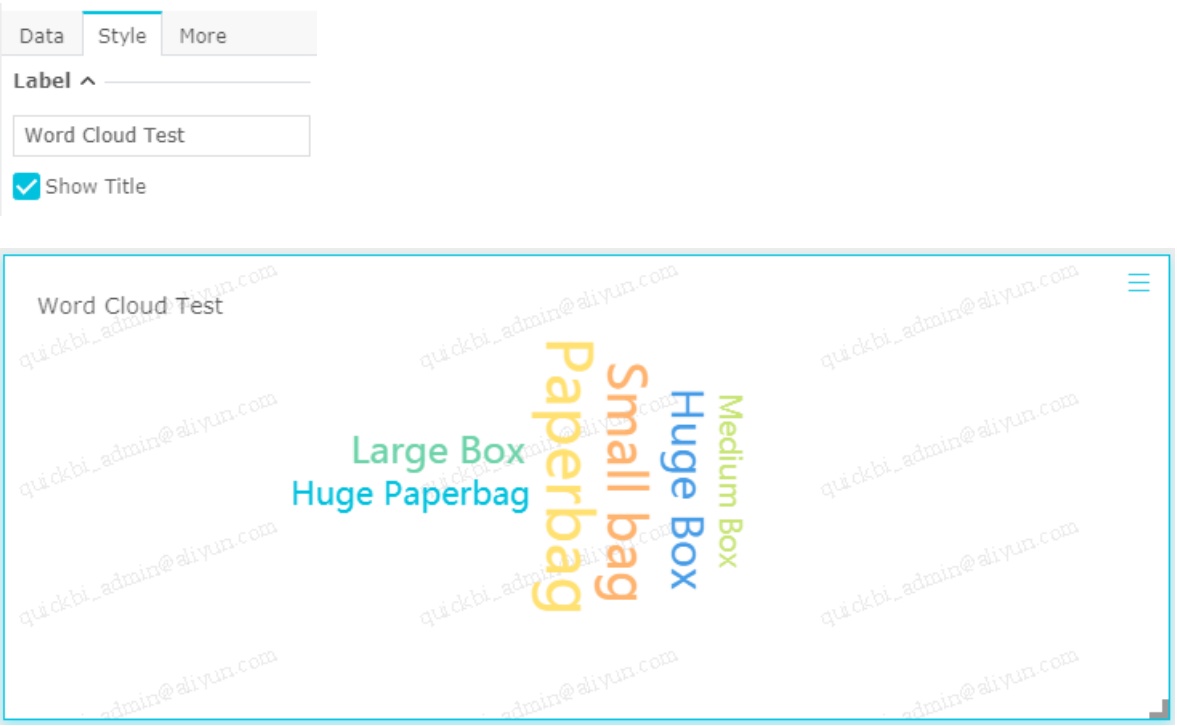
   For instructions on how to change the dimension field type, see *Edit a dimension field*.

Figure 26-131: Select word cloud fields



6. Click **Update** to update the chart.
7. On the **Style** tab, edit the title of the chart, as shown in *Figure 26-132: Word cloud after editing*.

Figure 26-132: Word cloud after editing



8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

26.4.4.15 Tornado chart

A tornado chart is a combination of a tornado chart and a funnel chart. A tornado chart can be used to compare the data of different indicators of two objects. For example, the salary and education degree of citizens in two cities. A funnel chart can be used to show the conversion rate

between each phase. It can also apply to analysis of complex business procedures. For example, a funnel chart can intuitively show the rate of visitors that finally buy any product after accessing the website.

Context

A tornado chart can be used to compare the migrant population ratio, employment rate, and transaction volume of commercial residential buildings in Beijing and Shanghai. If there is a funneling between compared items, you can use a tornado chart to show the data difference of each indicator between two cities and display the hierarchy of the compared items.

If there is no funneling effect between compared items, a common tornado chart is displayed. If there is a funneling effect between compared items, but these items belong to only one object, a common funnel chart is displayed.

A tornado chart consists of comparisons. Each comparison is determined by the data measurement, such as the order quantity or order amount.

Each comparison must have at least one measurement.

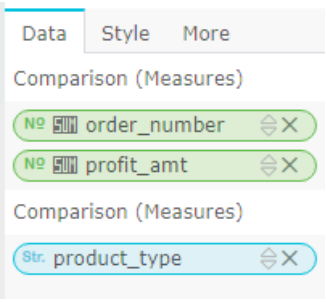
The following uses the company_sales_record dataset as an example to describe how to use a tornado chart to compare the order quantity, profit amount, and average profit of different types of products.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the file list dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **tornado** icon.
5. On the **Data** tab, select required measurement fields.

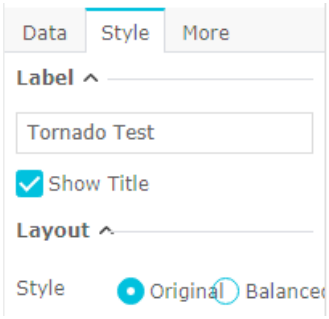
   Locate the order number, profit amt and add them to the comparison area in sequence, as shown in *Figure 26-133: Select tornado funnel chart fields*.

Figure 26-133: Select tornado funnel chart fields



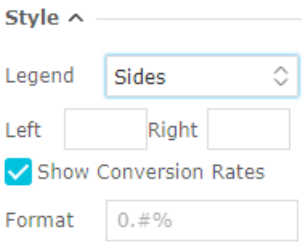
6. Click **Update** to update the chart.
7. On the **Style** tab, edit the chart title, layout, and legend positions. Then set the color scheme and whether to display the conversion rate.
   1. A tornado chart provides two chart layouts for your choice.

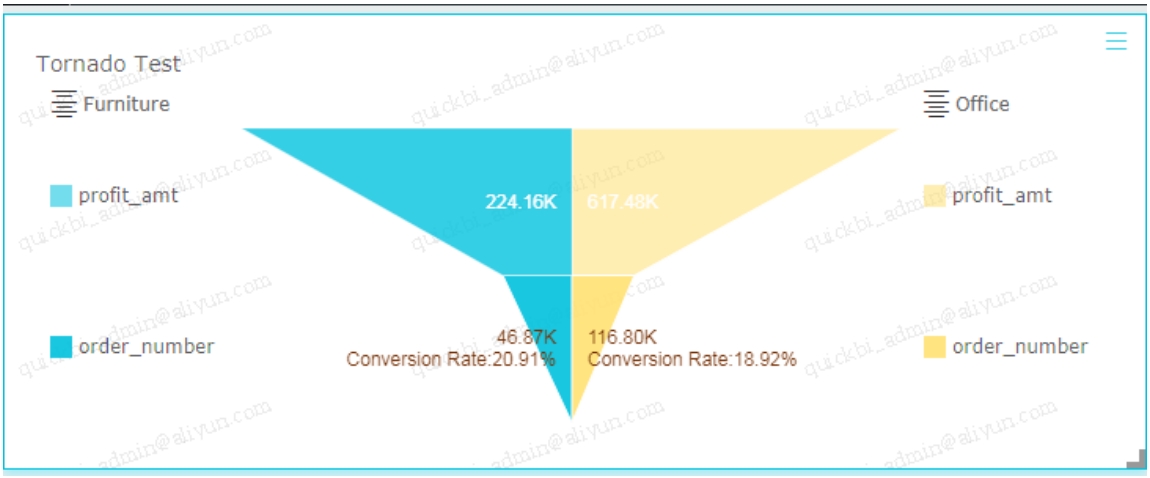
      Figure 26-134: Chart layouts

   2. You can change legend positions, set the color scheme of the chart, and set whether to display the conversion rate on the chart.

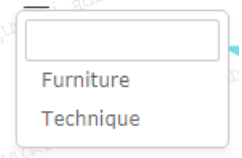
      Figure 26-135: Tornado funnel chart after editing



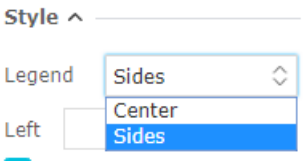
- (1) Place the cursor over the product type field, and select the expected product type, as shown in *Figure 26-136: Select a product type*.

  Figure 26-136: Select a product type



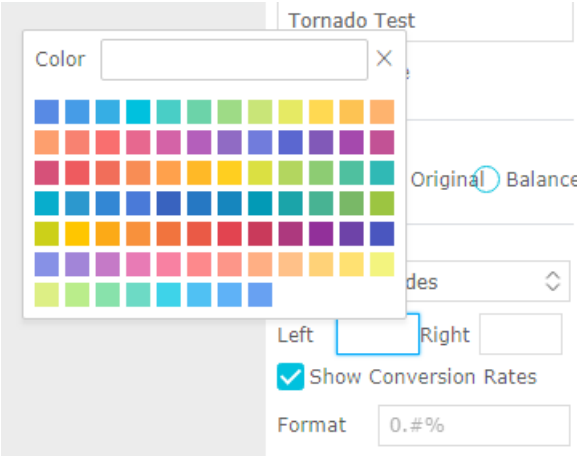
- (2) Adjust the legend position, as shown in *Figure 26-137: Legend position*.

  Figure 26-137: Legend position



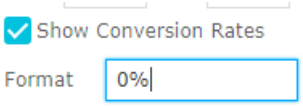
- (3) Click the left box and right box and select the expected color from the color menu, as shown in *Figure 26-138: Change the color scheme*.

Figure 26-138: Change the color scheme



- (4) Adjust the format of the conversion rate manually, as shown in *Figure 26-139: Conversion rate format*.

Figure 26-139: Conversion rate format



8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

26.4.4.16 Hierarchy chart

A hierarchy chart displays the organizational relationship of hierarchical data by using a tree structure and organizes objects using a parent-child structure. It can be used for enumeration. For example, if you want to view the income of each prefecture-level city in a province, you can use a hierarchy chart to show the parent-child relationship of the province and its prefecture-level cities. The hierarchy chart applies to analysis related to organizational structures, for example, staff structure of a company or department structure of a hospital.

Context

A hierarchy chart consists of node lable and node metrics. The node label is determined by the data dimension, such as the region or product type. The node metrics is determined by the data measurement, such as the order quantity or order amount.

This section provides the following two scenarios, including usage of the filter:

- Scenario 1: Compare the order quantity of different products in provinces in different regions.
- Scenario 2: View the average profit of different products in different municipalities.

Node labels must have at least two dimensions. The dimension fields have parent-child relationsh ip. Node metrics must have at least one measurement.

Scenario 1: The following uses the company_sales_record dataset as an example to describe how to use a hierarchy chart to compare the order quantity of different products in provinces in different regions.

Procedure

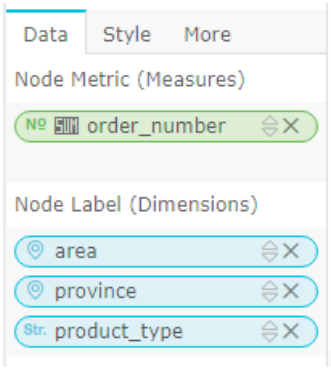
1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **Hierarchy** icon.
5. On the **Data** tab, select required dimension fields and measurement fields.

   In the dimension list, locate the region, province, and product type options and add them to the node label area in sequence. The sequence of these options is the parent-child relationship displayed on the chart. In the measurement list, locate the order quantity option and add it to the node metrics area, as shown in *Figure 26-140: Select hierarchy fields*.

   Note:
   Make sure that the dimension type of the region and province fields has been changed from string to geological information.

   For instructions on how to change the dimension field type, see *Edit a dimension field*.

   Figure 26-140: Select hierarchy fields

6. Click **Update** to update the chart.

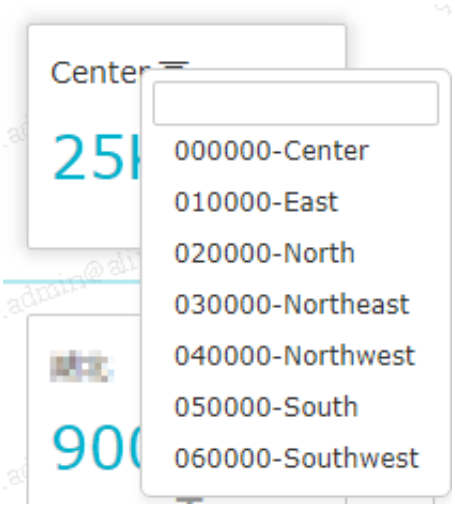
7. On the **Style** tab, you can set the title, layout, and design of the chart.

   1. A hierarchy chart supports three layouts. You can select the expansion mode (root nodes are merged by default) and display mode of parent and child nodes as needed. *Chart layouts* shows the display mode when **Straight Line** is selected for the layout.

      Figure 26-141: Chart layouts

      - (1) Place the cursor over the region field, and select the expected region name, as shown in *Figure 26-142: Select a region*.

Figure 26-142: Select a region



- (2) Click the minus sign or plus sign to collapse or expand the child node information.
- (3) If you select **Show Subtotals** in the Layout area, the chart automatically displays the aggregated value in the box.

2. You can edit the display hierarchy of the chart in the design menu and manually enter the number of hierarchy levels. You can select a main path through the corresponding field. The main path is displayed in a different color from other paths. You can load the toolbar to the chart so that you can edit the chart in preview mode or on the dashboard.

   In the following example, in the layout area, the order quantity is selected for the main path, the sorting order is set to ascending order, the toolbar is embedded into the chart, and **Curve** is selected in the Layout area.

8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

Scenario 2: View the average profit of different products in different municipalities.

Data modeling may be required in this scenario.For instructions on how to perform data modeling, see *Create a calculated field*.

10. In the dimension list, locate the province field and add it to the filter.

    Use the filter to filter out municipalities from provinces.

11. Click **Filter** and select **Enumeration**, as shown in the following figure.

    The system automatically lists all available options of the province field.

12.Select the expected municipality or manually enter the field name.

13.Click **OK** to complete filter condition settings.

14.In the dimension list, locate the city, product type, and product subtype options and add them to the node label area in sequence.

The sequence of these options is the parent-child relationship displayed on the chart. In the measurement list, locate the average profit option and add it to the node metrics area.

15.Click **Update** to update the chart.

16.On the **Style** tab, you can edit the title, layout, and design of the chart.

17.Click **Save** and enter a name for the dashboard.

18.Click **OK** to save the dashboard.

By default, the dashboard is saved to **My Dashboard**.

26.4.4.17 Conversion path

A conversion path illustrates the conversion rate of a webpage by comparing its page traffic or page views (PV) and the number of its unique visitors (UV). This allows you to know the overall operation status of the website and the final transaction volume of a type of products. Conversion path charts apply to analysis related to e-commerce or marketing. For example, you can use a source direction chart to analyze which products are the bestsellers and which time periods are traffic peaks.

Context

Currently, a conversion path only supports three levels of dimensions, which are the previous page, current page, and the next page. The measurements of a source direction chart are PV, UV , the path conversion rate, and the page bounce rate, among which PV and UV are collected from the previous, current, and next pages.

Each previous, current, and next pages in a source direction chart can have a maximum of one dimension. The dimension fields must be hierarchical. The sequence of dimensions fields determines their hierarchy displayed on the chart. PV (for the previous, current, and next page), UV (for the previous, current, and next page), the path conversion rate, and the page bounce rate can have a maximum of one measurement, respectively.

When you create a conversion path, the three dimensions for the previous, current, and next pages and two conversion rates (path conversion rate and page bounce rate) are mandatory.

However, you can select either the three PV items or three UV items. If an entered dimension or measurement field is incorrect, the system prompts you accordingly.

The following uses the page_source_target_day_stat dataset as an example to describe how to use a source direction chart to display the conversion rate and bounce rate between each page based on PV.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Dashboards** to go to the dashboard management page.
3. Choose **Create** > **Dashboard**. The dashboard editing page appears.
4. In the dashboard configuration area, double-click the **Conversion path** icon.
5. On the **Data** tab, select required dimension fields and measurement fields.

   In the dimension list, locate and enter the names of the previous, current, and next page in the corresponding fields in sequence. The sequence of these pages is the hierarchy displayed on the chart. In the measurement list, locate and enter the page path conversion rate and page bounce rate in the corresponding fields in sequence. Then locate the three PV values and/or UV values and enter them in the corresponding fields in sequence.
6. Click **Update** to update the chart.
7. On the **Style** tab, edit the title and layout of the chart.

   The conversion path provides three layouts, including highlight of the main path or prompted windows. For example, if you select **Highlight Main Path** for the layout, the main path is displayed in a different color on the chart.
8. Click **Save** and enter a name for the dashboard.
9. Click **OK** to save the dashboard.

   By default, the dashboard is saved to **My Dashboard**.

26.4.5 Query dashboards

Procedure

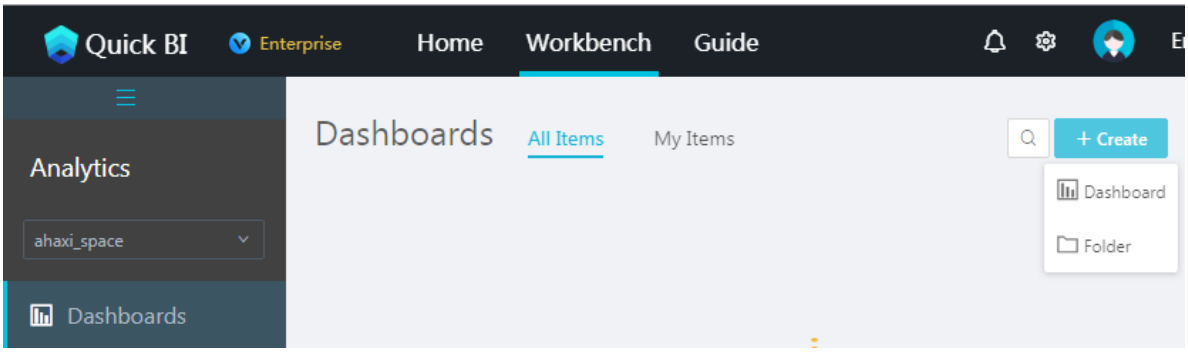
1. On the Quick BI console, choose the **Workbench** > **Dashboards**.
2. Enter a keyword in the search box.

26.4.6 Create a dashboard folder

Procedure

1. On the dashboard management page, click **Create**.
2. Select **Folder**, as shown in *Figure 26-143: Create a folder*.

Figure 26-143: Create a folder



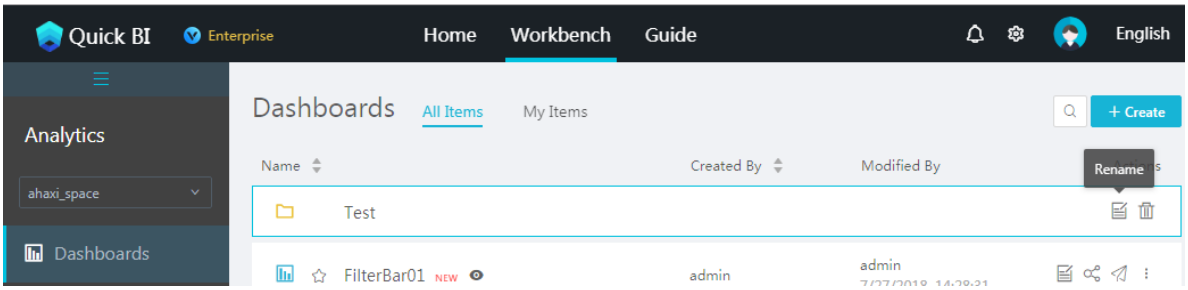
3. Enter a name for the folder and click a blank area.

26.4.7 Rename a dashboard folder

Procedure

1. In the dashboard management page, select a dashboard folder.
2. Click **Rename**, as shown in *Figure 26-144: Rename a dashboard folder*.

Figure 26-144: Rename a dashboard folder



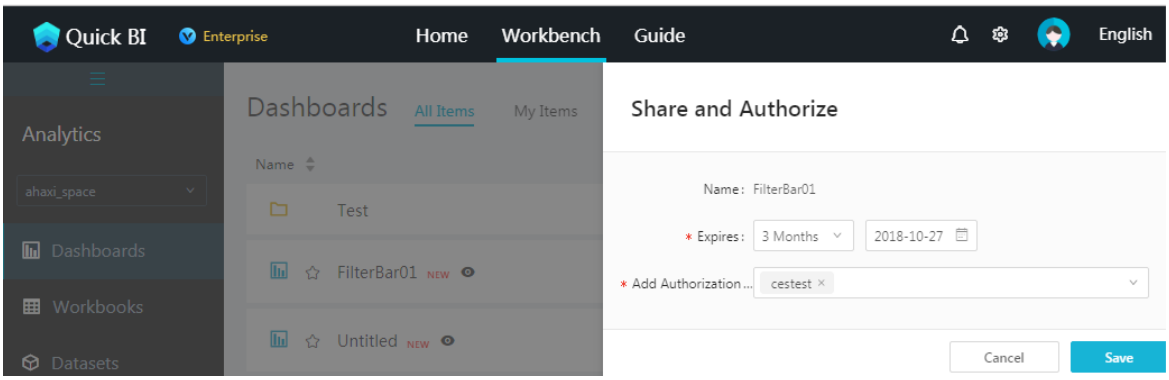
3. Change the folder name and click a blank area.

26.4.8 Share a dashboard

Procedure

1. On the Quick BI console, choose the **Workbench** > **Dashboards**
2. Select the dashboard that you want to share and click the **Share** icon.
3. Enter the account of the person you want to share the dashboard with. Then select an expiration date for the shared dashboard, as shown in *Figure 26-145: Share a dashboard*.

Figure 26-145: Share a dashboard



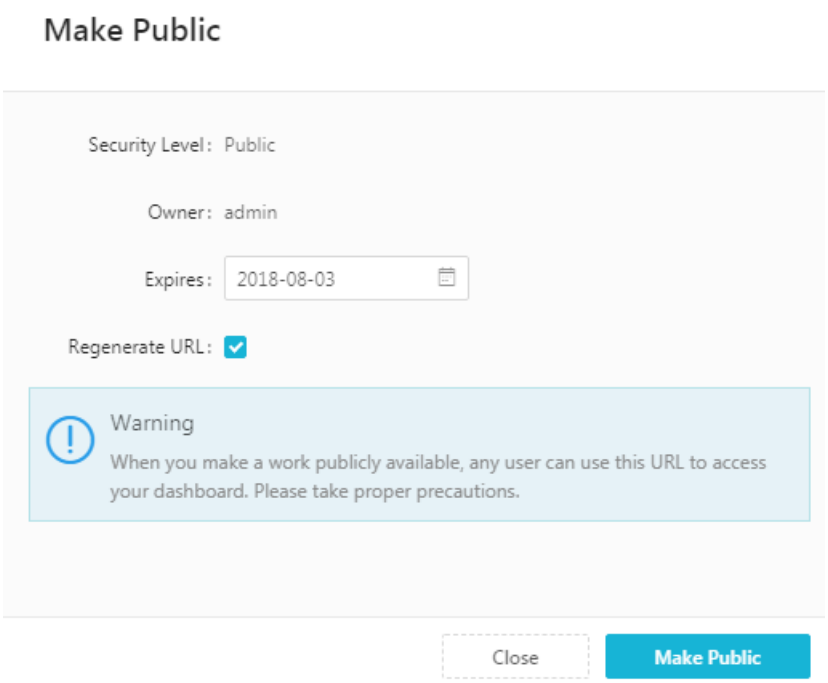
4. Click **Save**.

26.4.9 Make public a dashboard

Procedure

1. On the Quick BI console, choose the **Workbench** > **Dashboards**.
2. Select the dashboard that you want to make public and click the **Make Public**.
3. Select an expiration date for the published dashboard and allow to regenerate a URL.
4. Click **Make Public**, as shown in *Figure 26-146: Make public a dashboard*.

Figure 26-146: Make public a dashboard



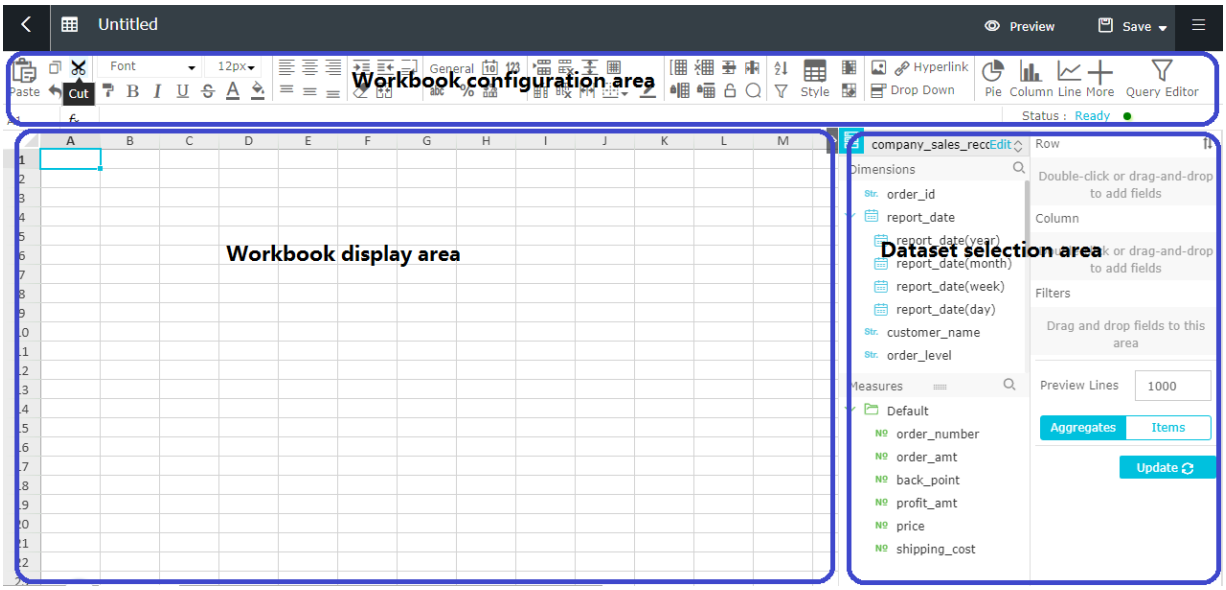
26.5 Use workbooks

This section describes how to use the workbook editing page to filter and query datasets and then use different data charts to visually display data.

The workbook editing page has three areas, as shown in *workbook editing page*.

- Dataset selection area
- Workbook configuration area
- Workbook display area

Figure 26-147: Workbook editing page



- Dataset selection area: In this area, you can select a dataset. The fields of the selected dataset are displayed in the dimension and measurement lists based on the data types preset in the system.You can select the required dimension and measurement fields from the lists according to the data elements provided by data charts.
- Workbook configuration area: In this area, you can select the expected chart type, and set the color, font, and data format of cells as needed.
- Workbook display area: In this area, you can display data in cells and reference data to complete data reprocessing.

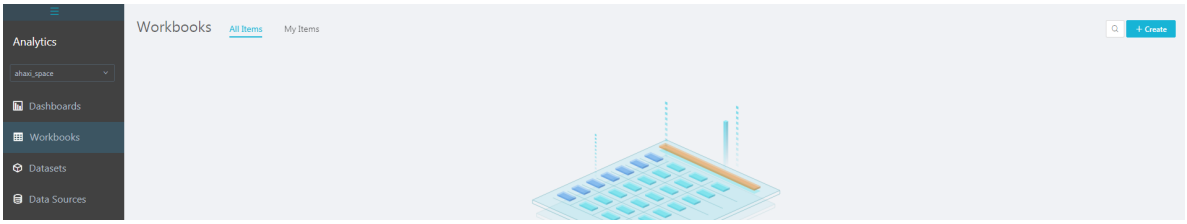
26.5.1 Create a workbook

Procedure

1. *Log on to the Quick BI console*.

2. On the left-side navigation bar, click **Workbooks**.
3. Click **Create**.
4. Click **Workbook**. The workbook editing page appears, as shown in *Figure 26-148: Create a workbook*.

Figure 26-148: Create a workbook



26.5.2 Change from the current dataset to another one

In the dataset selection area, you can select a dataset or change from the current dataset to another one.

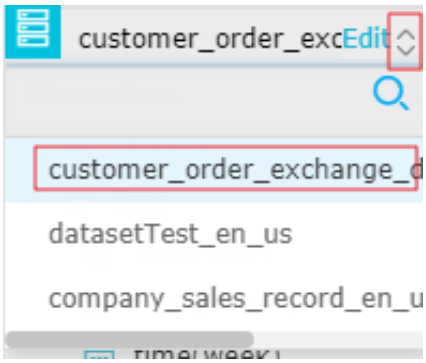
Context

If you cannot find the expected dataset, click **Dataset** on the left-side navigation bar. On the dataset list page that appears, check whether the dataset has been successfully created. For information about how to create a dataset, see *Create a dataset*.

Procedure

1. Click the dataset switchover icon.
2. In the drop-down list, select or search for the dataset to be analyzed, as shown in *Figure 26-149: Change from the current dataset to another one*.

Figure 26-149: Change from the current dataset to another one



26.5.3 Search for dimension and measurement fields

You can search for fields in the dimension and measurement lists.

Context

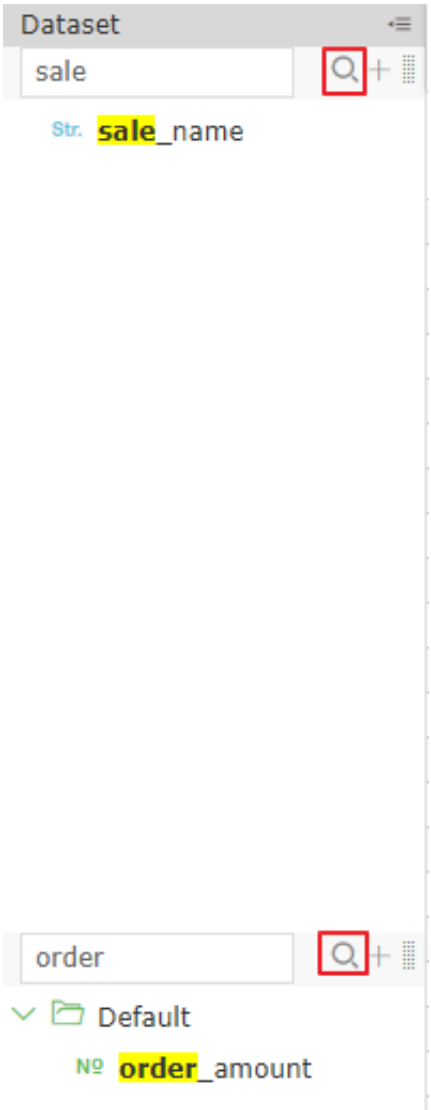
After you select a dataset, the system displays fields of different data types in the dimension and measurement lists.

For information about how to edit dimension and measurement fields, see *Edit a dimension field* and *Edit a measurement field*.

Procedure

1. Enter a keyword manually.
2. Click the magnifier icon to search for the fields that match the keyword.

Figure 26-150: Search for fields



26.5.4 Set font

The font setting function allows you to set the font of selected text, including the font style, font size, font color, and background color.

Procedure

Set the font style

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbook**.
3. Click a workbook name in the list. The workbook editing page appears.

   For information about how to create a workbook, see *Create a workbook*.
4. In the font area, click the drop-down arrow.
5. Select a proper font style from the drop-down list.

Set the font size

6. In the font size area, click the drop-down arrow.
7. Select a proper font size from the drop-down list.

   Workbooks support color inversion. If the font style is set to bold, then in the font setting box, marker "B" for the bold style is in reversed style. If the font color is set to red, then in the font setting box, the horizontal line below marker "A" for the font color is red.

26.5.5 Set the alignment mode

The alignment mode feature allows you to adjust the layout of text.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbook**.
3. Click a workbook name in the list. The workbook editing page appears.

   For information about how to create a workbook, see *Create a workbook*.
4. Click the alignment icon to adjust the text layout of the workbook, as shown in *Figure 26-151: Alignment mode*.

   Figure 26-151: Alignment mode

26.5.6 Set the text format and numeric format

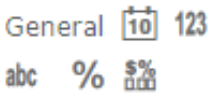
The format setting function allows you to set the display format of the text or numbers in a table.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbook**.
3. Click a workbook name in the list. The workbook editing page appears.

   For information about how to create a workbook, see *Create a workbook*.
4. Click "Format" to adjust the display format of text or numbers, as shown in *Figure 26-152: Display format*.

   Figure 26-152: Display format



- Regular: The regular cell format does not include any special numeric format.
- Date: The recommended date format is YYYY-MM-DD.
- Number: By default, numeric values are right-aligned with two decimal places shown and without the thousands separator, and the numbers exceeding the column width are replaced with ###. To show complete numbers, you can double-click the column width or adjust the column width.
- Text: By default, text is left-aligned. Text with characters exceeding the column width is displayed in the format of "abc…". Complete characters are shown when the cursor is placed over the text. To show complete characters, you can double-click the column width or adjust the column width.
- Percent: By default, percent values are right-aligned with two decimal places shown and without the thousands separator, and the numbers exceeding the column width are replaced with ###.To show complete numbers, you can double-click the column width or adjust the column width.

26.5.7 Set style, cell, and window

The style, cell, and window setting function allows you to adjust the gridline, border style, and cell height for a table.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbook**.
3. Click a workbook name in the list. The workbook editing page appears.

   For information about how to create a workbook, see *Create a workbook*.
4. Click the style, cell, and window icons to adjust the overall style of the workbook, as shown in *Figure 26-153: Style, cell, and window*.

   Figure 26-153: Style, cell, and window

   - Gridline: It is displayed by default. To hide the gridline, click **Gridline**.
   - Border: You can select upper border, lower border, left border, right border, no border, full border, outer border, and bold border. You can also set the border color.
   - Line color: You can set the color and type of any line.
   - The following operations are supported: insert row, delete row, insert column, delete column , and global Insert and Delete operations on a workbook.
   - Automatic column width: Double-click the column width icon, and the column width is adjusted automatically.
   - Automatic row height: Double-click the row height icon, and the row height is adjusted automatically.

26.5.8 Set image, link, and drop-down list

The image, link, and drop-down list features allow you to insert images, links, and drop-down lists to a table.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbook**.
3. Click a workbook name in the list. The workbook editing page appears.

   For information about how to create a workbook, see *Create a workbook*.

Insert an image

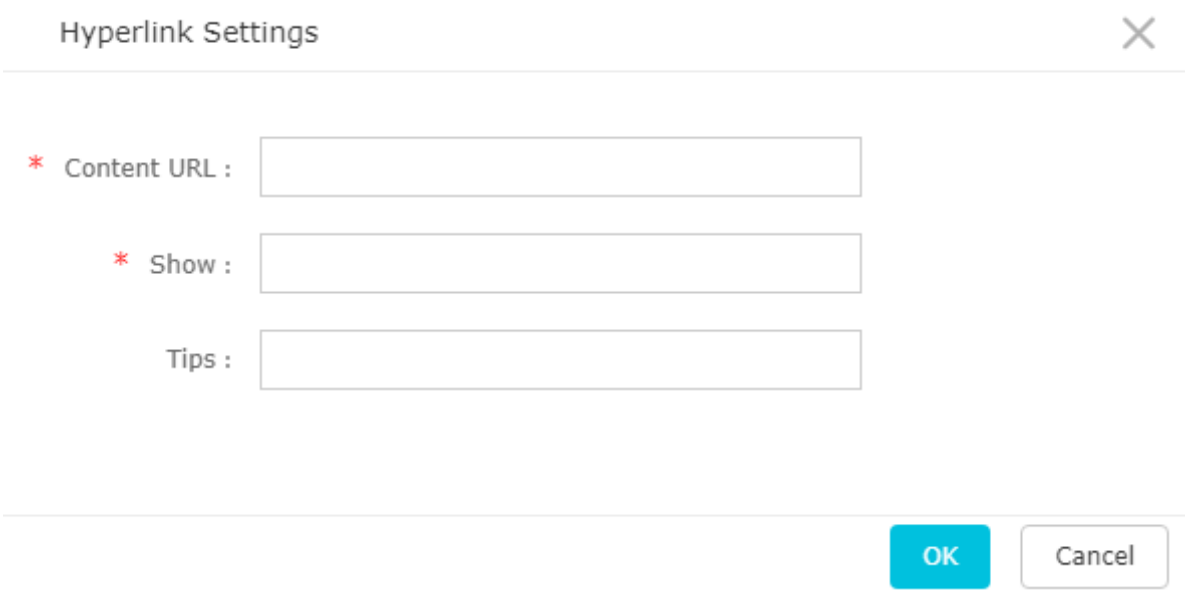
4. Click the image icon.

5. Click **Select File** and select an image.
6. Click **OK** to insert the image.

Insert a hyperlink

7. Click the link icon.
8. Enter a link address and a link display name, as shown in *Figure 26-154: Insert a hyperlink*.

Figure 26-154: Insert a hyperlink

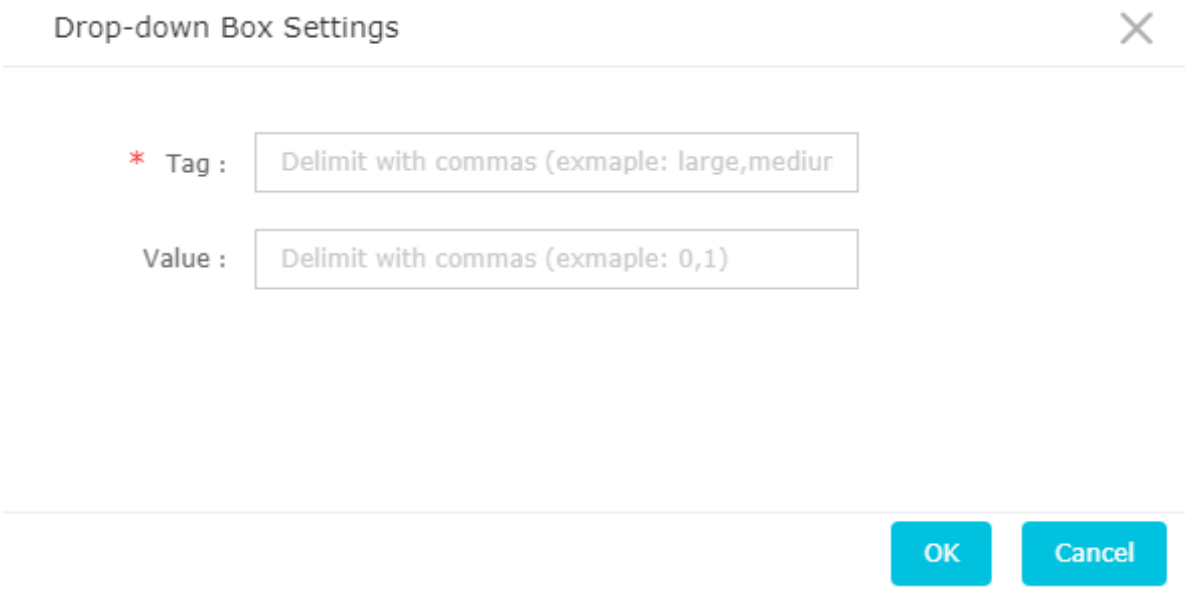


9. Click **OK** to insert the hyperlink.

Insert a drop-down list

10. Click the drop-down list icon.
11. Enter entry labels and entry values, as shown in *Figure 26-155: Set a drop-down list*.

Figure 26-155: Set a drop-down list



12.Click **OK** to insert the drop-down list.

26.5.9 Set the table format

The table format feature allows you to adjust the display format of a workbook.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbook**.
3. Click a workbook name in the list. The workbook editing page appears.

   For information about how to create a workbook, see *Create a workbook*.
4. Click **Table Format** to display the format list.
5. Click "Format" to select a proper table format, as shown in *Figure 26-156: Table format*.

Figure 26-156: Table format

26.5.10 Set condition rules

The condition rule feature allows you to set the content of a table. For example, you can highlight certain numeric values and add arrows to numeric values to indicate increase or decrease.

Procedure

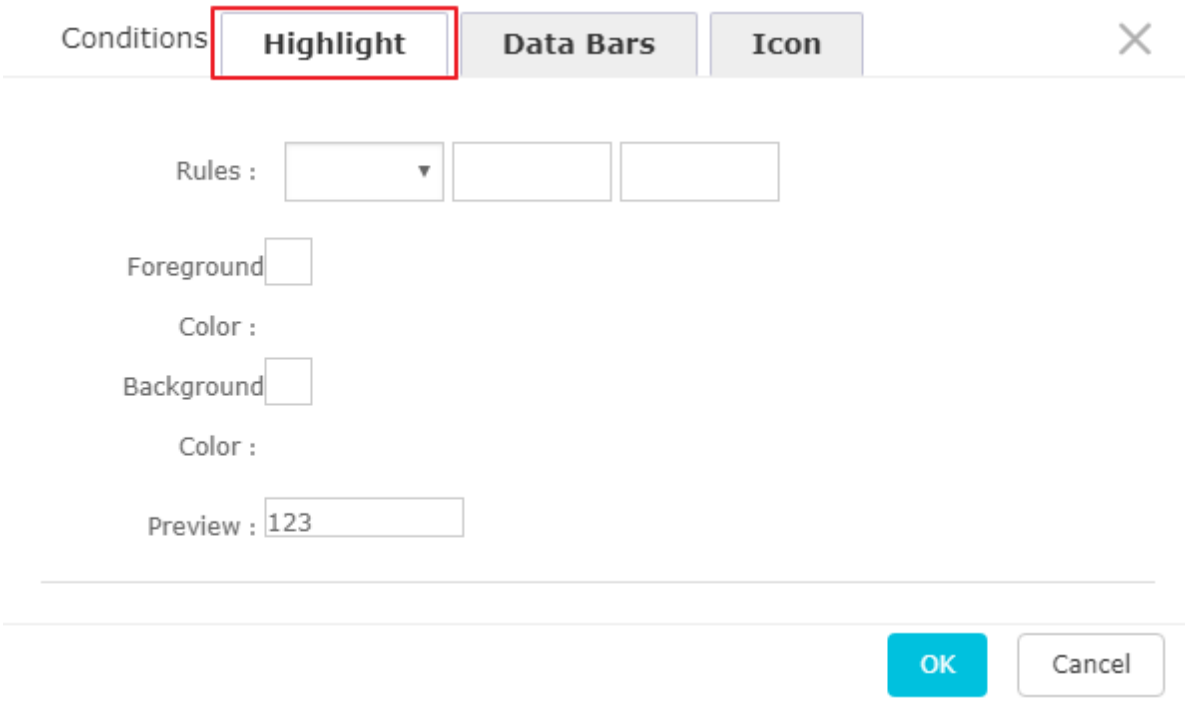
1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbook**.
3. Click a workbook name in the list. The workbook editing page appears.

   For information about how to create a workbook, see *Create a workbook*.

Highlight

4. Go to the "Condition Rule" menu.
5. Select the **Highlight** tab, as shown in *Figure 26-157: Highlight*.

Figure 26-157: Highlight



- Rule: Click the drop-down arrow, select the expected highlighting rule, and manually enter a number in the box next to the rule.
- Foreground color: Click "Color Block" and select the expected foreground color.
- Background color: Click "Color Block" and select the expected background color.
- Preview: Manually enter the content to be previewed to preview the highlight effect.

6. After the preceding settings, click **OK**.

Data column

7. Select the **Data Bars** tab, as shown in *Figure 26-158: Data column*.

Figure 26-158: Data column

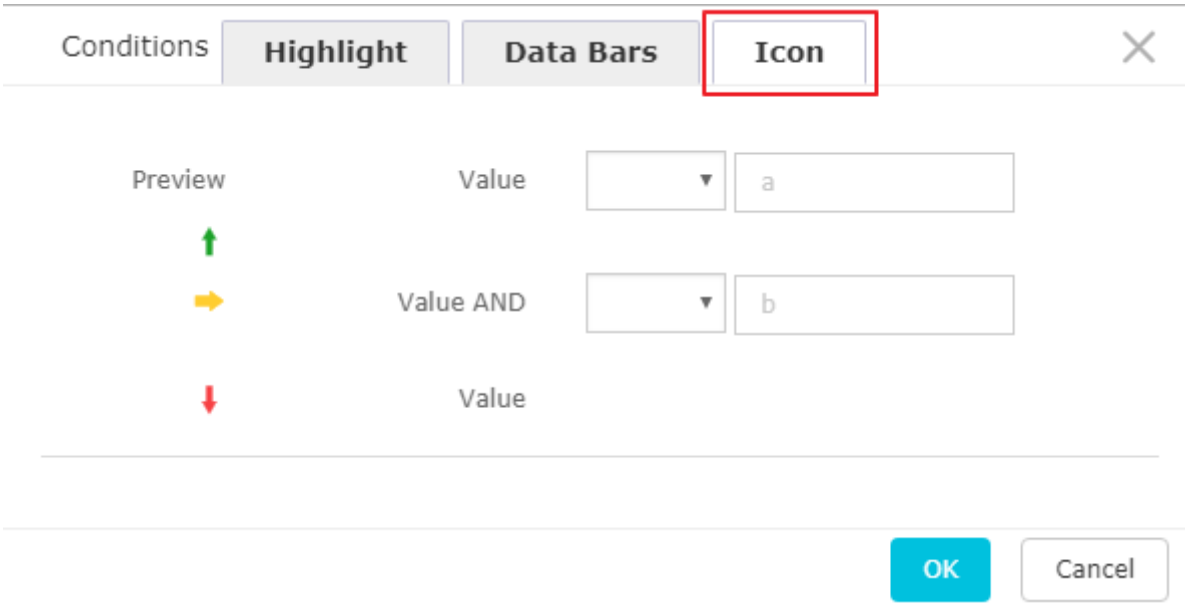
- Min. value: Manually enter the minimum value.
- Max. value: Manually enter the maximum value.
- Color: Click "Color Block" to select the expected color.

8. After the preceding settings, click **OK**.

Icon

9. Select the **Icon** tab, as shown in *Figure 26-159: Icon*.

Figure 26-159: Icon



Click the drop-down arrow, select the expected numeric character, and manually enter a number in the box next to the numeric character. Three arrows respectively in green, yellow, and red appear next to the number within the specified range.

10. After the preceding settings, click **OK**.

26.5.11 Query workbooks

Procedure

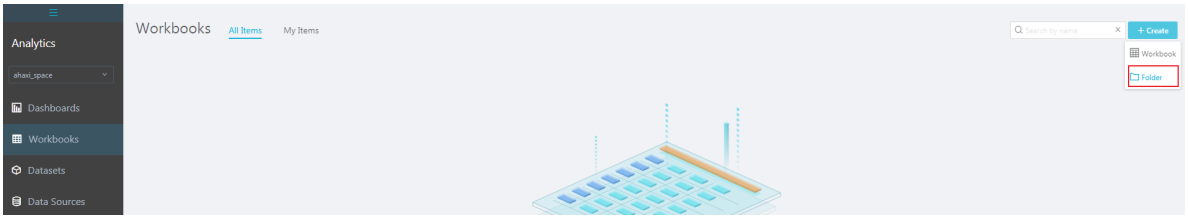
1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbook**.
3. Enter a keyword in the search box.
4. Click "Search" to search for the workbooks that match the keyword.

26.5.12 Create a workbook folder

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbooks**.
3. Choose **Create** > **Folder**, as shown in *Figure 26-160: Create a folder*.

Figure 26-160: Create a folder



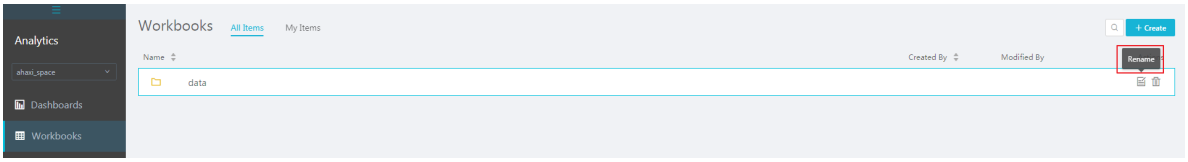
4. Enter a name for the folder and click a blank area.

26.5.13 Rename a workbook folder

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbooks**.
3. Select a workbook from the list.
4. Click **Rename**, as shown in *Figure 26-161: Rename a folder*.

Figure 26-161: Rename a folder



26.5.14 Transfer workbooks

You can transfer existing workbooks to other users.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbook**.
3. Select a workbook from the list.
4. Click **Transfer**.
5. Manually enter the account of the transferee and click **Transfer**.

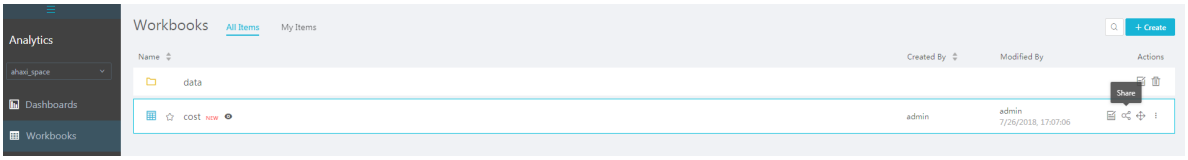
26.5.15 Share workbooks

You can share existing workbooks with other users.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbooks**.
3. Select a workbook from the list.
4. Click **Share**, as shown in *Figure 26-162: Share workbooks*.

Figure 26-162: Share workbooks



5. Enter the account of the person you want to share the workbook with.
6. Select an expiration date for the shared workbook.
7. Click **Save**.

26.5.16 Publish workbooks

You can publish existing workbooks to allow other users to access them.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, click **Workbook**.
3. Select a workbook from the list.
4. Click **Publish**.
5. Select an expiration date for the published workbook.
6. Select **Generate Link** and click **Publish**.

26.6 Build data portals

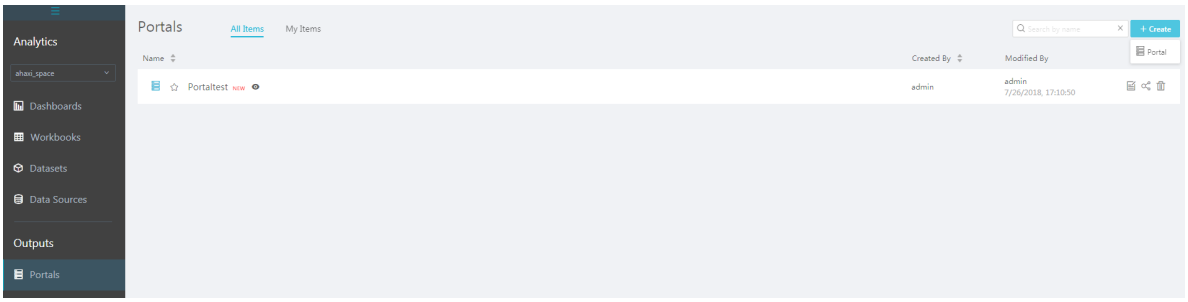
A data portal is also called a data product. It is a set of dashboards that contain menus. Through a data portal, you can create complex topical analyses with navigation menus.

26.6.1 Create a data portal

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, choose **Data Portal** > **Portal**.
3. On the data portal list page, choose **Create** > **Portal**, as shown in *Figure 26-163: Create a data portal*.

Figure 26-163: Create a data portal



26.6.2 Set a template

On the "Template Setting" tab, you can set the template title, upload a logo, and edit the footer.

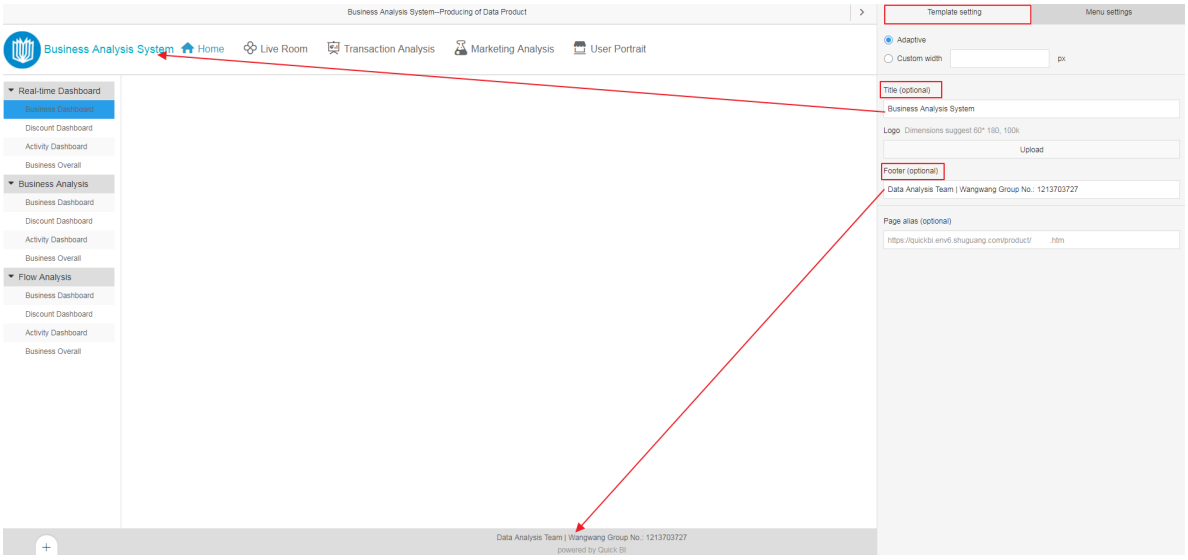
Procedure

1. *Log on to the Quick BI console*.

2. On the left-side navigation bar, choose **Portals**.
3. On the data portal list page, click a data portal name.

   For information about how to create a data portal, see *Create a data portal*.
4. Select the **Template Setting** tab and edit the portal template, as shown in *Figure 26-164: Template settings*.

Figure 26-164: Template settings



26.6.3 Set a menu

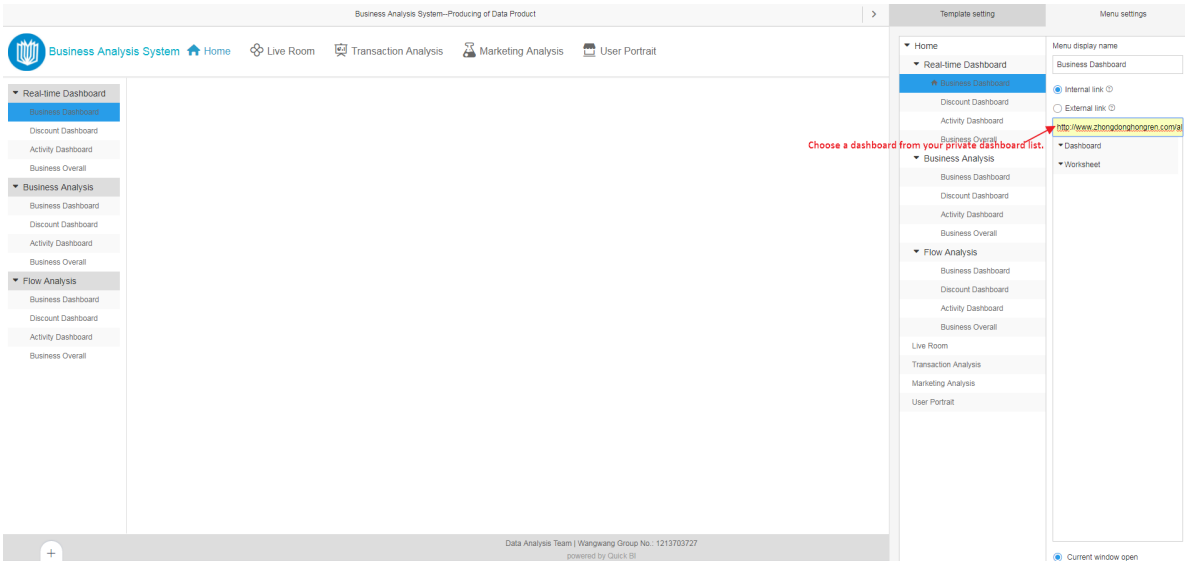
On the "Menu Setting" tab, you can set the menu name and the URL that is clicked to open a menu.

Procedure

1. *Log on to the Quick BI console*.
2. On the left-side navigation bar, choose **Portals**.
3. On the data portal list page, click a data portal name.

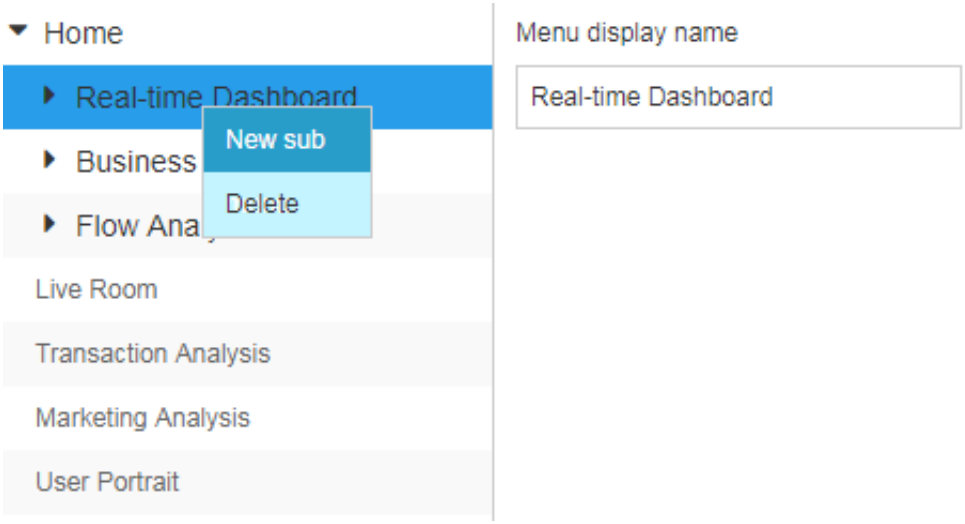
   For information about how to create a data portal, see *Create a data portal*.
4. Select the **Menu settings** tab and edit the menu, as shown in *Figure 26-165: Menu settings*.

Figure 26-165: Menu settings



- Right-click the menu name and choose "Edit" from the shortcut menu. The menu editing list appears.

Figure 26-166: Edit the menu structure



- You can reference dashboards and workbooks.

26.7 Organizational unit management

This section describes how to create and manage organizational units and group spaces.

An organizational unit generally refers to a small- or medium-sized enterprise, public institution, school faculty, or a department of a large-sized company.

If you have many organizational unit members, some of whom need to collaborate to perform data analysis, and you have high requirements on data security, you can use Quick BI to:

- Enable different departments to access different or the same report.
- Enable employees with different roles to access different contents.

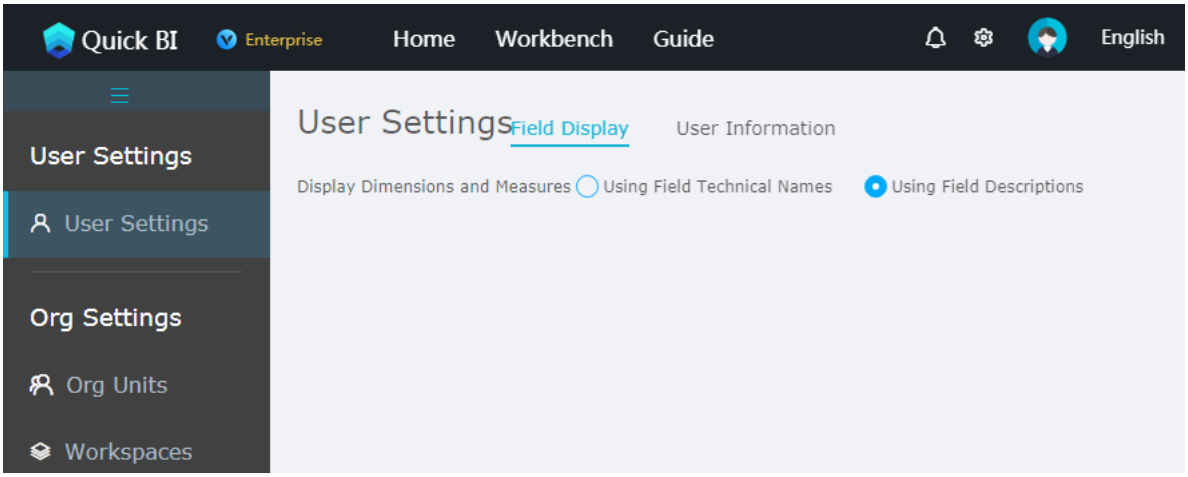
There are only two types of organizational unit members: the organizational unit administrator and common organizational unit members.

26.7.1 Create an organizational unit

Procedure

1. On the Quick BI console, click **Manage** to go to the management page.
2. In organizational unit permission setup, select **Organizational Unit Management** to go to the organizational unit management page.
3. Select **I Agree** and click **Activate Organizational Unit**.
4. On the "Create Organizational Unit" page, enter an organizational unit name, as shown in *Figure 26-167: Create an organizational unit*.

Figure 26-167: Create an organizational unit



Note:

- Only a primary Alibaba Cloud account that has not activated or joined any organizational unit can create an organizational unit.
- An Alibaba Cloud account can activate or join only one organizational unit.

26.7.2 Modify organizational unit information

Context

An organizational unit administrator can modify organizational unit attributes after creating the organizational unit, as shown in Figure.

The organizational unit administrator adds Alibaba Cloud accounts that need to work collaborat ively to the organizational unit.

In group space management, the organizational unit administrator adds members in the organizati onal unit to different group spaces based on their work scope and responsibilities. Group spaces can correspond to actual business departments of the organizational unit. The organizational unit administrator can manage group spaces based on actual business requirements. For example, if the organizational unit has a sales department and an HR department, the administrator creates a sales group and an HR group accordingly. The administrator then adds employees in the sales department to the sales group, and employees in the HR department to the HR group.

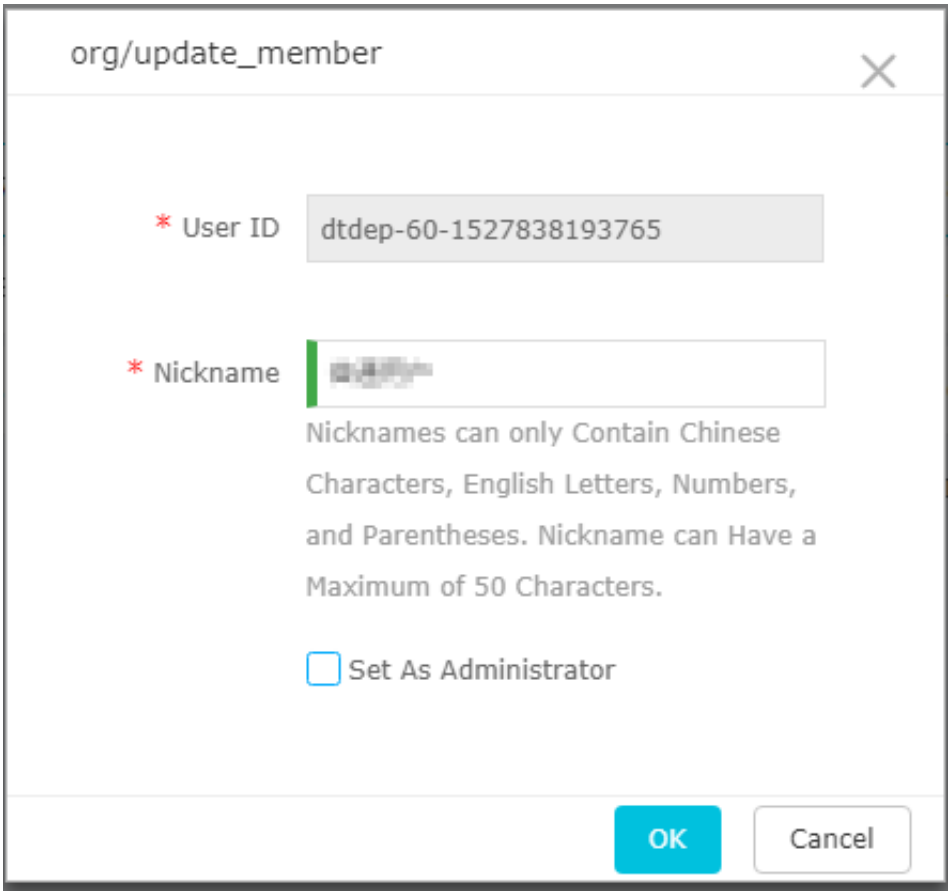
Only the organizational unit administrator has the permission to use the organizational unit member management module. The creator of an organizational unit becomes the organizational unit administrator by default.

There are only two types of organizational unit members: the organizational unit administrator and common organizational unit members.

Procedure

1. Go to the organizational unit management page.
2. Select the **Org Units** tab.
3. Click the icon of organizational unit attribute editing, as shown in *Figure 26-168: Modify organizational unit attributes*.

Figure 26-168: Modify organizational unit attributes

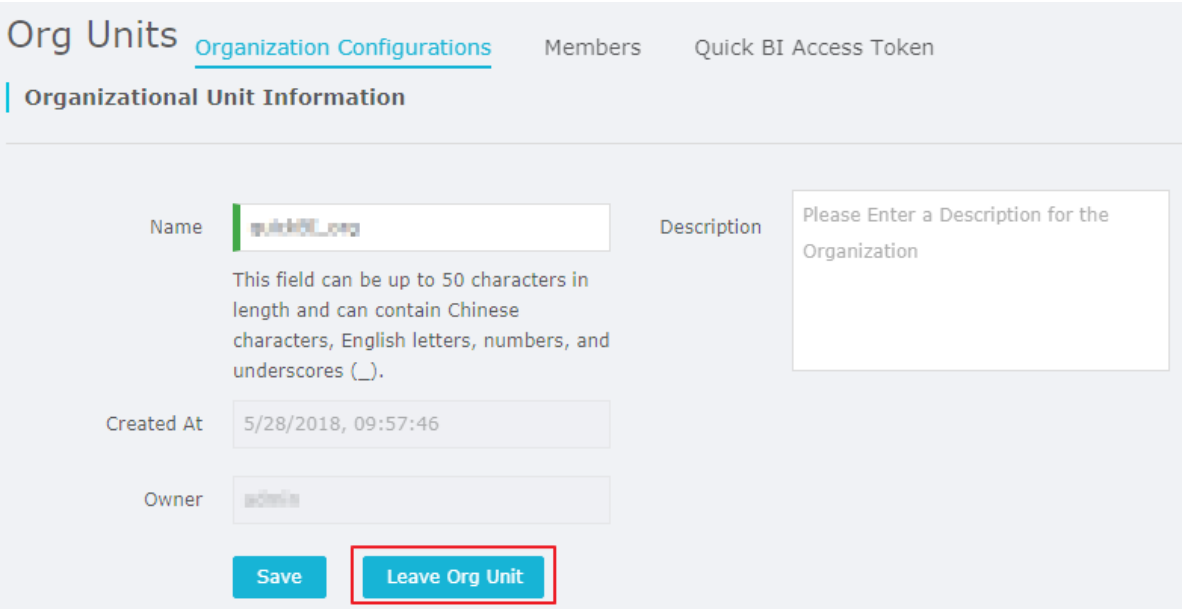


26.7.3 Withdraw from an organizational unit

Procedure

1. Log on to the organizational unit management page.
2. Click **Leave Org Unit** to withdraw from the current organizational unit, as shown in *Figure 26-169: Withdraw from an organizational unit*.

Figure 26-169: Withdraw from an organizational unit



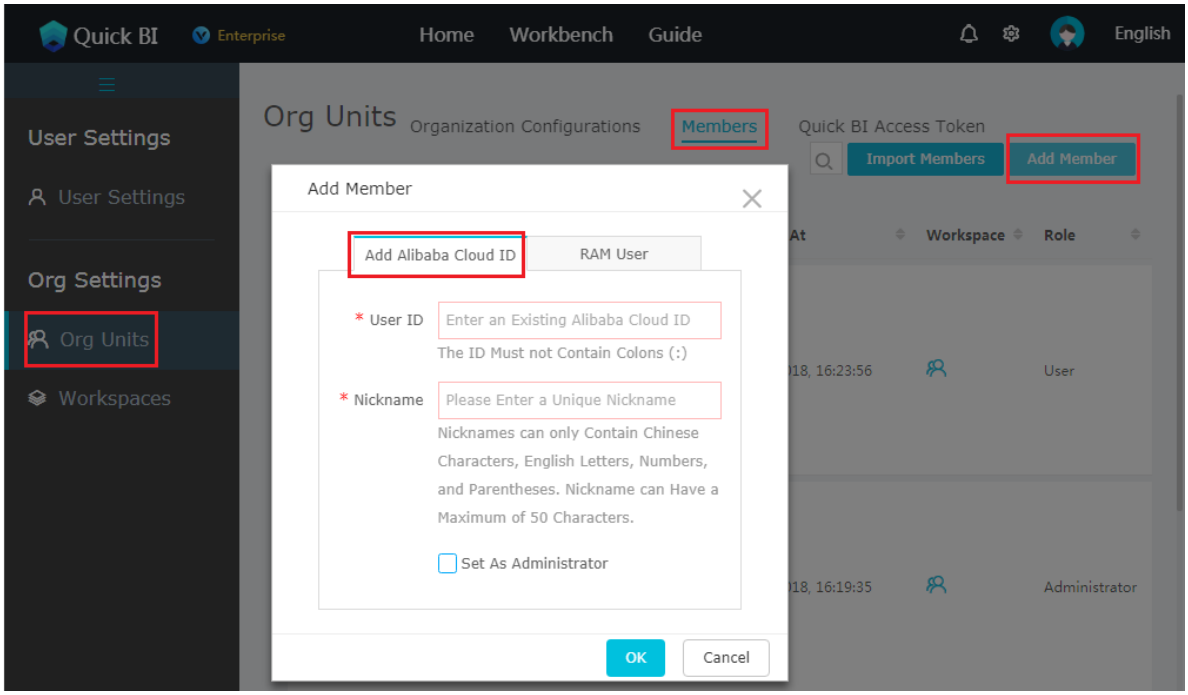
26.7.4 Add an organizational unit member

Procedure

Add an organizational unit member using a primary Alibaba Cloud account.

1. Go to the organizational unit management page.
2. Select the **Org Units** tab.
3. Click **Add Member**.
4. Select the **Add Alibaba Cloud ID** tab.
5. Enter the primary Alibaba Cloud account of the user you want to add to the organizational unit, as shown in *Figure 26-170: Add an organizational unit member*.

Figure 26-170: Add an organizational unit member

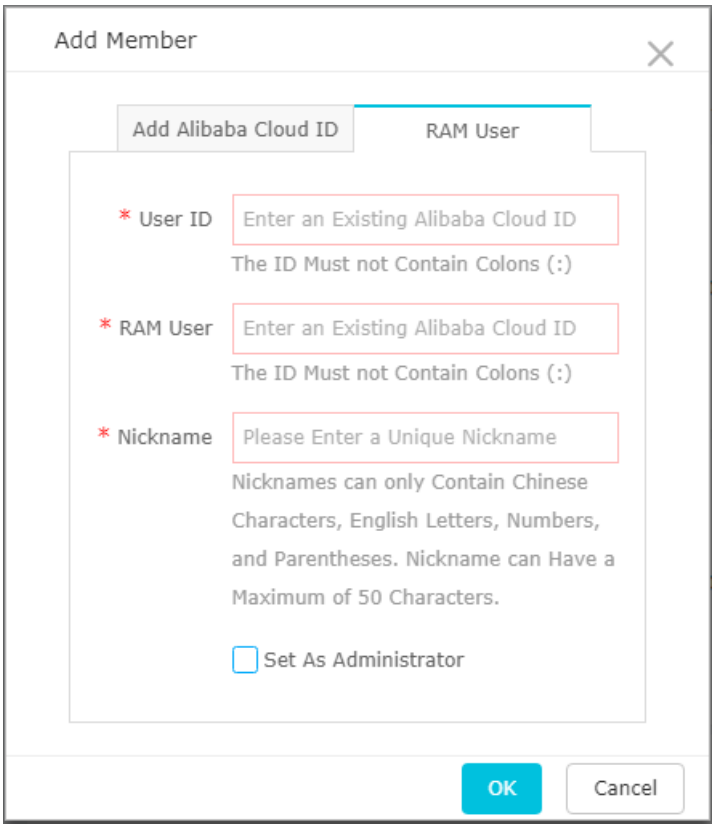


6. Click **OK** to add the organizational unit member.

Add an organizational unit member using a RAM subaccount.

7. Select the **RAM User** tab.
8. Enter the RAM subaccount of the user you want to add to the organizational unit, as shown in *Figure 26-171: Add an organizational unit member*.

Figure 26-171: Add an organizational unit member



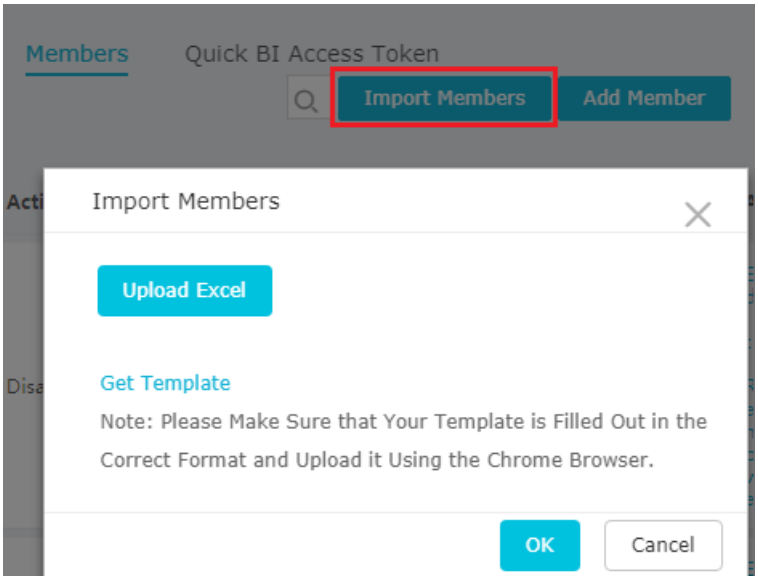
9. Click **OK** to add the organizational unit member.

   If the account has joined an organizational unit, a message is displayed, indicating that the user cannot be added to the current organizational unit.

Batch add organizational unit members

10. Click **Import Members**.
11. Upload a list of users that you want to add to the organizational unit from your computer, as shown in *Figure 26-172: Batch add organizational unit members*.

Figure 26-172: Batch add organizational unit members



12.Click **OK** to add the organizational unit members.

26.7.5 Modify an organizational unit member

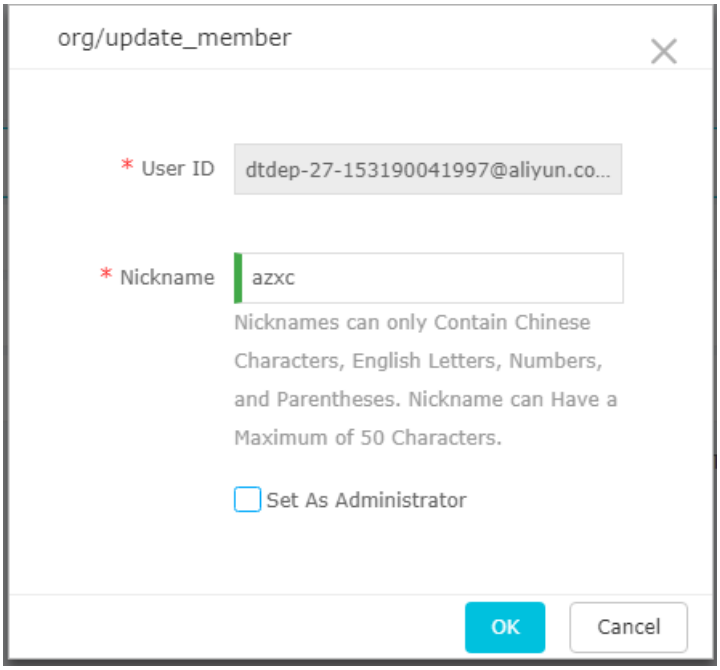
Context

You can set roles (organizational unit administrator or common user) for members in an organizational unit, and set nicknames for members for easy member search.

Procedure

1. Go to the organizational unit management page.
2. Select the **Org Units** tab.
3. Select an organizational unit member and click **Edit** next to it.
4. Modify information about the organizational unit member, as shown in *Figure 26-173: Modify an organizational unit member*.

Figure 26-173: Modify an organizational unit member



5. After you finish modification, click **OK**.

26.7.6 Remove a member from an organizational unit

Context

The administrator of an organizational unit has the permission to remove members from the organizational unit. When you remove a member from an organizational unit, if the member is still included in a group space, you must first remove the member from the group space; otherwise, a message is displayed, indicating that the member cannot be removed from the organizational unit.

Procedure

1. Go to the organizational unit management page.
2. Select the **Org Units** tab.
3. Select an organizational unit member and click **Remove** next to it, as shown in Figure .
4. Click **OK** to remove the organizational unit member.

26.7.7 View the group space a user belongs to

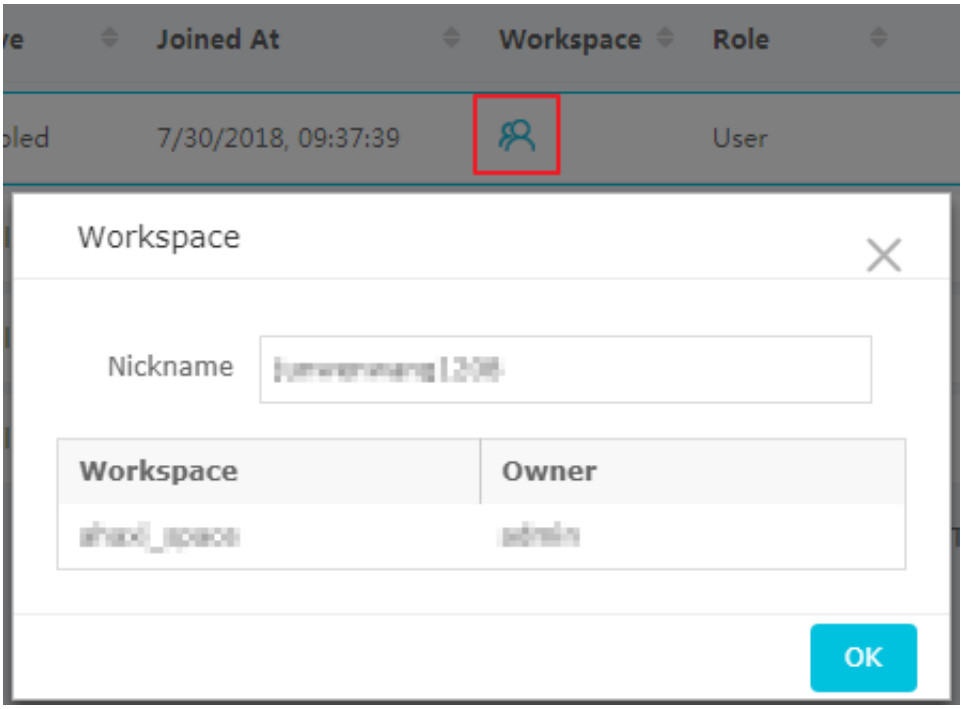
Context

You can view the group space an organizational unit member belongs to.

Procedure

1. Go to the organizational unit management page.
2. Select the **Org Units** tab.
3. Select an organizational unit member and click "Workspace" next to it, as shown in *Figure 26-174: View the group space a user belongs to*.

Figure 26-174: View the group space a user belongs to



4. Click **OK** to close the dialog box.

26.7.8 Query organizational unit members

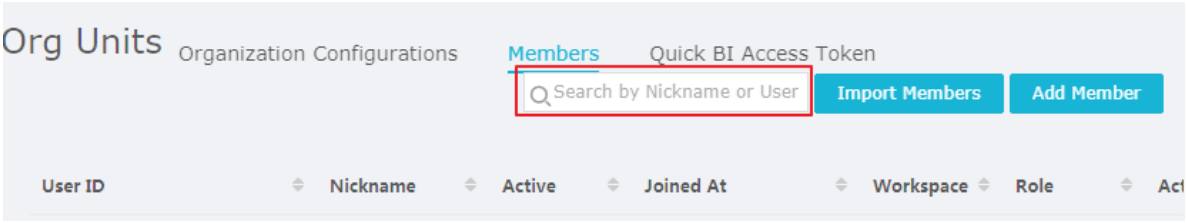
Context

You can query organizational unit members by nickname or account.

Procedure

1. Go to the organizational unit management page.
2. Select the **Org Units** tab.
3. Enter a nickname or account and click **Search**, as shown in *Figure 26-175: Query organizational unit members*.

Figure 26-175: Query organizational unit members



26.7.9 Group space management

A group space is managed by the group space administrator, who is designated by the organizational unit administrator who creates the group. A group space administrator can designate another member as the group space administrator.

Group space management includes:

- Create a group space
- Modify a group space
- Set a default group space

26.7.9.1 What is a group space

A group space is a workspace where organizational unit members perform collaborative development. In a group space, group members can collaborate to create and modify data sources, datasets, worksheets, dashboards, and data portals based on their roles. These data objects exist in the group space they belong to. Different group spaces have different data objects.

A group space has the following attributes:

- Group space name
- Group space description
- Function permissions: Whether worksheets can be exported – Worksheets can be exported by default. If this function is disabled, data in the group space cannot be exported and downloaded . Whether dashboards are public – Dashboards are public by default. If this function is disabled , dashboards in the group space cannot be used by all users. Whether files can be shared – Files can be shared by default. If this function is disabled, files in the group space cannot be shared to users out of this group.
- Preference settings: Whether to use physical fields or field annotations as the dimension and measurement name. After this attribute is set, it takes effect for the default name setting when

new datasets in the group space are created. The name setting of old datasets are not affected .

You can enter member accounts to search for users in the organizational unit in fuzzy match mode, and assign roles to the users. Different roles have different viewpoints and permissions. A user can have different roles and must have at least one role.

The following roles are available: group administrator, developer, analyst, and viewer.

List of mappings between roles and permissions

The mapping between a role and its permissions is fixed and cannot be modified. When authorizing a user, you can only set the role of the user. Mappings between roles and permissions are listed as follows:

- Function navigation entry

Table 26-3: Function navigation entry

Permission	Developer	Analyst	Viewer
Data	Yes	No	No
Worksheet	Yes	Yes	Yes
Dashboard	Yes	Yes	Yes
Data portal	Yes	Yes	Yes

- Data

Table 26-4: Data

Permission	Developer	Analyst	Viewer
Create data sources	Yes	No	No
Modify data sources	Only data sources of the current user can be modified.	No	No
Delete a data source	Only data sources of the current user can be deleted.	No	No
Transfer data sources	Only data sources of the current user can be transferred.	No	No

Permission	Developer	Analyst	Viewer
Use data sources	Yes	No	No
Create a dataset	Yes	No	No
Modify datasets	Only datasets of the current user can be modified.	No	No
Delete a dataset	Only datasets of the current user can be deleted.	No	No
Transfer a dataset	Only datasets of the current user can be transferred.	No	No
Use datasets	Yes	Yes	No

- Workbook

Table 26-5: Workbook

Permission	Developer	Analyst	Viewer
Create workbooks	Yes	Yes	No
Modify workbooks	Only workbooks of the current user can be modified.	Only workbooks of the current user can be modified.	No
Delete workbooks	Only workbooks of the current user can be deleted.	Only workbooks of the current user can be deleted.	No
Preview workbooks	Yes	Yes	Yes
Transfer workbooks	Only workbooks of the current user can be transferred.	Only workbooks of the current user can be transferred.	No
Share workbooks	Only workbooks of the current user can be shared.	Only workbooks of the current user can be shared.	No
Reference workbooks	Yes	Yes	No

- Dashboard

Table 26-6: Dashboard

Permission	Developer	Analyst	Viewer
Create a dashboard	Yes	Yes	No
Modify dashboards	Only dashboards of the current user can be modified.	Only dashboards of the current user can be modified.	No
Delete dashboards	Only dashboards of the current user can be deleted.	Only dashboards of the current user can be deleted.	No
Preview dashboards	Yes	Yes	Yes
Transfer a dashboard	Only dashboards of the current user can be transferred.	Only dashboards of the current user can be transferred.	No
Share a dashboard	Only dashboards of the current user can be shared.	Only dashboards of the current user can be shared.	No
Reference dashboards	Yes	Yes	No
Publish dashboards	Only dashboards of the current user can be published.	Only dashboards of the current user can be published.	No

- Data portal

Table 26-7: Data portal

Permission	Developer	Analyst	Viewer
Create data portals	Yes	Yes	No
Modify data portals	Only data portals of the current user can be modified.	Only data portals of the current user can be modified.	No
Delete data portals	Only data portals of the current user can be deleted.	Only data portals of the current user can be deleted.	No
Preview data portals	Yes	Yes	Yes

Permission	Developer	Analyst	Viewer
Transfer data portals	Only data portals of the current user can be transferred.	Only data portals of the current user can be transferred.	No
Share data portals	Only data portals of the current user can be shared.	Only data portals of the current user can be shared.	No

26.7.9.2 Differences between a personal space and a group space

In Quick BI Basic, a user's workspace is called a personal space. Main differences between a personal space and a group space are as follows:

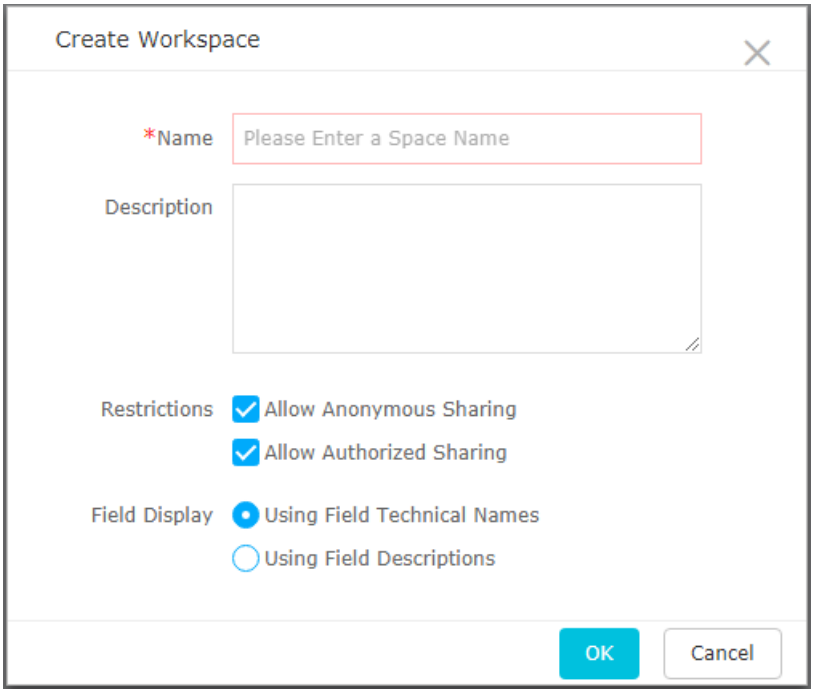
- The personal space is automatically created when a user logs on to the group for the first time. A group space must be manually created by the organizational unit administrator.
- The personal space cannot be deleted, and a new personal space cannot be created.
- Other users cannot be added to a personal space of a user. Therefore, the personal space cannot be shared.
- Data objects in a personal space can be transferred to or shared with any Alibaba Cloud Quick BI user. Data objects in a group space can be transferred within the group space and shared with members in the organizational unit.

26.7.10 Create a group space

Procedure

1. Go to the organizational unit permission management page.
2. Choose **Workspaces** > **Create Workspace**.
3. On the page that appears, enter a name for the group space, as shown in *Figure 26-176: Create a group space*.

Figure 26-176: Create a group space



4. Click **OK** to create the group space.

26.7.11 Modify a group space

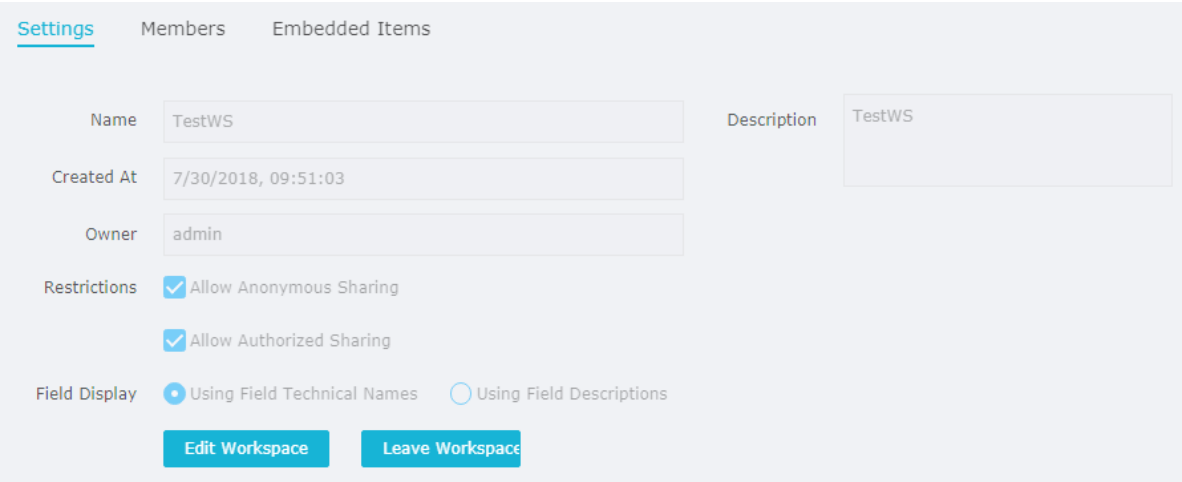
Context

Only the personal space owner can modify parameter settings of a personal space. Only the group space administrator can modify the parameter settings of a group space.

Procedure

1. Go to the organizational unit permission management page.
2. Select the **Workspaces** tab.
3. Modify group space information as follows:

Figure 26-177: Modify a group space



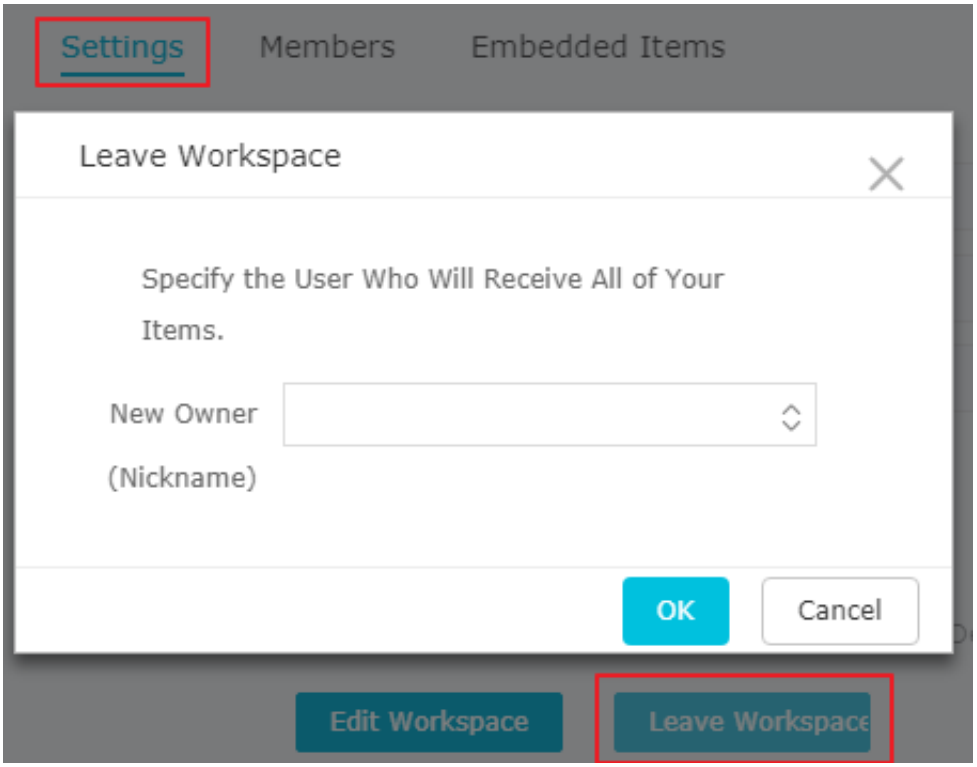
4. After you finish modification, click **OK**.

26.7.12 Withdraw from a group space

Procedure

1. Go to the organizational unit permission management page.
2. Select the **Workspaces** tab.
3. Select a group space and select the **Settings** tab.
4. Click **Leave Workspace** to withdraw from the selected group space, as shown in *Figure 26-178: Withdraw from a group space*.

Figure 26-178: Withdraw from a group space



26.7.13 Transfer a group space

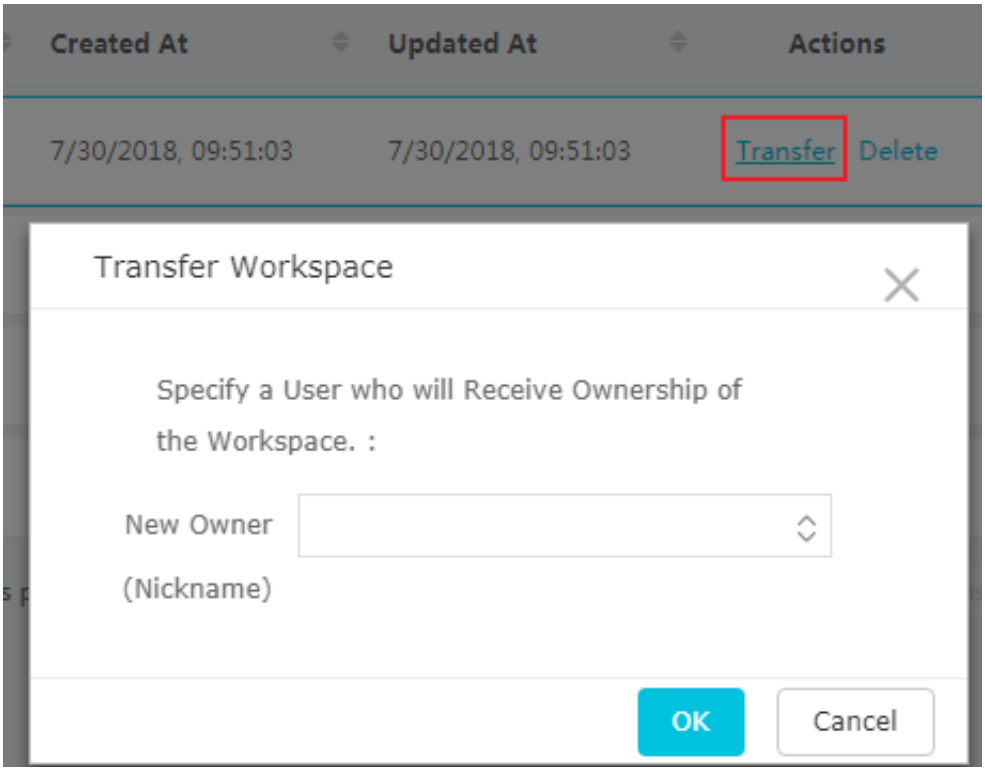
Context

To remove the creator of a group space from an organizational unit, you can transfer that group space to another member in the organizational unit. The transferee does not have to be the organizational unit administrator. Any common member can become the owner of the group space .

Procedure

1. Go to the organizational unit permission management page.
2. Select the **Workspaces** tab.
3. Select a group space and click **Transfer**.
4. Enter the nickname of the transferee and click **OK**, as shown in *Figure 26-179: Transfer a group space*.

Figure 26-179: Transfer a group space

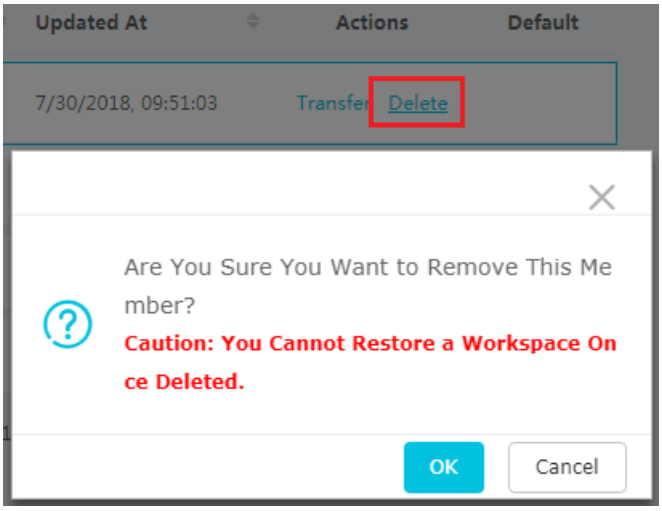


26.7.14 Delete a group space

Procedure

1. Go to the organizational unit permission management page.
2. Select the **Workspaces** tab.
3. Select a group space and click **Delete**, as shown in *Figure 26-180: Delete a group space*.

Figure 26-180: Delete a group space

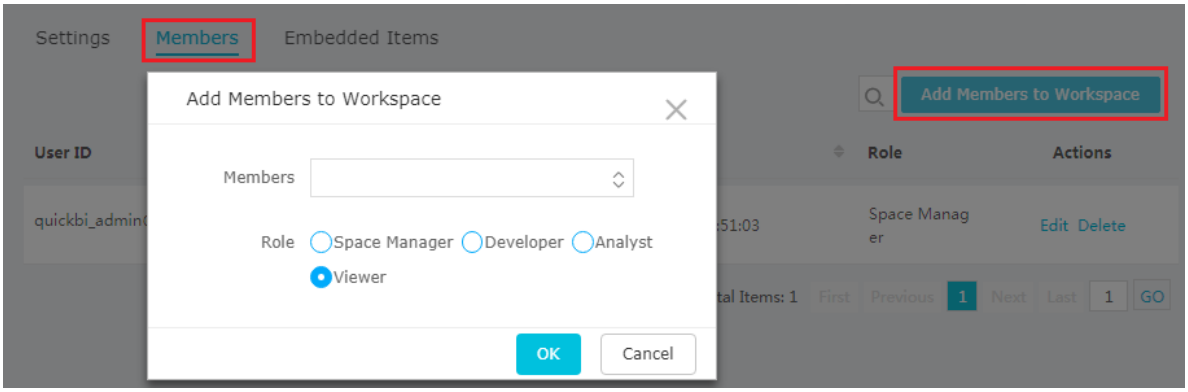


26.7.15 Add a group space member

Procedure

1. Go to the organizational unit permission management page and select a workspace.
2. Select the **Members** tab.
3. Click **Add Members to Workspace**.
4. Enter an account and assign a role to that account, as shown in *Figure 26-181: Add a group space member*.

Figure 26-181: Add a group space member



26.7.16 Modify a group space member

Procedure

1. Go to the organizational unit permission management page.

2. Select the **Workspaces** tab.
3. Select a group space member and click **Edit**.
4. After you finish modification, click **OK**.

26.7.17 Delete a group space member

Procedure

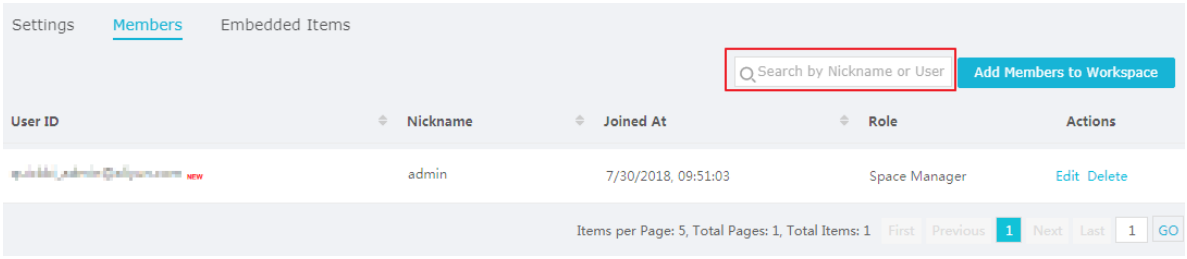
1. Go to the organizational unit permission management page.
2. Select the **Workspaces** tab.
3. Select a group space member and click **Delete**.
4. Click **OK** to delete the selected member.

26.7.18 Query group space members

Procedure

1. On the group space management page, select a group space.
2. Select the **Group Space Member** tab, as shown in *Figure 26-182: Query group space members*.

Figure 26-182: Query group space members



You can also enter a keyword in the search box to search for a specific group space member.

26.8 Permission management

Permission management includes data object management and row-level permission management.

Data objects include data sources, datasets, workbooks, dashboards, and data portals. Data object management can be classified into management of data objects in a personal space and in a group space.

26.8.1 Manage data objects

Data objects include data sources, datasets, workbooks, dashboards, and data portals.

Data object management can be classified into management of data objects in a personal space and in a group space.

Transfer data objects in a group space

Data sources, datasets, workbooks, dashboards, and data portals can be transferred.

- Only the creator of a file and the group administrator can transfer data objects.
- Data objects can be transferred only within the group space.

Share data objects in a group space

Workbooks, dashboards, and data portals can be shared. Shared files can be accessed by other users in read-only mode but cannot be modified, deleted, or saved as another files.

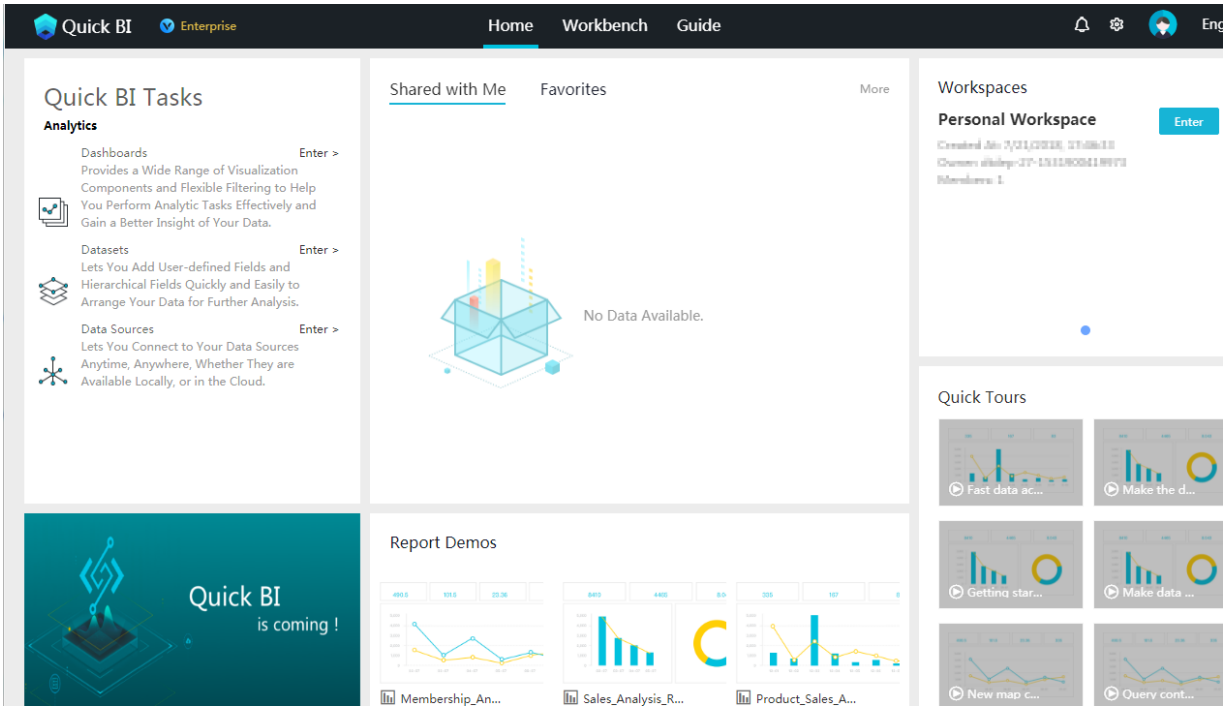
- Only the creator of a file and the group administrator can share data objects.
- If the share function is disabled, all files in the group space cannot be shared.
- Share scope: Data objects can be shared only within the group space. Currently, data objects cannot be shared with Alibaba Cloud accounts out of the organizational unit.

By default, all files in a group space can be viewed by group members in the group space.

Files in a group space can be shared to an organizational unit member who is not in the group space.

The member to whom the file is shared can view the file on the **Home** page, as shown in *Figure 26-183: View shared files*.

Figure 26-183: View shared files



Dashboard objects in a group space can be shared to all people. A public dashboard can be accessed by all people. Therefore, we do not recommend that you publish dashboards that contain business data.

26.8.2 Row-level authorization

For relevant information, see *Dataset row-level permission control*.

26.8.3 Manage data objects in a personal space

Only the creator can modify or delete the data objects in his/her own personal space. Other users have no modification or deletion permission.

Transfer data objects in a personal space

Datasets, workbooks, dashboards, and data portals can be transferred.

- Only the creator of the data objects can transfer the data objects.
- Transfer scope: Alibaba Cloud Quick BI users

Currently, datasets in AnalyticDB data sources cannot be transferred.

Share data objects in a personal space

Workbooks, dashboards, and data portals can be shared. Shared files can be accessed by other users in read-only mode but cannot be modified, deleted, or saved as another files.

- Only the creator of the data objects can share the data objects.
- Share scope: Alibaba Cloud Quick BI users

The member to whom the file is shared can view the file on the **My** page.

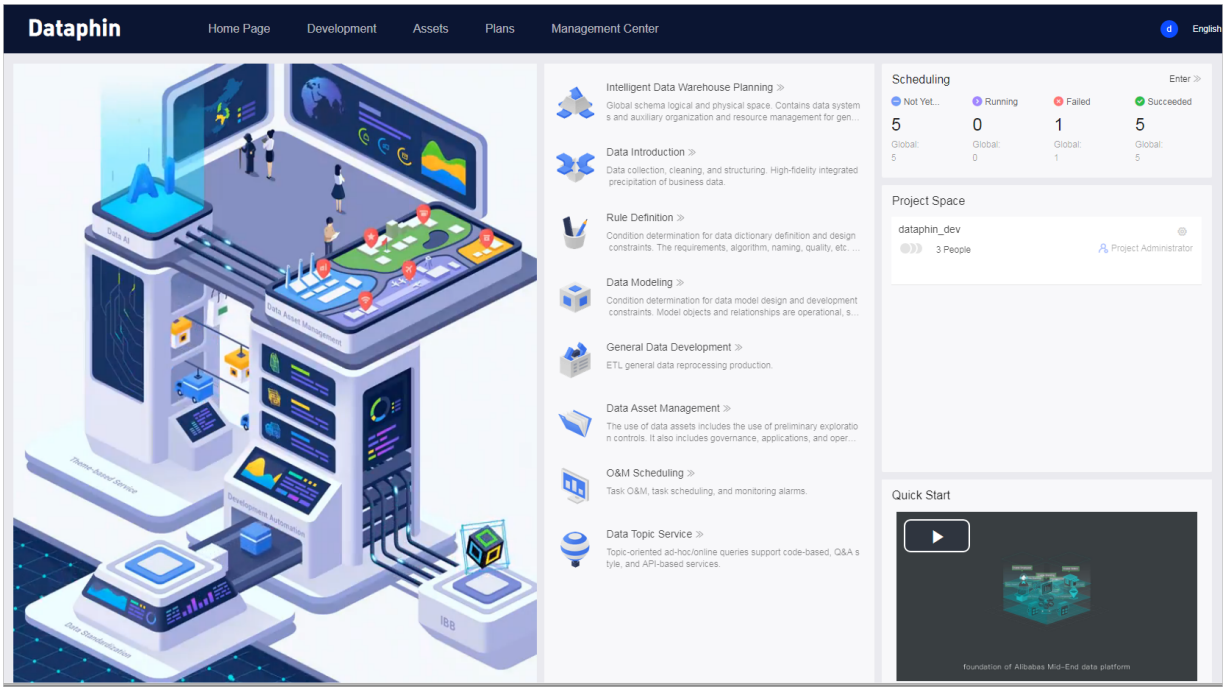
Dashboard objects in a personal space can be shared to all people. A public dashboard can be accessed by all people. Therefore, we do not recommend that you publish dashboards that contain business data.

27 Dataphin

27.1 What is Dataphin

Dataphin is an intelligent data construction and management engine that has been designed for utilization in multiple industries. Dataphin applies the OneData, OneID, and OneService data construction technology that has been tested by business in Alibaba Group for 10 years. Dataphin provides an end-to-end intelligent data construction and management service, which includes data importing, data standardization, data modeling, data development, data distilling, asset management, and other data services. These features can be used to help the government and enterprises build an intelligent data system that includes standardization, integration, assets, services, and closed-loop optimization.

Dataphin aims to support different compute and storage environments. By using Dataphin, you can quickly import data, construct standardized data, and build data models. The service also allows you to create a tag system using customer and product data to gain business knowledge , create data assets, and resolve business issues. Dataphin also provides multiple types of data services including data table search and intelligent voice search.



27.2 Before you start

To guarantee stable operation of the system, observe the following constraints or recommendations when using Dataphin.

Operation	Constraint for RDS use
Computing engine type - selection	Select a computing engine type (Hadoop or MaxCompute) based on the condition of your computing cluster. The system completes data creation based on a single type of computing engine.
Data source management - new	We recommend that you set an AccessKey with the administrative privilege for data source management. Do not set data sources on the same physical database (exactly the same configuration).
Project management - project name	For a MaxCompute data source, the project name must be the same as the project name on MaxCompute. The project name cannot start with LD_/ld_; otherwise, it may conflict with a business section name, causing unavailability of the query function.
Project management - data source configurat ion	The following configurations are not recommended: 1. Use a non-Dataphin console to add, delete, or modify data in a physical database that has been configured as the data source of a project. 2. Configure data sources in different clusters for a project.
Research and development workbench - code management	Data cannot be read from data sources of a project in different clusters.
Development Workbench - data management	We recommend that you exercise caution when naming standard definitions and logical table objects, and set the names in lowercase for ease of reading, because the names may be unchangeable due to dependency of downstream nodes.
Development Workbench - temporary query	The name of a logical table must be prefixed with the associated business section name. The name of a physical table used for multiple project must be prefixed with the projects names.
Data distillation - ID center	We recommend that you set IDs based on user information to enable precise ID connection.

27.3 Quick start

Dataphin enables one-stop data creation and consists of multiple sub-products and subsystems. This chapter describes how to log on to Dataphin and use sub-products or subsystems on the corresponding product pages.

27.3.1 Notice for system administrators

This topic is for system administrators. Before you using Dataphin, you must have Dataphin system administrators and deployers ensure a prepared environment and create related roles.

Procedure

1. Ensure a prepared hardware environment:
   - Alibaba Could Monitoring System is deployed and DTCenter is accessible.
   - Database resources, such as MaxCompute, OSS, Dauth, SLB, ECS, three physical machines, and PostgreSQL are prepared.

     Note:
     For the PostgreSQL database resource, you must have a PostgreSQL database of network interoperability. You may encounter the following two situations:
     - PostgreSQL on RDS: 6U model is required for Apsara Stack v3.7.0 and earlier versions , otherwise 7U model is required. The sale availability of miniRDS is gradually not supported. You can ignore miniRDS.
     - PostgreSQL on ECS: Configure Apsara Stack v3.7.0 and later Dataphin built-ins according to 3.7.0 and later baseline, otherwise the sales must contact the Dataphin team to assign the specified ECS resources to deploy PostgreSQL in a non-standard way.
2. Obtain computer clusters:
   - MaxCompute endpoint: for calculation engine settings.
   - AccessID and AccessKey of MaxCompute Project [Dataphin_Meta]: for metadata calculatio n storage.

     Note:
     When the environment of Apsara Stack and MaxCompute are deployed, a Project will be generated to obtain MaxCompute metadata. System administrators and deployers need to confirm that Project [Dataphin_Meta] exists in MaxCompute and obtains the corresponding

MaxCompute AccessID and AccessKey. If you encounter any problems, you must contact deployers to create a Project manually and empower MaxCompute.

3. Generate accounts:

   When the deployment is complete, Apsara Stack accounts have the following three role types:

   - Super operation and maintenance administrator: Belongs to a system-independent metadata tenant. Obtains and parses metadata of customer clusters. One system has only one of this role. Check with deployers for details on the account and password.

     Note:
     Ensure system administrators to carefully store super operation and maintenance administrators' account and password.

   - Super administrator: Belongs to a customer research and development tenant. Completes tenant management, top-level design, and construction of core business data systesm. One Apsara Stack department account (or the master account) corresponds to one of this role.
   - Common user: Belongs to a department research and development tenant. Completes the detailed design of specific business data systems. Sub-accounts can be added to the Apsara Stack department account . One sub-account is the user from the tenants of super administrators corresponding the department account. This user can be added as a tenant.

   Note:
   Apsara Stack does not currently support multiple tenants, so each role belongs to only one tenant. Apsara Stack does not support roles to join a tenant across department. It is recommended that only one department (that is one tenant) be created in one system.

27.3.2 Log on to the Dataphin console

Prerequisites

Before logging on to the Apsara Stack console, make sure that:

- You have obtained the IP address or domain name address of the Apsara Stack console from the deployment personnel.

  The access address of the Apsara Stack console is http://*x.x.x.x*/manage, where *x.x.x.x* represents the IP address or domain name address.
- You have upgraded your Chrome browser to 42.0.0 or later versions.

Context

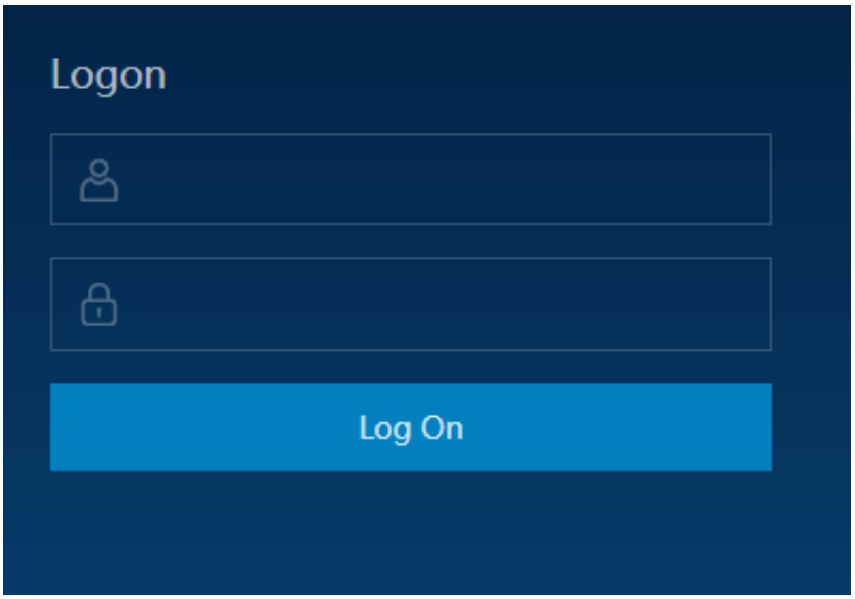
After logging on to Dataphin console, each sub-product included in the product can be used.

Procedure

1. Open your Chrome browser.
2. In the address bar, enter the access address of the Apsara Stack console in the format of http://*IP address or domain name address of the Apsara Stack console*/manage, and then press Enter.

   The logon interface appears, as shown in *Figure 27-1: Log on to the Apsara Stack console*.

   Figure 27-1: Log on to the Apsara Stack console

3. Enter the correct username and password.
   - The system has a default administrator, with the username super and password super. The administrator can create other system users and notify them of their default passwords by SMS or email.
   - You must modify the password of your username as instructed when you log on to the Apsara Stack console for the first time. To improve security, the password must meet minimum complexity requirements, that is, 8-20 characters long and containing at least two types of the following characters: English uppercase or lowercase letters (A-Z or a-z), numbers (0-9), or special characters (such as exclamation marks (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)).
4. Click **Log On** to go to the **Home** page.

5. Select **Console** > **Big Data** > **Dataphin** to enter the **Dataphin** console entrance.
6. Select **`Region`** and **`Department`**, and then click **Dataphin** to enter the **Dataphin** console.

27.3.3 Management Center

you can manage account and set system in the Management Center.

Select **Account Management** > **Manage Members**, you can add a member, delete members batchly, synchronize account system and search for a member.

Select **System Settings** > **Computation Settings**, you can configure computing engine, cluster and metadata, and you can also verify the Endpoint. For more information, see *Management Center*.

27.3.4 Data warehouse planning

Data warehouse planning is a key step of top-level design in data creation.

To complete this step, define logical spaces (business sections, data domains, business spaces, and public definitions) based on business features. Then, create physical projects based on the development cooperation and management model. For the physical projects, set project and member information, and register the required bottom-layer data sources (physical databases). For more information, see *Data warehouse planning*.

27.3.5 Data modeling and development

Dataphin provides systematic modeling and development functions to achieve in-depth application and development of the data warehouse theory with tools.

It can quickly create enterprise dimensions and business processes in a top-down mode, and refine the development dimension tables, fact tables, summary tables, and application layer to develop a public layer of enterprise data. This facilitates data application in a business hierarchy and optimizes the computing and storage functions. For more information, see *Data modeling and development*.

27.3.6 O&M center

The O&M center completes operation and maintenance for all tasks.

Rule-based and ordered submission of R&D code relies on a publish scheduling system. The O&M center completes operation and maintenance for all tasks. Its core function is to provide directed acyclic graphs (DAGs) of task scheduling, which help you obtain and adjust task

execution progresses to improve stability of data creation. After data modeling and distillation, the coding module can automatically control the O&M system.

The coding module automatically generates globally optimal code to reduce the coding cost. You only need to determine the computing logic and do not need to worry about storage and computing. On the O&M center, you can see the script and code submission tasks configured by yourself and the code tasks created by the system. For more information, see *O&M center*.

27.3.7 Data assets

Dataphin can check and assess the data assets of an enterprise according to the standard and methodology of enterprise data asset management.

The data asset module can manage all data and APIs created by modeling, development, and distillation. You can find and use the data and APIs using the semantic search or data catalog function. The data asset module can provide a data asset dashboard to help executives of an enterprise discover and understand the value of data assets. It also supports automatic metadata extraction and analysis. This module provides data asset check and analysis throughout the entire data creation chain covering data computing, storage, security, and application. When problems are found, it provides optimization solutions for data governance and implements the solutions intelligently, to help enterprises reduce data maintenance cost and improve data analysis efficiency effectively.

The data asset module displays created data in graphs and detailed data tables. For more information, see *Data assets*.

27.3.8 Data service

The data service supports real-time query of data produced during development or in the computing engine, and Presto-based interactive and hybrid storage query.

It is worth noting that, you can drill down in the logical tables created by standardized data modeling and distillation following the recommendations provided by the code editor to obtain data quickly. The drill-down can be performed in a star or snow flake schema. You can also use SQL queries to obtain all data fields if there are no redundant model fields.

This function can be implemented using temporary query for modeling and development . For more information, see *Data service*.

27.4 Management Center

System administrators can do **Account Management** and **System Settings** through **Management Center**, which is a prerequisite for using Dataphin. This section will show you how to set data warehouse and computing type by using operation and maintenance super administrator account and super administrator account.

For more information about operation and maintenance super administrator and super administrator, see *Notice for system administrators*.

27.4.1 Members Management

You can search, add, delete a member and synchronize the account system with the member management function.

Procedure

1. *Log on to the Dataphin console*.
2. Select **Management Center** > **Manage Members**.
3. Enter an account name or nickname in the search box on the right side of the page to search for a member.
4. Click the **Add Member** to add a member.
5. Select one or more members, click **Delete** to delete one or more members.

27.4.2 Configure computing type

After the metadata initialization is complete, you can use the super administrator account to log on to the Dataphin console for the calculation type setting. This section takes the computing type MaxCompute as an example.

Procedure

1. Using the super administrator account to *Log on to the Dataphin console*. For more information about super administrator account, see *Notice for system administrators*.
2. Select **Management Center** > **Computation Settings**.
3. Enter **Endpoint** and click **Verify**.
4. After the verification is successful , click **Confirm and Start Data Constrution**.

   After the computation type is set, you can continue to configure computing engine using the super administrator account, for more information, see *Computing engine source*.

27.5 Data warehouse planning

Data warehouse planning is a key step of top-level design in data creation. To complete this step, define logical spaces (business sections, data domains, business spaces, and public definitions) based on business features. Then, create physical projects based on the development cooperation and management model. For the physical projects, set project and member information, and register the required bottom-layer data sources (physical databases).

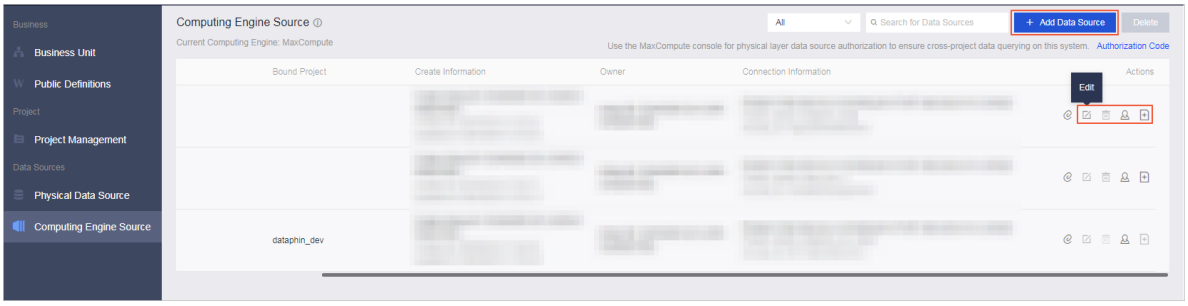
27.5.1 Computing engine source

Context

Currently, Dataphin does not support computing between heterogeneous databases. Therefore, you must specify a computing type for the entire system to ensure transfer of underlying data and applications.

Procedure

1. *Log on to the Dataphin console* as the super administrator.
2. Choose **Intelligent Data Warehouse Planning** > **Computing Engine Source**, and select a computing engine and click the **Edit** icon.
3. Click **Add Data Source**, and enter the basic information of the data source to add a data source.
4. Click the icon buttons below the **Actions**, you can **Connection Test**, **Edit**, **Delete**, **Transfer Owner**, and **Create Project** for a data source.



27.5.2 Physical data source

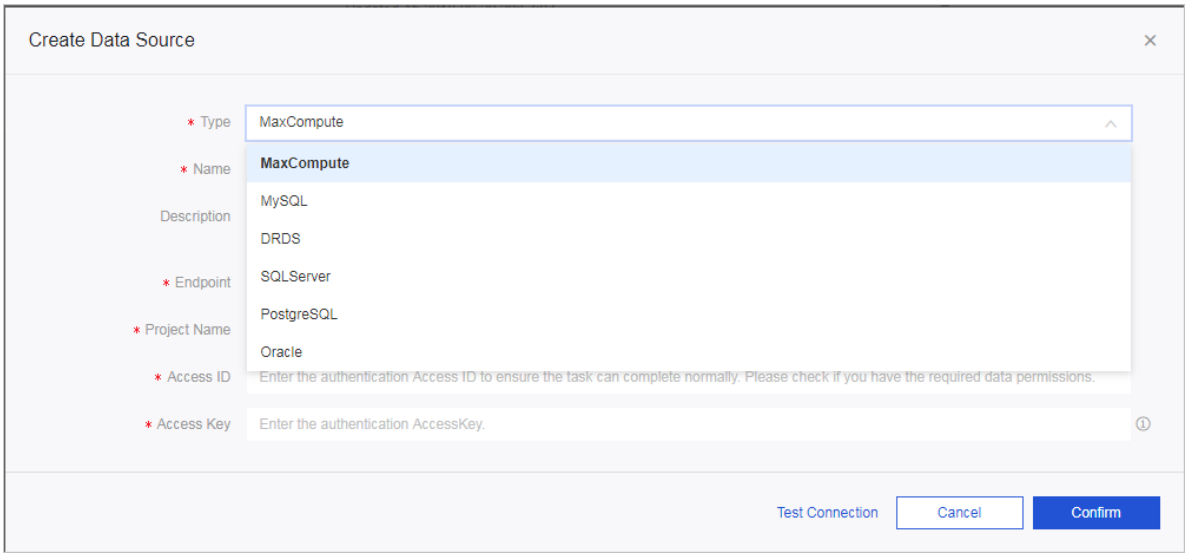
Register physical databases as bottom-layer sources of data for the project and origins of data synchronization.

Procedure

1. *Log on to the Dataphin console*, and select **Intelligent Data Warehouse Planning** > **Physical Data Source**.

2. Click **Add Data Source**.
3. Select **MaxCompute** as the **Type**, as shown in the *Figure 27-2: Data source type*.

Figure 27-2: Data source type



4. Enter the name and description of the data source. The name can be the same as the project name.
5. If data synchronization is required, repeat the preceding steps.

   To import data, determine the data required based on the big data top-level design, and then use tools to complete data collection, cleaning, structured conversion, integration, and synchronization. Dataphin supports data synchronization between databases used as data sources on the Development Workbench.

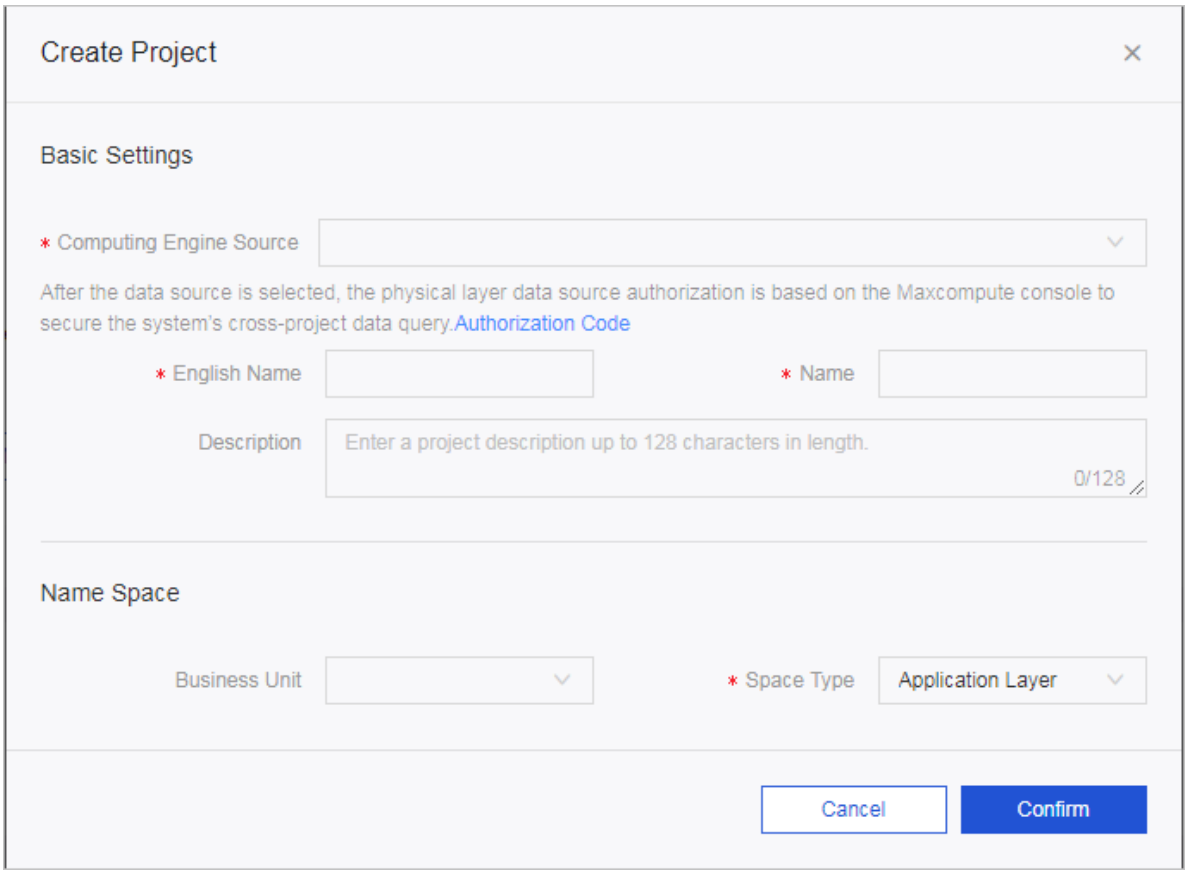
   Now, you have finished the first step for the project , that is global data planning.

27.5.3 Project management

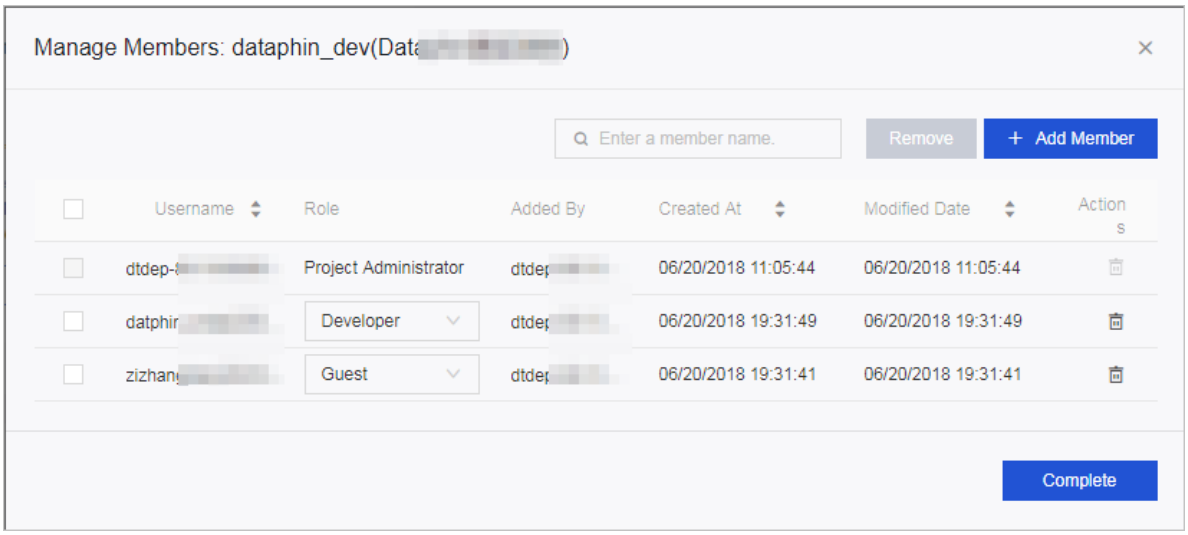
For project management, define physical spaces to isolate physical resources and group developers. After setting names for physical spaces, you can start data modeling and development.

Procedure

1. *Log on to the Dataphin console* and select**Intelligent Data Warehouse Planning** > **Project Management**.
2. Click **Create Project**.
3. Enter basic information about the project, including its name and description.



4. Click **Manage Members**, specify project members, and set roles for the members to control their operation permissions in the project.



27.5.4 Public definitions

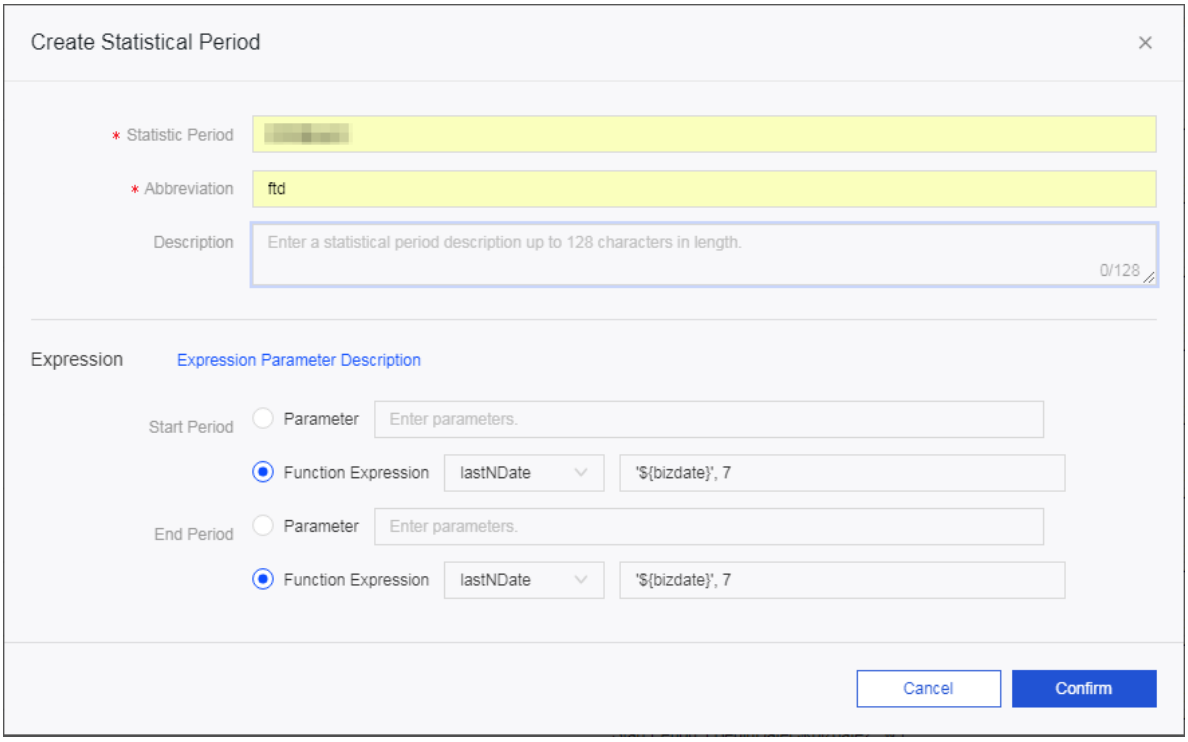
Public definitions ensure global consistency of object definitions and enable reference of objects.

Procedure

1. *Log on to the Dataphin console* and select **Intelligent Data Warehouse Planning** > **Public Definitions**.
2. Click **Create Statistical Period**.
3. Set the start and end dates, name, and expression of the statistic period, as shown in *Figure 27-3: Statistic period*.

   The system has initialized the statistic periods for commonly used indicators, and the names and expressions of these statistical periods.

   Figure 27-3: Statistic period



27.5.5 Business unit

Context

Define the data warehouse name and business space. In this use case, the group is engaged in retail business and does not require much isolation between systems; therefore, only one business section needs to be created.

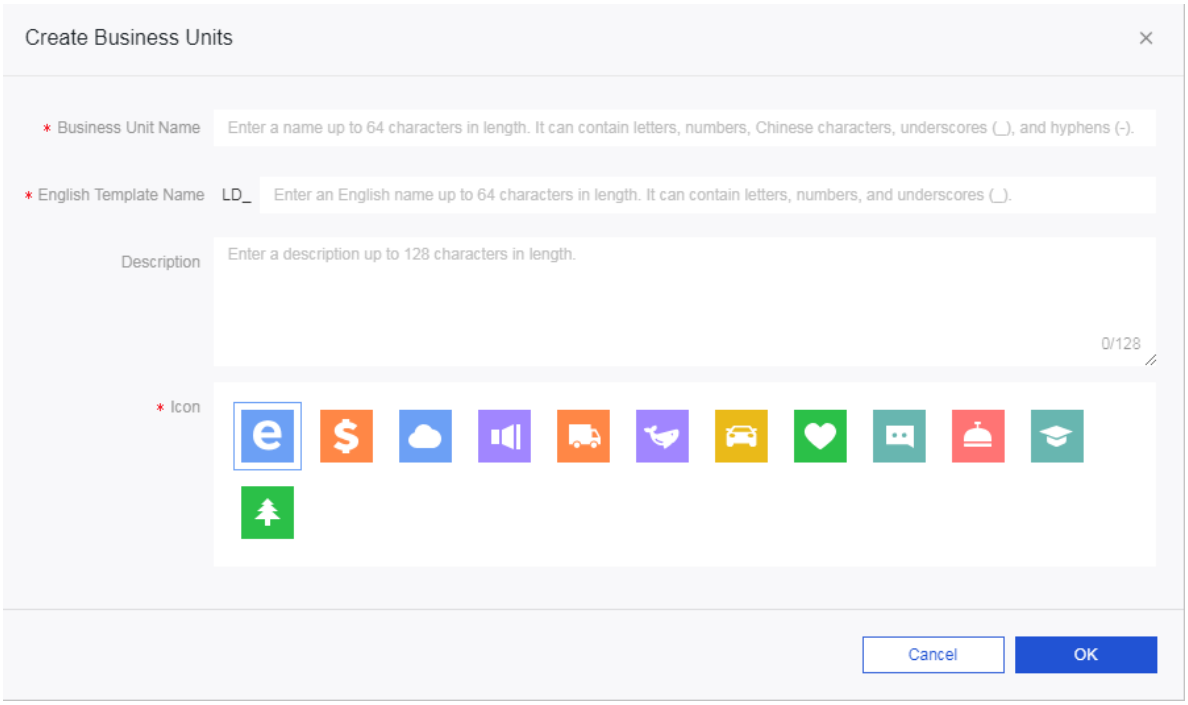
Procedure

1. *Log on to the Dataphin console* and select **Intelligent Data Warehouse Planning** > **Business Unit**.
2. Click **Create Business Units**.

3. Set the name and description of the business section, as shown in *Figure 27-4: Business section information*.

> **Note:**
> We recommend that you not change the prefix of the business section name.

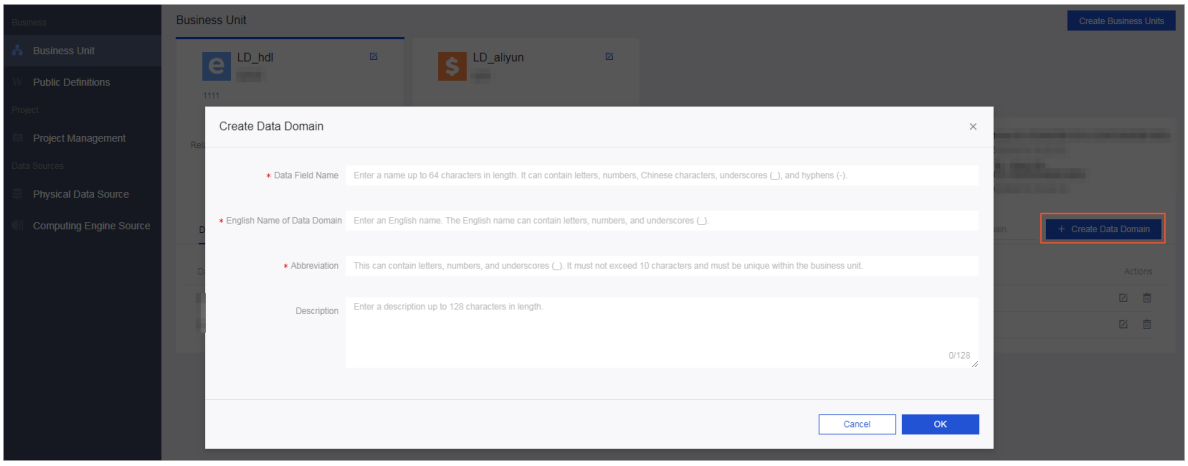
Figure 27-4: Business section information



4. Select a business name in the business unit, for example, **`LD_retail`**.
5. Click **Create Data Domain** to create a new data domain.

   Data domains are used to classify data in a business section for data management.

Figure 27-5: Data domains



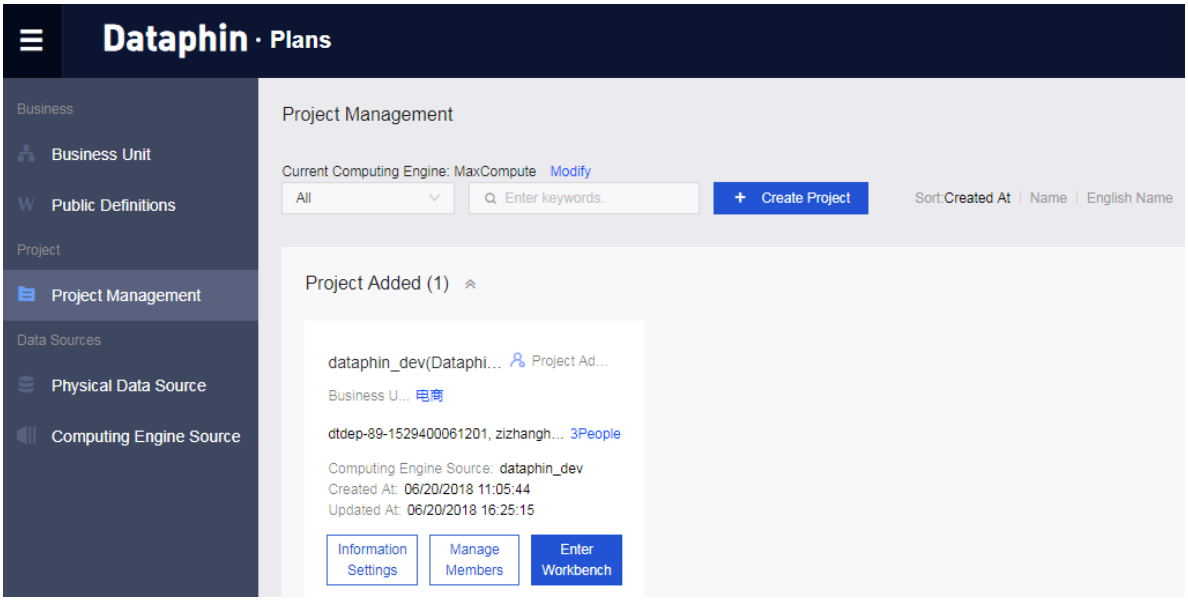
27.6 Data modeling and development

Dataphin provides systematic modeling and development functions to achieve in-depth application and development of the data warehouse theory with tools. It can quickly create enterprise dimensions and business processes in a top-down mode, and refine the development dimension tables, fact tables, summary tables, and application layer to develop a public layer of enterprise data. This facilitates data application in a business hierarchy and optimizes the computing and storage functions.

Procedure

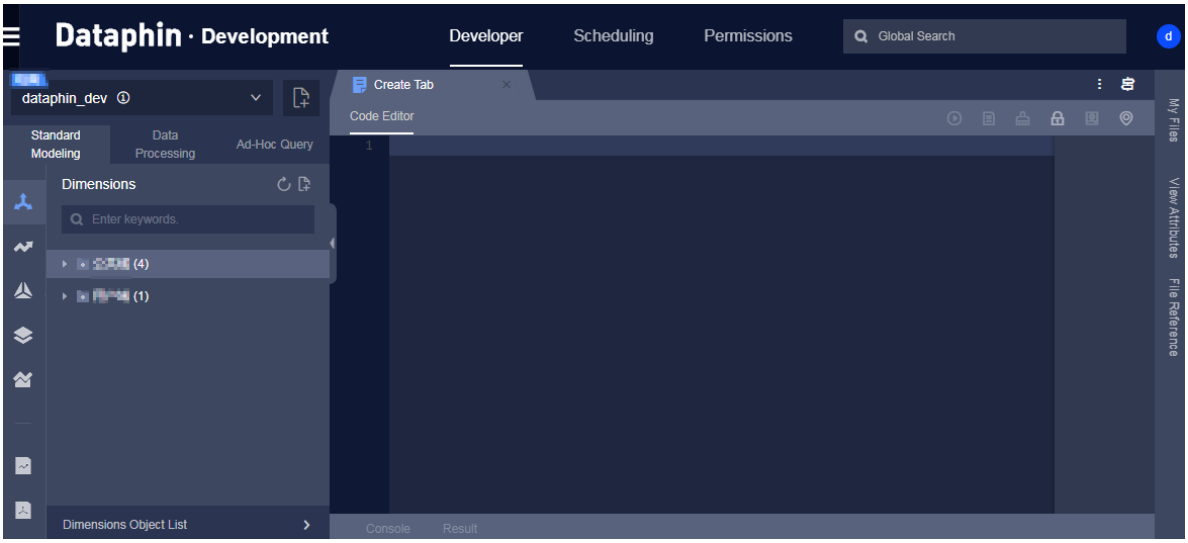
1. *Log on to the Dataphin console*.
2. On the Dataphin homepage, click **Plans**.
3. Click **Project Management** and select a project, such as retail, as shown in *Figure 27-6: Project management*.

Figure 27-6: Project management



4. Choose **Enter Workbench** > **Developer**to start project development, as shown in*Figure 27-7: Smart development*.

Figure 27-7: Smart development



27.6.1 Standard definition - dimension

A dimension is a statistical object, which is generally an existing entity. By creating a dimension, you can standardize a business subject (master data) at the top level and ensure uniqueness of the entity.

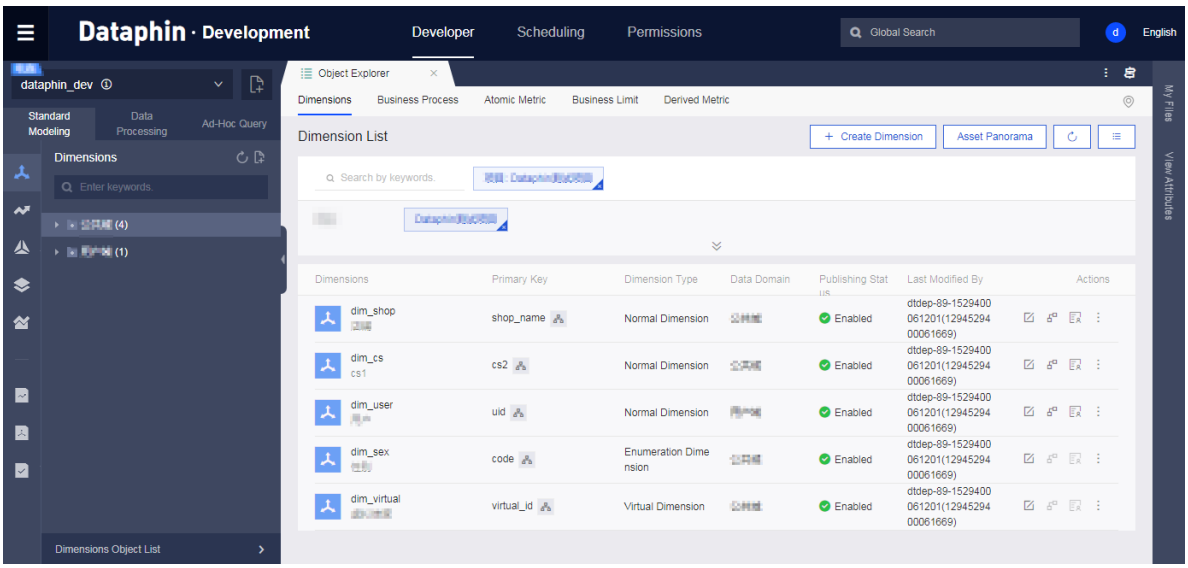
Context

The definitions should be made based on business development to avoid changes in the future.

Procedure

1. *Log on to the Dataphin console*.
2. On the Dataphin homepage, click **Development**.
3. Choose **Developer** > **Standard Modeling** > **Dimensions** > **Dimensions Object List**, as shown in *Figure 27-8: Dimension list*.

Figure 27-8: Dimension list



4. Click **Create Dimension** to create a dimension, such as a member or product dimension.
5. Click **Edit** to edit dimension information, including its name, type, primary key, computing logic, and parent-child relation.

27.6.2 Standard definition - business process

A business process is a collection of all events in a business activity. By creating a business process, you can standardize a type of transaction events in business and ensure its uniqueness.

Context

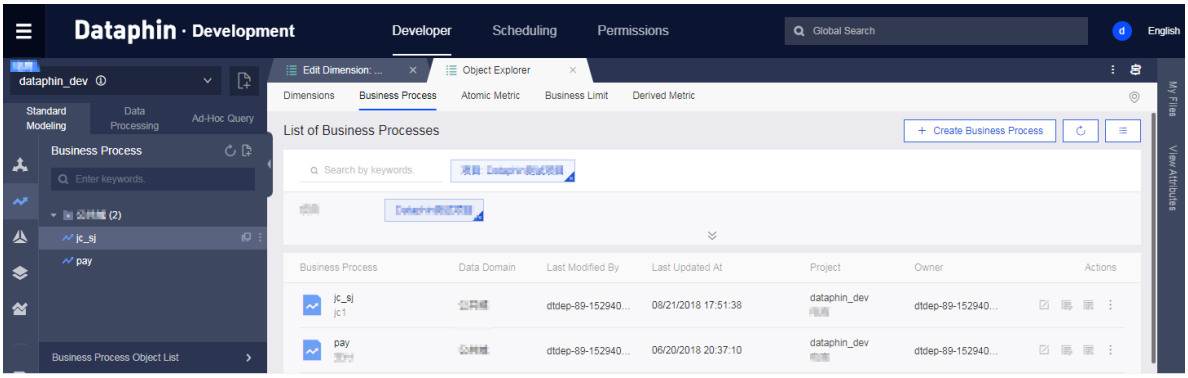
The definitions should be made based on business development to avoid changes in the future.

Procedure

1. *Log on to the Dataphin console*.
2. On the Dataphin homepage, click **Development**.

3. Choose **Developer** > **Standard Modeling** > **Business Process** > **Business Process Object List**, as shown in *Figure 27-9: Business process list*.

Figure 27-9: Business process list



4. Click **Create Business Process** to create a business process, such as transaction payment, order placement, or return of goods.
5. Click **Edit** to edit the business process information, including its name.

27.6.3 Logical table - dimension logical table

Each dimension has a dimension logical table to describe attributes of the dimension. Dimension logical tables are used to filter public object detailed data, so that detailed data of objects can be retrieved from business data.

Context

Abstract the inherent and persistent attributes of a dimension without redundant information.

Procedure

1. *Log on to the Dataphin console*.
2. On the Dataphin homepage, click **Development**.
3. Choose **Developer** > **Standard Modeling** > **Logical Table** > **Dimension Logical Table**.
4. Click **Edit** to edit the logical table, as shown in *Figure 27-10: Dimension logical table*.

Figure 27-10: Dimension logical table

27.6.4 Logical table - fact logical table

The fact logical table of a business process provides detailed information about the business process. Fact logical tables are used to filter public transaction detailed data, so that detailed data of transactions can be retrieved from business data.

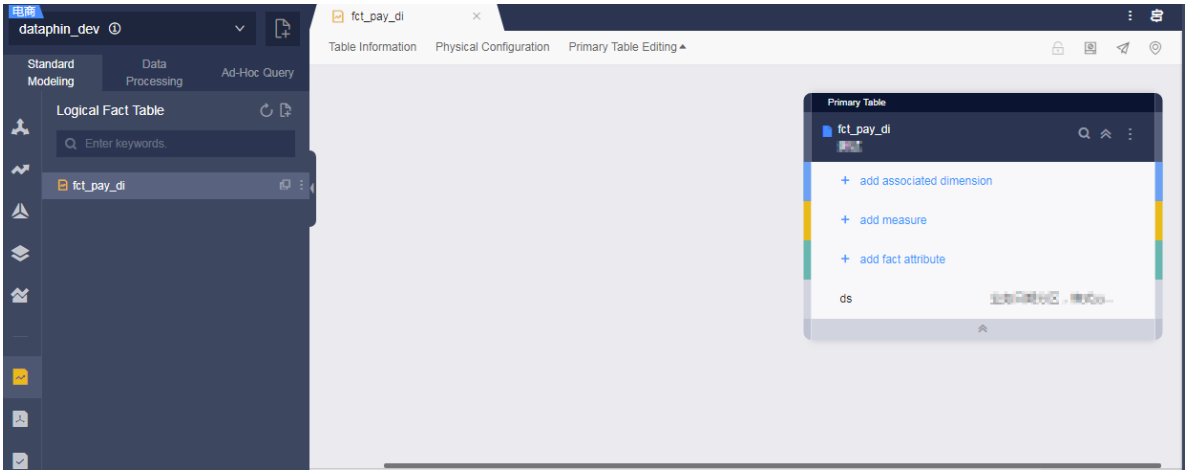
Context

Abstract descriptive information about transactions. Basic information in a fact logical table can be modified by adding supplementary information such as primary table content and filtering logic to the business process information.

Procedure

1. *Log on to the Dataphin console*.
2. On the Dataphin homepage, click **Development**.
3. Choose **Developer** > **Logical Table** > **Logical Fact Table**, as shown in *Figure 27-11: Fact logical table*.

Figure 27-11: Fact logical table



4. Create a new fact logical table or edit an existing one as required.

   Now, you have finished the first half of the second step – general detailed data model standardization.

27.6.5 Standard definition - atomic metric and business limit

An atomic metric and business limit are the computing logic and attributive limitation commonly used in business. They are expressions formulated based on fields in a logical table to abstract universal data among summarized data and enable reuse of the universal data.

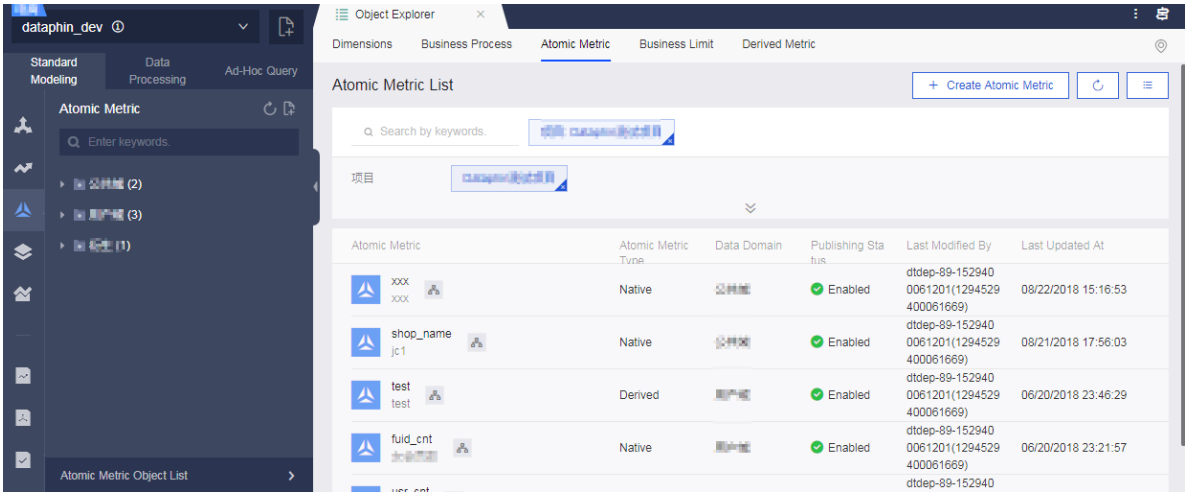
Context

Abstract commonly used atomic computing logic and limitation conditions to fully reuse universal data.

Procedure

1. *Log on to the Dataphin console*.
2. On the Dataphin homepage, click **Development**.
3. Choose **Developer** > **Standard Modeling** > **Atomic Metric**, as shown in *Figure 27-12: Atomic metric list*.

Figure 27-12: Atomic metric list



4. Click**Create Atomic Metric**to create an atomic metric.
5. Choose **Business Limit** > **Create Business Limit**to create a new business limit.

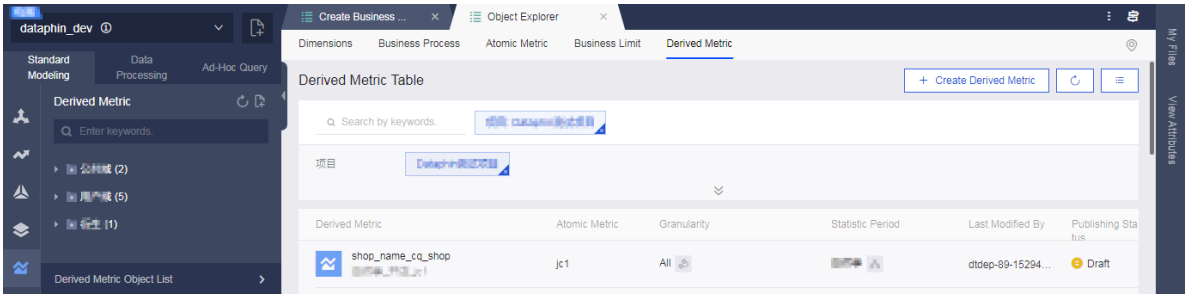
27.6.6 Standard definition - derived metric

A derived metric is a commonly used statistical metric, that is, summarized data of an object group in a certain range produced during a statistical period. Therefore, a derived metric is defined by the statistical period (time cycle), statistical object (statistics granularity), statistical range (business limitation), and statistical method (atomic metric). After specifying the preceding factors, you only need to edit and confirm the name of the derived metric. To meet the business requirement in this use case, define the coupon payment amount of members in a natural quarter as a metric, and add other conditions as required.

Procedure

1. *Log on to the Dataphin console*.
2. On the Dataphin homepage, click **Development**.
3. Choose **Developer** > **Standard Modeling** > **Derived Metric**, as shown in *Figure 27-13: Derived Metric*.

Figure 27-13: Derived Metric



4. Click **Create Derived Metric**to create a derived metric.
5. Click **Edit**to edit the derived metric.

27.7 O&M center

Rule-based and ordered submission of R&D code relies on a publish scheduling system. The O&M center completes operation and maintenance for all tasks. Its core function is to provide directed acyclic graphs (DAGs) of task scheduling, which help you obtain and adjust task execution progresses to improve stability of data creation. After data modeling and distillation, the coding module can automatically control the O&M system.

The coding module automatically generates globally optimal code to reduce the coding cost. You only need to determine the computing logic and do not need to worry about storage and computing. On the O&M center, you can see the script and code submission tasks configured by yourself and the code tasks created by the system.

27.7.1 Node

A node can be any object that is launched after submission of code or a script, and can automatically run at specified cycles or be triggered manually. It is also called a task.

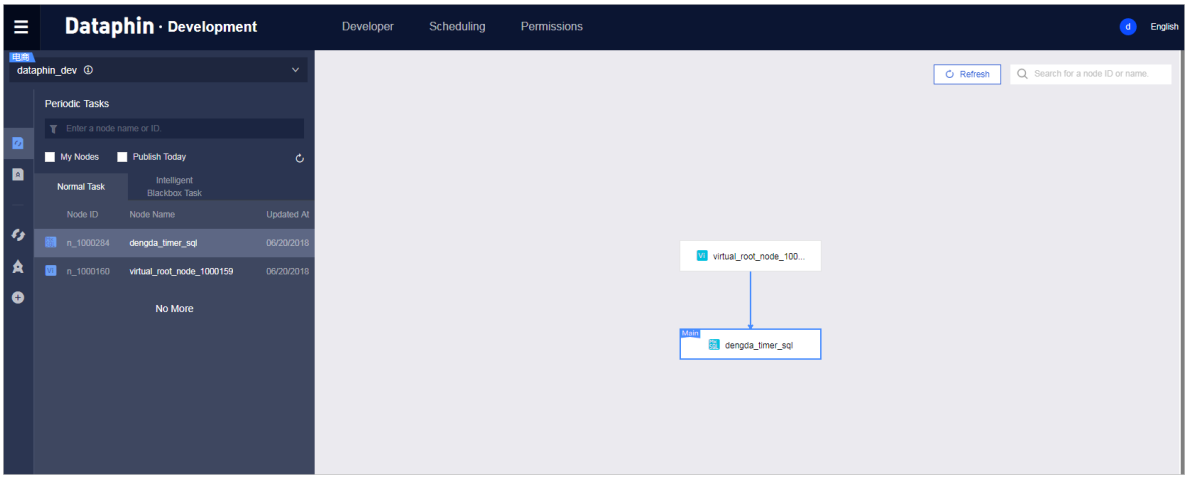
Context

Dataphin can automatically create tasks through data modeling and distillation, and provide intelligent maintenance of tasks.

Procedure

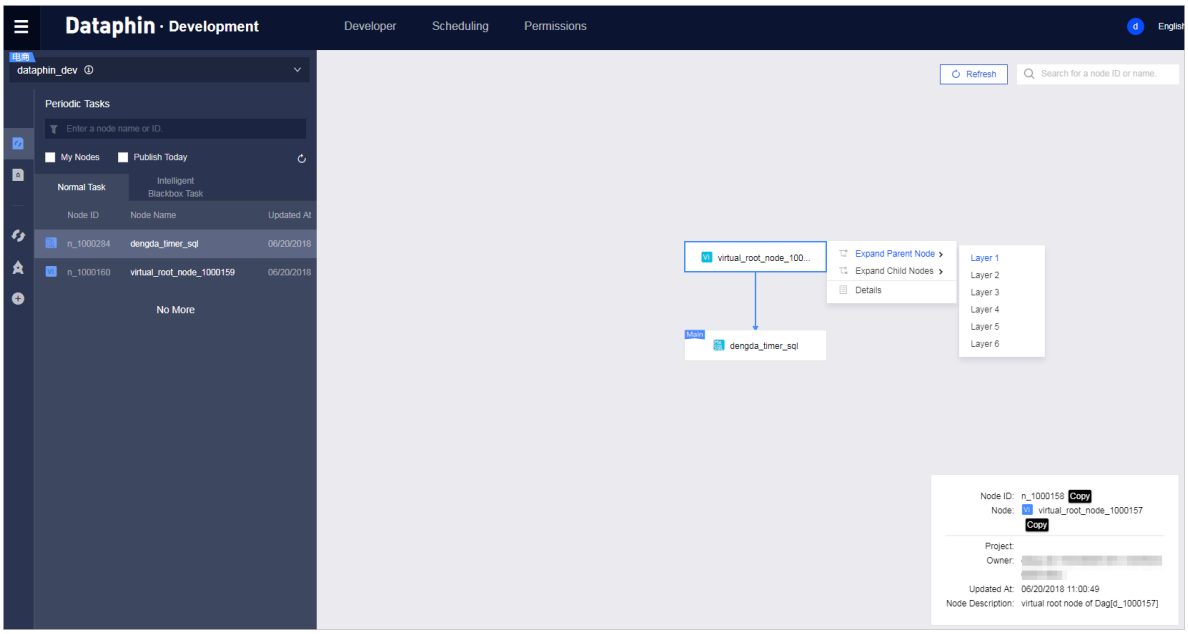
1. *Log on to the Dataphin console*.
2. Click **O&M Scheduling**, and select a project to check its nodes. You can select a node to view its running logic graph, as shown in *Figure 27-14: O&M center*.

Figure 27-14: O&M center



3. Select a node and right-click it to edit the node information, as shown in *Figure 27-15: Edit a node*.

Figure 27-15: Edit a node



27.7.2 Instance

A periodic node generates an instance every time it runs in a cycle. You can also create a data population instance for a node by supplementing data of specified dates to trigger passive data. Manual nodes on the R&D workbench are triggered by emails and create instances when they are running. An instance is a dynamic node with a running state.

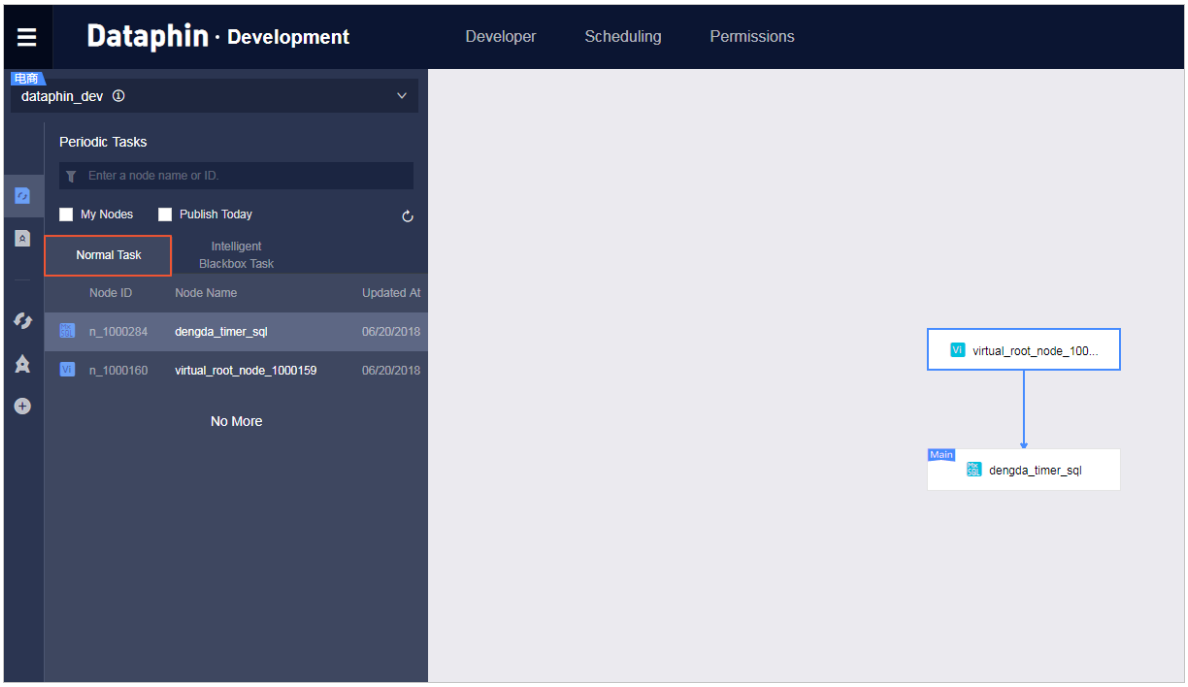
Context

Dataphin can automatically create tasks through data modeling and distillation, and provide intelligent O&M of tasks.

Procedure

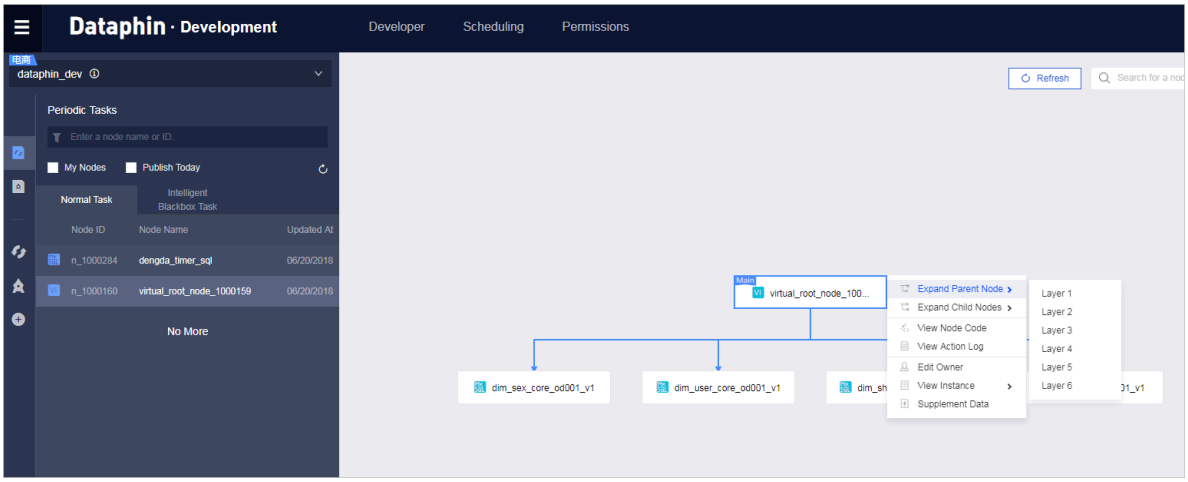
1. *Log on to the Dataphin console*.
2. Click **O&M Scheduling**, and select a project to check its instances. You can select an instance to view its running logic graph, as shown in *Figure 27-16: Instance*.

   Figure 27-16: Instance



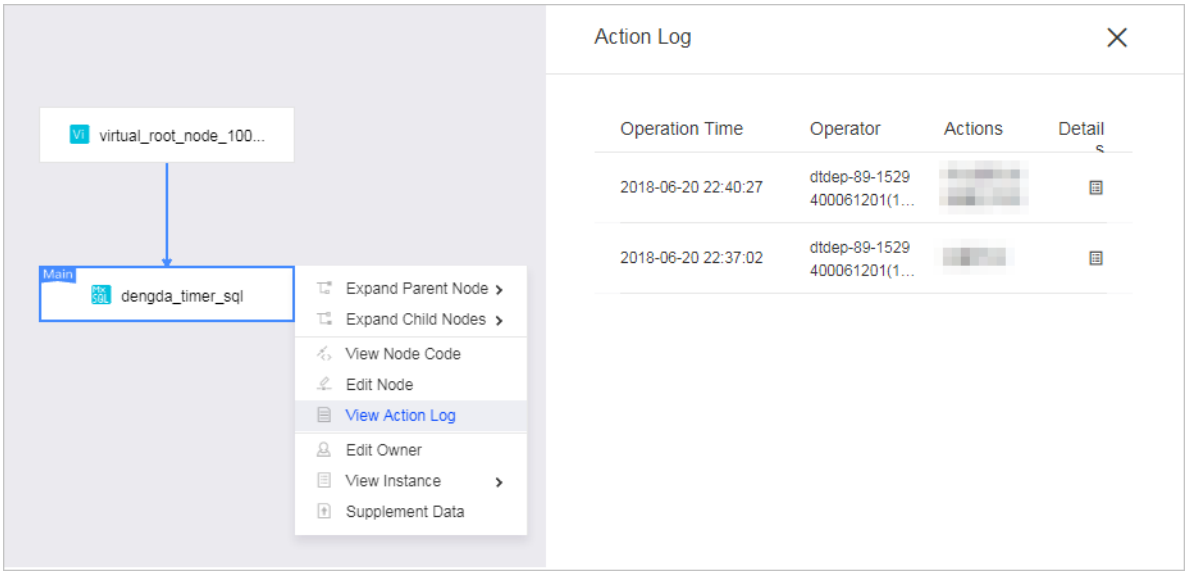
3. Select an instance and right-click it to edit the instance information, as shown in *Figure 27-17: Edit an instance*.

Figure 27-17: Edit an instance



You can view operation logs to check the code created through data modeling or distillation and the code running status, as shown in *Figure 27-18: Operation log*.

Figure 27-18: Operation log



27.8 Data assets

Dataphin can check and assess the data assets of an enterprise according to the standard and methodology of enterprise data asset management.

The data asset module can manage all data and APIs created by modeling, development, and distillation. You can find and use the data and APIs using the semantic search or data catalog function. The data asset module can provide a data asset dashboard to help executives of an

enterprise discover and understand the value of data assets. It also supports automatic metadata extraction and analysis. This module provides data asset check and analysis throughout the entire data creation chain covering data computing, storage, security, and application. When problems are found, it provides optimization solutions for data governance and implements the solutions intelligently, to help enterprises reduce data maintenance cost and improve data analysis efficiency effectively.

The data asset module displays created data in graphs and detailed data tables.

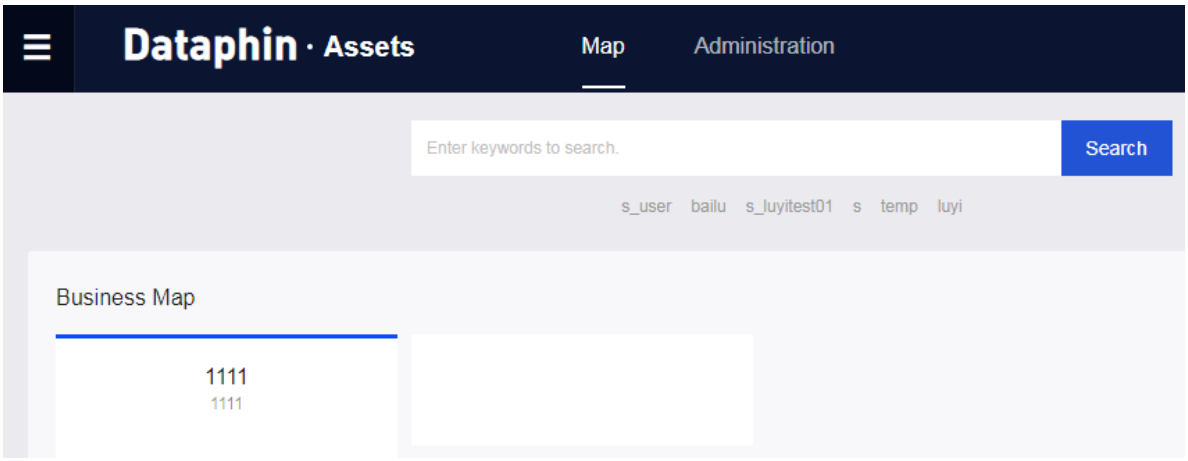
27.8.1 Map

A data asset map provides data catalogs and knowledge graphs through standard data creation, helping you quickly find required data.

Procedure

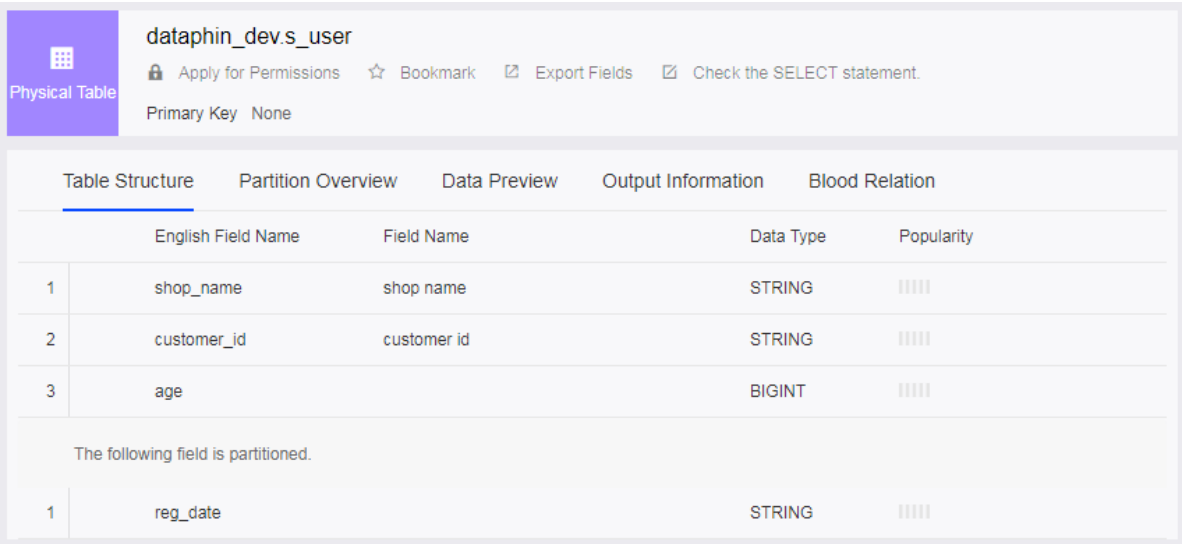
1. *Log on to the Dataphin console*.
2. On the Dataphin homepage, click **Assets**.
3. Click **Map**, and search for a data table or select it from the data catalog navigation, as shown in *Figure 27-19: Data table*.

   Figure 27-19: Data table



4. Click the table name to view the details, including the table structure and metadata, as shown in *Figure 27-20: Metadata*.

Figure 27-20: Metadata



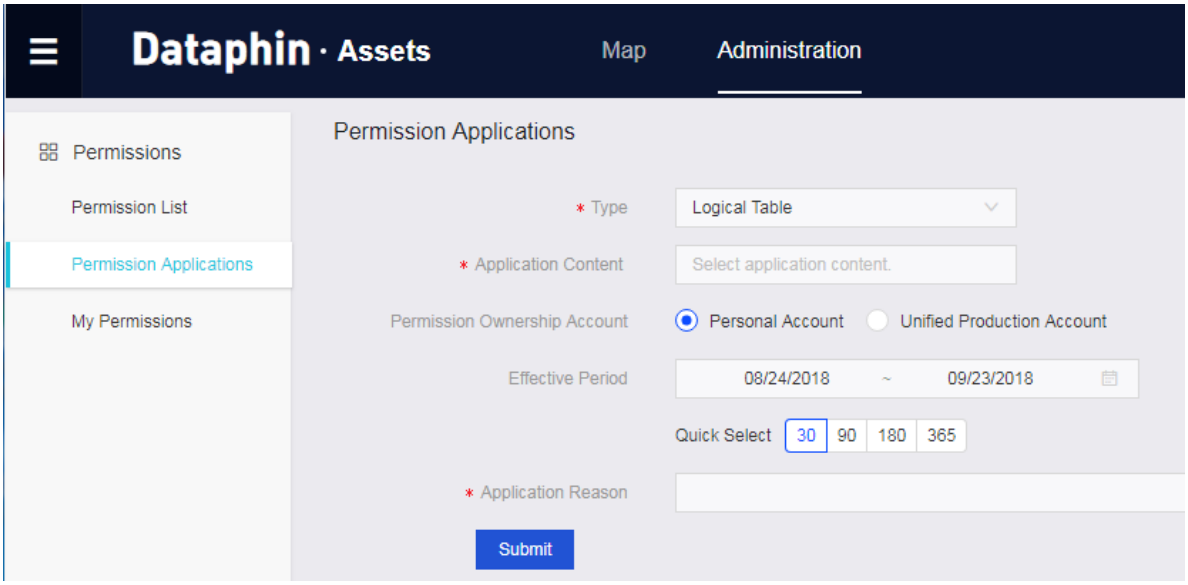
27.8.2 Management

All data resources are created in projects. To guarantee secure data use, you must apply for certain data use permission. Your application must specify the fields to use. You can use the related data only after your application is approved.

Procedure

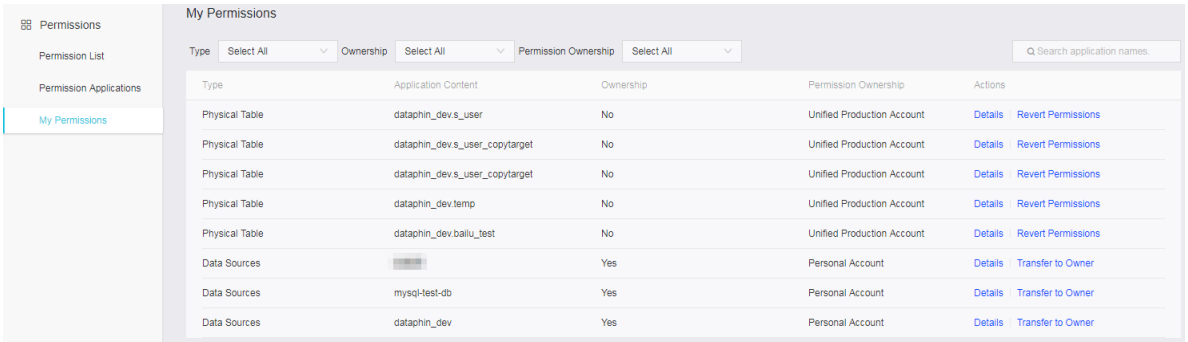
1. *Log on to the Dataphin console*, and click **Assets**.
2. Choose **Administration** > **Permission Applications**.
3. Apply for the permission to use data in the payment logical table and select the validity period of the permission, as shown in *Figure 27-21: Permission application*.

Figure 27-21: Permission application



After submitting the application, you can see this logical table in **My Permissions**. The application needs to be approved by the project administrator displayed in the record of the logical table. After your application is approved, you can view and use the data resources in the logical table, as shown in *Figure 27-22: My Permissions*.

Figure 27-22: My Permissions



27.9 Data service

The data service supports real-time query of data produced during development or in the computing engine, and Presto-based interactive and hybrid storage query.

It is worth noting that, you can drill down in the logical tables created by standardized data modeling and distillation following the recommendations provided by the code editor to obtain data quickly. The drill-down can be performed in a star or snow flake schema. You can also use SQL queries to obtain all data fields if there are no redundant model fields.

This function can be implemented using temporary query for modeling and development .

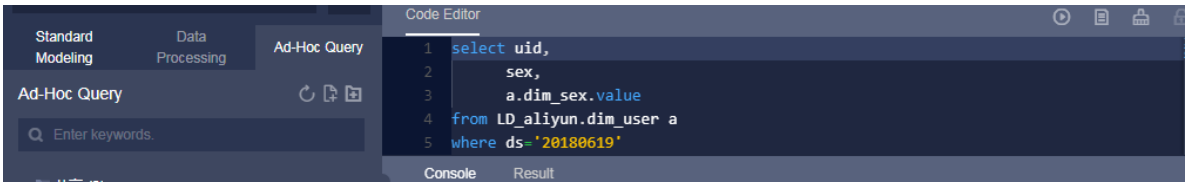
27.9.1 SQL query

The SQL query function enables you to quickly obtain required data.

Procedure

1. *Log on to the Dataphin console*.
2. On the Dataphin homepage, click **Development**.
3. Choose **Ad-Hoc Query** and you can create your query file.
4. Enter the query code to view payment details of members in the payment fact detailed table or obtain all member data including payment amounts and preferences of stars from the member summary logical table, as shown in *Figure 27-23: Member data*.

   Figure 27-23: Member data



5. After execution of the code, click **Result** to view the query result.