ALIBABA CLOUD

Alibaba Cloud

Apsara Stack Enterprise

Elastic Compute Service

User Guide

Product Version: v3.16.2

Document Version: 20220913

Alibaba Cloud

Legal disclaimer

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

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

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

3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.

4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides this document based on the "status quo", "being defective", and "existing functions" of its products and services. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, special, or punitive damages, including lost profits arising from the use or trust in this document (even if Alibaba Cloud has been notified of the possibility of such a loss).

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

6. Please directly contact Alibaba Cloud for any errors of this document.

Document conventions

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

Table of Contents

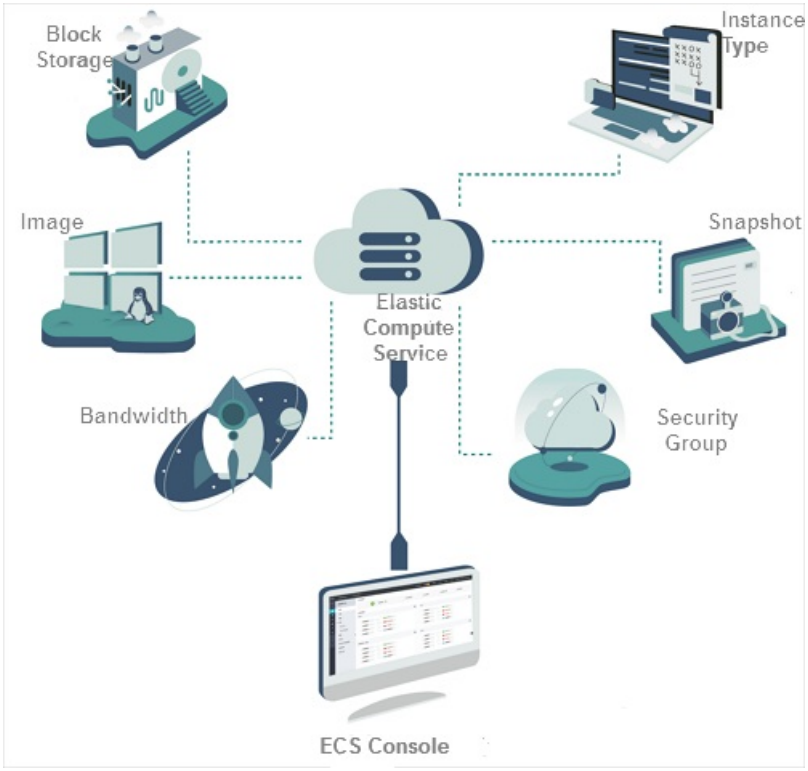
1.What is ECS?

1.1. Overview

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

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

ECS components



1.2. Instance lifecycle

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

The following table describes the instance states in the ECS console and their corresponding instance states in API responses.

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

The following figure shows the transitions between instance states in API responses.

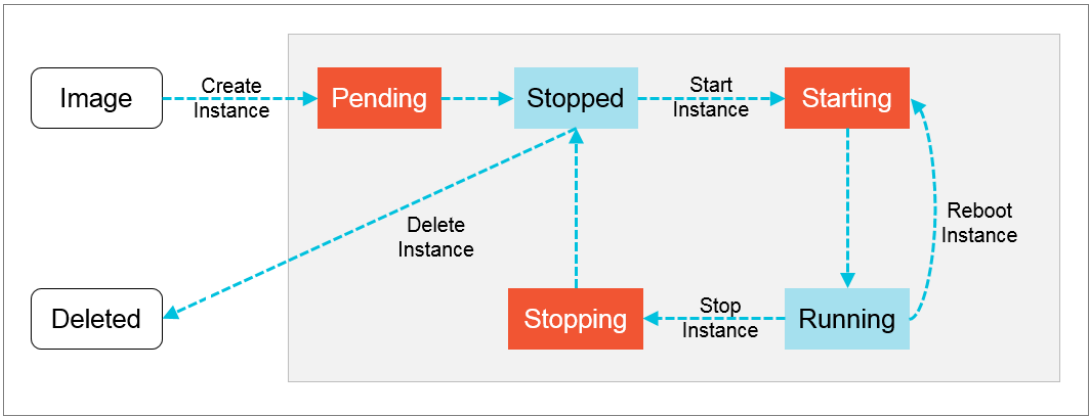


Image
Create Instance
Pending
Stopped
Start Instance
Starting
Reboot Instance
Delete Instance
Deleted
Stopping
Stop Instance
Running

2.Instructions

2.1. Restrictions

Before you perform operations on Elastic Compute Service (ECS) instances, learn about the following restrictions:

- Do not arbitrarily upgrade the kernel or operating system versions of instances.
- Do not enable SELinux on Linux systems except for CentOS and Red Hat.
- Do not uninstall PV drivers.
- Do not arbitrarily modify the media access control (MAC) addresses of network interface controllers (NICs).

2.2. Suggestions

Consider the following suggestions to use Elastic Compute Service (ECS) more efficiently:

- ECS instances with 4 GiB or higher memory must use a 64-bit operating system. A 32-bit operating system can be used for instances with a maximum of 4 GiB of memory.
- A 32-bit Windows operating system can support up to 4 CPU cores.
- To ensure service continuity and prevent service unavailability due to failovers, we recommend that you configure service applications to start automatically on system startup.

2.3. Limits

When you use Elastic Compute Service (ECS) instances, be aware of the following limits on instance families.

General limits

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

Instance family ga1

To create a ga1 instance, use one of the following images for which drivers are pre-installed:

- Ubuntu 16.04 for which an AMD GPU driver is pre-installed
- Windows Server 2016 English edition for which an AMD GPU driver is pre-installed
- Windows Server 2008 R2 English edition for which an AMD GPU driver is pre-installed

Take note of the following items:

- A ga1 instance uses an optimized driver provided by Alibaba Cloud and AMD. The driver is installed in images provided by Alibaba Cloud and is currently unavailable for download.

- If the GPU driver malfunctions due to improper removal of related components, you must replace the system disk to restore GPU related features.

  > **Note** This operation results in data loss.

- If the driver malfunctions because an improper image is selected, you must replace the system disk to reselect an image for which an AMD GPU driver is pre-installed.
- For Windows Server 2008 or earlier, you cannot connect to the VNC after the GPU driver takes effect. The VNC does not respond and a black screen appears or the system becomes stuck on the splash screen. You can use other methods such as Remote Desktop Protocol (RDP) to access the system.
- RDP does not support DirectX, OpenGL, or other related applications. You must install the VNC and a client, or use other supported protocols such as PCOIP and XenDesktop HDX 3D.

Instance families gn4, gn5i, and gn5

- Bandwidth: If you use a Windows Server 2008 R2 image for a gn4 instance, you cannot use the Connect to VNC feature in the ECS console to connect to the instance after the installed GPU driver takes effect. You must set the bandwidth to a non-zero value or attach an elastic IP address (EIP) to the created instance.
- Image: If an NVIDIA GPU driver is not required, you can select any image based on your business requirements, and then install the CUDA and GPU drivers for the instance. For more information, see Install the CUDA and GPU drivers for a Linux instance or Install the CUDA and GPU drivers for a Windows instance.

2.4. Notice for Windows users

Before using Windows-based ECS instances, you must consider the following points:

- Data loss may occur if a local disk is used as the data disk of an instance. We recommend that you use a cloud disk to create your instance if you are not sure about the reliability of the data architecture.
- Do not close the built-in shutdownmon.exe process. Otherwise, the server may require a longer time to restart.
- Do not rename, delete, or disable Administrator accounts or it may affect the use of the server.
- We do not recommend that you use virtual memory.
- When you modify your computer name, you must synchronize the following key values in the registry. Otherwise, the computer name cannot be modified, causing failure when installing 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
  ```

2.5. Notice for Linux users

Before using Linux-based ECS instances, you must consider the following points:

- Do not modify content of the default /etc/issue files under a Linux instance. Otherwise, the custom image created from the instance cannot be recognized, and instances created based on the image cannot start as expected.
- Do not arbitrarily modify the permissions of each directory in the partition where the root directory is

located, especially permissions of /etc, /sbin, /bin, /boot, /dev, /usr, and /lib directories. Improper modification of permissions can cause errors.

- Do not rename, delete, or disable Linux root accounts.
- Do not compile or perform any other operations on the Linux kernel.
- We do not recommend the use of Swap for partitioning.
- Do not enable the NetWorkManager service. This service conflicts with the internal network service of the system, causing network errors.

2.6. Notice on defense against DDoS attacks

You need to purchase Anti-DDoS Pro to defend against DDoS attacks. For more information, see *Apsara Stack Security Product Introduction*.

3.Quick start

3.1. Overview

This topic describes how to create and connect to an Elastic Compute Service (ECS) instance.

Perform the following operations:

1. Create a security group

   A security group acts as a virtual firewall to control inbound and outbound traffic for ECS instances. Each ECS instance must belong to at least one security group. When you create an instance, you must select a security group for the instance.

2. Create an instance

   An ECS instance is a virtual machine that contains the basic computing components of a server, such as CPU, memory, operating system, network configurations, and disks. After a security group is created, you can select an instance type based on your business needs to create instances in the security group.

3. Instance connecting overview

   Select a method to connect to an instance based on the network configurations and operating system of the instance and the operating system of your computer. After you connect to the instance, you can perform operations on it such as installing applications.

3.2. Log on to the ECS console

This topic describes how to log on to the Elastic Compute Service (ECS) console.

Prerequisites

- Before you log on to the Apsara Uni-manager Management Console, the endpoint of the console is obtained from the deployment staff.
- We recommend that you use Google Chrome.

Procedure

1. Enter the URL of the Apsara Uni-manager Management Console in the address bar and press the Enter key.
2. Enter your username and password.

   Obtain the username and password that you use to log on to the Apsara Uni-manager Management Console from the operations administrator.

> **Note** The first time that you log on to the Apsara Uni-manager Management Console, you must change the password of your account. For security purposes, your password must meet the minimum complexity requirements. The password must be 10 to 32 characters in length and must contain at least two of the following character types:
>
> - Uppercase or lowercase letters
> - Digits
> - Special characters, including exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)

3. Click **Log On**.
4. If your account has multi-factor authentication (MFA) enabled, perform corresponding operations in the following scenarios:
   - You log on to the Apsara Uni-manager Management Console for the first time after MFA is enabled by the administrator:
     a. On the Bind Virtual MFA Device page, bind an MFA device.
     b. Enter the username and password again as in Step 2 and click **Log On**.
     c. Enter a six-digit MFA verification code and click **Authenticate**.
   - You have enabled MFA and bound an MFA device:

     Enter a six-digit MFA verification code and click **Authenticate**.

     > **Note** For more information, see the *Bind a virtual MFA device to enable MFA* topic in *Apsara Uni-manager Management Console User Guide*.
5. In the top navigation bar, choose **Products > Elastic Computing > Elastic Compute Service**.

3.3. Create a security group

Security groups are an important means to implement network security isolation. They control network traffic to or from one or more Elastic Compute Service (ECS) instances.

Prerequisites

A virtual private cloud (VPC) is created. For more information, see the "Create a VPC" topic in *VPC User Guide*.

Context

Security groups determine whether instances in the same VPC, region, and account can communicate with each other. By default, if the instances belong to the same security group, they can communicate with each other over the internal network. If the instances belong to different security groups, you can allow mutual access between the security groups to allow the instances to communicate with each other over the internal network.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.

3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create Security Group**.
5. Configure the parameters described in the following table.

Section	Parameter	Required	Description
Area	Organization	Yes	Select an organization in which to create the security group. Make sure that the security group and the VPC belong to the same organization.
	Resource Set	Yes	Select a resource set in which to create the security group. Make sure that the security group and the VPC belong to the same resource set.
	Region	Yes	Select a region in which to create the security group. Make sure that the security group and VPC reside within the same region.
	Zone	Yes	Select a zone in which to create the security group.
Basic Configurations	Sharing Scope	Yes	Select the scope for which to share the security group. Valid values: Current Resource Set, Current Organization and Subordinate Organizations, and Current Organization.
	VPC	Yes	Select a VPC in which to create the security group.
	Security Group Name	Yes	Enter a name for the security group. The name must be 2 to 128 characters in length and start with a letter. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,). It cannot start with http:// or https://.
	Description	No	Enter a description for the security group for easy management. The description must be 2 to 256 characters in length and start with a letter. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,). It cannot start with http:// or https://.

6. Click **Submit**.

   After the security group is created, it is displayed on the Security Groups page.

What's next

- After the security group is created, it contains no security group rules. You can add security group rules to allow or deny access to or from the Internet or internal network for ECS instances within the security group. For more information, see Add a security group rule.
- Each ECS instance must belong to at least one security group. You can add an instance to one or more security groups. For more information, see Add an instance to a security group.

3.4. Create an instance by using the wizard

An Elastic Compute Service (ECS) instance is a virtual machine that contains the basic computing components of a server, such as the CPU, memory, operating system, network settings, and disks.

Prerequisites

- A virtual private cloud (VPC) and a vSwitch are created. For more information, see the "Create a VPC" and "Create a vSwitch" topics in *Apsara Stack VPC User Guide*.
- If you want to assign an IPv6 address to the instance that you want to create, make sure that the VPC and vSwitch are associated with IPv6 CIDR blocks. For more information, see the "*Create an IPv6 VPC*" topic in *Apsara Stack VPC User Guide*.
- A security group is created. For more information, see Create a security group.

Context

Some limits apply when you create GPU-accelerated instances. For more information, see Limits.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create Instance**.
5. Configure the basic settings.

   In the Basic Configurations step, you can configure the basic parameters and resources that are required to create an instance. The basic parameters include Organization, Resource Set, Region, and Zone. The basic resources include the instance type, image, and storage. After you complete the basic settings, click **Next**.

i. Configure the basic settings of the instance.

Parameter	Required	Description
Organization	Yes	Select an organization.
Resource Set	Yes	Select a resource set.
Region	Yes	Select a region.
Zone	Yes	Select a zone. Zones are physical locations in the same region. Each zone has separate power supplies and networks. The internal networks of zones are connected. Faults in one zone are isolated from the other zones. To increase the availability of your applications, we recommend that you create instances in different zones.

ii. Configure instance settings, such as the instance family and instance type.

Parameter	Required	Description
Architecture	Yes	Select an architecture type. Valid values: ■ x86_64 ■ i386 ■ arm64
Support Hot Configuration Changes	No	If you turn on the switch, you can change the instance type without the need to stop the instance. For more information, see Perform a hot configuration change on an instance to change the instance type.
Instance Generation	Yes	Select an instance family. After you select an instance family, you must select an instance type.
Instance Specifications	Yes	Select an instance type. Information such as the CPU, memory, and instance family level are displayed in the Instance Specifications list. Select an instance type based on your business requirements. Instance types that have specific CPU and memory combinations do not support Windows Server images. For more information, see the "*Limits*" topic in *ECS Product Introduction*.

iii. Configure the image that is used by the instance.

Parameter	Required	Description
Image Type	Yes	Select an image type. Valid values: **Public Image**, **Custom Image**, and **Shared Custom Image**.
Public Image	Subject to the image type	Select a public image. Public images provided by Alibaba Cloud (excluding licensed operating system images) deliver the high security and stability. Public images including Windows Server images and major Linux images are provided. This parameter is required if you set Image Type to **Public Image**. When you use an image that supports Dynamic Host Configuration Protocol version 6 (DHCPv6) to create an instance, an IPv6 address is automatically assigned to the instance. The instance can use this IPv6 address to communicate over the internal network. When you use an image that does not support DHCPv6 to create an instance, you must manually assign an IPv6 address to the instance. The following images support DHCPv6: ■ Linux images: ■ CentOS 7.6 IPV6 64Bit ■ CentOS 6.10 64Bit ■ SUSE Linux Enterprise Server 12 SP4 64Bit ■ Windows Server images **Note** To use an IPv6 address to communicate over the Internet, you must enable public bandwidth for the IPv6 address. For more information, see the "*Enable Internet connectivity for an IPv6 address*" topic in *Apsara Stack VPC User Guide*.
Custom Image	Subject to the image type	Select a custom image. Custom images are created from instances or snapshots or imported from your computer. This parameter is required if you set Image Type to **Custom Image**.

Parameter	Required	Description
Shared Custom Images	Subject to the image type	Select a custom image that is shared by another Apsara Stack tenant. This parameter is required if you set Image Type to **Shared Custom Images**.

iv. Configure the storage settings of the instance.

Parameter	Required	Description
Creation Method	Yes	Select a method that is used to create a disk. Valid values: Disk Creation and Storage Set Creation. If you want to create a disk in a partition of a storage set, select **Storage Set Creation**. **Notice** ■ If only Disk Creation is displayed, storage sets are not supported in the current environment. Configure storage sets first. ■ Before you create a disk in a partition of a storage set, make sure that the storage set is created and the partition is configured. For more information, see the "*Create storage sets*" topic in *CDS User Guide*.
Storage Set	Yes	Select the created storage set. This parameter is required only if you set Creation Method to **Storage Set Creation**.
Partitions	Yes	Specify the number of partitions. This parameter is required only if you set Creation Method to **Storage Set Creation**.

Parameter	Required	Description
System Disk	Yes	Specify the system disk on which the operating system is installed. Different Elastic Block Storage (EBS) clusters support different disk categories. ■ Newly deployed EBS clusters in Cloud Defined Storage (CDS) support **premium performance disks** and **standard performance disks**. ■ EBS clusters in CDS that were created in Apsara Stack V3.15.0 and earlier support **ultra disks**, **standard SSDs**, **premium performance disks**, and **standard performance disks**. ■ Existing EBS clusters continue to provide shared ultra disks and shared standard SSDs. The system disk capacity must range from 20 GiB to 500 GiB.
Data Disk	No	Different EBS clusters support different disk categories. ■ Newly deployed EBS clusters in CDS support **premium performance disks** and **standard performance disks**. ■ EBS clusters in CDS that were created Apsara Stack V3.15.0 and earlier support **ultra disks**, **standard SSDs**, **premium performance disks**, and **standard performance disks**. ■ Existing EBS clusters continue to provide shared ultra disks and shared standard SSDs. A maximum of 16 data disks can be attached to an instance. The maximum capacity of each data disk is 32 TiB. You can select **Release with Instance** and **Encryption** for each data disk. To encrypt a data disk, configure the following parameters: ■ **Encryption Method**: Select **AES256** or **SM4**. ■ **Encryption Key**: You can select a key created in **Key Management Service**. You can also add data disks after the instance is created. For more information, see Attach a disk.

v. Configure the deployment sets of the instance.

A deployment set is a policy that controls the distribution of ECS instances. You can use deployment sets to design how to implement disaster recovery and service availability when you create ECS instances. You can use deployment sets to disperse or aggregate the instances used in your business.

6. Configure the network settings.

You can make network and security group configurations to allow the instance to communicate with the Internet and other resources and to safeguard the instance on the network. After you complete the network settings, click **Next**.

Parameter	Required	Description
VPC	Yes	Select a VPC.
VSwitch	Yes	Select a vSwitch.
Private IP Address	No	Specify a private IPv4 address for the instance. The private IPv4 address must be within the CIDR block of the selected vSwitch. If you do not specify a private IP address, the system allocates a private IP address to the instance.
IPv6	No	Specify whether to assign an IPv6 address to the instance.
Security Group	Yes	Select a security group.

7. Configure the system settings.

The system settings include the logon password, hostname, and user data of the instance. You can view the specified instance settings in the console. After you complete the system settings, click **Next: Confirm Information**.

i. Configure logon credentials.

Logon credentials are used to log on to the instance. For more information about how to connect to an instance, see Instance connecting overview.

Notice A logon password cannot be specified for an ECS instance created from the arm_neokylin_7u6_64_20G_20210205.raw or arm_centos_7_6_64_20G_20210205.raw image. You can use a key pair to log on to the instance.

Parameter	Description

Parameter	Description
Key Pair	Select an existing key pair or click **Create Key Pair** to create a key pair. After a key pair is created, go back to the ECS instance creation wizard and click the icon to query the most recent key pair list. For more information, see Create a key pair. **Note** Key pairs can be used to log on only to Linux instances.
	Enter and confirm a password. When you log on to an instance by using a username and a password, the default username is `root` for Linux and `administrator` for Windows. The password must be 8 to 30 characters in length. The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. Supported special characters include `() ' ~ ! @ # $ % ^ & * - _ + = | {} [] : ; ' <> , . ? /` . **Notice** If you use an image of the NFS, UOS, or Kylin type to create an ECS instance, the password used to log on to the instance must meet the following requirements based on the image type: ■ NFS: If you use the NFS_V4_G195_x86_20G image, the password must be at least 8 characters in length. The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary. You can run the `echo "xxx" | cracklib-check` command to check whether the password meets the requirements. Supported special characters include `() ' ~ ! @ # $ % ^ & * - _ + = | {} [] : ; ' <> , . ? /` .

Parameter	Description
Password	■ UOS: ■ If you use the uos_server20_1032d_x64_20G and arm_uos_server20_1032d_20G images, the password must be 8 to 512 characters in length. The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary. You can run the `echo "xxx" | cracklib-check` command to check whether the password meets the requirements. The monotonous character sequences in the password cannot exceed 4 bits in length, such as 1234. Supported special characters include `() ' ~ ! @ # $ % ^ & * - _ + = | {} [] : ; ' <> , . ? / .` ■ If you use the uos_server20_1040d_x64_20G and arm_uos_server20_1040d_20G images, the password must be 8 to 512 characters in length. The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary. You can run the `echo "xxx" | cracklib-check` command to check whether the password meets the requirements. The monotonous character sequences in the password cannot exceed 3 bits in length, such as 123. The number of consecutively identical characters in the password cannot exceed 3, such as aaa. Supported special characters include `() ' ~ ! @ # $ % ^ & * - _ + = | {} [] : ; ' <> , . ? / .` ■ Kylin: If you use an image of the Kylin type, the password must be at least 8 characters in length. The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary. You can run the `echo "xxx" | cracklib-check` command to check whether the password meets the requirements. Supported special characters include `() ' ~ ! @ # $ % ^ & * - _ + = | {} [] : ; ' <> , . ? /`
Later	After the instance is created, bind the key pair or reset the instance password.

ii. Configure the name of the instance.

If you want to create multiple instances, you can set sequential instance names and hostnames to facilitate management.

Parameter	Required	Description
Quantity	Yes	You can create a maximum of 100 instances at a time by using the wizard. In addition, the number of your instances cannot exceed the quota. The specific quota is displayed on the homepage of the Apsara Uni-manager Management Console.
Instance Name	Yes	Enter the name of the instance. The name must be 2 to 128 characters in length. The name must start with a letter but cannot start with http:// or https://. The name can contain letters, digits, colons (:), underscores (_), periods (.), and hyphens (-). When you create multiple instances, their names are automatically suffixed with incremental three-digit numbers in order. By default, the incremental suffixes can range from 001 to 999. Examples: LocalHost001 and LocalHost002, and MyInstance001 and MyInstance002.
Instance Description	No	Enter the description of the instance. The description must be 2 to 256 characters in length, and cannot start with http:// or https://. The description can contain letters, digits, full-width characters, commas (,), periods (.), underscores (_), and hyphens (-).

Parameter	Required	Description
Hostname	No	Enter the hostname displayed in the operating system. Take note of the following items: ■ The hostname cannot start or end with a period (.) or hyphen (-). The hostname cannot contain consecutive periods (.) or hyphens (-). ■ For Windows instances, the hostname must be 2 to 15 characters in length and cannot contain periods (.) or contain only digits. The hostname can contain letters, digits, and hyphens (-). ■ For Linux instances, you can specify the hostname based on the following requirements: ■ The hostname must be 2 to 64 characters in length. You can use periods (.) to separate a hostname into multiple segments. Each segment can contain letters, digits, and hyphens (-). ■ You can use the `${instance_id}` placeholder to pass instance IDs into the hostname specified by the Hostname parameter. For example, if you set `HostName` to k8s-${instance_id} and the instance is assigned an ID of `i-123abc****`, the hostname of the instance is `k8s-i-123abc****`.
Release Protection	No	You can use the release protection feature to prevent ECS instances from being manually released, which can effectively minimize loss caused by unintended operations or lack of timely communication among team members.

iii. Configure advanced options.

User data can be run as scripts on instance startup to automate instance configurations, or can be passed into instances as common data. For more information, see Customize instance data.

In the User Data field, enter the user data that you prepared. If the user data is already encoded in Base64, select **Based64 Encoded Data**.

8. On the Confirm Information page, check the basic configurations, network configurations, and system configurations of the instance, and then click **Submit**.

Result

The new instance appears in the instance list. When the instance is being created, it is in the **Preparing** state. After the instance is created, it enters the **Running** state.

3.5. Connect to an instance

3.5.1. Overview

After an instance is created, you can connect to the instance to perform operations, such as installing applications.

You can use one of the following methods to connect to an instance:

- Use remote connection tools to connect to instances that use public IP addresses. For more information, see the following topics:
  - Connect to a Linux-based instance by using SSH commands in Linux or Mac OS X
  - Connect to a Linux-based instance by using remote connection tools in Windows
  - Connect to a Windows-based instance by using RDP
- Use the VNC feature in the Elastic Computer Service (ECS) console. For more information, see Connect to an instance by using a VNC management terminal.

The username of a Windows instance is Administrator.The username of a Linux instance is root.

3.5.2. Connect to a Linux instance by using SSH commands in Linux or Mac OS X

This topic describes how to use SSH commands to connect to a Linux instance.

Prerequisites

- The instance and the security group are created.
- The instance is in the **Running** state.
- A logon password is set for the instance.
- An Elastic IP address (EIP) is bound with the instance.
- An inbound security group rule is added to the security group to allow the SSH port.

Rule direction	Authorization policy	Protocol type	Port range	Priority	Authorization type	Authorization object
Inbound	Accept	TCP	22/22	1	IPv4 CIDR block	0.0.0.0/0

Procedure

1. Enter the following command and press the Enter key.

   ```
   ssh root@instance IP
   ```

2. Enter the instance password of the root user and press the Enter key.

3.5.3. Connect to a Linux instance by using a remote connection tool in Windows

This topic describes how to connect to a Linux instance by using the PuTTY tool.

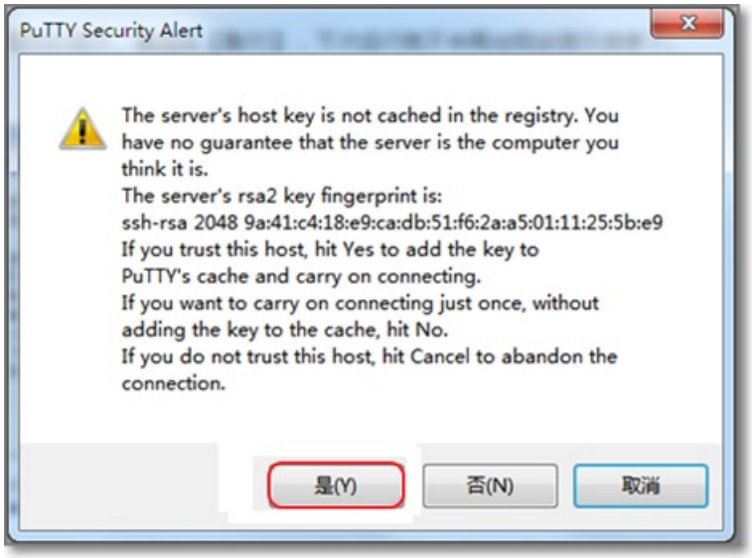
Prerequisites

All remote connection tools are designed based on similar logic. In this example, PuTTY is used to connect to a Linux instance. To download PuTTY, go to the Putty official website.

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 elastic IP address of the instance to which you want to connect.
   - Port: Specify the default port 22.
   - Connection Type: Select SSH.
   - Saved Session: Enter the name of the session, and then click **Save**. After the settings are saved, PuTTY records the name and IP address of the instance.
3. Click **Open** to connect to the instance.

   The first time you connect to the instance, a PuTTY security alert is displayed. Click **Yes** to proceed.



4. Enter username root and press the Enter key.
5. Enter the password of the instance and press the Enter key.

   If a message similar to the following one appears, a connection to the instance is established:

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

3.5.4. Connect to a Windows instance by using RDC

This topic describes how to connect to a Windows instance by using Remote Desktop Connection (RDC).

Prerequisites

- A security group and a Windows instance are created.

- The instance is in the **Running** state.
- A logon password is set for the instance.
- An Elastic IP address is associated with the instance.
- An inbound security group rule is added to the security group to allow traffic on the RDP port.

Rule direction	Action	Protocol	Port range	Priority	Authorization type	Authorization object
Inbound	Allow	tcp	3389/3389	1	IPv4 addresses	0.0.0.0/0

Procedure

1. Use one of the following methods to enable RDC:
   - Click **Start**, enter *mstsc* in the search box, and click **mstsc** in the search result.
   - Press the Windows logo key+R. In the **Run** dialog box that appears, enter *mstsc* and click **OK**.
2. In the **Remote Desktop Connection** dialog box, enter the Elastic IP address of the instance and click **Show Options**.
3. Enter the username.

   The default username is administrator.
4. (Optional)If you do not want to enter the password upon subsequent logons, select **Allow me to save credentials**.
5. Click **Connect**.
6. In the **Windows Security** dialog box that appears, enter the password corresponding to the username you entered and click **OK**.

Result

If the Windows desktop appears, a connection to the Windows instance is established.

If an error message is returned indicating that an authentication error has occurred and the function requested is not supported, install CredSSP updates and try again. Follow these steps to install the updates:

1. Connect to an ECS instance by using the VNC.
2. Choose **Start > Control Panel**.
3. Click **System and Security**.
4. Click **Check for updates** in the **Windows Updates** section.
5. If updates are available, click **Install updates**.
6. Restart the instance.

3.5.5. Connect to an instance by using a VNC management terminal

If other remote connection tools such as PuTTy, Xshell, and SecureCRT are not installed or do not work properly, you can access your instances by using a VNC management terminal in the ECS console.

Prerequisites

- The instance to which you want to connect is in the **Running** state.
- The root certificate is imported to your web browser. For more information, see Install the certificate for VNC in Windows.
- The VNC password is reset if it is your first time to connect to the instance after the instance is created. For more information, see Change the VNC password.

Context

The VNC password is used to log on to a VNC management terminal in the ECS console, whereas the instance password is used to log on to the instance.

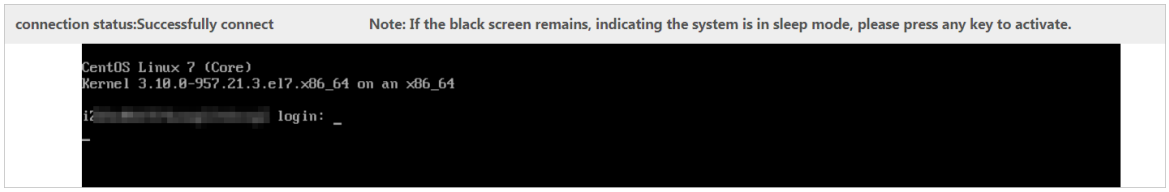
You can use a VNC management terminal to connect to an instance to solve specific issues. The following table lists some of the issues.

Issue	Solution
The instance starts slowly due to self-check on startup.	Check the self-check progress.
The firewall of the operating system is enabled by mistake.	Disable the firewall.
Abnormal processes appear and consume large amounts of CPU or bandwidth resources.	Troubleshoot and terminate the abnormal processes.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance to which you want to connect and click **Remote Connection** in the **Actions** column.
5. Enter the VNC password and click **OK**.

   After you are logged on to the VNC management terminal, a logon page similar to the following one appears.

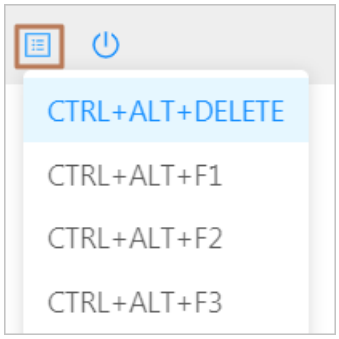
   

6. Enter your username and password.
   - Fo a Linux instance, enter the username *root* and the logon password.

     > **Note** Passwords in Linux are not displayed as you type. Press the Enter key after you enter the password.

   - For a Windows instance, to use a key combination such as Ctrl+Alt+Delete, click the List icon in

the upper-right corner of the VNC page and select the corresponding key combination from the drop-down list.



Enter the username and password as prompted, and click the Log On icon such as  .

4.Instances

4.1. Overview

An Elastic Compute Service (ECS) instance is a virtual server that includes basic components such as CPUs, memory, an operating system (OS), network configurations, and disks. You can use management tools provided by Alibaba Cloud such as the ECS console and ECS API to create and manage ECS instances. You can manage the status of ECS instances and their deployed applications in the same manner as you would do with local servers. You can also upgrade the capabilities (such as compute and storage capabilities) of your ECS instances as your requirements increase.

Basic instance configurations

The following basic configurations of ECS instances determine the basic resources that the instances require:

- Instance types

  Instance types define the basic attributes of ECS instances, such as compute capacity, storage capacity, and networking capacity. Instance types must be used together with images, Elastic Block Storage (EBS) devices, and network resources to create ECS instances that serve different purposes.

  ECS provides a variety of instance families for typical use scenarios. Each instance family consists of multiple instance types that have different compute capabilities and are suited to different scenarios and requirements.

- Images

  Images contain the information necessary to run ECS instances, such as OSs and initialization data of applications. Alibaba Cloud provides ready-to-use images for Windows Server and several mainstream Linux OSs. You can also create or import your own images to save time in making repeated configurations.

- Storage

  ECS instances use their attached system disks and data disks for storage. Each instance must have a system disk attached. The first time the instance starts, the OS is installed and instance configurations are initialized based on the system disk image.

  Cloud disks can be used as system disks or data disks. Local disks can be used only as data disks and are available only for specific instance types, such as instance types with local HDDs. If you want your instances to have more storage space, you can resize their attached cloud disks or attach more cloud disks after the instances are created. For more information, see Create a disk and Attach a disk.

  Business data is an important asset. Cloud disks use a triplicate mechanism to ensure the durability of data. To ensure that your data remains available, we recommend that you back up your data on a regular basis. You can create snapshots of cloud disks to back up disk data. If you are using local disks, you must implement data redundancy at the application layer to ensure data availability.

In addition to basic configurations, you can also configure network configurations, security groups, and OSs for instances. For more information, see Create an instance.

Security suggestions

When you use Apsara Stack services, we recommend that you follow security suggestions to improve resource security, such as the following ones:

- Suggestions for permission control: Use the Enterprise Permissions features to control which users can manage resources such as instances and what permissions to grant to the users.
- Suggestions for security features: Use security features such as security hardening and cloud disk encryption to ensure the security of data and runtime environments.
- Suggestions for network security: Use virtual private clouds (VPCs) to isolate services of different security levels. Use security groups to control inbound and outbound traffic for instances and allow instances access to the Internet only when required to minimize the attack surface area of resources.

4.2. Create an instance by using the wizard

An Elastic Compute Service (ECS) instance is a virtual machine that contains the basic computing components of a server, such as the CPU, memory, operating system, network settings, and disks.

Prerequisites

- A virtual private cloud (VPC) and a vSwitch are created. For more information, see the "Create a VPC" and "Create a vSwitch" topics in *Apsara Stack VPC User Guide*.
- If you want to assign an IPv6 address to the instance that you want to create, make sure that the VPC and vSwitch are associated with IPv6 CIDR blocks. For more information, see the "*Create an IPv6 VPC*" topic in *Apsara Stack VPC User Guide*.
- A security group is created. For more information, see Create a security group.

Context

Some limits apply when you create GPU-accelerated instances. For more information, see Limits.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create Instance**.
5. Configure the basic settings.

   In the Basic Configurations step, you can configure the basic parameters and resources that are required to create an instance. The basic parameters include Organization, Resource Set, Region, and Zone. The basic resources include the instance type, image, and storage. After you complete the basic settings, click **Next**.

i. Configure the basic settings of the instance.

Parameter	Required	Description
Organization	Yes	Select an organization.
Resource Set	Yes	Select a resource set.
Region	Yes	Select a region.
Zone	Yes	Select a zone. Zones are physical locations in the same region. Each zone has separate power supplies and networks. The internal networks of zones are connected. Faults in one zone are isolated from the other zones. To increase the availability of your applications, we recommend that you create instances in different zones.

ii. Configure instance settings, such as the instance family and instance type.

Parameter	Required	Description
Architecture	Yes	Select an architecture type. Valid values: ■ x86_64 ■ i386 ■ arm64
Support Hot Configuration Changes	No	If you turn on the switch, you can change the instance type without the need to stop the instance. For more information, see Perform a hot configuration change on an instance to change the instance type.
Instance Generation	Yes	Select an instance family. After you select an instance family, you must select an instance type.
Instance Specifications	Yes	Select an instance type. Information such as the CPU, memory, and instance family level are displayed in the Instance Specifications list. Select an instance type based on your business requirements. Instance types that have specific CPU and memory combinations do not support Windows Server images. For more information, see the "*Limits*" topic in *ECS Product Introduction*.

iii. Configure the image that is used by the instance.

Parameter	Required	Description
Image Type	Yes	Select an image type. Valid values: **Public Image**, **Custom Image**, and **Shared Custom Image**.
Public Image	Subject to the image type	Select a public image. Public images provided by Alibaba Cloud (excluding licensed operating system images) deliver the high security and stability. Public images including Windows Server images and major Linux images are provided. This parameter is required if you set Image Type to **Public Image**. When you use an image that supports Dynamic Host Configuration Protocol version 6 (DHCPv6) to create an instance, an IPv6 address is automatically assigned to the instance. The instance can use this IPv6 address to communicate over the internal network. When you use an image that does not support DHCPv6 to create an instance, you must manually assign an IPv6 address to the instance. The following images support DHCPv6: ■ Linux images: ■ CentOS 7.6 IPV6 64Bit ■ CentOS 6.10 64Bit ■ SUSE Linux Enterprise Server 12 SP4 64Bit ■ Windows Server images **Note** To use an IPv6 address to communicate over the Internet, you must enable public bandwidth for the IPv6 address. For more information, see the "*Enable Internet connectivity for an IPv6 address*" topic in *Apsara Stack VPC User Guide*.
Custom Image	Subject to the image type	Select a custom image. Custom images are created from instances or snapshots or imported from your computer. This parameter is required if you set Image Type to **Custom Image**.

Parameter	Required	Description
Shared Custom Images	Subject to the image type	Select a custom image that is shared by another Apsara Stack tenant. This parameter is required if you set Image Type to **Shared Custom Images**.

iv. Configure the storage settings of the instance.

Parameter	Required	Description
Creation Method	Yes	Select a method that is used to create a disk. Valid values: Disk Creation and Storage Set Creation. If you want to create a disk in a partition of a storage set, select **Storage Set Creation**. **Notice** ■ If only Disk Creation is displayed, storage sets are not supported in the current environment. Configure storage sets first. ■ Before you create a disk in a partition of a storage set, make sure that the storage set is created and the partition is configured. For more information, see the "*Create storage sets*" topic in *CDS User Guide*.
Storage Set	Yes	Select the created storage set. This parameter is required only if you set Creation Method to **Storage Set Creation**.
Partitions	Yes	Specify the number of partitions. This parameter is required only if you set Creation Method to **Storage Set Creation**.

Parameter	Required	Description
System Disk	Yes	Specify the system disk on which the operating system is installed. Different Elastic Block Storage (EBS) clusters support different disk categories. ■ Newly deployed EBS clusters in Cloud Defined Storage (CDS) support **premium performance disks** and **standard performance disks**. ■ EBS clusters in CDS that were created in Apsara Stack V3.15.0 and earlier support **ultra disks**, **standard SSDs**, **premium performance disks**, and **standard performance disks**. ■ Existing EBS clusters continue to provide shared ultra disks and shared standard SSDs. The system disk capacity must range from 20 GiB to 500 GiB.
Data Disk	No	Different EBS clusters support different disk categories. ■ Newly deployed EBS clusters in CDS support **premium performance disks** and **standard performance disks**. ■ EBS clusters in CDS that were created Apsara Stack V3.15.0 and earlier support **ultra disks**, **standard SSDs**, **premium performance disks**, and **standard performance disks**. ■ Existing EBS clusters continue to provide shared ultra disks and shared standard SSDs. A maximum of 16 data disks can be attached to an instance. The maximum capacity of each data disk is 32 TiB. You can select **Release with Instance** and **Encryption** for each data disk. To encrypt a data disk, configure the following parameters: ■ **Encryption Method**: Select **AES256** or **SM4**. ■ **Encryption Key**: You can select a key created in **Key Management Service**. You can also add data disks after the instance is created. For more information, see Attach a disk.

v. Configure the deployment sets of the instance.

A deployment set is a policy that controls the distribution of ECS instances. You can use deployment sets to design how to implement disaster recovery and service availability when you create ECS instances. You can use deployment sets to disperse or aggregate the instances used in your business.

6. Configure the network settings.

You can make network and security group configurations to allow the instance to communicate with the Internet and other resources and to safeguard the instance on the network. After you complete the network settings, click **Next**.

Parameter	Required	Description
VPC	Yes	Select a VPC.
VSwitch	Yes	Select a vSwitch.
Private IP Address	No	Specify a private IPv4 address for the instance. The private IPv4 address must be within the CIDR block of the selected vSwitch. If you do not specify a private IP address, the system allocates a private IP address to the instance.
IPv6	No	Specify whether to assign an IPv6 address to the instance.
Security Group	Yes	Select a security group.

7. Configure the system settings.

The system settings include the logon password, hostname, and user data of the instance. You can view the specified instance settings in the console. After you complete the system settings, click **Next: Confirm Information**.

i. Configure logon credentials.

Logon credentials are used to log on to the instance. For more information about how to connect to an instance, see Instance connecting overview.

Notice A logon password cannot be specified for an ECS instance created from the arm_neokylin_7u6_64_20G_20210205.raw or arm_centos_7_6_64_20G_20210205.raw image. You can use a key pair to log on to the instance.

Parameter	Description

Parameter	Description
Key Pair	Select an existing key pair or click **Create Key Pair** to create a key pair. After a key pair is created, go back to the ECS instance creation wizard and click the icon to query the most recent key pair list. For more information, see Create a key pair. **Note** Key pairs can be used to log on only to Linux instances.
	Enter and confirm a password. When you log on to an instance by using a username and a password, the default username is `root` for Linux and `administrator` for Windows. The password must be 8 to 30 characters in length. The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. Supported special characters include `() ' ~ ! @ # $ % ^ & * - _ + = | {} [] : ; ' <> , . ? /` . **Notice** If you use an image of the NFS, UOS, or Kylin type to create an ECS instance, the password used to log on to the instance must meet the following requirements based on the image type: ■ NFS: If you use the NFS_V4_G195_x86_20G image, the password must be at least 8 characters in length. The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary. You can run the `echo "xxx" | cracklib-check` command to check whether the password meets the requirements. Supported special characters include `() ' ~ ! @ # $ % ^ & * - _ + = | {} [] : ; ' <> , . ? /` .

Parameter	Description
Password	■ UOS: ■ If you use the uos_server20_1032d_x64_20G and arm_uos_server20_1032d_20G images, the password must be 8 to 512 characters in length. The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary. You can run the `echo "xxx" | cracklib-check` command to check whether the password meets the requirements. The monotonous character sequences in the password cannot exceed 4 bits in length, such as 1234. Supported special characters include `() ' ~ ! @ # $ % ^ & * - _ + = | {} [] : ; ' <> , . ? /` . ■ If you use the uos_server20_1040d_x64_20G and arm_uos_server20_1040d_20G images, the password must be 8 to 512 characters in length. The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary. You can run the `echo "xxx" | cracklib-check` command to check whether the password meets the requirements. The monotonous character sequences in the password cannot exceed 3 bits in length, such as 123. The number of consecutively identical characters in the password cannot exceed 3, such as aaa. Supported special characters include `() ' ~ ! @ # $ % ^ & * - _ + = | {} [] : ; ' <> , . ? /` . ■ Kylin: If you use an image of the Kylin type, the password must be at least 8 characters in length. The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary. You can run the `echo "xxx" | cracklib-check` command to check whether the password meets the requirements. Supported special characters include `() ' ~ ! @ # $ % ^ & * - _ + = | {} [] : ; ' <> , . ? /`
Later	After the instance is created, bind the key pair or reset the instance password.

ii. Configure the name of the instance.

If you want to create multiple instances, you can set sequential instance names and hostnames to facilitate management.

Parameter	Required	Description
Quantity	Yes	You can create a maximum of 100 instances at a time by using the wizard. In addition, the number of your instances cannot exceed the quota. The specific quota is displayed on the homepage of the Apsara Uni-manager Management Console.
Instance Name	Yes	Enter the name of the instance. The name must be 2 to 128 characters in length. The name must start with a letter but cannot start with http:// or https://. The name can contain letters, digits, colons (:), underscores (_), periods (.), and hyphens (-). When you create multiple instances, their names are automatically suffixed with incremental three-digit numbers in order. By default, the incremental suffixes can range from 001 to 999. Examples: LocalHost001 and LocalHost002, and MyInstance001 and MyInstance002.
Instance Description	No	Enter the description of the instance. The description must be 2 to 256 characters in length, and cannot start with http:// or https://. The description can contain letters, digits, full-width characters, commas (,), periods (.), underscores (_), and hyphens (-).

Parameter	Required	Description
Hostname	No	Enter the hostname displayed in the operating system. Take note of the following items: ■ The hostname cannot start or end with a period (.) or hyphen (-). The hostname cannot contain consecutive periods (.) or hyphens (-). ■ For Windows instances, the hostname must be 2 to 15 characters in length and cannot contain periods (.) or contain only digits. The hostname can contain letters, digits, and hyphens (-). ■ For Linux instances, you can specify the hostname based on the following requirements: ■ The hostname must be 2 to 64 characters in length. You can use periods (.) to separate a hostname into multiple segments. Each segment can contain letters, digits, and hyphens (-). ■ You can use the `${instance_id}` placeholder to pass instance IDs into the hostname specified by the Hostname parameter. For example, if you set `HostName` to k8s-${instance_id} and the instance is assigned an ID of `i-123abc****`, the hostname of the instance is `k8s-i-123abc****`.
Release Protection	No	You can use the release protection feature to prevent ECS instances from being manually released, which can effectively minimize loss caused by unintended operations or lack of timely communication among team members.

iii. Configure advanced options.

User data can be run as scripts on instance startup to automate instance configurations, or can be passed into instances as common data. For more information, see Customize instance data.

In the User Data field, enter the user data that you prepared. If the user data is already encoded in Base64, select **Based64 Encoded Data**.

8. On the Confirm Information page, check the basic configurations, network configurations, and system configurations of the instance, and then click **Submit**.

Result

The new instance appears in the instance list. When the instance is being created, it is in the **Preparing** state. After the instance is created, it enters the **Running** state.

4.3. Connect to an instance

4.3.1. Overview

After an instance is created, you can connect to the instance to perform operations, such as installing applications.

You can use one of the following methods to connect to an instance:

- Use remote connection tools to connect to instances that use public IP addresses. For more information, see the following topics:
  - Connect to a Linux-based instance by using SSH commands in Linux or Mac OS X
  - Connect to a Linux-based instance by using remote connection tools in Windows
  - Connect to a Windows-based instance by using RDP
- Use the VNC feature in the Elastic Computer Service (ECS) console. For more information, see Connect to an instance by using a VNC management terminal.

The username of a Windows instance is Administrator.The username of a Linux instance is root.

4.3.2. Connect to a Linux instance by using SSH commands in Linux or Mac OS X

This topic describes how to use SSH commands to connect to a Linux instance.

Prerequisites

- The instance and the security group are created.
- The instance is in the **Running** state.
- A logon password is set for the instance.
- An Elastic IP address (EIP) is bound with the instance.
- An inbound security group rule is added to the security group to allow the SSH port.

Rule direction	Authorization policy	Protocol type	Port range	Priority	Authorization type	Authorization object
Inbound	Accept	TCP	22/22	1	IPv4 CIDR block	0.0.0.0/0

Procedure

1. Enter the following command and press the Enter key.

   ```
   ssh root@instance IP
   ```

2. Enter the instance password of the root user and press the Enter key.

4.3.3. Connect to a Linux instance by using a remote connection tool in Windows

This topic describes how to connect to a Linux instance by using the PuTTY tool.

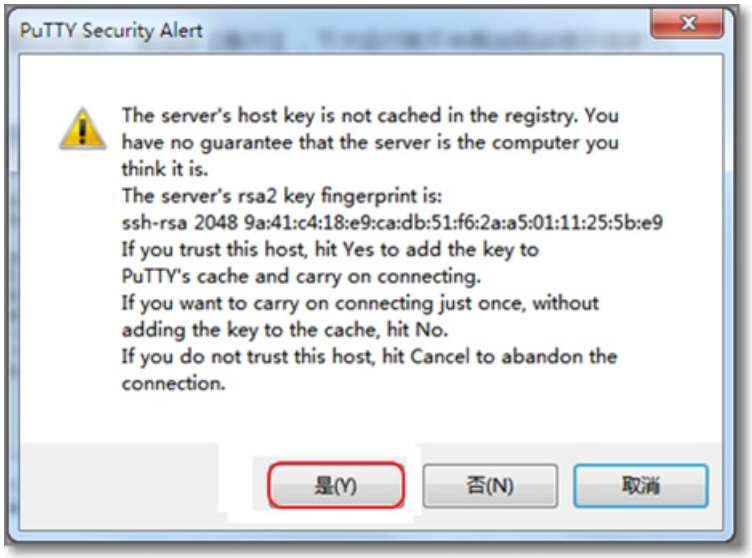
Prerequisites

All remote connection tools are designed based on similar logic. In this example, PuTTY is used to connect to a Linux instance. To download PuTTY, go to the Putty official website.

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 elastic IP address of the instance to which you want to connect.
   - Port: Specify the default port 22.
   - Connection Type: Select SSH.
   - Saved Session: Enter the name of the session, and then click **Save**. After the settings are saved, PuTTY records the name and IP address of the instance.
3. Click **Open** to connect to the instance.

   The first time you connect to the instance, a PuTTY security alert is displayed. Click **Yes** to proceed.



4. Enter username root and press the Enter key.
5. Enter the password of the instance and press the Enter key.

   If a message similar to the following one appears, a connection to the instance is established:

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

4.3.4. Connect to a Windows instance by using RDC

This topic describes how to connect to a Windows instance by using Remote Desktop Connection (RDC).

Prerequisites

- A security group and a Windows instance are created.

- The instance is in the **Running** state.
- A logon password is set for the instance.
- An Elastic IP address is associated with the instance.
- An inbound security group rule is added to the security group to allow traffic on the RDP port.

Rule direction	Action	Protocol	Port range	Priority	Authorization type	Authorization object
Inbound	Allow	tcp	3389/3389	1	IPv4 addresses	0.0.0.0/0

Procedure

1. Use one of the following methods to enable RDC:
   - Click **Start**, enter *mstsc* in the search box, and click **mstsc** in the search result.
   - Press the Windows logo key+R. In the **Run** dialog box that appears, enter *mstsc* and click **OK**.
2. In the **Remote Desktop Connection** dialog box, enter the Elastic IP address of the instance and click **Show Options**.
3. Enter the username.

   The default username is administrator.
4. (Optional)If you do not want to enter the password upon subsequent logons, select **Allow me to save credentials**.
5. Click **Connect**.
6. In the **Windows Security** dialog box that appears, enter the password corresponding to the username you entered and click **OK**.

Result

If the Windows desktop appears, a connection to the Windows instance is established.

If an error message is returned indicating that an authentication error has occurred and the function requested is not supported, install CredSSP updates and try again. Follow these steps to install the updates:

1. Connect to an ECS instance by using the VNC.
2. Choose **Start > Control Panel**.
3. Click **System and Security**.
4. Click **Check for updates** in the **Windows Updates** section.
5. If updates are available, click **Install updates**.
6. Restart the instance.

4.3.5. Install the certificate for VNC in Windows

Before you log on to the Virtual Network Computing (VNC) management terminal, you must export the relative certificate from the site such as the Apsara Uni-manager Management Console and install the certificate in a browser on your computer.

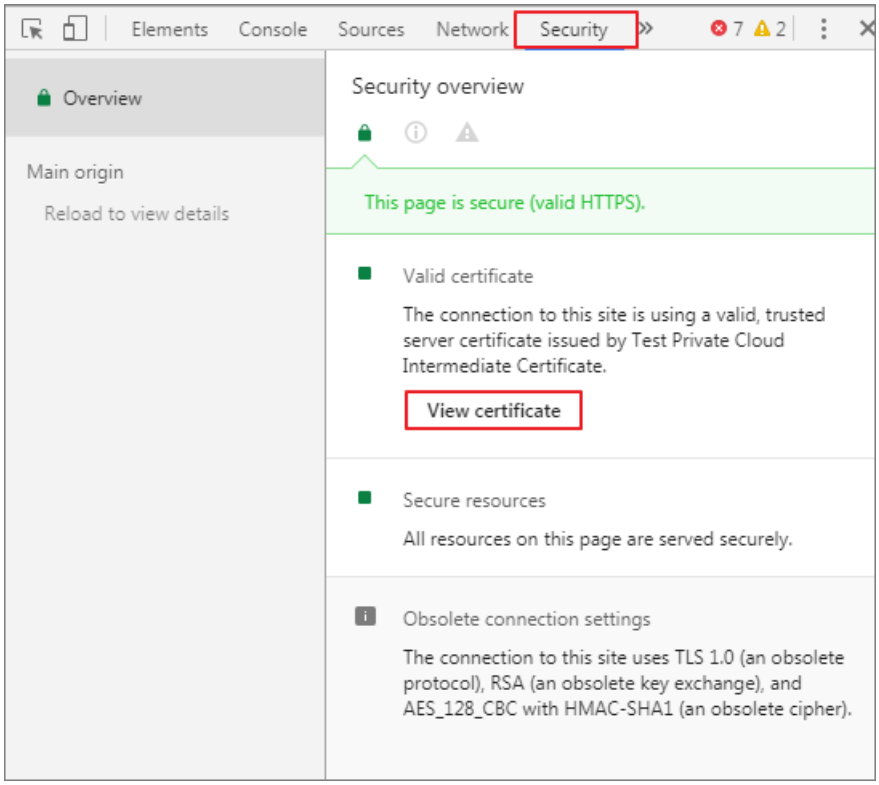
Context

The VNC feature is provided by the VNC proxy service. The VNC proxy service uses a different certificate than that of Apsara Infrastructure Management Framework. The certificate of the VNC proxy service must be imported manually.

Procedure

1. Export the certificate from the Apsara Uni-manager Management Console.
   i. Log on to the Apsara Uni-manager Management Console. Press the **F12** key or **Fn+F12** to view and select the certificate.

      In this example, the Chrome browser is used. Press the **F12** key to open Chrome DevTools and click View certificate on the Security tab, as shown in the following figure.



   ii. In the **Certificate** dialog box, click the **Certificate Path** tab, select the root certificate, and then click **View Certificate**.
   iii. In the **Certificate** dialog box, click the **Details** tab and then click **Copy to File**.
   iv. In the **Certificate Export Wizard** dialog box, click **Next**.
   v. Select **DER encoded binary X.509 (.CER)** as the format and then click **Next**.
   vi. Click **Browse**, select the location where to store the certificate, enter a file name, and then click **Save**.
   vii. Click **Next**.
   viii. Click **OK**.
   ix. Click **OK**.
2. Install the certificate in a browser on your computer.
   i. Double-click the certificate.
   ii. In the **Certificate** dialog box, click **Install Certificate**.

iii. In the **Certificate Import Wizard** dialog box, click **Next**.

iv. Select **Place all certificates in the following store** and click **Browse**.

v. In the **Select Certificate Store** dialog box, select **Trusted Root Certificate Authority** and then click **OK**.

vi. In the **Certificate Import Wizard** dialog box, click **Next**.

vii. Click **OK**.

viii. If a security warning message is displayed, click **Yes**.

3. Restart your browser and log on to the Apsara Uni-manager Management Console.

   After the certificate is installed, the security warning message is no longer displayed on the left of the URL when you log on to the Apsara Uni-manager Management Console.

4.3.6. Connect to an instance by using a VNC management terminal

If other remote connection tools such as PuTTy, Xshell, and SecureCRT are not installed or do not work properly, you can access your instances by using a VNC management terminal in the ECS console.

Prerequisites

- The instance to which you want to connect is in the **Running** state.
- The root certificate is imported to your web browser. For more information, see Install the certificate for VNC in Windows.
- The VNC password is reset if it is your first time to connect to the instance after the instance is created. For more information, see Change the VNC password.

Context

The VNC password is used to log on to a VNC management terminal in the ECS console, whereas the instance password is used to log on to the instance.

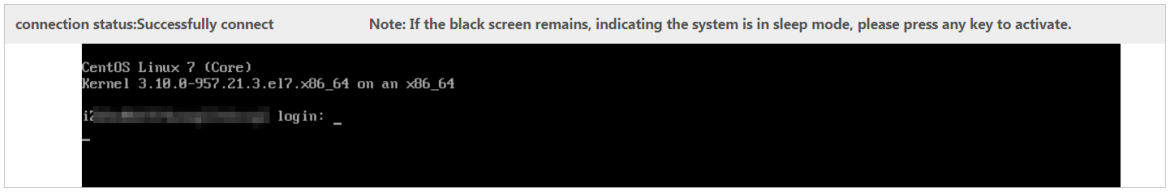
You can use a VNC management terminal to connect to an instance to solve specific issues. The following table lists some of the issues.

Issue	Solution
The instance starts slowly due to self-check on startup.	Check the self-check progress.
The firewall of the operating system is enabled by mistake.	Disable the firewall.
Abnormal processes appear and consume large amounts of CPU or bandwidth resources.	Troubleshoot and terminate the abnormal processes.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance to which you want to connect and click **Remote Connection** in the **Actions** column.
5. Enter the VNC password and click **OK**.

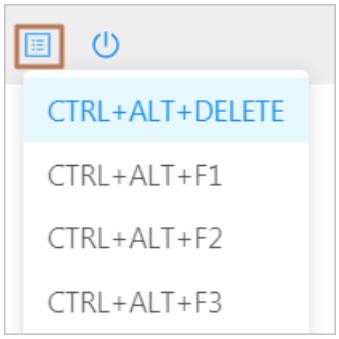
   After you are logged on to the VNC management terminal, a logon page similar to the following one appears.



6. Enter your username and password.
   - Fo a Linux instance, enter the username *root* and the logon password.

     > **Note** Passwords in Linux are not displayed as you type. Press the Enter key after you enter the password.

   - For a Windows instance, to use a key combination such as Ctrl+Alt+Delete, click the List icon in the upper-right corner of the VNC page and select the corresponding key combination from the drop-down list.



   Enter the username and password as prompted, and click the Log On icon such as  .

4.4. Manage instance status

4.4.1. Stop instances

You can stop the instances that you no longer use. The stop operation interrupts services that are running on the instances. Exercise caution when you perform this operation.

Prerequisites

The instance that you want to stop is in the **Running** state.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Use one of the following methods to stop instances:
   - To stop a single instance, find the instance and choose **Instance Status > Stop Instance** in the **Actions** column.
   - To stop one or more instances at a time, select the instances and click **Stop** in the lower-left corner of the Instances page.
5. In the message that appears, click **Stop Instance**.

Result

When the instance is being stopped, its state in the **Status** column changes from **Running** to **Stopping**. After the instance is stopped, its state changes to **Stopped**.

4.4.2. Start an instance

If an Elastic Compute Service (ECS) instance is in a state (such as Stopped) in which it cannot provide services, you must start the instance before you can use it. This topic describes how to start an instance in the ECS console.

Prerequisites

The instance is in the **Stopped** state.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Use one of the following methods to start an instance:
   - To start a single instance, find the instance and choose **Instance Status > Start Instance** in the **Actions** column.
   - To start one or more instances at a time, select the instances and click **Start** in the lower-left corner of the Instances page.
   - In the message that appears, click **Start Instance**.

Result

When the instance is being started, its state in the **Status** column changes from **Stopped** to **Starting**. When the instance is started, its state changes to **Running**.

4.4.3. Restart instances

You must restart instances after you change their logon passwords or install system updates for the instances.

Prerequisites

The instance that you want to restart is in the **Running** state.

> **Warning** The restart operation stops the instance for a short period of time and interrupts services that are running on the instance. Exercise caution when you perform this operation.

Procedure

1. Log on to the ECS console
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Use one of the following methods to restart instances:
   - To restart a single instance, find the instance and choose **Instance Status > Restart Instance** in the **Actions** column.
   - To restart one or more instances at a time, select the instances and click **Restart** in the lower-left corner of the Instances page.
   - In the message that appears, select whether to forcibly restart the instances.
     - If you select **Force Restart**, the instances are forcibly restarted. This may result in the loss of unsaved data.
     - If you do not select **Force Restart**, the instances are restarted.
   - Click **Restart Instance**.

4.4.4. Delete an instance

You can delete Elastic Compute Service (ECS) instances that are no longer needed to release their resources. This topic describes how to delete an instance.

Prerequisites

The instance that you want to delete is in the **Stopped** state.

Context

Deleted instances cannot be recovered. We recommend that you back up data before you delete instances. If data disks are released along with the instances, the disk data cannot be recovered. For more information about how to create snapshots of disks to back up disk data, see Create a snapshot.

To ensure business continuity, we recommend that you select **Move to Recycle Bin** when you delete an instance. This way, the deleted instance is moved to the recycle bin and retained there for a specified period of time. Then, you can make sure that your business is not affected by the instance delete operation before you permanently delete the instance.

> **Note** Instances that use local disks cannot be moved to the recycle bin.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.

4. Use one of the following methods to delete an instance.
   - To delete a single instance at a time, perform the following operations:
     a. Find the instance and choose **Instance Status > Delete** in the **Actions** column.
     b. In the dialog box that appears, select **Move to Recycle Bin** as needed. Then, click **Delete**.
   - To delete one or more instances at a time, perform the following operations:
     a. Select the instances that you want to delete and click **Delete** in the lower part of the Instances page.
     b. In the dialog box that appears, select **Move to Recycle Bin** as needed. Then, click **Delete**.

> **Note** When deleted instances are moved to the recycle bin, the instances enter the retention period, which is three days by default. The computing resources (vCPUs and memory) of the instances are released, their elastic IP addresses are disassociated, and their storage resources are retained during the retention period. If you do not restore resources from the recycle bin during the retention period, the resources are automatically deleted when the retention period expires, and cannot be restored. You can restore resources from the recycle bin anytime during the retention period. For more information, see Restore an instance.

4.5. Manage instance attributes

4.5.1. View instance information

You can view the list of Elastic Compute Service (ECS) instances in your account and the details of individual instances. The details of an instance include the basic configurations, disks, snapshots, security groups, and elastic network interfaces (ENIs).

View the information of instances on the Overview page

By default, the **Overview** page is displayed when you log on to the ECS console.

The **Overview** page consists of the following sections:

- **Instance Overview**

  This section shows statistics about resources in the current region that belong to all resource sets under the current organization in your account. These statistics include the total number of instances, number of running instances, number of disks, number of security groups, CPU quota, memory quota, disk quota, and usage of these quotas.

- **Top 10 Organizations by Quota Usage**

  This section shows the 10 organizations that have the highest quota usage. You can click **By CPU Quota Usage**, **By Memory Quota Usage**, or **By Storage Quota Usage** to list the 10 organizations that have the highest CPU, memory, or disk quota usage and view the recent quota usage rankings and trends of the organizations.

- **Top 20 Instances by Resource Usage**

  This section shows the 20 instances that have the highest resource usage. You can click **By CPU Utilization** or **By Memory Usage** to list the 20 instances that have the highest CPU utilization or memory usage and check whether the resource usage of these instances meets your expectations.

- **Monitoring & Maintenance**

This section shows organizations that have high and low quota usage and instances that have high and low resource usage and provides suggestions for optimizing resource allocation to improve resource utilization.

View the information of a single instance on the Instance Details page

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Select a filter option from the drop-down list to search for an instance.
   - Select a filer option from the drop-down list and enter relevant information in the search box. Then, the system searches for instances based on your specified filter condition.
   - Click **Advanced Filter**, specify multiple filter options, and then click **Search**.

     The following table describes the filter options in the Advanced Filter section.

Filter option	Description
Instance Name	Enter an instance name to search for the instance.
Instance ID	Enter an instance ID to search for the instance.
IP Address	Enter the IP address of an instance to search for the instance.
VPC ID	Enter a VPC ID to search for instances that belong to the VPC.
Image ID	Enter an image ID to search for instances that use the image.
Security Group ID	Enter a security group ID to search for instances that belong to the security group.
Operating System	Enter the name of operating system to search for instances that use the operating system.
Instance Type	Enter an instance type to search for instances of the instance type.
Deployment Set ID	Enter a deployment set ID to search for the instances that belong to the deployment set.
Zone	Enter a zone to search for instances that reside in the zone.
Key Pair Name	Enter a key pair name to search for instances that have the key pair bound.
Tag	Click the Tag field and then select tag keys in the **Tag key** section or tag values in the **Tag value** section to search for instances that have the specified tags added.

   - You can use one of the following methods to go to the Instance Details page of an instance and view the basic information and configurations of the instance:
     - On the Instances page, click the instance ID in the **Instance ID/Name** column

- On the Instances page, click the instance ID in the **Instance ID/Name** column.
- On the Instances page, click **Manage** in the **Actions** column.

4.5.2. Modify the properties of an instance

After you create an instance, you can modify its name, description, and user data any time to suit your business requirements.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance that you want to modify, click the ... icon in the **Actions** column, and then choose **Instance Settings > Modify Instance Properties**.
5. In the **Modify Instance Properties** dialog box, modify the following parameters of the instance:
   - **Instance Name**: The name of the instance. The name must be 2 to 128 characters in length. It must start with a letter and cannot start with `http://` or `https://`. It can contain letters, digits, periods (.), underscores (_), hyphens (-), and colons (:).
   - **Hostname**: The hostname that is displayed inside the operating system. Take note of the following items:
     - The hostname cannot start or end with a period (.) or hyphen (-). It cannot contain consecutive periods (.) or hyphens (-).
     - For Windows instances, the hostname must be 2 to 15 characters in length and cannot contain periods (.) or contain only digits. It can contain letters, digits, and hyphens (-).
     - For Linux instances, the hostname must be 2 to 64 characters in length. Separate a hostname into multiple segments with periods (.). Each segment can contain letters, digits, and hyphens (-).
   - **Instance Description**: The description of the instance. The description must be 2 to 256 characters in length, and cannot start with http:// or https://. It can contain letters, digits, commas (,), periods (.), underscores (_), and hyphens (-).
   - **User Data**. The user data of the instance. For more information, see Customize instance data.
6. Click **OK**.

4.5.3. Reset the logon password of an instance

If you did not set a logon password when you created an Elastic Compute Service (ECS) instance or if you have forgotten the password, you can reset the password.

Context

The logon password of ECS instances created from the arm_neokylin_7u6_64_20G_20210205.raw or arm_centos_7_6_64_20G_20210205.raw image cannot be reset. You can bind key pairs to these instances for logons. For more information, see Bind a key pair to an instance.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance that you want to manage and use one of the following methods to reset its logon password:
   - On the Instance Details page, reset the logon password.
     - On the Instances page, click the instance ID in the **Instance ID/Name** column. In the upper-right corner of the Instance Details page, choose **More > Reset Instance Password**.
     - On the Instances page, click **Manage** in the **Actions** column. In the upper-right corner of the Instance Details page, choose **More > Reset Instance Password**.
   - Click the ... icon in the **Actions** column and choose **Password/Key Pair > Reset Instance Password**.
   - In the Reset Password dialog box, enter and confirm the new password and click **OK**.

     The password must be 8 to 30 characters in length and contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. Supported special characters include: `( ) ' ~ ! @ # $ % ^ & * - _ + = | { } [ ] : ; ' < > , . ? /` .

     The passwords of Windows instances cannot start with a forward slash (/).

     Notice If instances run Unity Operating System (UOS), NFS, or Kylin operating systems, the following limits on passwords apply:

     - NFS: For an instance that runs an NFS operating system and uses the NFS_V4_G195_x86_20G image, the password must be at least 8 characters in length and contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary, and you can run `echo "xxx" | cracklib-check` to check whether the password fulfills this requirement. Supported special characters include: `( ) ' ~ ! @ # $ % ^ & * - _ + = | { } [ ] : ; ' < > , . ? /` .

- UOS:
  - For an instance that runs a UOS operating system and uses the uos_server20_1032d_x64_20G or arm_uos_server20_1032d_20G image, the password must be 8 to 512 characters in length and contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary, and you can run `echo "xxx" | cracklib-check` to check whether the password fulfills this requirement. In addition, the password can contain an ascending or descending sequence of up to four consecutive numbers such as 1234. Supported special characters include: `( ) ' ~ ! @ # $ % ^ & * - _ + = | { } [ ] : ; ' < > , . ? /` .
  - For an instance that runs a UOS operating system and uses the uos_server20_1040d_x64_20G or arm_uos_server20_1040d_20G image, the password must be 8 to 512 characters in length and contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary, and you can run `echo "xxx" | cracklib-check` to check whether the password fulfills this requirement. In addition, the password can contain an ascending or descending sequence of up to three consecutive numbers such as 123 and contain up to three consecutive identical characters such as aaaa. Supported special characters include: `( ) ' ~ ! @ # $ % ^ & * - _ + = | { } [ ] : ; ' < > , . ? /` .
- For an instance that runs a Kylin operating system, the password must be at least 8 characters in length and contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The password cannot contain words in the cracklib dictionary, and you can run `echo "xxx" | cracklib-check` to check whether the password fulfills this requirement. Supported special characters include: `( ) ' ~ ! @ # $ % ^ & * - _ + = | { } [ ] : ; ' < > , . ? /` .

- Restart the instance by using the ECS console or by calling an API operation for the new password to take effect.

  For more information, see Restart an instance or the "RebootInstance" topic in *ECS Developer Guide*.

4.5.4. Change the VNC password

If you use Virtual Network Computing (VNC) to connect to an Elastic Compute Service (ECS) instance for the first time or if you forget the VNC password, you can change the VNC password.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the ECS instance whose VNC password you want to change and use one of the following methods to change the VNC password:
   - On the Instance Details page, change the VNC password

- On the Instance Details page, change the VNC password.
  - On the Instances page, click the ID of the instance in the **Instance ID/Name** column. In the upper-right corner of the Instance Details page, choose **More > Reset VNC Password**.
  - On the Instances page, click **Manage** in the **Actions** column. In the upper-right corner of the Instance Details page, choose **More > Reset VNC Password**.
- On the Instances page, change the VNC password. Click the ... icon in the **Actions** column and choose **Password/Key Pair > Reset VNC Password**.

5. In the **Reset VNC Password** dialog box, enter and confirm the new password and then click **OK**.

   The VNC password must be 8 characters in length and must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. Supported special characters include: `! @ ~ ` # $ % ^ * ( ) _ - + = { [}] : ; , . ? / | \`.

   > **Note** If your instance is a non-I/O optimized instance, you must restart the instance in the ECS console for the new password to take effect.

4.5.5. Enable and disable release protection for instances

You can enable release protection for Elastic Compute Service (ECS) instances to prevent manual release. This topic describes how to enable and disable release protection for ECS instances and how to check whether release protection is enabled.

Context

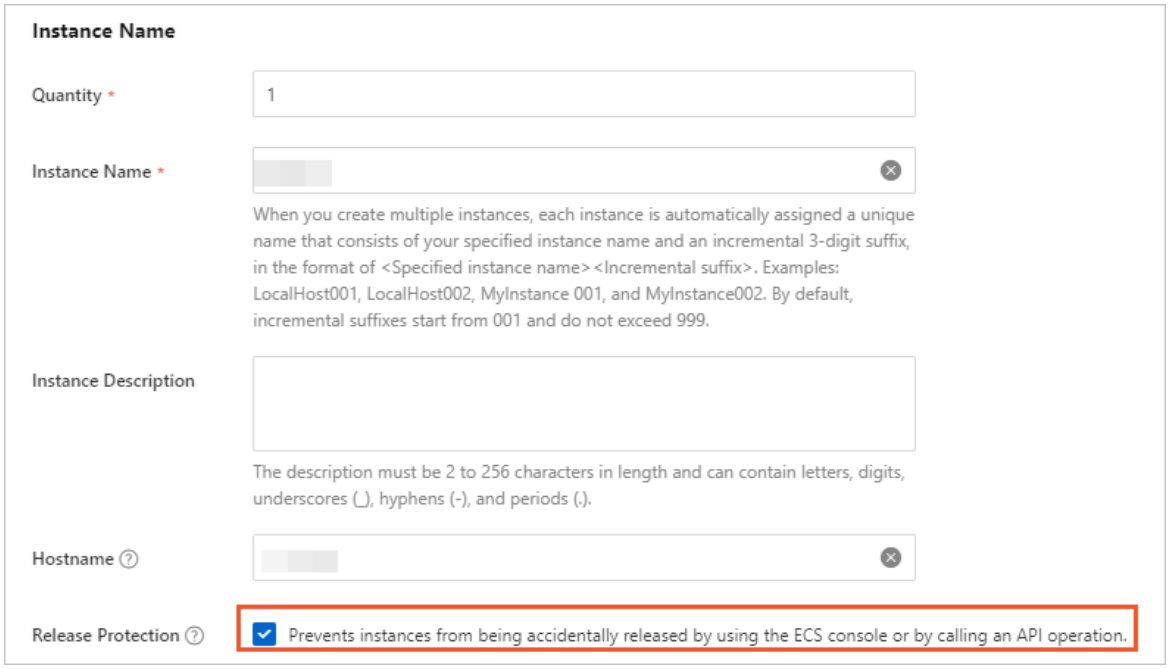
The release protection feature cannot prevent the automatic release of an instance in normal scenarios such as the following scenarios:

- The automatic release time that you set for the instance has arrived.
- The instance does not comply with the applicable security compliance policies.
- The instance was automatically created by Auto Scaling and is removed during subsequent scale-in events.

Enable release protection when you create an instance

This section describes how to configure release protection settings when you create an instance. For more information about how to create an instance, see Create an instance.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Instances** page, click **Create Instance**.
5. In the **Basic Configurations** step, configure parameters and click **Next**.
6. In the **Networking** step, configure parameters and click **Next**.
7. In the **System Configurations** step, select **Release Protection**, configure other parameters, and then click **Confirm Information**.



8. Configure other parameters based on your business requirements and click **Submit**.

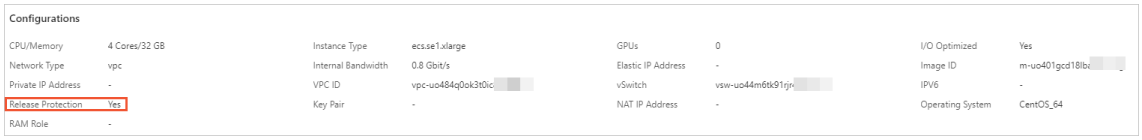
Change the release protection settings

You can also enable or disable release protection for an instance by modifying the attributes of the instance.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance that you want to modify, click the ... icon in the **Actions** column, and then choose **Instance Settings > Change Release Protection Settings**.
5. In the dialog box that appears, select or clear **Release Protection**.
6. Click **OK**.

Check whether release protection is enabled

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Instances** page, use one of the following methods to view details of an instance:
   - In the **Instance ID/Name** column, click the ID of the instance.
   - Find the instance and click **Manage** in the **Actions** column.
   - On the **Instance Details** tab, check the **Release Protection** parameter in the **Configurations** section. If the value is Yes, Release Protection is enabled. If the value is no, Release Protection is disabled.



4.5.6. Instance user data

The user data feature provided by Elastic Compute Service (ECS) allows you to customize the startup behaviors of instances and pass data into instances.

Context

This feature is applicable to both Windows and Linux instances. You can use this feature to perform the following operations:

- Run scripts during instance startup.
- Pass user data as common data into an ECS instance for future reference.

Usage notes

- **Limits**

  The user data feature can be used only when instances meet the following requirements:

  - Network type: Virtual Private Cloud (VPC)
  - Image: a system image or a custom image that is derived from the system image
  - Operating system: one of the following supported operating systems. For more information about the supported operating systems, see .

Supported operating systems

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 configure user data scripts, enter user data based on the operating system type and script type.

  > **Note** Only English characters are supported.

  - If your data is Base64-encoded, select **Enter Base64 Encoded Information**.

    > **Note** The size of the user data script cannot exceed 16 KB before the data is Base64-encoded.

  - For Linux instances, the script format must meet the requirements described in Types of user data scripts of Linux instances.
  - For Windows instances, the scripts must start with `[bat]` or `[powershell]`.

- After an instance is started, run a command to view the following information:
  - Execution result of the user data script
  - Data passed into the instance
- **Console**: You can modify the user data of instances in the console. Whether the modified user data script needs to be re-executed depends on the script type. For example, if the `bootcmd` script in Cloud Config is modified for Linux instances, the script is automatically executed each time the instances are restarted.
- **API**: You can also call API operations to work with instance user data. For more information, see the **CreateInstance** and **ModifyInstanceAttribute** sections in ECS Developer Guide.

User data scripts of Linux instances

User data scripts for Linux instances are implemented based on the cloud-init architecture. The scripts are used to complete automated configurations of Linux instances. User data scripts are compatible with cloud-init.

Description of user data scripts of Linux instances

- User data scripts of Linux instances are executed after the instances are started but before `/etc/init` is executed.
- By default, user data scripts of Linux instances can only be executed with root permissions.

Types of user data scripts of Linux instances

- **User-Data Script**
  - Description: the script that is used to customize configurations, such as a shell script.
  - Format: The first line must start with `#!`. Example: `#!/bin/sh`.
  - Limit: The script including the first line is limited to 16 KB in raw form before it is Base64-encoded.
  - Frequency: The script is executed only when instances are 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: the predefined data that is used to configure services such as specifying yum repositories or importing SSH keys.
- Format: The first line must be `#cloud-config`.
- Limit: The script including the first line is limited to 16 KB in raw form before it is Base64-encoded.
- Frequency: The script execution frequency varies based on the specific service.
- Example:

```
#cloud-config
apt:
primary:
- arches: [default]
uri: http://us.archive.ubuntu.com/ubuntu/
```

- **Include**
  - Description: The configuration content in the script can be saved as a text file and passed into cloud-init as a URL.
  - Format: The first line must be `#include`.
  - Limit: The script including the first line is limited to 16 KB in raw form before it is Base64-encoded.
  - Frequency: The script execution frequency varies based on the script type in the text file.
  - Example:

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

- **GZIP format**
  - Description: cloud-init limits the size of each user data script to 16 KB. If the size of a file exceeds 16 KB, you can compress the file before you pass it into the user data script.
  - Format: The .gz file is passed into the user data script as a URL in `#include`.
  - Frequency: The script execution frequency varies based on the script content contained in the GZIP file.
  - Example:

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

View the user data of a Linux instance

To view the user data of a Linux instance, run the following command in the instance:

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

User data scripts of Windows instances

User data scripts for Windows instances are developed by Alibaba Cloud and can be used to initialize Windows instances.

The following types of user data scripts for Windows instances are available:

- Batch processing program: The first line starts with `[bat]`. The script is limited to 16 KB in raw form before it is Base64-encoded.

- PowerShell script: The first line starts with `[powershell]`. The script is limited to 16 KB in raw form before it is Base64-encoded.

View the user data of a Windows instance

To view the user data of a Windows instance, run the following PowerShell command in the instance:

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

4.6. Manage the instance recycle bin

4.6.1. Set the retention period

When you put Elastic Compute Service (ECS) instances into the recycle bin, the instances enter a retention period. This topic describes how to set the retention period.

Prerequisites

Only the admin account can set the retention period.

Context

When an instance enters a retention period, its computing resources (CPU and memory) are released, its elastic IP address is disassociated, and its storage resources are retained. When the retention period expires, the ECS instance and its storage resources are completely destroyed and cannot be recovered.

The default retention period is three days. You can set an appropriate retention period of up to 30 days and 23 hours based on your business needs.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Recycle Bin**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Recycle Bin page, click **Set Retention Period**.
5. In the **Set Retention Period** dialog box, specify a period of time and click **OK**.

4.6.2. Restore an instance

When you place Elastic Compute Service (ECS) instances into the recycle bin, the instances enter a retention period. During the retention period, you can restore the instance anytime based on your business requirements.

Context

When instances are restored from the recycle bin, computing resources (vCPUs and memory) are automatically re-assigned to the instances. If these instances require elastic IP addresses (EIPs), you must manually associate EIPs with the instances. For more information, see the "Associate an EIP with an ECS instance" topic in the *EIP User Guide*.

> **Note** When you delete instances to the recycle bin, their computing resources are automatically reclaimed. Your attempt to restore instances from the recycle bin may fail due to insufficient quotas or computing resources. Before you restore instances, make sure that your quotas and computing resources are sufficient.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Recycle Bin**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Restore an instance from the recycle bin.

   Use one of the following methods to restore the instance from the recycle bin:

   - To restore a single instance, find the instance in the recycle bin and click **Restore** in the Actions column. In the message that appears, click **OK**.
   - To restore one or more instances at a time, select the instances and click **Restore** in the lower-left corner of the page. In the message that appears, click **OK**.

4.6.3. Permanently delete instances

You can permanently delete the instances that you no longer need from the recycle bin. This topic describes how to permanently delete instances.

Context

The instances that are permanently deleted cannot be restored. Make sure that your business is not interrupted before you permanently delete instances.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Recycle Bin**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Instances page, view information about all ECS instances in the retention period, including the instance ID, name, instance type, deletion time, and automatic release time of the instances.
5. Permanently delete instances from the recycle bin.

   You can use one of the following methods to permanently delete instances:

   - To delete a single instance, click **Permanently Delete** in the Actions column corresponding to the instance. In the message that appears, click **Delete**.
   - To delete one or more instances at a time, select all instances that you want to delete, and then click **Permanently Delete** in the lower-left corner of the Instances page. In the message that appears, click **Delete**.

4.7. Change the instance type

4.7.1. Upgrade or downgrade the instance type of an instance

When the instance type of your Elastic Compute Service (ECS) instance does not meet application requirements, you can change it (including vCPUs, memory, and internal bandwidth). This topic describes how to upgrade or downgrade the instance type of an instance.

Prerequisites

The instance whose instance type you want to upgrade or downgrade is in the **Stopped** state.

> **Note** Service interruptions may occur when you stop instances. We recommend that you stop instances during off-peak hours.

Context

The following limits apply when you upgrade or downgrade the instance type of an instance:

- When you specify a new instance type for the instance, the vCPU, memory, and internal bandwidth specifications of the instance are changed together. These specifications cannot be separately changed.
- You must wait at least 10 minutes between two consecutive changes to the instance type of a single instance.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance whose instance type you want to upgrade or downgrade, click the ... icon in the **Actions** column, and then click **Upgrade/Downgrade**.
5. On the **Change Configuration of ECS instance** page, select a desired instance type and click **Submit**.
6. Restart the instance by using the ECS console or by calling an API operation for the new instance type to take effect.

   For more information, see Stop an instance or the "*StartInstance*" topic in *ECS Developer Guide*.

4.7.2. Perform a hot configuration change on an instance to change the instance type

If the instance type of your Elastic Compute Service (ECS) instance does not suit application requirements, you can change it (including vCPUs, memory, and internal bandwidth). This topic describes how to perform a hot configuration change on an instance to change its instance type.

Prerequisites

The instance supports hot configuration changes and **Support Hot Configuration Changes** was selected when the instance was created in the ECS console. For more information, see Create an instance.

> **Note** Instance types that support hot configuration changes are displayed in the ECS console.

Context

Compared with the instance type change feature, the hot configuration change feature can be used to change instance types without stopping the instances.

When you perform a hot configuration change on an instance to change the instance type, the following limits apply:

- Windows instances do not support hot configuration changes.
- When you specify a new instance type for the instance, the vCPU, memory, and internal bandwidth specifications of the instance are changed together. These specifications cannot be separately changed.
- You must wait at least 10 minutes between two consecutive hot configuration changes to the instance type of a single instance.
- You can perform a maximum of five consecutive hot configuration changes to the instance type of an instance. After the instance is restarted, stopped, or started, the change times are cleared and recalculated. If a hot configuration change on the instance fails, hot configuration changes cannot be performed on the instance.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance whose instance type you want to change, click the ... icon in the **Actions** column, and then click **Hot Configuration Change**.
5. In the **Hot Configuration Change** dialog box, select a desired instance type and click **OK**.

4.8. View the monitoring information of an instance

You can view monitoring charts in the CloudMonitor console to learn about the running conditions of Elastic Compute Service (ECS) instances. This topic describes how to go to the CloudMonitor console to view the monitoring information of an ECS instance.

Context

CloudMonitor provides real-time monitoring, alerting, and notification services for resources to protect your services and business. For more information, see the "CloudMonitor overview" topic in *Apsara Uni-manager Management Console User Guide*.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the ECS instance whose monitoring information you want to view and click the  icon in the **Monitoring** column.
5. On the **Monitoring Charts** page, view the monitoring information of the ECS instance.

   For more information, see the "CloudMonitor overview" topic in *Apsara Uni-manager Management Console User Guide*.

4.9. Add an instance to a security group

You can add created instances to security groups and configure security group rules to manage network access for the instances.

Context

Security groups are used as virtual firewalls to provide security isolation and implement network access control for instances.

Security groups determine whether the instances in the same account that are deployed in the same virtual private cloud (VPC) and region can connect to each other over the internal network. By default, if the instances belong to the same security group, they can connect to each other over the internal network. If the instances belong to different security groups, you can authorize mutual access between the security groups to allow the instances to connect to each other over the internal network.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the security group to which you want to add an instance. In the **Actions** column, click the  icon and then select **Manage Instances**.
5. Click **Add Instance**.
6. In the Add Instance dialog box, select an instance and click **OK**.

   An instance can belong to up to five security groups. After an instance is added to a security group, the rules of the security group automatically take effect on the instance.

4.10. Change the private IP address of an instance

Each instance is assigned a private NIC and associated with a private IP address. You can change the private IP address of an instance. The new private IP address must be within the CIDR block of the vSwitch to which the instance is connected and must not be in use by another instance.

Prerequisites

The instance is in the **Stopped** state.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance for which you want to change the IP address, click the ... icon in the **Actions** column. Then, choose **Network and Security Group > Manage Secondary Private IP Addresses**.
5. Enter a new private IP address.

   The new private IP address must be within the CIDR block of the vSwitch to which the instance is connected. This IP address must not be in use by another instance or be reserved for specific purposes.

   For example, if the CIDR block of the vSwitch is 192.168.1.0/24, you can use an IP address within the range from 192.168.1.1 to 192.168.1.254. The first address 192.168.1.0 is reserved to identify the subnet itself, and the last address 192.168.1.255 is reserved as the broadcast address. Neither of the first address and the last address can be used.
6. Click **OK**.

4.11. Assign an IPv6 address to an ECS instance

Compared with IPv4 addresses, IPv6 addresses are more sufficient and allow more types of devices to access the Internet. If your network environment supports IPv6, you can assign IPv6 addresses for exiting Elastic Compute Service (ECS) instances. This topic describes how to assign an IPv6 address to an ECS instance.

Prerequisites

- The vSwitch and virtual private cloud (VPC) of the ECS instance to which you want to assign an IPv6 address are associated with an IPv6 CIDR block. For more information, see the "Create a VPC with an IPv6 CIDR block" topic in *Apsara Stack VPC User Guide*.
- The instance family of the ECS instance supports IPv6.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.

4. On the Instances page, find the instance to which you want to assign an IPv6 address. Click the ... icon in the **Actions** column and choose **Network and Security Group > Manage Secondary Private IP Addresses**.
5. In the **Manage Secondary Private IP Addresses** dialog box, click **Enable** in the **IPv6 Addresses** section.

   > **Note** Specific instance families do not support IPv6.

6. On the Switch details page, check whether IPv6 is enabled for the vSwitch.

   If IPv6 is not enabled for the vSwitch, click **Open IPv6** in the **IPv6 network segment** section.
7. In the **Manage Secondary Private IP Addresses** dialog box, click **OK**.

Result

After the configuration is complete, you can click the ID of the instance to go to the **Instance Details** page and view the **IPv6** parameter to check whether an IPv6 address is assigned to the instance.

4.12. Install the CUDA and GPU drivers for a Linux instance

You must install a GPU driver on GPU instances to use the GPU. If the image you use does not contain a pre-installed GPU driver, you must manually install the CUDA and GPU drivers for the instance.

Prerequisites

If your instance cannot connect to the Internet, the installation file cannot be downloaded. You can install an FTP client on the instance to transfer the installation file to the instance.

Context

When installing NVIDIA drivers, you must install the kernel package that contains the kernel header file before you install the CUDA and GPU drivers on the instance.

Procedure

1. Install the kernel package.
   i. Run the **uname -r** command to view the current kernel version.

   A similar output is displayed:

   - CentOS: `3.10.0-862.14.4.el7.x86_64`
   - Ubuntu: `4.4.0-117-generic`

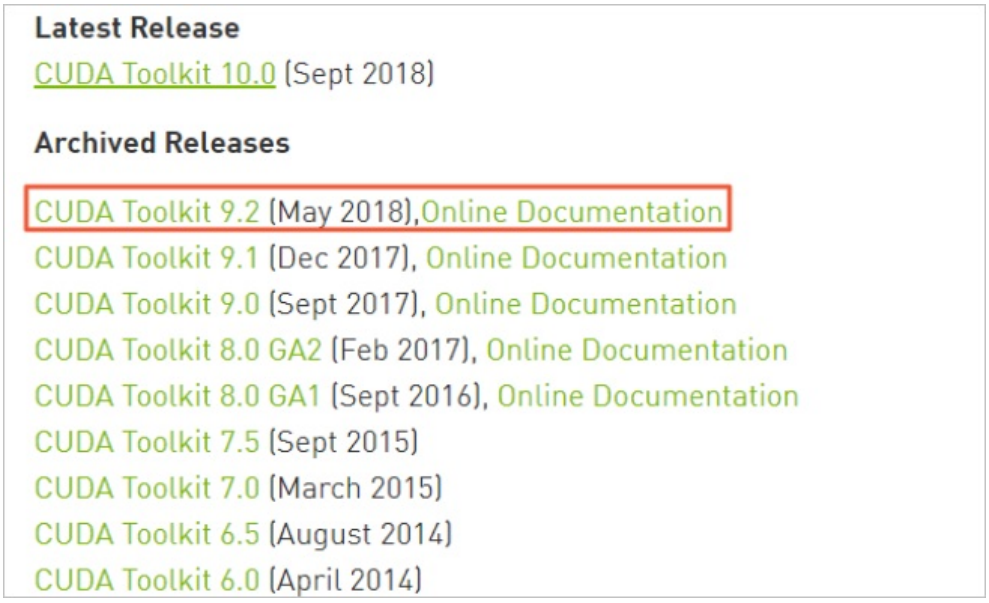
   ii. Copy the kernel package of the corresponding version to the instance and install the package.

   - CentOS: Copy the RPM package of the `kernel-devel` component and run the **rpm -ivh 3.10.0-862.14.4.el7.x86_64.rpm** command to install the package. *3.10.0-862.14.4.el7.x86_64.rpm* is used as an example. Replace it with the actual package name.
   - Ubuntu: Copy the DEB package of the `linux-headers` component and run the **dpkg -i 4.4.0-117-generic.deb** command to install the package. *4.4.0-117-generic.deb* is used as an example. Replace it with the actual package name.

2. Download the CUDA Toolkit.
   i. Access the official CUDA download page. Choose the version based on the GPU application requirements for CUDA.

      This example uses CUDA Toolkit 9.2.

      Download the CUDA Toolkit

   ii. Choose a platform based on your operating system. Select **Installer Type** to **runfile (local)** and click **Download**.

      NVIDIA drivers are already included in the CUDA Toolkit.

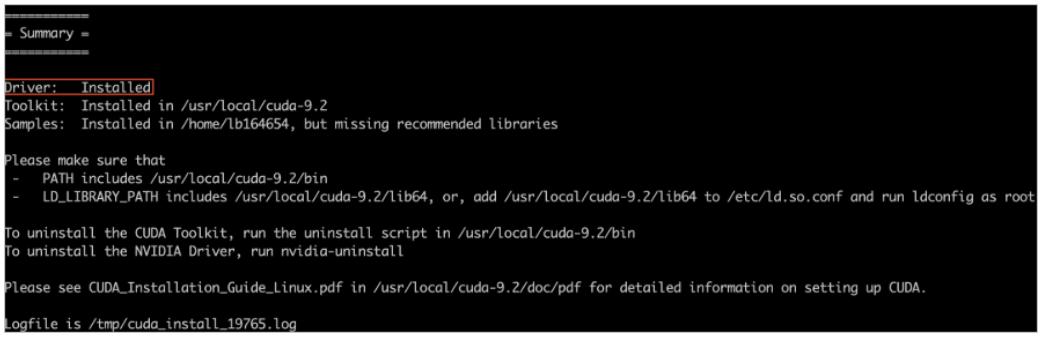
      Download the drivers

3. Copy the downloaded *cuda_9.2.148_396.37_linux.run* file to the instance. *cuda_9.2.148_396.37_linux.run* is used as an example. Replace it with the actual file name.
4. Run the **sudo sh ./cuda_9.2.148_396.37_linux.run --silent --verbose --driver --toolkit --**

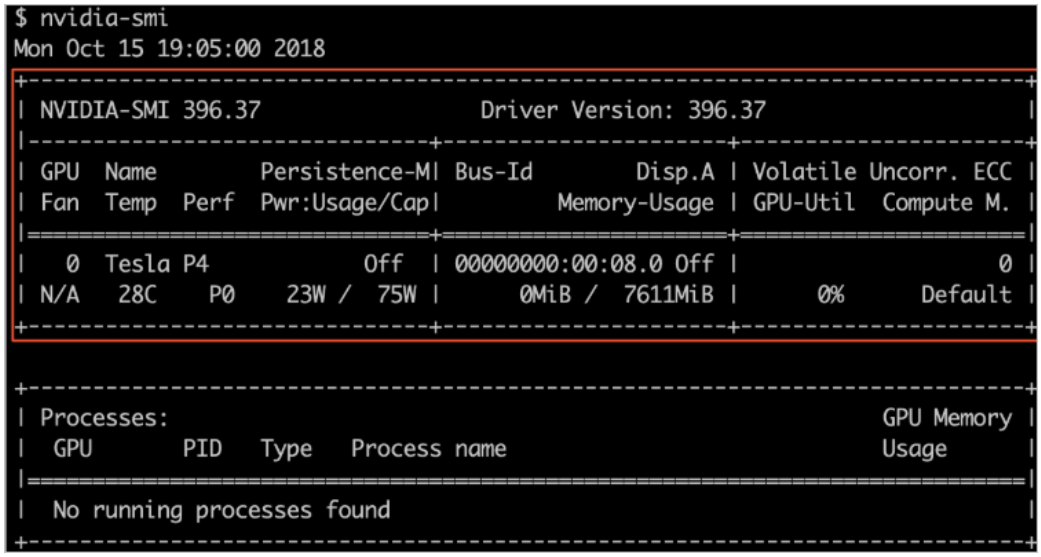
samples command to install the CUDA driver. *cuda_9.2.148_396.37_linux.run* is used as an example. Replace it with the actual file name.
The installation takes about 10 to 20 minutes. When `Driver: Installed` is displayed, the installation is successful.

View the CUDA installation result



5. Run the **nvidia-smi** command to view the GPU driver status.
   If the output displays the details of the GPU driver, the driver is running properly.

   View the GPU driver status



What's next

If you want to run the OpenGL program, you must first purchase the licenses and install the GRID drivers. For information about the installation procedure, see the official NVIDIA documentation.

4.13. Install the CUDA and GPU drivers for a Windows instance

You must install a GPU driver on GPU instances to use the GPU. If the image you use does not contain a pre-installed GPU driver, you must manually install the CUDA and GPU drivers for the instance.

Prerequisites

- If your instance cannot connect to the Internet, the installation file cannot be downloaded. You can install an FTP client on the instance to transfer the installation file to the instance.

- To compile CUDA programs, first install a Windows compiling environment, such as Visual Studio 2015. If you do not need to compile CUDA programs, ignore it.

Procedure

1. Download the CUDA Toolkit.
   i. Access the official CUDA download page. Choose the version based on the GPU application requirements for CUDA.

      This example uses CUDA Toolkit 9.2.

   ii. Choose a platform based on your operating system. Set **Installer Type** to **exe (local)** and click **Download**.

      NVIDIA drivers are already included in the CUDA Toolkit.
2. Copy the downloaded *cuda_9.2.148_windows.exe* file to the instance. *cuda_9.2.148_windows.exe* is used as an example. Replace it with the actual file name.
3. Double-click *cuda_9.2.148_windows.exe* and follow the installation wizard to install the CUDA driver. *cuda_9.2.148_windows.exe* is used as an example. Replace it with the actual file name. The installation takes about 10 to 20 minutes. When `Installed: - Nsight Monitor and HUD Launcher` is displayed, the driver is installed.
4. Press *Win + R* and enter **devmgmt.msc**.
   The NVIDIA device is displayed in **Display Adapter**.
5. Press *Win + R*, enter **cmd**, and run the **"C:\Program Files\NVIDIA Corporation\NVSMI\nvidia-smi"** command.
   If the output displays the details of the GPU driver, the driver is running properly.

What's next

If you want to run the OpenGL and DirectX programs, you must first purchase the licenses and install the GRID drivers. For information about the installation procedure, see the official NVIDIA documentation.

5.Disks

5.1. Overview

This topic describes the categories of disks and operations that can be performed on disks.

Categories of disks

Disks are block-level storage devices provided by Apsara Stack for Elastic Compute Service (ECS) instances. Disks can be classified based on their performance or purposes.

- **Performance-based classification**

  Disks can be classified into ultra disks, shared ultra disks, standard SSDs, shared SSDs, standard performance disks, and premium performance disks based on their performance.

  - Ultra disks and shared ultra disks are ideal for medium I/O load scenarios and deliver up to 5,000 random IOPS.
  - Standard SSDs and shared SSDs are ideal for I/O-intensive scenarios and deliver up to 25,000 random IOPS.
  - Standard performance disks and premium performance disks are ideal for online transaction processing (OLTP) databases and NoSQL databases and deliver up to 25,000 random IOPS.

  > **Notice** Different Cloud Defined Storage (CDS)-Elastic Block Storage (EBS) clusters support different disk categories.
  >
  > - New CDS-EBS clusters support premium performance disks and standard performance disks.
  > - CDS-EBS clusters that were created in Apsara Stack V3.15.0 and earlier support ultra disks, standard SSDs, premium performance disks, and standard performance disks.
  > - Existing EBS clusters continue to provide shared ultra disks and shared SSDs.

  The following table compares the performance of different disk categories.

Category	Standard SSD and shared SSD	Ultra disk and shared ultra disk	Standard performance disk	Premium performance disk
Maximum capacity per disk (GiB)	32,768 GiB	32,768 GiB	32,768 GiB	32,768 GiB
Maximum IOPS	25,000	5,000	5,000	25,000
Maximum throughput (MB/s)	300 MB/s	140 MB/s	140 MB/s	300 MB/s
Formula for calculating the IOPS per disk	min{1,800 + 30 × Capacity, 25,000}	min{1,800 + 8 × Capacity, 5,000}	min{1,800 + 8 × Capacity, 5,000}	min{1,800 + 30 × Capacity, 25,000}

Category	Standard SSD and shared SSD	Ultra disk and shared ultra disk	Standard performance disk	Premium performance disk
Formula for calculating the throughput per disk (MB/s)	min{120 + 0.5 × Capacity, 300}	min{100 + 0.15 × Capacity, 140}	min{100 + 0.15 × Capacity, 140}	min{120 + 0.5 × Capacity, 300}
API parameter value	cloud_ssd	cloud_efficiency	cloud_sperf	cloud_pperf
Use scenario	Small and medium-sized development and testing environments that require high data durability	○ Development and testing applications ○ System disks	○ OLTP databases: relational databases such as MySQL, PostgreSQL, Oracle, and SQL Server databases ○ NoSQL databases: non-relational databases such as MongoDB, HBase, and Cassandra databases ○ Elasticsearch distributed logs: Elasticsearch, Logstash, and Kibana (ELK) log analysis	

- **Purpose-based classification**

  Disks can be classified into system disks and data disks based on their purposes.

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

Related operations

You can perform operations on disks based on your business needs. The following table describes the operations that can be performed on disks.

Operation	Description	References
Create a disk	You can create an empty disk to use as a data disk.	Create a disk
Attach a data disk	Separately created disks can be attached to instances within the same zone only as data disks.	Attach a data disk
Partition and format a disk	Before you can use a separately created disk that is attached to a Linux or Windows instance, you must partition and format the disk on the instance.	• Format a data disk for a Linux instance • Format a data disk of a Windows instance

Operation	Description	References
View disks	You can view the list of created disks and the details of a single disk.	View disks
Roll back a disk by using a snapshot	If you have created snapshots of a disk, you can use a snapshot to roll back the disk to the point in time when the snapshot was created.	Roll back a disk by using a snapshot
Modify the properties of a disk	You can modify the properties of a created disk, including the Release Disk with Instance and Release Automatic Snapshots with Disk properties.	Modify the properties of a disk
Modify the name and description of a disk	You can modify the name and description of a created disk.	Modify the name and description of a disk
Resize a disk	You can resize the system disk or data disks of an instance online or offline. After you resize a disk of an instance offline, you must restart the instance for the resize operation to take effect.	Resize disks
Enable multi-attach	When you create a premium performance disk, you can enable the multi-attach feature for the disk.	Enable the multi-attach feature
Encrypt a disk	You can encrypt a new disk to improve security in a simple and secure manner.	• Encrypt a system disk • Encrypt a data disk
Re-initialize a disk	You can re-initialize a system disk or data disk so that the disk can return to the state it was in when the disk was created.	• Re-initialize a system disk • Re-initialize a data disk
Detach a data disk	You can detach a data disk from an instance when the instance no longer needs the disk. You can also detach a data disk from an instance and then attach the disk to another instance within the same zone.	Detach a data disk
Release a data disk	You can manually release disks that are no longer needed. When a disk is released, all data stored on the disk is deleted.	Release a data disk

5.2. Create a disk

You can create a data disk separately and then attach it to an Elastic Compute Service (ECS) instance to increase the storage space of the instance. This topic describes how to create an empty data disk.

Limits

Determine the number and sizes of data disks that you need before you create them. Take note of the following limits:

> **Note** System disks cannot be created separately.

- The maximum number of data disks that can be attached to a single instance varies based on the instance type.
- A single premium performance disk that has the multi-attach feature enabled can be attached to up to 16 ECS instances at the same time.
- Each ultra disk, shared ultra disk, standard SSD, shared SSD, premium performance disk, or standard performance disk can be up to 32 TiB in size.
- Disks cannot be combined in ECS. They are independent of each other. You cannot combine multiple disks into one by formatting them.

A snapshot can back up data only of a single disk. We recommend that you do not use Logical Volume Manager (LVM) to create logical volumes across disks. If you use LVM to create a logical volume across multiple existing disks, data discrepancy may occur when you restore disks from snapshots.

Create a disk on the Disks page

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create Disk**.
5. Configure the parameters described in the following table.

Section	Parameter	Required	Description
Area	Organization	Yes	Select an organization in which to create the disk.
	Resource Set	Yes	Select a resource set in which to create the disk.
	Region	Yes	Select a region in which to create the disk.
	Zone	Yes	Select a zone in which to create the disk.

Section	Parameter	Required	Description
	Creation Method	Yes	Select a method used to create the disk. Valid values: ○ **Disk Creation** ○ **Storage Set Creation** **Note** Storage Set is a service that provides block storage resources in clusters. Each storage set is physically isolated and allows its owner exclusive access to its resources. You can use storage sets to improve the security and O&M efficiency of your business data storage. For more information about storage sets, see the *Create a storage set* topic in *Cloud Defined Storage (CDS) User Guide*.
	Storage Set	Yes	Select a storage set to use to create the disk. **Note** This parameter is required when you set **Creation Method** to **Storage Set Creation**.
	Partitions	Yes	Specify the number of a partition from the specified storage set to use to create the disk. A storage set must contain two or more partitions. **Note** This parameter is required when you set **Creation Method** to **Storage Set Creation**.
	Name	Yes	Enter a name for the disk. The name must be 2 to 128 characters in length. It must start with a letter but cannot start with http:// or https://. The name can contain letters, digits, colons (:), underscores (_), and hyphens (-).

Section	Parameter	Required	Description
Basic Configurations	Quantity	Yes	Specify the number of disks that you want to create.
	Specifications	Yes	Select a disk category. Different CDS-Elastic Block Storage (EBS) clusters support different disk categories. ○ New CDS-EBS clusters support the following disk categories: ■ **Premium Performance Disk** ■ **Standard Performance Disk** ○ CDS-EBS clusters that were created in Apsara Stack V3.15.0 and earlier support the following disk categories: ■ **Ultra Disk** ■ **Standard SSD** ■ **Premium Performance Disk** ■ **Standard Performance Disk** ○ Existing EBS clusters continue to provide shared ultra disks and shared SSDs. **Notice** If you have used shared ultra disks and shared SSDs, you can continue to use them. If you have never used shared ultra disks and shared SSDs, you cannot use disks of these categories.
	Disk Size	Yes	Specify a disk size in the range of 20 GiB to 32,768 GiB.
	Multi-attach	No	Specify whether to enable multi-attach for the disk. For more information, see Overview of disks that support NVMe. **Note** This parameter is required when you set **Specifications** to **Premium Performance Disk**.
	Encryption	No	Specify whether to encrypt the disk.

Section	Parameter	Required	Description
	Encryption Method	No	Select an encryption algorithm. This parameter is required when you set **Encryption** to **Yes**. Valid values: ○ **AES256** ○ **SM4**
	Encryption Key	No	Select a key to use to encrypt the disk. This parameter is required when you set **Encryption** to **Yes**. **Note** If no keys are available, create a key in **Key Management Service (KMS)**.
	Use Snapshot	No	Specify whether to create the disk from a snapshot. If you select **Yes**, you must specify a snapshot. The size of the disk is determined based on the size of the selected snapshot. ○ If the disk size that you specify is greater than the snapshot size, the disk is created with the size that you specified. ○ If the disk size that you specify is smaller than the snapshot size, the disk is created with the snapshot size.

6. Click **Submit**.

Create a disk on the Instance Details page

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Instances page, click the instance ID in the **Instance ID/Name** column.
5. On the **Instance Details** page, click the **Disks** tab.
6. Click **Create Disk**.
7. Configure the parameters described in the following table.

Section	Parameter	Required	Description
	Organization	Yes	Select an organization in which to create the disk.

Section	Parameter	Required	Description
Area	Resource Set	Yes	Select a resource set in which to create the disk.
	Region	Yes	Select a region in which to create the disk.
	Zone	Yes	Select a zone in which to create the disk.
	Creation Method	Yes	Select a method used to create the disk. Valid values: ○ **Disk Creation** ○ **Storage Set Creation** **Note** Storage Set is a service that provides block storage resources in clusters. Each storage set is physically isolated and allows its owner exclusive access to its resources. You can use storage sets to improve the security and O&M efficiency of your business data storage. For more information about storage sets, see the *Create a storage set* topic in *Cloud Defined Storage (CDS) User Guide*.
	Storage Set	Yes	Select a storage set to use to create the disk. **Note** This parameter is required when you set **Creation Method** to **Storage Set Creation**.
	Partitions	Yes	Specify the number of a partition from the specified storage set to use to create the disk. A storage set must contain two or more partitions. **Note** This parameter is required when you set **Creation Method** to **Storage Set Creation**.

Section	Parameter	Required	Description
Basic Configurations	Name	Yes	Enter a name for the disk. The name must be 2 to 128 characters in length. It must start with a letter but cannot start with http:// or https://. The name can contain letters, digits, colons (:), underscores (_), and hyphens (-).
	Quantity	Yes	Specify the number of disks that you want to create.
	Specifications	Yes	Select a disk category. Different CDS-Elastic Block Storage (EBS) clusters support different disk categories. ◦ New CDS-EBS clusters support the following disk categories: ■ **Premium Performance Disk** ■ **Standard Performance Disk** ◦ CDS-EBS clusters that were created in Apsara Stack V3.15.0 and earlier support the following disk categories: ■ **Ultra Disk** ■ **Standard SSD** ■ **Premium Performance Disk** ■ **Standard Performance Disk** ◦ Existing EBS clusters continue to provide shared ultra disks and shared SSDs. **Notice** If you have used shared ultra disks and shared SSDs, you can continue to use them. If you have never used shared ultra disks and shared SSDs, you cannot use disks of these categories.
	Disk Size	Yes	Specify a disk size in the range of 20 GiB to 32,768 GiB.

Section	Parameter	Required	Description
	Multi-attach	No	Specify whether to enable multi-attach for the disk. For more information, see Overview of disks that support NVMe. **Note** This parameter is required when you set **Specifications** to **Premium Performance Disk**.
	Encryption	No	Specify whether to encrypt the disk.
	Encryption Method	No	Select an encryption algorithm. This parameter is required when you set **Encryption** to **Yes**. Valid values: ○ **AES256** ○ **SM4**
	Encryption Key	No	Select a key to use to encrypt the disk. This parameter is required when you set **Encryption** to **Yes**. **Note** If no keys are available, create a key in **Key Management Service (KMS)**.
	Use Snapshot	No	Specify whether to create the disk from a snapshot. If you select **Yes**, you must specify a snapshot. The size of the disk is determined based on the size of the selected snapshot. ○ If the disk size that you specify is greater than the snapshot size, the disk is created with the size that you specified. ○ If the disk size that you specify is smaller than the snapshot size, the disk is created with the snapshot size.

8. Click **Submit**.

Result

The disk that you created is displayed in the disk list and is in the **Available** state.

What's next

After the disk is created, you can attach the disk and partition and format it. For more information, see the following topics:

- Attach a disk
- Format a data disk for a Linux instance
- Format a data disk of a Windows instance

5.3. Attach a data disk

You can attach a separately created disk as a data disk to an instance that resides in the same zone.

Prerequisites

- The disk resides in the same zone as the instance to which you want to attach the disk.
- The instance is in the **Running** (*Running*) or **Stopped** (*Stopped*) state.
- The disk is in the **Available** (*Available*) state.

Context

Before you attach a disk, take note of the following items:

- Disks that were created along with instances are already attached to the instances.
- A disk can only be attached to an instance that is located in the same zone and region as the disk.
- Each disk can be attached only to a single instance at the same time.
- Each Shared Block Storage device can be attached to up to four instances at the same time.

Attach a disk on the Instance Details page

If you want to attach multiple disks to an instance, go to the details page of the instance.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Instances page, find the instance to which you want to attach disks and click the instance ID.
5. On the **Instance Details** page, click the **Disks** tab.
6. Click **Attach Disk**.
7. In the **Attach Disk** dialog box, select a disk from the **Disk** drop-down list.
8. Click **OK**.
9. Create partitions and file systems on the disk after the disk is attached to the instance.

> **Note** After the disk is attached to the instance, the disk cannot be used on the instance. For example, if the disk is attached to a Linux instance, you cannot view the mount information of the disk by running the df -h command. Perform the following operations based on the operating system type of the instance to make the disk usable on the instance.

- Linux: Partition and format the disk on the instance. For more information, see Format a data disk for a Linux instance.
- Windows Server: Partition and format the disk on the instance. For more information, see Format a data disk of a Windows instance.

Attach a disk on the Disks page

If you want to attach multiple disks to different instances, go to the Disks page.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find a disk for which the **Available** is displayed in the **Status** column, click the ••• icon in the **Actions** column, and then click **Attach**.
5. In the Attach dialog box, configure the parameters described in the following table.

Parameter or option	Description
Instance	Select the instance to which you want to attach the disk.
Release Disk with Instance	If you select this option, the disk is automatically released when its associated instance is released. If you do not select this option, the disk is retained when its associated instance is released. **Note** If the disk that you want to attach is a system disk that was detached from another instance, the instance specified by **Release Disk with Instance** refers to the instance from which the disk was detached, not the current instance.

6. Click **OK**.
   When the disk is attached, its state changes to **Running**.
7. Create partitions and file systems on the disk after the disk is attached to the instance.

   Note After the disk is attached to the instance, the disk cannot be used on the instance. For example, if the disk is attached to a Linux instance, you cannot view the mount information of the disk by running the df -h command. Perform the following operations based on the operating system type of the instance to make the disk usable on the instance.

   - Linux: Partition and format the disk on the instance. For more information, see Format a data disk for a Linux instance.
   - Windows Server: Partition and format the disk on the instance. For more information, see Format a data disk of a Windows instance.

5.4. Partition and format disks

5.4.1. Format a data disk for a Linux instance

Data disks created separately are not partitioned or formatted. This topic describes how to partition and format a data disk of a Linux instance.

Prerequisites

The disk has been attached to the instance.

Procedure

1. Connect to the instance.
2. Run the **fdisk -l** command to view all data disks attached to the ECS instance.

   If */dev/vdb* is not displayed in the command output, the ECS instance does not have a data disk. Check whether the data disk is attached to the instance.

```
[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
```

3. Create partitions for the data disk.
   i. Run the **fdisk /dev/sdb** command.
   ii. Enter *n* to create a new partition.
   iii. Enter *p* to set the partition as the primary partition.
   iv. Enter a partition number and press the Enter key. In this example, *1* is entered to create Partition 1.
   v. Enter the number of the first available sector. This example uses the default value. This is done by pressing the Enter key. You can also enter a value from 1 to 41610 and press the Enter key.
   vi. Enter the number of the last sector. This example uses the default value. This is done by pressing the Enter key. You can also enter a value from 1 to 11748 and press the Enter key.
   vii. (Optional)Optional. To create multiple partitions, repeat steps b through f until all four primary partitions are created.

viii. 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!
```

4. Run the **fdisk -l** command to view the partitions.

   If */dev/vdb1* is displayed in the command output, new partition *vdb1* is 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
```

5. Format the new partition. In this example, the new partition is formatted as ext3 after you run the **mkfs.ext3 /dev/vdb1** command.

   The time required for formatting depends on the disk size. You can also format the new partition to another file system. For example, you can run the **mkfs.ext4 /dev/vdb1** command to format the partition as ext4.

   Compared with ext2, ext3 only adds the log function. Compared with ext3, ext4 improves on some important data structures. ext4 provides better performance and reliability, and more 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.
```

6. Run the `echo '/dev/vdb1 /mnt ext3 defaults 0 0' >> /etc/fstab` command to write the information of the new partition to the /etc/fstab file. You can run the `cat /etc/fstab` command to view the new partition information.

   Ubuntu 12.04 does not support barriers. To write the information of the new partition into the /etc/fstab file, you must run the `echo '/dev/vdb1 /mnt ext3 barrier=0 0 0' >> /etc/fstab` command.

   In this example, the partition information is added to the ext3 file system. You can also modify the ext3 parameter to add the partition information to another type of file system.

   To attach the data disk to a specific folder, for example, to store web pages, modify the /mnt part of the preceding command.

```
[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-bc96-057dd64f**** / 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
```

7. Mount the new partitions. Run the **mount -a** command to mount all the partitions listed in */etc/fstab* and run the **df -h** command to view the result.

If the following information is displayed, the new partitions are mounted and 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
```

5.4.2. Format a data disk of a Windows instance

Data disks created separately are not partitioned or formatted. This topic describes how to partition and format a data disk of a Windows instance. This example uses Windows Server 2008.

Prerequisites

The disk has been attached to an instance.

Procedure

1. In the lower-left corner of the screen, click the **Server Manager** icon.
2. In the left-side navigation pane of the Server Manager window, choose **Storage > Disk Management**.
3. Right-click an empty partition and select **New Simple Volume** from the shortcut menu.

   If the disk status is **Offline**, change it to **Online**.
4. Click **Next**.
5. Set the size of the simple volume, which is the partition size. Then click **Next**.

   The default value is the maximum value of the disk space. You can specify the partition size as needed.
6. Specify the drive letter and then click **Next**.
7. Specify the formatting options and then click **Next**.

   We recommend that you format the partition with the default settings provided by the wizard.
8. When the wizard prompts that the partition has been completed, click **Finish** to close the wizard.

5.5. View disks

You can view the list of created disks and the details of a single disk.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Use one of the following methods to search for disks:
   - Select a filter option from the drop-down list and enter relevant information in the search box. Then, the system searches for disks based on your specified filter condition.
   - Click **Advanced Filter**, specify multiple filter options, and then click **Search**. Eligible disks are

displayed in the disk list.

> **Note** You can use the Advance Filter feature and specify multiple filter options to narrow down search results.

Filter option	Description
Disk Name	Enter the name of the disk. **Note** If the disk is created along with an instance, the name of the disk complies with the following naming conventions: ○ If the disk is a system disk, the name of the disk is the <Instance name>_systemDisk format. Example: ecstest_systemDisk. ○ If the disk is a data disk, the name of the disk is in the <Instance name>_DataDisk_<Serial number> format. Example: ecstest_DataDisk_001 and ecstest_DataDisk_002.
Disk ID	Enter the ID of the disk.
Encryption Key ID	Enter the ID of the encryption key used to encrypt the disk.
Storage Set	Enter the storage set to which the disk belongs.
Partitions	The serial number of the storage set partition to which the disk belongs.
Instance ID	Enter the ID of the instance to which the disk is attached.
Snapshot Policy ID	Enter the ID of an automatic snapshot policy that applies to the disk.
Tag	Enter the key or value of a key of the disk in the tag filter.

5. Click the ID of the disk in the **Disk ID/Name** column.

   In the panel that appears, the attributes and attach information of the disk are displayed.

5.6. Roll back a disk by using a snapshot

If you have created a snapshot of a disk, you can use the snapshot to roll back the disk to the point in time when the snapshot was created. The disk rollback operation is irreversible. After the disk is rolled back, the data that was stored on the disk before the rollback operation is performed is lost and cannot be recovered. Exercise caution when you perform this operation.

Prerequisites

Before you roll back a disk by using a snapshot, make sure that the following requirements are met:

- Snapshots have been created for the disk, and no snapshots are currently being created for the disk. For more information, see Create a snapshot.
- The disk is not released.
- The disk is attached to an instance that is in the **Stopped** state.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Snapshots**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Disk Snapshots** tab.
5. Find a snapshot of the disk that you want to use to roll back and click **Roll Back Disk** in the **Actions** column.

   > **Note** You can roll back only one disk on an instance at a time. Other disks that are attached to the instance are not affected. After the rollback operation is complete, the entire disk (not a partition or a directory) is restored to the state that the disk was in when the snapshot was created.

6. In the message that appears, click **Roll Back Disk**.

What's next

- After the disk is rolled back, the hosts configuration file and the configurations such as the hostname, SSH key pair, passwords, network settings, operating system repository settings, and clock source are initialized. You must reconfigure the file and configurations
- If you have resized the disk after you created the snapshot for the disk, you must log on to the instance to resize the file systems on the disk again after the disk is rolled back. For more information, see Expand a disk.

5.7. Modify the properties of a disk

You can modify the properties of a created disk, such as the configurations of the Release Disk with Instance and Release Automatic Snapshots with Disk options.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the disk whose information you want to modify, click the ... icon in the **Actions** column, and then select **Modify Disk Properties**.
5. In the **Modify Disk Properties** dialog box, modify the Release Mode settings. You can specify one of the following release modes:
   - **Release Disk with Instance**: If this option is selected, the disk is released when the instance to which the disk is attached is deleted. If this option is not selected, the disk is retained and enters

the **Pending** state when the instance to which the disk is attached is deleted.

- **Release Automatic Snapshots with Disk**: If this option is selected, the automatic snapshots created for the disk are released when the disk is deleted. If this option is not selected, the automatic snapshots are retained when the disk is deleted.

6. Click **OK**.

5.8. Modify the name and description of a disk

You can modify the names and descriptions of disks.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the disk whose information you want to modify, click the ... icon in the **Actions** column, and then click **Modify Disk Description**.
5. Modify the name and description of the disk.
   - The disk name must be 2 to 128 characters in length. It must start with a letter and cannot start with http:// or https://. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,).
   - The description must be 2 to 256 characters in length and cannot start with http:// or https://.
6. Click **OK**.

5.9. Resize disks

You can resize your disks to store a larger volume of data as your storage requirements increase. You can resize the system disk or data disks of an instance online or offline. After you resize a disk of an instance offline, you must restart the instance for the resize operation to take effect.

Prerequisites

- To prevent data loss, we recommend that you create a snapshot to back up disk data before you resize a disk. For more information, see Create snapshots.
- No snapshot is being created for the disk that you want to resize.
- The following requirements are met:
  - If the disk is a system disk, the associated instance is in the **Running** state.
  - If the disk is a data disk, one of the following requirements is met:
    - The disk is in the **Pending** state.
    - If the disk is attached to an instance, the instance is in the **Running** state.
  - If the disk is a Shared Block Storage device, the disk is in the **Pending** state.

Context

You can use one of the following methods to increase the storage capacity of a single instance:

> **Note** When you resize a disk, only the storage capacity of the disk is increased. The sizes of partitions and file systems do not change.

- Resize an existing disk. You can resize the existing partitions of the disk or create more partitions on the disk.

  The following table describes the methods that can be used to resize an existing disk.

Method	Usage notes
Resize a disk online	○ The instance is in the **Running** state. ○ After you resize the disk, the new size automatically takes effect without the need to restart the associated instance.
Resize a disk offline	○ The instance is in the **Running** or **Stopped** state. ○ After you resize the disk, you must restart the associated instance in the Elastic Compute Service (ECS) console or by calling the RebootInstance operation for the new size to take effect.

- Create a disk, attach the disk to an ECS instance as a data disk, and then partition and format the disk.
- Replace the system disk of an instance and specify a larger size for the new system disk.

Limits

Before you resize disks, take note of the following items.

Item	Limits

Item	Limits

Disk category	Different CDS-EBS clusters support different disk categories.
	• New CDS-EBS clusters support the following disk categories:
	  ◦ **Premium performance disk**
	  ◦ **Standard performance disk**
	• CDS-EBS clusters that were created in Apsara Stack V3.15.0 and earlier support the following disk categories:
	  ◦ **Ultra disk**
	  ◦ **Standard SSD**
	  ◦ **Premium performance disk**
	  ◦ **Standard performance disk**
	• Existing Elastic Block Storage (EBS) clusters continue to provide shared ultra disks and shared standard SSDs.
	Notice If you have used shared ultra disks and shared SSDs, you can continue to use them. If you have never used shared ultra disks and shared standard SSDs, you cannot use disks of these categories.

Operating system	The system disks of Windows Server 2003 instances cannot be resized.
Partitioning mode	If a data disk uses the MBR partition format, the data disk cannot be resized to more than 2 TiB. If you want to resize a data disk to more than 2 TiB and the data disk uses the MBR partition format, we recommend that you create and attach another data disk. Then, format a GPT partition and copy the data in the MBR partition to the GPT partition.
File system	For Windows instances, only disks that use NTFS file systems can be resized.

Capacity	• System disk:
	  ◦ The new capacity must be greater than the current capacity.
	  ◦ The new capacity must be smaller than or equal to 2,048 GiB.
	For example, the system disk of a CentOS instance is 35 GiB in size. When you resize this system disk, the specified new size must be larger than 35 GiB but cannot exceed 2,048 GiB.
	• Data disk:
	  ◦ The new capacity must be greater than the current capacity.
	  ◦ The new capacity must be smaller than or equal to 32,768 GiB.
	For example, a data disk of a CentOS instance is 35 GiB in size. When you resize this data disk, the specified new size must be larger than 35 GiB but cannot exceed 32,768 GiB.

Item	Limits
Operation	• When you resize a disk, only the storage capacity of the disk is increased. The sizes of partitions and file systems do not change. After the disk is resized, you must manually re-allocate the storage space on the disk. • You cannot roll back a resize operation on a disk to decrease the storage capacity of the disk.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the disk that you want to resize, click the ... icon in the **Actions** column, and then select **Resize Disk**.
5. In the **Resize Disk** dialog box, select a resizing method and specify the new capacity.
   - Resizing methods:
     - Offline resizing: After you resize the disk, you must restart the associated instance in the ECS console or by calling the RebootInstance operation for the new size to take effect.
     - Online resizing: After you resize the disk, the new size automatically takes effect without the need to restart the associated instance.

       For more information, see Restart an instance or the "RebootInstance" topic in *ECS Developer Guide*.
   - Maximum new capacity:
     - System disk: 2,048 GiB
     - Data disk: 32,768 GiB
6. Click **OK**.

What's next

After a disk is resized, you must manually re-allocate the storage space on the disk.

5.10. Support the NVMe protocol and multi-attach feature

5.10.1. Overview of disks that support NVMe

Non-Volatile Memory Express (NVMe) is a host controller interface protocol used to accelerate the transfer of data from non-volatile memory. Alibaba Cloud premium performance disks support NVMe. Each premium performance disk can be simultaneously attached to multiple Elastic Compute Service (ECS) instances that support NVMe for data sharing. This topic describes premium performance disks that support NVMe, limits on attaching this type of disks, and operations related to this type of disks.

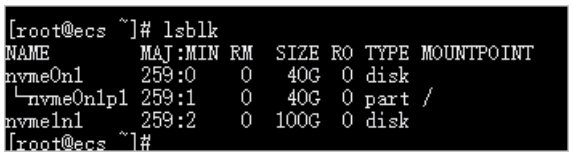
Premium performance disks that support NVMe

Premium performance disks can be attached to multiple ECS instances. After premium performance disks are attached to multiple instances, the disks support concurrent read and write access from these ECS instances and provide high reliability, high concurrency, and high performance. Premium performance disks provide the multi-attach and I/O blocking features.

- After the multi-attach feature is enabled for a premium performance disk, the disk can be attached to up to 16 ECS instances at a time.
- You can run NVMe commands to manage the permissions of ECS instances on premium performance disks. For more information about NVMe commands, see NVM Express Base Specification.

The preceding features improve service availability without compromising data reliability. If a single point of failure (SPOF) occurs, you can use a premium performance disk to quickly schedule and restore data. Data sharing among multiple ECS instances greatly reduces storage costs and improves service flexibility. Premium performance disks are suitable for high-availability databases and distributed database clusters that each consist of one write node and multiple read-only nodes.

Premium performance disks can be attached to ECS instances that support NVMe. For example, after premium performance disks are attached to Linux instances based on NVMe, you can run the **lsblk** command to check the device names and partition names of the disks, as shown in the following figure.



Description of the command output:

- The device names of the premium performance disks are displayed in the `/dev/nvmeXn1` format. Examples: */dev/nvme0n1*, */dev/nvme1n1*, and */dev/nvme2n1*.
- The partition names of the premium performance disks are displayed in the `<Device name of the disk>p<Partition number>` format. Examples: */dev/nvme0n1p1*, */dev/nvme1n1p1*, and */dev/nvme1n1p2*.

Premium performance disks provide the multi-attach feature and are used by enterprises to migrate high-availability services to the cloud. For more information, see Enable the multi-attach feature.

Limits

Before you attach premium performance disks to an ECS instance based on NVMe, the resources of the instance must meet the limits described in the following table.

Item	Limits

Item	Limits
Instance family	By default, the instance family must support NVMe. The following instance families support NVMe: • ecs.ebmg7s-se-x25-c1m8 • ecs.ebmg7m-se-x25-c1m8 • ecs.ebmg7x-se-x25-c1m8 • ecs.ebmg7s-se-numaoff-x25-c1m8 • ecs.ebmg7m-se-numaoff-x25-c1m8 • ecs.ebmg7x-se-numaoff-x25-c1m8 • ecs.g7x-se-x25 **Note** You can call the DescribeInstanceTypes operation to query instance families and check whether the instance family supports NVMe based on the NvmeSupport parameter in the response. For more information, see the *DescribeInstanceTypes* topic in *ECS Developer Guide*.
Image	The image must contain the NVMe driver. The NVMe driver is installed in the following public images. **Note** Only some public Linux images support the NVMe driver. • CentOS 7: CentOS 7.6 and later • CentOS 8: CentOS 8.0 and later

Disk	• Disk category: premium performance disk
	• Creation method:
	  ◦ Create premium performance disks when you create instances that support NVMe.
	  ◦ When you create disks, select premium performance disks as the disk category and enable the multi-attach feature.

Related operations

You can perform the following operations on premium performance disks that support NVMe:

- Create premium performance disks when you create ECS instances that support NVMe. In this case, the created premium performance disks support NVMe. For more information, see Create an instance by using the wizard.
- Enable the multi-attach feature when you separately create premium performance disks. For more information, see Enable the multi-attach feature.
- Format premium performance disks and create file systems. For more information, see Format a data disk for a Linux instance.

> **Notice** The preceding operations are applicable only to premium performance disks that are created together with instances but not to premium performance disks for which the multi-attach feature is enabled.

5.10.2. Enable the multi-attach feature

When you create a premium performance disk, you can enable the multi-attach feature for the disk. After multi-attach is enabled for a premium performance disk, the disk can be attached to up to 16 Elastic Compute Service (ECS) instances that support the Non-Volatile Memory Express (NVMe) protocol within the same zone to allow concurrent read and write access from the instances.

Benefits

This feature is suitable for high-availability databases and distributed database clusters that each consist of one write node and multiple read-only nodes. This feature provides the following benefits:

- Usage of NVMe commands: NVMe commands can be used to manage permissions of ECS instances on premium performance disks. This helps improve service availability without compromising data durability. For more information about NVMe commands, see NVM Express Base Specification.
- Cross-instance data sharing: This feature enables data sharing across multiple ECS instances to reduce storage costs and improve service flexibility.
- Disaster recovery: This feature allows quick scheduling of services to normal ECS instances to ensure service continuity in single-point-of-failure (SPOF) scenarios.

Limits

The following limits apply to the multi-attach feature:

- The following list provides instance families that support this feature.
  - ecs.ebmg7s-se-x25-c1m8
  - ecs.ebmg7m-se-x25-c1m8
  - ecs.ebmg7x-se-x25-c1m8
  - ecs.ebmg7s-se-numaoff-x25-c1m8
  - ecs.ebmg7m-se-numaoff-x25-c1m8
  - ecs.ebmg7x-se-numaoff-x25-c1m8
  - ecs.g7x-se-x25
- Data disks are supported but system disks are not supported.
- The multi-attach feature can be enabled only when you create premium performance disks. This feature cannot be enabled or disabled after premium performance disks are created.
- After the multi-attach feature is enabled for a premium performance disk, we recommend that you use cluster file systems such as Oracle Cluster File System version 2 (OCFS2), Global File System 2 (GFS2), Veritas Cluster File System (Veritas CFS), Oracle Automatic Storage Management Cluster File System (Oracle ACFS), and Databricks File System (DBFS) on the disk.

> **Warning** When a premium performance disk for which the multi-attach feature is enabled is attached to multiple ECS instances, data cannot be synchronized among the instances and data inconsistency may occur if file systems such as EXT2, EXT3, EXT4, XFS, and New Technology File System (NTFS) are used.

- The performance of premium performance disks is limited. When a premium performance disk is attached to multiple ECS instances, the total performance of the disk on all instances cannot exceed the maximum performance that can be provided by the disk.

The following table describes the limits that apply to the features provided by premium performance disks for which the multi-attach feature is enabled.

Feature	Limits
Disk attaching	A single premium performance disk can be attached to up to 16 instances that support the NVMe protocol.
Release of disks with instances	Not supported.
Disk re-initialization	Not supported.
Disk category change	Not supported.
Disk resizing	Only offline resizing is supported for disks. For more information, see Resize disks.
Snapshot-consistent group	Not supported.

Procedure

Perform the following steps to use the multi-attach feature:

1. Create a premium performance disk for which the multi-attach feature is enabled. For more information, see Step 1: Create a premium performance disk for which the multi-attach feature is enabled.
2. Attach the disk to multiple ECS instances that support the NVMe protocol. For more information, see Step 2: Attach the premium performance disk to multiple instances that support the NVMe protocol.

Step 1: Create a premium performance disk for which the multi-attach feature is enabled

To use the multi-attach feature, enable the feature when you create a premium performance disk.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create Disk**.
5. Configure the parameters described in the following table for the premium performance disk that you create.

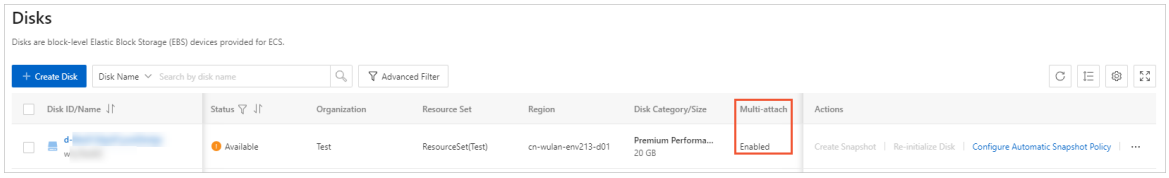
Type	Parameter	Required	Description
Area	Organization	Yes	Select the organization to which the disk belongs.
	Resource Set	Yes	Select the resource set to which the disk belongs.
	Region	Yes	Select the region in which you want to deploy the disk.
	Zone	Yes	Select the zone in which you want to deploy the disk.
	Creation Method	Yes	Select the method that you want to use to create the disk. Valid values: Disk Creation and Storage Set Creation. **Note** Storage Set is a service that provides block storage resources in clusters. Each storage set is physically isolated and allows its owner exclusive access to its resources. Storage sets can be used to improve the security and O&M efficiency of your business data storage. For more information about storage sets, see the *Create a storage set* topic in *Cloud Defined Storage (CDS) User Guide*.
	Storage Set	Yes	Select the storage set that you want to use to create the disk. **Note** This parameter is required if you set **Creation Method** to **Storage Set Creation**.
	Partition	Yes	Select the number of partitions in a storage set. The value of this parameter must be greater than or equal to 2. **Note** This parameter is required if you set **Creation Method** to **Storage Set Creation**.

Type	Parameter	Required	Description
Basic Configurations	Name	Yes	Enter a name for the disk that you want to create. The name must be 2 to 128 characters in length. It must start with a letter and cannot start with http:// or https://. The name can contain letters, digits, colons (:), underscores (_), and hyphens (-).
	Quantity	Yes	Enter the number of disks that you want to create.
	Specifications	Yes	Select **Premium Performance Disk**.
	Disk Size	Yes	Specify a disk size in the range of 20 GiB to 32,768 GiB.
	Multi-Attach	Yes	Select whether to enable the multi-attach feature for the disk.
	Encryption	No	Select whether to encrypt created disks.
	Encryption Method	No	Select the encryption algorithm that you want to use to encrypt the disk. This parameter is required if you set **Encryption** to **Yes**. Valid values: ○ **AES256** ○ **SM4**
	Encryption Key	No	Select the key that you want to use to encrypt the disk. This parameter is required if you set **Encryption** to **Yes**. **Note** If no keys are available, create a key in **Key Management Service**.

Type	Parameter	Required	Description
	Use Snapshot	No	Select whether to create the disk from a snapshot. If you select **Yes**, you must specify a snapshot. In this case, the size of the created disk may be affected by the snapshot size. ○ If the disk size that you specify is greater than the snapshot size, the disk is created with the size that you specify. ○ If the disk size that you specify is smaller than the snapshot size, the disk is created with the snapshot size.

> **Note** For more information about how to create disks, see Create a disk.

6. Confirm the configurations and click **Submit**.
7. In the message that appears, click **OK**.
   After the disk is created, you can go to the **Disks** page to view the created disk. The state of the disk is **Available**, and **Enabled** is displayed in the Multi-attach column.



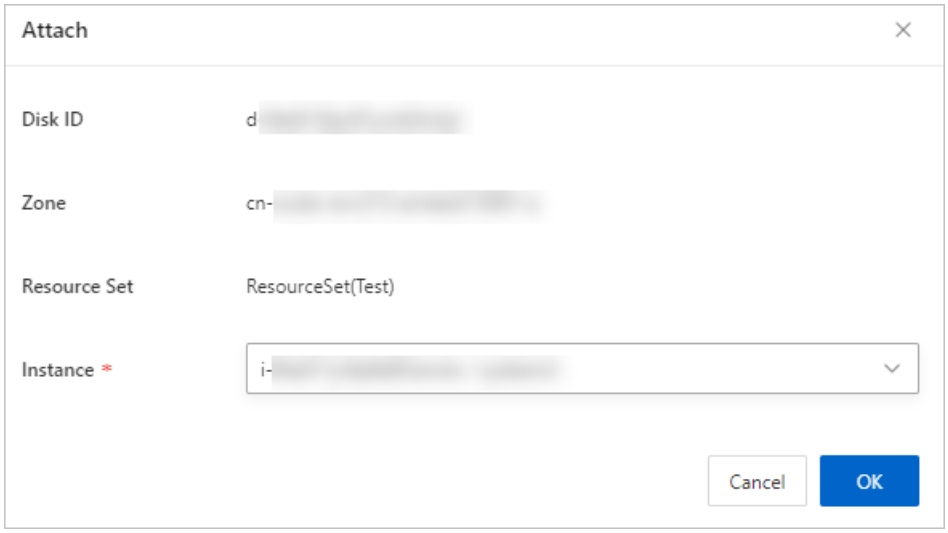
Step 2: Attach the premium performance disk to multiple instances that support the NVMe protocol

Before you attach the premium performance disk to instances, make sure that the following requirements are met:

- The premium performance disk and the instances reside in the same zone.
- The instance families and images of the instances comply with the NVMe protocol. For more information, see Limits.

1. Obtain the ID of the instance to which you want to attach the premium performance disk.
   In the left-side navigation pane, choose **Instances & Images > Instances**. On the Instances page, view and copy the instance ID in the instance list.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. Use the instance ID to search for the premium performance disk that you want to attach to the instance. Then, click the ... icon in the **Actions** column corresponding to the disk, and select **Attach**.
4. In the **Attach** dialog box, select an instance ID from the **Instance** drop-down list and click **OK**.
   You must select the instance ID that you obtained in Step 1 from the **Instance** drop-down list.

> **Note** You can attach the premium performance disk to only a single instance each time you perform the preceding steps. To attach the premium performance disk to multiple instances, repeat the preceding steps.



- You can go to the **Disks** page to view the state of the premium performance disk. The state of the disk is displayed as **Running** on this page.
- You can click the name of the premium performance disk to go to the disk detail panel. In the **Attachment Information** section of this panel, you can view information about the premium performance disk that is attached to a single instance or multiple instances.

What to do next

After the premium performance disk is attached to multiple instances, we recommend that you create cluster file systems based on your business requirements. Common cluster file systems include OCFS2, GFS2, Veritas CFS, Oracle ACFS, and DBFS.

> **Warning** When a premium performance disk for which the multi-attach feature is enabled is attached to multiple ECS instances, data cannot be synchronized among the instances and data inconsistency may occur if file systems such as EXT2, EXT3, EXT4, XFS, and New Technology File System (NTFS) are used.

5.11. Encrypt a disk

5.11.1. Overview

Disk encryption is a simple and secure method provided by Elastic Compute Service (ECS) to encrypt new disks.

Disk encryption eliminates the need to create or maintain your own key management infrastructure, to modify existing applications and maintenance procedures, and to perform additional encryption operations. Disk encryption does not have negative impacts on your business. The following types of data can be encrypted in disk encryption operations:

- Data stored on disks.

- Data transmitted between disks and instances. Data within instance operating systems is not encrypted.
- All snapshots created from encrypted disks. These snapshots are encrypted snapshots.

Data transmitted from instances to disks is encrypted on the hosts where the instances reside.

Ultra disks, standard SSDs, premium performance disks, and standard performance disks can be encrypted by using disk encryption.

> **Note** By default, shared SSDs and shared ultra disks cannot be encrypted by using disk encryption. To create and encrypt shared SSDs or shared ultra disks, submit a ticket.

5.11.2. Encrypt a system disk

In scenarios that require data security and regulatory compliance, you can encrypt disks to secure your data stored on the disks. To encrypt system disks, you can encrypt custom images and then use the encrypted custom images to create instances. The system disks of the created instances are automatically encrypted.

Context

You can encrypt system disks only by encrypting custom images.

Procedure

To encrypt a system disk, perform the following steps:

1. Step 1: Create a custom image
2. Step 2: Encrypt the created custom image
3. Step 3: Use the encrypted custom image to create an instance

If you have created a custom image and encrypted the image, you can directly use the encrypted image to create an instance.

Step 1: Create a custom image

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance from which you want to create a custom image, click the ... icon in the **Actions** column, and then choose **Disk and Image > Create Custom Image**.
5. Configure the parameters described in the following table.

Parameter	Description
Image Name	Enter a name for the custom image. The name must be 2 to 128 characters in length and can contain letters, digits, periods (.), underscores (_), hyphens (-), and colons (:). It must start with a letter.

Parameter	Description
Sharing Scope	Select the scope for which to share the custom image. Valid values: ○ **Current Organization and Subordinate Organizations** ○ **Current Resource Set** ○ **Current Organization**
Image Description	Enter a description of the custom image. The description must be 2 to 256 characters in length and cannot start with http:// or https://.

6. Click **OK**.

Step 2: Encrypt the created custom image

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Images**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Images page, click the **Custom Images** tab.
5. On the Custom Images tab, find the image that you want to encrypt. Click the ... icon in the **Actions** column and click **Encrypt Image**.
6. In the **Encrypt Image** dialog box, configure the parameters described in the following table.

Parameter	Description
Image ID	The system automatically obtains the ID of the image to be encrypted. You do not need to configure this parameter.
Name	Enter a name for the new encrypted custom image. The name must be 2 to 128 characters in length and start with a letter. It can contain underscores (_), periods (.), and hyphens (-).
Description	Enter a description for the new encrypted custom image. The description must be 2 to 256 characters in length and cannot start with http:// or https://.

7. Click **OK**.

Step 3: Use the encrypted custom image to create an instance

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create Instance**.
5. On the Create Instance page, configure the parameters.

For more information about how to configure the parameters, see Create an instance

In the **Image** section, set Image Type to **Custom Image**. Then, select the encrypted image described in Step 2 from the **Custom Image** drop-down list. For more information about the image, see Step 2: Encrypt the created custom image.

6. Click **Submit**.

Result

After the instance is created, you can click its ID to go to the **Instance Details** page. Then, you can click the **Disks** tab and check the value in the **Encrypted** column corresponding to the system disk. If the value is **Yes**, the system disk is encrypted.

5.11.3. Encrypt a data disk

In scenarios that require data security and regulatory compliance, you can encrypt disks to secure your data stored on the disks. After a data disk is created, its encryption state cannot be changed. If you want to encrypt a data disk, enable encryption for the disk when you create it.

Context

Determine the number and sizes of data disks that you need before you create them. Take note of the following limits:

> **Note** System disks cannot be created separately.

A snapshot can back up data only of a single disk. We recommend that you do not use Logical Volume Manager (LVM) to create logical volumes across disks. If you use LVM to create a logical volume across multiple existing disks, data discrepancy may occur when you restore disks from snapshots.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create Disk**.
5. On the Create Disk page, configure the parameters.

   When you encrypt the disk, take note of the following parameters:

   - **Encryption**: Select **Yes**.
   - **Encryption Method**: Select an encryption algorithm. Valid values:
     - **AES256**
     - **SM4**
   - **Encryption Key**: Select an encryption key.

   For more information about parameters that you can configure, see Create a disk.

6. Click **Submit**.

5.12. Re-initialize disks

5.12.1. Re-initialize a system disk

This topic describes how to re-initialize a system disk. After a system disk is re-initialized, it is restored to the state that it was in when it was created.

Prerequisites

- The system disk that you want to re-initialize is in the **Running** state.
- The Elastic Compute Service (ECS) instance to which the system disk is attached is in the **Stopped** state.
- After a disk is re-initialized, the data stored on the disk is lost and cannot be recovered. Exercise caution when you perform this operation. We recommend that you back up disk data or create snapshots before you re-initialize a disk. For more information, see Create a snapshot.

Context

The result of disk re-initialization depends on the disk type and how the disk was created.

For the system disk that is attached to an instance, you can re-initialize the disk to restore it to the state that it was in when it was created. The following changes take place after the system disk is re-initialized:

- The system disk is restored to the state that it was in when it was created.

  Warning When the system disk is re-initialized, all data stored in the disk is deleted. We recommend that you create snapshots for the disk to back up data before you re-initialize the disk.

- The automatic snapshot policy applied to the system disk before the re-initialization remains valid.
- The IP addresses and disk IDs of the instance remain unchanged.
- The automatic and manual snapshots of the system disk remain available. You can use these snapshots to roll back the disk. For more information, see Restore a disk.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the system disk that you want to re-initialize and click **Reinitialize Disk** in the **Actions** column.
5. In the **Reinitialize Disk** dialog box, configure the parameters described in the following table.

Parameter	Description
Password	Reset the password used to log on to the instance. You can retain the original password or specify a new password.
Confirm Password	
Instance Startup Policy	If you select **Start Instance After Re-initializing Disk**, the instance automatically starts after you re-initialize the system disk.

6. Click **OK**.

Result

While the disk is being re-initialized, the disk is in the **Initializing** state. When the disk is re-initialized, it enters the **Running** state.

What's next

- If data disks are attached to a Linux instance before its system disk is re-initialized, you must re-create mount points for the partitions of the data disks and mount file systems to these partitions.

  > **Note** After you re-initialize the system disk of a Linux instance, data stored in the data disks attached to the instance remains unchanged, but the mount information of the data disks is lost. You must re-create mount points for the partitions of the data disks and mount file systems to these partitions.

- After the system disk is re-initialized, you must redeploy applications and configurations to restore your business.
- If you have created a snapshot for the system disk before the system disk is re-initialized, you can use the snapshot to create a data disk and attach the data disk to the instance to obtain data that was originally stored in the system disk. For more information, see Create a snapshot.

5.12.2. Re-initialize a data disk

When a data disk is attached to an Elastic Compute Service (ECS) instance, you can re-initialize the disk to restore it to the state when it is created.

Prerequisites

- Snapshots are created for your data disk. For more information, see Create a snapshot.
- The data disk is attached to an ECS instance. For more information, see Attach a disk.
- The ECS instance is in the **Stopped** state.
- For a Linux instance, if you create an empty data disk and add a command in the */etc/fstab* file to mount partitions of the data disk on system startup, the command is not executed and the instance cannot start as expected after you re-initialize the data disk. We recommend that you comment out the command in the */etc/fstab* file. Perform the following steps:
  i. Connect to the instance. For more information, see Instance connecting overview.
  ii. Run the `vim /etc/fstab` command.
  iii. Press the `I` key to enter the edit mode.
  iv. Find the command used to mount data disk partitions and comment out the command by using a number sign (#), as shown in the following command line.

  ```
  # /dev/vdb1 /InitTest ext3 defaults 0 0
  ```

  > **Note** In this example, */dev/vdb1* is a partition and */InitTest* is a mount point. You can modify the command based on your business requirements.

  v. Press the `Esc` key to exit the edit mode, and enter *:wq* to save the file and exit.

Context

The state of a data disk after it is re-initialized varies based on its original state when it was created and the operating system that the instance runs:

- The data disk is restored to the initial state when it was created:
  - The data disk becomes an empty disk if it was originally an empty disk.
  - The data disk stores the data recorded in the source snapshot if the data disk was created from a snapshot.
- For a Windows instance, after you re-initialize a data disk, the data disk is ready for use without the need for additional operations regardless of its original state.
- For a Linux instance:
  - If a data disk was created from a snapshot, the data disk stores only the data recorded in the source snapshot after the data disk is re-initialized. You do not need to re-mount the partitions, but all the data generated after the disk was created is lost.
  - If a data disk was created as an empty disk, all the data and file systems on the disk are lost after the data disk is re-initialized. You must reformat and partition the disk, and then re-mount the partitions.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the data disk that you want to re-initialize, and click **Re-initialize Disk** in the **Actions** column.
5. In the **Re-initialize Disk** dialog box, click **OK**.

What's next

- For a Linux instance, if the data disk was created as an empty disk, you must format the data disk after you re-initialize it. For more information, see Format a data disk for a Linux instance.
- After the data disk is re-initialized, you must redeploy the applications and reconfigure the parameters on the disk to restore your business at the earliest opportunity.

5.13. Detach a data disk

If a data disk is no longer needed by an Elastic Compute Service (ECS) instance, you can detach the disk from the instance. You can also detach a data disk from an instance and then attach the disk to another instance within the same zone. This topic describes how to detach a data disk.

Prerequisites

Before you detach a data disk, make sure that the following conditions are met:

- The disk is attached to an instance and **Running** is displayed in the **Status** column that corresponds to the disk in the ECS console.
- To prevent data loss and ensure data integrity, we recommend that you stop read and write operations on the disk before you detach it.

Context

Perform the following steps to detach a data disk.

> **Note** Local disks that are used as data disks cannot be detached.

1. If file systems are mounted to the partitions of the data disk, unmount the data disk from within the operating system of the instance to which the data disk is attached.

   For more information, see the Step 1: Unmount the data disk from within the operating system of the instance to which the disk is attached section in this topic.

2. Detach the data disk on the Disks page.

   For more information, see the Step 2: Detach the data disk on the Disks page section in this topic.

Step 1: Unmount the data disk from within the operating system of the instance to which the disk is attached

If the data disk is partitioned and has file systems mounted, perform the following operations to unmount the data disk from within the operating system of the instance to which the disk is attached:

For a Linux instance, perform the following procedure.

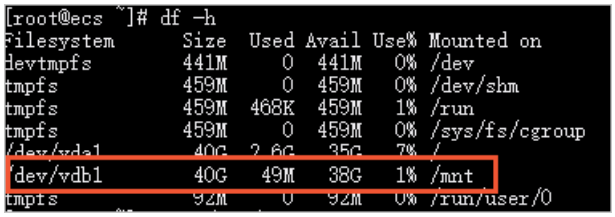
1. Connect to the instance.

   For more information, see Instance connecting overview.

2. Run the following command to view the mount information of the data disk:

   ```
   df -h
   ```

   A command output similar to the following one is displayed. In this example, the /dev/vdb1 partition of the data disk is used. In actual scenarios, look up the command output for the data disk partitions whose mount information you want to view.

3. Run the umount command to unmount the file systems from the data disk partitions.

   For example, you can run the following command to unmount the file system from the /dev/vdb1 partition of the data disk:

   ```
   umount /dev/vdb1
   ```

4. Run the following command to view the UUIDs of the data disk partitions:

   ```
   blkid
   ```

   The following command output shows the UUID of the /dev/vdb1 partition of the data disk.

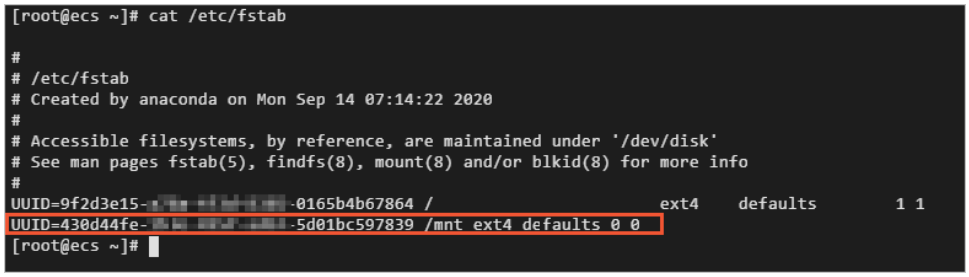
   

5. Run the following command to check whether automatic mounting is configured in the */etc/fstab*

file for the file systems in the data disk partitions:

```
cat /etc/fstab
```

Find the UUID obtained in the previous step in the command output. The file system mounted to the /dev/vdb1 partition is configured in */etc/fstab*, as shown in the following figure.



6. Delete the automatic mounting configurations of the file systems in the data disk partitions from */etc/fstab*.

   > **Note** If you do not delete the automatic mounting configurations of the file systems in the data disk partitions from */etc/fstab*, the instance cannot be restarted after you detach the data disk from the instance in the ECS console.

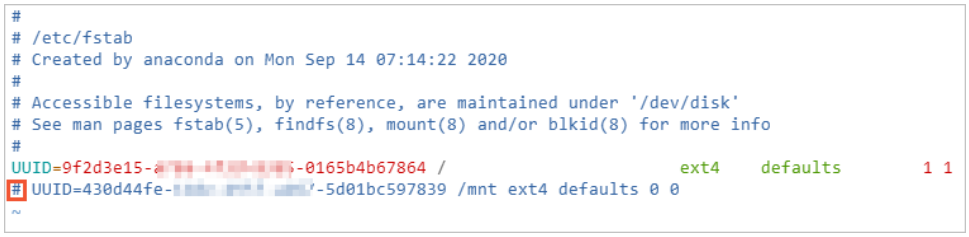
   i. Run the following command to edit */etc/fstab*:

   ```
   vim /etc/fstab
   ```

   ii. Press the `I` key to enter the edit mode.

   iii. Delete or comment out the automatic mounting configurations of the file systems in the data disk partitions.

   For example, you can add the number sign (`#`) before the line of the automatic mounting configuration to comment it out, as shown in the following figure.



   iv. Press the Esc key, enter `:wq` , and then press the Enter key to save your changes and exit the edit mode.

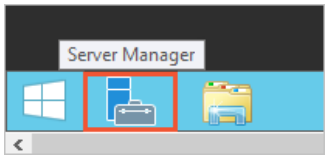
For a Windows instance, perform the following procedure.

> **Note** In this example, Windows Server 2012 R2 is used.

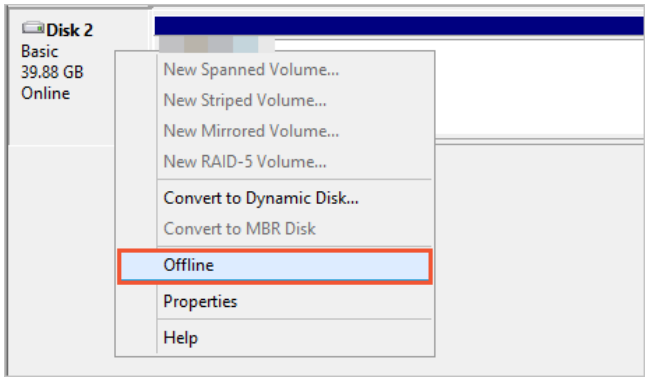
1. Connect to the instance.

   For more information, see Instance connecting overview.

2. On the Windows Server desktop, click the **Server Manager** icon in the lower-left corner.



3. In the upper-right corner of the **Server Manager** window, choose **Tools > Computer Management**.
4. In the left-side navigation pane, choose **Computer Management (Local) > Storage > Disk Management**.
5. Right-click the disk that you want to unmount and select **Offline**.



Step 2: Detach the data disk on the Disks page

You can detach disks from instances on the Disks or Instances page in the ECS console. This section describes how to detach a data disk on the Disks page.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the data disk that you want to detach, click the ... icon in the **Actions** column, and then click **Detach**.
5. In the message that appears, click **Detach**.

Alternatively, you can go to the Instance Details page of the associated instance and detach the data disk from the instance on the **Disks** tab.

Result

After you perform the preceding steps to detach the data disk, you can choose **Storage & Snapshots > Disks** to find the disk on the Disks page. If **Available** is displayed in the **Status** column, the disk is detached.

What's next

- You can attach the data disk to another instance within the same zone. For more information, see Attach a disk.
- When the data disk is no longer needed, you can back up data stored on the disk and then release the disk. For more information, see Create a snapshot and Release a data disk.

5.14. Release a data disk

You can manually release disks that are no longer needed. When a disk is released, all data stored on the disk is deleted. This topic describes how to release a data disk on the Disks page.

Prerequisites

- A snapshot is created for the data disk that you want to release to back up disk data. For more information, see Create a snapshot.
- The data disk is in the **Available** state. If the data disk is attached to an instance, you must detach the disk before you can release it. For more information, see Detach a data disk.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the data disk that you want to release, click the ... icon in the **Actions** column, and then click **Release**.
5. In the message that appears, click **Release**.

6.Images

6.1. Overview

An image is a template for running environments within Elastic Compute Service (ECS) instances. An image includes an operating system and pre-installed applications.

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

Image types

ECS images are classified into public images, custom images, and shared custom images based on image sources.

Image description

Type	Description

Public image	Public images provided by Apsara Stack support the following Windows Server operating systems and mainstream Linux operating systems. Different platforms support different images.
	• Intel x86
	  ◦ Windows
	  ◦ CentOS
	  ◦ Debian
	  ◦ FreeBSD
	  ◦ OpenSUSE
	  ◦ SUSE Linux
	  ◦ Ubuntu
	  ◦ Anolis OS
	  ◦ Alibaba Cloud Linux 2
	• Hygon x86
	  ◦ UOS
	  ◦ Kylin
	  ◦ NFS
	• ARM
	  ◦ UOS
	  ◦ Kylin
	  ◦ CentOS
	  ◦ Anolis OS
	  ◦ Aliyun Linux 2

Type	Description
Custom image	Custom images are created from ECS instances or snapshots or imported from your computer. Custom images can contain applications and data. You can use custom images to create instances that have identical configurations. This eliminates the need to make repeated configurations.
Shared custom image	Shared custom images are shared to your organization by other Alibaba Cloud accounts. Shared images do not count against the image quotas of organizations to which the images are shared. For more information, see Share a custom image.

Find an image

You can find an image based on its type, name, ID, or the ID of the snapshot from which the image was created. Then, you can use the image to create instances or perform other operations. For more information, see View images.

Image formats

ECS supports images in the VHD, RAW, and QCOW2 formats. Images in other formats must be converted to the supported formats before they can be run in ECS. For more information, see Convert the image file format.

6.2. Create a custom image

You can create a custom image and use it to create identical instances or replace the system disks of existing instances.

Create a custom image from a snapshot

You can create a custom image from a system disk snapshot to load the operating system and data environment of the snapshot to the image. Before you perform this operation, make sure that the snapshot to use is a system disk snapshot. Custom images cannot be created only from data disk snapshots.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Snapshots**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find a snapshot for which **System Disk** is displayed in the **Disk Type (All)** column, and click **Create Custom Image** in the **Actions** column.
5. Configure the parameters described in the following table.

Parameter	Description
Sharing Scope	Select the scope for which to share the custom image. Valid values: ○ **Current Organization and Subordinate Organizations** ○ **Current Resource Set** ○ **Current Organization**

Parameter	Description
Image Name	Enter a name for the custom image. The name must be 2 to 128 characters in length and can contain letters, digits, periods (.), underscores (_), hyphens (-), and colons (:). It must start with a letter.
Image Description	Enter a description of the custom image. The description must be 2 to 256 characters in length and cannot start with http:// or https://.

6. Click **OK**.

Create a custom image from an instance

You can create a custom image from an instance to replicate the data of the system disk and data disks on the instance.

> **Note** To prevent data security risks, we recommend that you delete sensitive data from an instance before you use the instance to create a custom image.

When you create a custom image from an instance, a snapshot is automatically generated for each disk on the instance, and all the snapshots constitute a complete custom image.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance from which you want to create a custom image, click the ... icon in the **Actions** column, and then choose **Disk and Image > Create Custom Image**.
5. Configure the parameters described in the following table.

Parameter	Description
Image Name	Enter a name for the custom image. The name must be 2 to 128 characters in length and can contain letters, digits, periods (.), underscores (_), hyphens (-), and colons (:). It must start with a letter.
Sharing Scope	Select the scope for which to share the custom image. Valid values: ○ **Current Organization and Subordinate Organizations** ○ **Current Resource Set** ○ **Current Organization**

Parameter	Description
Image Description	Enter a description of the custom image. The description must be 2 to 256 characters in length and cannot start with http:// or https://.

6. Click **OK**.

6.3. Find an image

You can find an image based on its type, name, ID, or snapshot ID. After you find an image, you can use the image to create an instance or perform other operations. This topic describes how to find an image.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Images**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Select a tab based on the type of the image that you want to view.

   You can select the **Custom Images** or **Public Images** tab.
5. Click Advanced Filter, specify one or more options as filter conditions, and then enter the corresponding information. The image list displays images that match the specified conditions.

Filter option	Description
Image ID	Enter an image ID to search for the image.
Image Name	Enter an image name to search for the image.
Snapshot ID	Enter a snapshot ID to search for images associated with the snapshot. This option is not available for public images.
Tag	Enter the key or value of a tag to search for the images that use the tag. This option is not available for public images.

What's next

After you find an image that matches the preceding filter options, you can perform the following operations:

- Use the image to create an instance. For more information, see Create an instance.
- Share the image. For more information, see Share a custom image.
- Encrypt the image. For more information, see Encrypt a custom image.
- Export the image. For more information, see Export a custom image.
- Delete the image. For more information, see Delete a custom image.
- Modify the description of the image. For more information, see Modify the description of a custom image.

6.4. View instances related to an image

You can view the instances that use a specified image.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Images**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Select a tab based on the type of the image that you want to view.

   You can select the **Custom Images** or **Public Images** tab.
5. Find the image and click **Related Instances** in the **Actions** column.

Result

The Instances page appears and displays the instances that use the image. You can perform operations on these instances, such as updating the image.

6.5. Modify the description of a custom image

You can modify the description of a created custom image in the Elastic Compute Service (ECS) console to manage the image.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Images**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the custom image that you want to modify and click **Edit Description** in the **Actions** column.
5. In the **Edit Description** dialog box, modify the description of the custom image.

   The description must be 2 to 256 characters in length and cannot start with http:// or https://.
6. Click **OK**.

6.6. Share a custom image

You can share a created custom image to the organizations that you manage. Then, the organizations can use the shared image to create instances that have identical environments.

Context

Only custom images can be shared. Shared images do not count against the image quotas of the organizations to which the images are shared.

The organizations to which images are shared can use the shared images to create instances or replace the system disks of existing instances.

You can delete shared images. After a shared image is deleted, the image is no longer visible to the organizations to which the image was shared and the system disks of the instances that were created from the shared image can no longer be re-initialized.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Images**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the image that you want to share, click the ... icon in the **Actions** column, and then click **Share Image**.
5. Select one or more organizations from the **Organization** drop-down list.

   > **Note** If no organization list is available due to lack of permissions, you can enter the name of a level-1 organization in the Organization field.

6. Click **OK**.

6.7. Encrypt a custom image

This topic describes how to encrypt a custom image to generate a new identical encrypted custom image.

Prerequisites

The custom image are in the **Available** state.

Context

To meet compliance requirements for data security, you can use encrypted custom images to create instances. The system disks of the created instances are automatically encrypted.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Images**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Images page, click the **Custom Images** tab.
5. On the Custom Images tab, find the image that you want to encrypt. Click the ... icon in the **Actions** column and click **Encrypt Image**.
6. In the **Encrypt Image** dialog box, configure the parameters described in the following table.

Parameter	Description
Image ID	The system automatically obtains the ID of the image to be encrypted. You do not need to configure this parameter.

Parameter	Description
Name	Enter a name for the new encrypted custom image. The name must be 2 to 128 characters in length and start with a letter. It can contain underscores (_), periods (.), and hyphens (-).
Description	Enter a description for the new encrypted custom image. The description must be 2 to 256 characters in length and cannot start with http:// or https://.

7. Click **OK**.

Result

After you encrypt the custom image, a new identical encrypted custom image is generated and displayed on the Custom Images tab.

6.8. Import custom images

6.8.1. Limits on importing images

This topic describes the limits on importing images. You must understand the limits to ensure that the imported images are available and make the import operation more efficient.

When you import images, take note of the limits described in the following sections:

- Import Linux images
- Import Windows images

Import Linux images

When you import Linux images, take note of the following limits:

- Multiple network interfaces are not supported.
- IPv6 addresses are not supported.
- Each password must be 8 to 30 characters in length. It must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters.
- Firewalls must be disabled. By default, port 22 is enabled.
- The Linux system disk size must range from 40 GiB to 500 GiB.
- Dynamic Host Configuration Protocol (DHCP) must be enabled in the images.
- SELinux must be disabled.
- The Kernel-based Virtual Machine (KVM) driver must be installed.
- We recommend that you install cloud-init to configure hostnames, NTP repositories, and YUM repositories.

Limits

Item	Standard operating system image	Non-standard operating system image
Definition	The following standard 32-bit and 64-bit operating systems are supported: • CentOS • Ubuntu • SUSE • OpenSUSE • RedHat • Debian • CoreOS • Aliyun Linux **Note** Support for standard operating systems may be subject to version changes. You can log on to the ECS console to view the latest supported operating systems.	The following operating systems are regarded as non-standard operating systems: • Operating systems that are not supported by Alibaba Cloud • Standard operating systems that do not meet the requirements on critical system configuration files, basic system environments, or applications To use non-standard operating system images, select Others Linux when you import images. If the images that you import are non-standard operating system images, Alibaba Cloud does not process the instances created from these images. After you create an instance from a non-standard operating system image, you must connect to the instance by using the Virtual Network Computing (VNC) feature in the ECS console and then configure the IP address, route, and password of the instance.

Item	Standard operating system image	Non-standard operating system image
Critical system configuration file	• Do not modify `/etc/issue*`. Otherwise, the version of the operating system cannot be identified, which leads to a failure to create the system. • Do not modify `/boot/grub/menu.lst`. Otherwise, the system fails to start. • Do not modify `/etc/fstab`. Otherwise, partitions cannot be loaded, which causes the system to fail to start. • Do not change `/etc/shadow` to read-only. Otherwise, the password file cannot be modified, which leads to a failure to create the system. • Do not modify `/etc/selinux/config` to enable SELinux. Otherwise, the system fails to start.	Requirements for standard operating system images are not met.
Basic system environment	• Do not adjust the system disk partitions. Only system disks with a single root partition are supported. • Make sure that the system disk has sufficient free space. • Do not modify critical system files such as `/sbin`, `/bin`, and `/lib*`. • Before you import an image, confirm the integrity of the file system. • Only ext3 and ext4 file systems are supported.	
Application	Do not install `qemu-ga` on a custom image. Otherwise, some of the services that Alibaba Cloud uses may be unavailable.	

Item	Standard operating system image	Non-standard operating system image
Image file format	Only images in the RAW, VHD, or QCOW2 format can be imported. To import images in other formats, use a tool to convert the images to a supported format. We recommend that you import VHD images that have smaller sizes.	
Image file size	We recommend that you configure the system disk size based on the virtual disk size rather than the image size. The configured system disk size must be at least 40 GiB.	

Import Windows images

When you import Windows images, take note of the following limits:

- Each password must be 8 to 30 characters in length. It must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters.
- Imported Windows images do not provide the Windows Activation Service.
- Firewalls must be disabled. Otherwise, remote logons are not supported. Port 3389 must be enabled.
- The Windows system disk size must range from 40 GiB to 500 GiB.

Limits

Item	Description

Item	Description

Operating system version	Alibaba Cloud allows you to import the following 32-bit and 64-bit versions of Windows operating system images:
	• Microsoft Windows Server 2016
	• Microsoft Windows Server 2012:
	  ◦ Microsoft Windows Server 2012 R2 (Standard Edition)
	  ◦ Microsoft Windows Server 2012 (Standard Edition and Datacenter Edition)
	• Microsoft Windows Server 2008:
	  ◦ Microsoft Windows Server 2008 R2 (Standard Edition, Datacenter Edition, and Enterprise Edition)
	  ◦ Microsoft Windows Server 2008 (Standard Edition, Datacenter Edition, and Enterprise Edition)
	• Microsoft Windows Server 2003:
	  ◦ Microsoft Windows Server 2003 R2 (Standard Edition, Datacenter Edition, and Enterprise Edition)
	  ◦ Microsoft Windows Server 2003 (Standard Edition, Datacenter Edition, and Enterprise Edition) or later, including Service Pack 1 (SP1)
	• Microsoft Windows 7:
	  ◦ Microsoft Windows 7 (Professional Edition)
	  ◦ Microsoft Windows 7 (Enterprise Edition)
	Note Support for operating systems may be subject to version changes. You can log on to the ECS console to view the latest supported operating systems.

Basic system environment	• Multi-partition system disks are supported. • Make sure that the system disk has sufficient free space. • Do not modify critical system files. • Before you import an image, confirm the integrity of the file system. • Disks can be partitioned to the MBR format and formatted to NTFS file systems.
Application	Do not install qemu-ga on an imported image. Otherwise, some of the services that Alibaba Cloud uses may be unavailable.

Item	Description
Supported image file format	• RAW • VHD • qcow2 We recommend that you configure the system disk size based on the virtual disk size rather than the image size. The configured system disk size must range from 40 GiB to 500 GiB. **Note** We recommend that you import images in the VHD format, which has a smaller transmission footprint.

6.8.2. Convert the image file format

You can only import image files in the RAW, VHD, and qcow2 formats to ECS. If you want to import images in other formats, you must convert the image into a supported format. This topic describes how to convert the image format in Windows and Linux.

Context

You can use the qemu-img tool to convert an image from VMDK, VDI, VHDX, qcow1, or QED to RAW, VHD, or qcow2, or implement conversion between RAW, VHD, and qcow2.

Note We recommend that you use the qcow2 format if your application environment supports this format.

Windows

1. Download qemu.

   Visit QEMU Binaries for Windows (64 bit) to download the qemu tool. Select a qemu version based on your operating system.

2. Install qemu.

   The installation path in this example is *C:\Program Files\qemu*.

3. Configure the environment variables for qemu.

   i. Choose **Start > Computer**, right-click Computer, and choose **Properties** from the shortcut menu.

   ii. In the left-side navigation pane, click **Advanced System Settings**.

   iii. In the **System Properties** dialog box that appears, click the **Advanced** tab and then click **Environment Variables**.

   iv. In the **Environment Variables** dialog box that appears, find the **Path** variable from the **System variables** section.

   - If the **Path** variable exists, click **Edit**.
   - If the **Path** variable does not exist, click **New**.

v. Add a system variable value.

- In the Edit System Variable dialog box that appears, add *C:\Program Files\qemu* to the **Variable value** field, separate different variable values with semicolons (;), and then click **OK**.
- In the New System Variable dialog box that appears, enter *Path* in the **Variable name** field, enter *C:\Program Files\qemu* in the **Variable value** field, and then click **OK**.

4. Open Command Prompt in Windows and run the `qemu-img --help` command. If a successful response is displayed, the tool is installed.
5. In the Command Prompt window, run the `cd [Directory of the source image file]` command to switch to a new file directory,

   for example, `cd D:\ConvertImage`.
6. In the Command Prompt window, run the `qemu-img convert -f raw -O qcow2 centos.raw centos.qcow2` command to convert the image file format.

   The parameters are described as follows:

   - The `-f` parameter is followed by the source image format.
   - The `-O` parameter (case-sensitive) is followed by the destination image format, source file name, and destination file name.

   After the conversion is complete, the destination file appears in the directory of the source image file.

Linux

1. Install the qemu-img tool.
   - For Ubuntu, run the `apt install qemu-img` command.
   - For CentOS, run the `yum install qemu-img` command.
   - Run the `qemu-img convert -f raw -O qcow2 centos.raw centos.qcow2` command to convert the image file format.

     The parameters are described as follows:

     - The `-f` parameter is followed by the source image format.
     - The `-O` parameter (case-sensitive) is followed by the destination image format, source file name, and destination file name.

6.8.3. Import an image

After you upload an image from your computer to an Object Storage Service (OSS) bucket, you can import the image as a custom image to Elastic Compute Service (ECS).

Prerequisites

- An image is made. It meets the limits and requirements for image import and is in the RAW, VHD, or QCOW2 format. For more information, see Limits on importing custom images and Convert the image file format.
- You are granted the permissions to import images. For more information, see the "RAM" chapter in *Aliyun Uni-manager Management Console User Guide*.
- An image is uploaded from your computer to a bucket by using the OSS console or by calling an OSS

API operation. For more information, see the "Upload objects" topic in *OSS User Guide* or the "PutObject" topic in *OSS Developer Guide*.

> **Note** Make sure that the bucket resides in the region to which you want to import the image as a custom image.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Images**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Import Image**.
5. In the **Import Image** panel, configure the parameters described in the following table.

Parameter	Required	Description
Organization	Yes	The organization in which to use the custom image.
Resource Set	Yes	The resource set to which to assign the custom image.
Region	Yes	The region to which to import the image as a custom image.
OSS Bucket Name	Yes	The name of the OSS bucket where the image to be imported is stored.
OSS Object Name	Yes	The URL of the object as which the image to be imported is stored in the OSS bucket. For information about how to obtain the URL of an OSS object, see the "Obtain object URLs" topic in *OSS User Guide*.
Image Name	Yes	The name of the custom image. The name must be 2 to 128 characters in length and can contain letters, digits, periods (.), underscores (_), hyphens (-), and colons (:). It must start with a letter.
Sharing Scope	Yes	The scope for which to share the custom image. Valid values: ○ **Current Organization and Subordinate Organizations** ○ **Current Resource Set** ○ **Current Organization**
Operating System	Yes	The operating system. Valid values: **Linux** and **Windows**.

Parameter	Required	Description
System Disk Size (GB)	Yes	The size of the system disk on an instance. Unit: GiB.
Architecture	Yes	Valid values: **x86_64** and **i386**.
Platform	Yes	Linux: ○ **CentOS** ○ **Ubuntu** ○ **SUSE** ○ **OpenSUSE** ○ **Debian** ○ **CoreOS** ○ **Aliyun** ○ **Others Linux** Windows: ○ **Windows Server 2003** ○ **Windows Server 2008** ○ **Windows Server 2012**
Image Description	No	The description of the custom image.
Add Data Disk Image	No	Imports other images that contain data from data disks. If you select Add Data Disk Image, you must click Add and specify parameters including OSS Bucket Name, OSS Object Name, Image Format, Device Name, and Disk Capacity to add images.

6. Click **OK**.

Result

You can go to the Images page to view the creation progress of the custom image. For more information, see View images. When the custom image is created, 100% is displayed in the Progress column.

6.9. Export a custom image

You can export custom images to Object Storage Service (OSS) buckets and download the images to your on-premises device.

Prerequisites

- OSS is activated and an OSS bucket is created. For more information, see the "Create buckets" topic in *OSS User Guide*.
- You are authorized to export images. For more information, see the "RAM" chapter in *Apsara Uni-manager Management Console User Guide*.

Context

You can export custom images to the RAW, VHD, or QCOW2 format. After a custom image is exported to an OSS bucket, you can download the image to your on-premises device. For more information, see the "Obtain object URLs" topic in *OSS User Guide*.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Images**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Images page, find the image that you want to export. Click the ... icon in the **Actions** column and click **Export Image**.
5. Configure the parameters described in the following table.

Parameter	Required	Description
Region	Yes	Select the region where the custom image resides.
OssBucket	Yes	Select an OSS bucket that belongs to the same region as the custom image.
OSS Prefix	Yes	Set the prefix of the object name for the custom image. For example, if you set OSS Prefix to Demo, the exported image is named Demo-[Automatically generated object name].

6. Click **OK**.

6.10. Delete a custom image

You can delete custom images that are no longer needed.

Context

Before you delete a custom image, take note of the following items:

- After a custom image is deleted, the image can no longer be used to create instances.
- After a custom image is deleted, the instances that use this image remain available but cannot have their system disks re-initialized.
- Before you can delete a shared image, you must unshare the image. After a shared image is unshared and then deleted, the following results occur:
  - The accounts from which the image is unshared can no longer query the image by using the Elastic Compute Service (ECS) console or by calling an API operation.
  - The accounts from which the image is unshared can no longer use the image to create instances or replace system disks.
  - The system disks of the instances that were created from the shared image cannot be re-initialized.

- If you delete a custom image, snapshots contained in the image are not deleted. If you do not want to keep the snapshots, navigate to the Snapshots page and delete the snapshots.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Images**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Use one of the following methods to delete a custom image:
   - To delete a single custom image, find the image and click **Delete Image** in the **Actions** column.
   - To delete one or more custom images at a time, select the images and click **Delete** in the lower-left corner of the image list.
5. In the message that appears, click **Delete**.

7.Snapshots

7.1. Overview

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

Scenarios

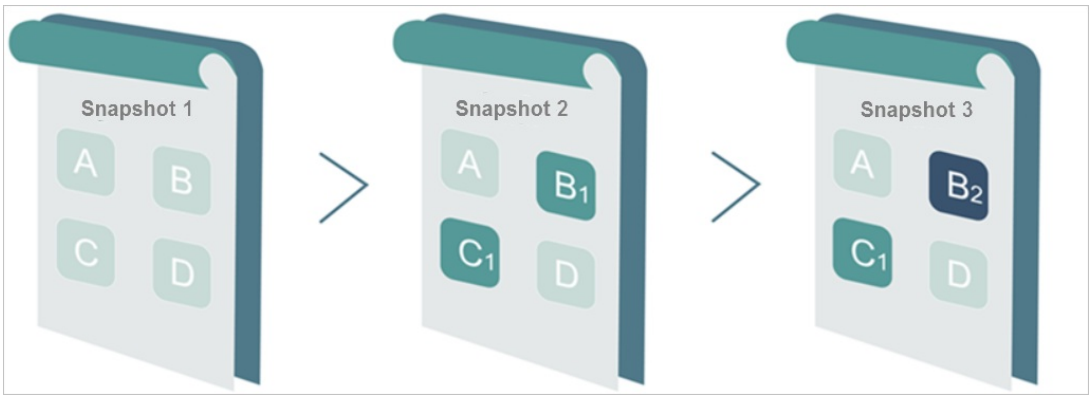
You can use snapshots in scenarios such as environment replication and disaster recovery:

- You may want to use the data of a disk as the source data for write operations or save data to a different disk. In this case, you can create a snapshot for a disk and then create another disk from the snapshot. The new disk contains the basic data of the original disk.
- Issues such as application errors or malicious read operations and write operations may occur on disks that store data. As a result, these disks require additional protection mechanisms. You can create snapshots at regular intervals to restore data to a previous point in time when data errors occur.

How snapshots work

Snapshots are created incrementally. This way, only the changed data between two snapshots is copied, as shown in Snapshots.

Snapshots



In this example, Snapshot 1, Snapshot 2, and Snapshot 3 are the first, second, and third snapshots of a disk. When a snapshot is created, the file system checks each block of data stored on the disk, and only copies the blocks of data that are different from those on the previous snapshots. The following section describes the changes between snapshots in the preceding figure:

- All data on the disk is copied to Snapshot 1 because it is the first disk snapshot.
- The changed blocks B1 and C1 are copied to Snapshot 2. Blocks A and D are referenced from Snapshot 1.
- The changed block B2 is copied to Snapshot 3. Blocks A and D are referenced from Snapshot 1, and block C1 is referenced from Snapshot 2.
- When you want to restore the disk to the state of Snapshot 3, blocks A, B2, C1, and D are copied to the disk. This way, the disk is restored to the state of Snapshot 3.
- If Snapshot 2 is deleted, block B1 in the snapshot is deleted, but block C1 is retained because it is referenced by other snapshots. When you roll back a disk to Snapshot 3, block C1 is restored.

> **Note** Snapshots are stored in Object Storage Service (OSS), but are hidden from users. Snapshots do not occupy storage space in OSS buckets. Snapshots can be managed only in the ECS console or by using API operations.

7.2. Create snapshots

You can use snapshots to back up data, restore Elastic Compute Service (ECS) instances that are accidentally released, and create custom images. You can create snapshots of disks to improve fault tolerance before you roll back a disk, modify key system files, or change the operating system of an instance.

Prerequisites

- The associated instance of the disk for which you want to create a snapshot is in the **Running** or **Stopped** state.
- The disk is in the **Running** state.

Context

Up to 64 snapshots can be retained for each disk.

Snapshots can be used in the following scenarios:

- Roll back data on disks

  For more information, see Restore a disk.

- Create a custom image

  For more information, see Create a custom image from a snapshot. Data disk snapshots cannot be used to create custom images.

- Create a data disk from a data disk snapshot

  To create a data disk from a snapshot, set Use Snapshot to Yes and then specify a snapshot on the Create Disk page. For more information, see Create a disk. The size of the created disk is determined based on the size of the specified snapshot and cannot be changed. When you re-initialize a data disk created from a snapshot, the disk is restored to the state of the snapshot.

When you create a snapshot, take note of the following items:

- For each disk, the first snapshot is a full snapshot and subsequent snapshots are incremental snapshots. A longer period of time is required to create the first snapshot. The period of time that is required to create an incremental snapshot varies based on the volume of data that is modified after the point in time when the previous snapshot is created. If a large volume of data is modified, a long period of time is required to create the snapshot.

  > **Note** If you want to use a snapshot immediately after the snapshot is created, you can enable the instant access feature.

- We recommend that you create snapshots during off-peak hours.

Procedure

1. Log on to the ECS console.

2. In the left-side navigation pane, choose **Storage & Snapshots > Disks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the disk for which you want to create a snapshot and click **Create Snapshot** in the **Actions** column.
5. In the **Create Snapshot** dialog box, configure the parameters described in the following table and click **OK**.

Parameter	Description
Snapshot Name	The name of the snapshot. The name must be 2 to 128 characters in length. The name must start with a letter and cannot start with http:// or https://. The name can contain letters, digits, colons (.), underscores (_), and hyphens (-). **Note** The name cannot start with auto because snapshots whose names start with auto are recognized as automatic snapshots.
Instant Access	The instance access feature can accelerate the snapshot creation process. You can use the instant access feature to create snapshots within seconds. ○ **Enable Instant Access**: You can select this check box to enable the instant access feature. By default, this feature is disabled. **Note** The instance access feature can be enabled for standard performance disks and premium performance disks. ○ **Duration of Instant Access**: After you enable the instant access feature, you can specify a validity duration for the feature. The default validity duration is one day. The instant access feature is automatically disabled when the specified duration expires.
Snapshot Description	The description of the snapshot. The description must be 2 to 256 characters in length and cannot start with http:// or https://.

You can go to the Snapshots page to check the creation progress of the snapshot. For more information, see View snapshots. After the snapshot is created, 100% is displayed in the Progress column.

7.3. View snapshots

You can view the list of created snapshots.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Snapshots**.

3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Disk Snapshots** tab.
5. Use one of the following methods to search for snapshots:
   - Select a filter option from the drop-down list and enter relevant information in the search box. Then, the system searches for snapshots based on your specified filter condition.
   - Click **Advanced Filter**, specify multiple filter options, and then click **Search**. Eligible snapshots are displayed in the snapshot list.

> **Note** You can use the Advance Filter feature and specify multiple filter options to narrow down search results.

Filter option	Description
Snapshot Name	Enter the name of the snapshot.
Snapshot ID	Enter the ID of the snapshot.
Disk ID	Enter the ID of the disk for which the snapshot were created.
Creation Time	Select a time range to search for snapshots that were created during the time range.
Tag	Enter the key or value of a key of the snapshot in the tag filter.

7.4. Roll back a disk by using a snapshot

If you have created a snapshot of a disk, you can use the snapshot to roll back the disk to the point in time when the snapshot was created. The disk rollback operation is irreversible. After the disk is rolled back, the data that was stored on the disk before the rollback operation is performed is lost and cannot be recovered. Exercise caution when you perform this operation.

Prerequisites

Before you roll back a disk by using a snapshot, make sure that the following requirements are met:

- Snapshots have been created for the disk, and no snapshots are currently being created for the disk. For more information, see Create a snapshot.
- The disk is not released.
- The disk is attached to an instance that is in the **Stopped** state.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Snapshots**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Disk Snapshots** tab.

5. Find a snapshot of the disk that you want to use to roll back and click **Roll Back Disk** in the **Actions** column.

   > **Note** You can roll back only one disk on an instance at a time. Other disks that are attached to the instance are not affected. After the rollback operation is complete, the entire disk (not a partition or a directory) is restored to the state that the disk was in when the snapshot was created.

6. In the message that appears, click **Roll Back Disk**.

What's next

- After the disk is rolled back, the hosts configuration file and the configurations such as the hostname, SSH key pair, passwords, network settings, operating system repository settings, and clock source are initialized. You must reconfigure the file and configurations
- If you have resized the disk after you created the snapshot for the disk, you must log on to the instance to resize the file systems on the disk again after the disk is rolled back. For more information, see Expand a disk.

7.5. Create a custom image from a snapshot

You can create a custom image from a snapshot that contains the operating system and data of an Elastic Compute Service (ECS) instance. Then, you can use the custom image to create multiple identical instances.

Prerequisites

A system disk snapshot is created. For more information about how to create a system disk snapshot, see Create a snapshot.

Context

Before you create custom images from snapshots, take note of the following items:

- Notes on snapshots used to create custom images:
  - A custom image can be created from a system disk snapshot and one or more data disk snapshots. You cannot use only data disk snapshots to create custom images.
  - Both encrypted and unencrypted snapshots can be used to create custom images.
- Custom images cannot be used across regions.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Snapshots**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Disk Snapshots** tab.
5. Find a snapshot whose property is displayed as **System Disk** in the **Disk Type (All)** column, and then click **Create Custom Image** in the **Actions** column.

6. Configure the parameters described in the following table to create a custom image.

Parameter	Description
Sharing Scope	Select the scope in which you want to share the custom image. ○ **Current Organization and Subordinate Organizations** ○ **Current Resource Set** ○ **Current Organization**
Image Name	Enter a name for the custom image. The name must be 2 to 128 characters in length and can contain letters, digits, periods (.), underscores (_), hyphens (-), and colons (:). It must start with a letter.
Image Description	The description of the custom image. The description must be 2 to 256 characters in length and cannot start with http:// or https://.

7. Click **OK**.

7.6. Delete snapshots

You can delete snapshots that are no longer needed. Deleted snapshots cannot be recovered. System disk snapshots that have been used to create custom images cannot be deleted.

Prerequisites

To delete snapshots that have been used to create custom images, you must first delete the custom images. For more information, see Delete a custom image.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Snapshots**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Disk Snapshots** tab.
5. Use one of the following methods to delete snapshots:
   - To delete a single snapshot, find the snapshot and click **Delete** in the **Actions** column.
   - To delete one or more snapshots at a time, select the snapshots and click **Delete** below the snapshot list.
   - In the message that appears, click **Delete**.

8.Snapshot-consistent groups

8.1. Create a snapshot-consistent group

You can create a snapshot-consistent group to simultaneously create snapshots for one or more disks attached to an Elastic Compute Service (ECS) instance. When a business system spans multiple disks, you can create a snapshot-consistent group to ensure a consistent write order and crash consistency of business system data.

Context

You can create a snapshot-consistent group to simultaneously create snapshots for multiple disks attached to an ECS instance. Snapshot-consistent groups are applicable to cluster services. Example scenarios that snapshot-consistent groups support:

- A business system is deployed on multiple disks attached to an ECS instance, and point-in-time consistency and crash-consistency are required across databases or enterprise-level applications. For example, a MySQL cluster is built on ECS instances, a single Logical Volume Manager (LVM) logical volume is created across multiple volumes, or Oracle or SAP HANA clusters are migrated to the cloud.
- Snapshots need to be created simultaneously for multiple nodes of a distributed application system such as a large-scale website or a multi-application collaborative system.
- Disks on an ECS instance need to be batch backed up with high point-in-time consistency.

Precautions

Before you use the snapshot-consistent group feature, take note of the following items:

- When you create a snapshot-consistent group based on an instance, take note of the following items:
  - To ensure that services are not affected, we recommend that you create the snapshot-consistent group during off-peak hours.
  - The instance must be in the **Running** or **Stopped** state.
  - The disks for which to create a snapshot-consistent group can only be standard performance disks and premium performance disks and must be in the **Running** state. If the instance has disks of other categories attached, you can select only standard performance disks and premium performance disks on the instance to create a snapshot-consistent group.
  - A single snapshot-consistent group can contain snapshots of up to 16 disks including the system disk and data disks and cannot exceed 32 TiB in size.
  - Snapshots that you created are retained until they are deleted. We recommend that you delete unnecessary snapshots on a regular basis to prevent snapshot capacity from exceeding the limit. For more information, see Delete a snapshot.
- For the amount of time required to create a snapshot-consistent group, take note of the following items:

The amount of time required to create a snapshot-consistent group depends on the sizes of the selected disks for which to create snapshots. The first snapshot of each disk is a full snapshot and takes longer to create than subsequent snapshots. Subsequent snapshots are incremental snapshots and do not take as long to be created as the first snapshot. The amount of time required varies based on the amount of data changed since the previous snapshot. If you want to use a snapshot while it is being created, you can enable the instant access feature.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Snapshots**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Snapshots page, click the **Snapshot-consistent Groups** tab. Then, click **Create Snapshot-consistent Group**.
5. In the **Create Snapshot-consistent Group** dialog box, configure the parameters described in the following table.

Parameter	Description
Data Source	Select the instances based on which to create the snapshot-consistent group.
Disks	Select standard performance disks and premium performance disks for which you want to create a snapshot-consistent group. Up to 16 disks can be selected, including the system disk and data disks. The total size of the selected disks cannot exceed 32 TiB.
Group Name	Enter a name for the snapshot-consistent group for easy management. The name must be 2 to 128 characters in length and must start with a letter. It can contain letters, digits, periods (.), underscores (_), hyphens (-), and colons (:) and cannot start with http:// or https://.
Description	Enter a description for the snapshot-consistent group for easy management. The description must be 2 to 256 characters in length and cannot start with http:// or https://.
Instant Access	The instance access feature can accelerate the process of snapshot creation. You can use the instant access feature to create snapshots within seconds. ○ **Enable Instant Access**: You can select this check box to enable the instant access feature. By default, this feature is disabled. ○ **Duration of Instant Access**: After you enable the instant access feature, you can specify a validity duration for the feature. The default validity duration is one day. The instant access feature is automatically disabled when the specified duration expires.

What's next

After the snapshot-consistent group is created, you can use it to roll back disks as needed. For more information, see Roll back disks by using a snapshot-consistent group.

8.2. Roll back disks by using a snapshot-consistent group

After a snapshot-consistent group is created, you can use it to roll back one or more disks in the event of system failures or data errors caused by accidental operations.

Prerequisites

Before you use a snapshot-consistent group to roll back disks on an Elastic Compute Service (ECS) instance, make sure that the following requirements are met:

- A snapshot-consistent group is created based on the disks. For more information, see Create a snapshot-consistent group.

  > **Notice** The rollback operation cannot be reversed. If data changes were made to the disk from the time the snapshot-consistent group was created until the disk is rolled back, all these data changes are lost. To prevent data loss caused by accidental operations, we recommend that you create a snapshot-consistent group to back up disk data before you roll back disks.

- The instance to which the disks to be rolled back are attached is in the **Stopped** state. For more information, see Stop an instance.
- The disks that correspond to the snapshots contained in the snapshot-consistent group have not been released or detached or do not have snapshots being created.
- The operating system of the instance has not been replaced since the snapshot-consistent group was created.
- The snapshots of the disks have not been deleted from the snapshot-consistent group. If the snapshot of a disk has been deleted from the snapshot-consistent group, the disk cannot be rolled back by using the snapshot-consistent group. The snapshot-consistent group can be used to roll back only other disks whose snapshots are contained in the group.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Snapshots**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Snapshot-consistent Groups** tab.
5. Find the snapshot-consistent group that you want to use and click **Roll Back** in the **Actions** column.
6. In the **Roll Back** dialog box, perform the following operations:
   i. Read the precautions about the rollback operation. Select **Start Instance Immediately After Rollback** based on your business needs.
   ii. In the **Disk Snapshots** section, select the snapshots of the disks that you want to roll back.
   iii. Click **Roll Back**.
   After the selected disks are rolled back, the `Rolled back.` message is displayed.

Result

After disks are rolled back, you can log on to their associated instances to check whether the disk data has been reverted to the state it was in when the snapshots were created.

8.3. Delete a snapshot-consistent group

You can delete snapshot-consistent groups that are no longer needed.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Snapshots**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Snapshot-consistent Groups** tab.
5. Find the snapshot-consistent group that you want to delete and click **Delete** in the **Actions** column.
6. In the message that appears, click **Delete**.

9.Automatic snapshot policies

9.1. Create an automatic snapshot policy

Automatic snapshot policies can be applied to system disks and data disks to create snapshots for the disks on a periodic basis. You can use automatic snapshot policies to improve data security and tolerance against accidental operations.

Context

Automatic snapshot policies can effectively eliminate the following risks associated with manual snapshots:

- When applications such as personal websites or databases deployed on an Elastic Compute Service (ECS) instance encounter attacks or system vulnerabilities, you may be unable to manually create snapshots. In this case, you can use the latest automatic snapshots to roll back the affected disks to restore your data and reduce loss.
- You can also specify an automatic snapshot policy to create snapshots before regular system maintenance tasks are performed. This eliminates the need to manually create snapshots and ensures that snapshots are always created before maintenance.

You can retain up to 64 snapshots for each disk. If the maximum number of snapshots for a disk has been reached while a new snapshot is being created, the system deletes the oldest automatic snapshot.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Automatic Snapshot Policies**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create Policy**.
5. In the **Create Policy** dialog box, configure the parameters described in the following table.

Parameter or option	Required	Description
Organization	Yes	The organization in which to apply the automatic snapshot policy.
Resource Set	Yes	The resource set in which to apply the automatic snapshot policy.
Region	Yes	The ID of the region in which to apply the automatic snapshot policy.

Parameter or option	Required	Description
Policy Name	Yes	The name of the automatic snapshot policy. The name must be 2 to 128 characters in length and cannot start with a digit or a special character. It can contain the following special characters: periods (.), underscores (_), hyphens (-), and colons (:).
Sharing Scope	Yes	The scope in which the automatic snapshot policy can be shared. Valid values:Current Organization and Subordinate Organizations, Current Resource Set, and Current Organization.
Creation Time	Yes	The time of the day at which to create an automatic snapshot. Valid values: 00:00 to 23:00 (the start of each hour). You can select multiple values. **Note** The default time zone for the snapshot policy is UTC+8. You can change the time zone to suit your business requirements. If the time scheduled for creating an automatic snapshot is due while a previous automatic snapshot is being created, the new snapshot creation task is skipped. This may occur when a disk contains a large volume of data. For example, assume that an automatic snapshot policy is applied to a disk that contains a large volume of data, and the system creates snapshots at 00:00, 01:00, and 02:00 based on the policy. If the system starts to create a snapshot at 00:00 and takes 70 minutes to complete the snapshot creation task, the system skips the automatic snapshot task scheduled for 01:00 and creates the next automatic snapshot at 02:00.
Frequency	Yes	The day of the week when to create automatic snapshots. The valid values range from Monday to Sunday. You can select multiple values.

Parameter or option	Required	Description
Retention Period	No	The retention period of the automatic snapshots. By default, automatic snapshot are retained for 30 days. Valid values: ○ **Custom Period**: You can select this option and then specify the number of days during which to retain the automatic snapshots. Valid values: 1 to 65536. ○ **Permanently**: You can select this option to retain the automatic snapshots until the maximum number of snapshots is reached.

6. Click **OK**.

What's next

After the automatic snapshot policy is created, you must apply it to a disk for snapshots to be automatically created. For more information, see Configure an automatic snapshot policy.

9.2. View automatic snapshot policies

You can view created automatic snapshot policies.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Automatic Snapshot Policies**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. View created automatic snapshot policies.

9.3. Modify an automatic snapshot policy

After an automatic snapshot policy is created, you can modify properties of the policy based on your business requirements, such as the policy name, creation time, execution frequency, and retention period.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Automatic Snapshot Policies**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the automatic snapshot policy that you want to modify and click **Modify Policy** in the **Actions** column.
5. In the **Modify Policy** dialog box, modify the original parameter settings of the policy.

For more information about the parameters that you can configure for a policy, see Create an automatic snapshot policy.

> **Note** Changes made to the retention period do not affect the existing snapshots and take effect only on subsequent snapshots.

6. Click **OK**.

9.4. Apply or cancel an automatic snapshot policy

You can apply automatic snapshot policies to disks. After an automatic snapshot policy is applied to a disk, snapshots are automatically created for the disk based on the policy.

Context

We recommend that you apply automatic snapshot policies to create automatic snapshots during off-peak hours. You can manually create snapshots for disks to which automatic snapshot policies are applied. When an automatic snapshot is being created for a disk, you must wait for the creation task to be completed before you can manually create a snapshot for the disk.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Storage & Snapshots > Automatic Snapshot Policies**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the automatic snapshot policy that you want to apply or cancel and click **Apply or Cancel Policy** in the **Actions** column.
5. In the **Apply or Cancel Policy** dialog box, perform the following operations based on your business requirements:
   - If you want to apply the automatic snapshot policy, click the **Disks Without Policy Applied** tab, find the disk to which you want to apply the policy, and then click **Apply Policy** in the Actions column.
   - If you want to cancel the automatic snapshot policy, click the **Disks With Policy Applied** tab, find the disk for which you want to cancel the policy, and then click **Cancel Policy** in the Actions column.

9.5. Delete an automatic snapshot policy

You can delete automatic snapshot policies that you no longer need. After you delete an automatic snapshot policy, the policy is automatically canceled for the disks to which the policy is applied.

Procedure

1. Log on to the ECS console.

2. In the left-side navigation pane, choose **Storage & Snapshots > Automatic Snapshot Policies**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the automatic snapshot policy that you want to delete and click **Delete Policy** in the **Actions** column.
5. In the message that appears, click **Delete**.

10.Security groups

10.1. Overview

A security group acts as a virtual firewall to control the inbound and outbound traffic of Elastic Compute Service (ECS) instances to improve security. Security groups provide Stateful Packet Inspection (SPI) and packet filtering capabilities. You can use security groups and security group rules to define security domains in the cloud.

Security groups and security group rules

A security group acts as a virtual firewall that is used to control access to and from one or more ECS instances. Each instance must belong to at least one security group. The following rules apply when you add instances to security groups:

- Each instance must belong to one or more security groups.
- The secondary ENIs that are bound to an instance can be assigned to security groups different from those of the instance.

The rules of a security group control the inbound or outbound traffic to or from the instances in the security group. You can add or modify security group rules based on your business needs to implement fine-grained access control.

New and modified rules are automatically applied to all instances within the security group. Security group rules can be used to control access to or from specific IP addresses, CIDR blocks, security groups, or prefix lists.

If an instance belongs to multiple security groups, the rules of all the security groups are applied to the instance. When an access request destined for the instance is detected, the request is matched against applied security group rules one by one based on the rule attributes such as protocol, port range, and priority. No sessions are established until an Allow rule matches the request.

Work with security groups

You can perform the following operations to use security groups to control traffic for instances:

1. Create security groups.
2. Add rules to the security groups.
3. Add instances to the security groups.
4. Manage existing security groups and security group rules based on your needs.

Operations in the ECS console

The following table describes the operations that you can perform in the ECS console to manage security groups.

Operation	Description	References
Create a security group	You can create a security group.	Create a security group

Operation	Description	References
Add a security group rule	After you create a security group, you can add or modify security group rules to control inbound or outbound network access.	Add a security group rule
Add an ECS instance to a security group	You can add an instance to a security group to control network access in a centralized manner.	Add an instance to a security group
Manage security groups	You can query, modify, and delete security groups as well as remove instances from security groups.	• View security groups • Modify a security group • Remove instances from a security group • Delete a security group
Manage security group rules	You can modify, clone, export, import, and delete security group rules.	• Modify a security group rule • Clone a security group rule • Export security group rules • Import security group rules • Delete a security group rule

10.2. Create a security group

Security groups are an important means to implement network security isolation. They control network traffic to or from one or more Elastic Compute Service (ECS) instances.

Prerequisites

A virtual private cloud (VPC) is created. For more information, see the "Create a VPC" topic in *VPC User Guide*.

Context

Security groups determine whether instances in the same VPC, region, and account can communicate with each other. By default, if the instances belong to the same security group, they can communicate with each other over the internal network. If the instances belong to different security groups, you can allow mutual access between the security groups to allow the instances to communicate with each other over the internal network.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.

3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create Security Group**.
5. Configure the parameters described in the following table.

Section	Parameter	Required	Description
Area	Organization	Yes	Select an organization in which to create the security group. Make sure that the security group and the VPC belong to the same organization.
	Resource Set	Yes	Select a resource set in which to create the security group. Make sure that the security group and the VPC belong to the same resource set.
	Region	Yes	Select a region in which to create the security group. Make sure that the security group and VPC reside within the same region.
	Zone	Yes	Select a zone in which to create the security group.
Basic Configurations	Sharing Scope	Yes	Select the scope for which to share the security group. Valid values: Current Resource Set, Current Organization and Subordinate Organizations, and Current Organization.
	VPC	Yes	Select a VPC in which to create the security group.
	Security Group Name	Yes	Enter a name for the security group. The name must be 2 to 128 characters in length and start with a letter. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,). It cannot start with http:// or https://.
	Description	No	Enter a description for the security group for easy management. The description must be 2 to 256 characters in length and start with a letter. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,). It cannot start with http:// or https://.

6. Click **Submit**.

   After the security group is created, it is displayed on the Security Groups page.

What's next

- After the security group is created, it contains no security group rules. You can add security group rules to allow or deny access to or from the Internet or internal network for ECS instances within the security group. For more information, see Add a security group rule.
- Each ECS instance must belong to at least one security group. You can add an instance to one or more security groups. For more information, see Add an instance to a security group.

10.3. Add a security group rule

You can use security group rules to manage access to and from the Elastic Compute Service (ECS) instances in a security group over the Internet and the internal network.

Prerequisites

The public or internal IP addresses from which you want to manage access to your instances are obtained.

Context

Security groups are used to manage access requests that are sent over the Internet or internal network. For security purposes, most security groups use Deny rules for inbound traffic.

This topic is provided for the following scenarios:

- If an application deployed on your instance initiates a request to connect to a network that is not managed by the security groups to which the instance belongs but the request remains in the waiting state, you must add a security group rule to allow this request.
- When attacks from some of the request sources are detected on running applications, you can add security group rules to block the malicious requests.

Before you add security group rules, take note of the following items:

- Both IPv4 and IPv6 addresses can be used as the authorization objects of security group rules.
- The total number of inbound and outbound rules within each security group cannot exceed 200.

For more information, see Overview.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Security Groups page, find the security group for which you want to configure rules and click **Manage Rules** in the **Actions** column.
5. Click **Create Rule**.
6. In the **Create Rule** dialog box, configure the parameters described in the following table.

Parameter	Required	Description
NIC Type	Yes	Valid value: **Internal NIC**. No public NICs are available for ECS instances that are deployed in virtual private clouds (VPCs). You can only add internal security group rules. However, the added security group rulesapply to both the Internet and the internal network.

Parameter	Required	Description
Rule Direction	Yes	◦ **Outbound**: access from the ECS instances in the current security group to other ECS instances on the internal network or to resources on the Internet. ◦ **Inbound**: access from other ECS instances on the internal network or from resources on the Internet to the ECS instances in the current security group.
Action	Yes	◦ **Allow**: allows access requests that are sent to specified ports. ◦ **Deny** :denies access requests and drops data packets without returning a response. If two security group rules use the same settings except for the action, the **Deny** action takes precedence over the **Allow** action.
Protocol Type	Yes	The protocol type of the security group rule. Valid values: ◦ **All**: This value can be used in scenarios in which requests are sent from trusted sources. ◦ **TCP**: This value can be used to allow or deny traffic on one or more consecutive ports. ◦ **UDP**: This value can be used to allow or deny traffic on one or more consecutive ports. ◦ **ICMP**: This value can be used when the `ping` command is used to test the status of network connection between instances. ◦ **ICMPv6**: This value can be used when the `ping6` command is used to test the status of network connection between instances. ◦ **GRE**: This value can be used for VPN.

Parameter	Required	Description
Port Range	Yes	The port range varies based on the protocol type. ◦ If you set Protocol Type to **All**, the value **-1/-1** is displayed, which indicates all ports. In this case, you cannot specify a port range. ◦ If you set Protocol Type to **TCP**, you can specify a port range in the *<start port>/<end port>* format. Valid values: 1 to 65535. For example, 80/80 indicates port 80, and 1/22 indicates ports 1 to 22. ◦ If you set Protocol Type to **UDP**, you can specify a port range in the *<start port>/<end port>* format. Valid values: 1 to 65535. For example, 80/80 indicates port 80, and 1/22 indicates ports 1 to 22. ◦ If you set Protocol Type to **ICMP**, the value **-1/-1** is displayed, which indicates all ports. In this case, you cannot specify a port range. ◦ If you set Protocol Type to **ICMPv6**, the value **-1/-1** is displayed, which indicates all ports. In this case, you cannot specify a port range. ◦ If you set Protocol Type to **GRE**, the value **-1/-1** is displayed, which indicates all ports. In this case, you cannot specify a port range.
Priority	Yes	The priority of the rule. Valid values: 1 to 100. The default value is 1, which indicates the highest priority.
Authorization Type	Yes	◦ **IPv4 CIDR Block**: IPv4 addresses or CIDR blocks. ◦ **IPv6 CIDR Block**: IPv6 addresses or CIDR blocks. ◦ **Security Group**: security groups. This authorization type takes effect only on the internal network. You can select another security group in the current account as the authorization object for the instances in the current security group. This way, you can manage access to or from the ECS instances in that security group over the internal network.

Parameter	Required	Description
Authorization Object	Yes	Authorization objects vary based on the authorization type. If you set Authorization Type to **IPv4 CIDR Block**, the following rules apply: ◦ You can enter a single IPv4 address or a CIDR block. Example: *192.0.2.1* or *192.0.2.0/24*. ◦ You can enter up to 10 authorization objects. Separate multiple objects with commas (,). ◦ If you enter *0.0.0.0/0*, all IPv4 addresses are allowed or denied based on the value of the Action parameter. Proceed with caution. If you set Authorization Type to **IPv6 CIDR Block**, the following rules apply: ◦ You can enter a single IPv6 address or a CIDR block. Example: *2001:db8:1:1:1:1:1:1* or *2001:db8::/32*. ◦ You can enter up to 10 authorization objects. Separate multiple objects with commas (,). ◦ If you enter *::/0*, all IPv6 addresses are allowed or denied based on the value of the Action parameter. Proceed with caution. If you set Authorization Type to **Security Group**, select a security group ID. The specified security group must be within the same VPC as the current security group.
Description	No	The description of the security group rule. To simplify future management operations, we recommend that you provide a specific description. The description must be 1 to 512 characters in length and cannot start with http:// or https://.

7. Click **OK**.

10.4. Add an ECS instance to a security group

You can add an existing Elastic Compute Service (ECS) instance to a security group within the same region. After the instance is added to the security group, the rules of the security group automatically apply to the instance.

Context

Security groups are an important means for security isolation. A security group can control access to or from the one or more ECS instances in it. An ECS instance must belong to one to five security groups.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.

4. On the Security Groups page, find the security to which you want to add the instance. Click the ... icon in the **Actions** column and click **Manage Instances**.
5. On the Security Group Details page, click **Add Instance**.
6. In the **Add Instance** dialog box, select the instance that you want to add to the security group from the Instance ID/Name drop-down list and click **OK**.

   After the ECS instance is added to the security group, the rules of the security group automatically apply to the instance.

10.5. Manage security groups

10.5.1. View security groups

This topic describes how to search for security groups within a region by using different methods and view the details of the security groups.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Use one of the following methods to search for security groups:
   - Select a filter option from the drop-down list and enter relevant information in the search box. Then, the system searches for security groups based on your specified filter condition.
   - Click **Advanced Filter**, specify multiple filter options, and then click **Search**. Eligible security groups are displayed in the security group list.

   > **Note** You can use the Advance Filter feature and specify multiple filter options to narrow down search results.

Filter option	Description
Security Group ID	Enter the ID of the security group.
Security Group Name	Enter the name of the security group.
VPC ID	Enter the ID of the virtual private cloud (VPC) in which the security group resides.
Tag	Enter the key or value of a key of the security group in the tag filter.

10.5.2. Modify a security group

You can modify the names and descriptions of created security groups.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the security group that you want to modify and click **Modify** in the **Actions** column.
5. In the **Modify Security Group** dialog box, modify the name and description of the security group.
   - The new name must be 2 to 128 characters in length and start with a letter. It can contain letters, digits, underscores (_), hyphens (-), and colons (:). It cannot start with http:// or https://.
   - The new description must be 2 to 256 characters in length and start with a letter. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,). It cannot start with http:// or https://.
6. Click **OK**.

10.5.3. Remove instances from a security group

You can remove Elastic Compute Service (ECS) instances from a security group based on your business requirements. Before you perform this operation, make sure that the instances belong to at least one security group.

Prerequisites

The instances that you want to remove belong to two or more security groups.

Context

After an instance is removed from a security group, the instance is isolated from the other instances in the security group. We recommend that you perform all tests in advance to ensure that services can continue to run as expected after you remove the instance from the security group.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the security group from which you want to remove instances, click the ... icon in the **Actions** column, and then select **Manage Instances**.
5. On the Instances page, select one or more instances that you want to remove and click **Remove from Security Group** in the lower-left corner.
6. In the message that appears, click **Remove from Security Group**.

10.5.4. Delete a security group

You can delete security groups that are no longer needed. When a security group is deleted, its rules are also deleted.

Prerequisites

- The security group that you want to delete does not contain Elastic Compute Service (ECS) instances.

If the security group contains ECS instances, you must remove the instances from the security group. For more information, see Remove instances from a security group.

- The security group is not referenced by rules of other security groups. If the security group is referenced by rules of other security groups, you must delete those rules as prompted. For more information, see Delete a security group rule.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Use one of the following methods to delete a security group:
   - To delete a single security group, find the security group and click **Delete** in the **Actions** column.
   - To delete one or more security groups at a time, select the security groups and click **Delete** in the lower-left corner of the Security Groups page.
   - In the message that appears, click **Delete**.

10.6. Manage security group rules

10.6.1. Modify security group rules

This topic describes how to modify security group rules. Improper configurations of security group rules can result in serious security risks. You can modify improper rules in a security group to ensure the network security of Elastic Compute Service (ECS) instances within the security group.

Prerequisites

A security group is created and security group rules are added. For more information, see the Create a security group and Add a security group rule.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Security Groups page, find the security group for which you want to modify security group rules and click **Manage Rules** in the **Actions** column.
5. On the Rules page, click the **Inbound** or **Outbound** tab.
6. Find the security group rule that you want to modify and click **Modify Properties** in the **Actions** column.
7. In the **Modify Properties** dialog box, modify the rule.

   For more information about the properties of security group rules, see Add a security group rule.
8. Click **OK**.

10.6.2. Clone a security group rule

You can clone a security group rule to quickly create a similar rule.

Context

You may need to clone a security group rule in the following scenarios:

- Assume that you have created Security Group Rule A to control TCP access and that you want to create Security Group Rule B to control UDP access. You can create Rule B by cloning Rule A and changing the protocol type to UDP.
- Assume that you have created Security Group Rule C to control TCP access over port 22 and that you want to create Security Group Rule D to control TCP access over port 3389. You can create Rule D by cloning Rule C and changing the port range to 3389/3389.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Security Groups page, find the security group in which you want to clone a rule and click **Manage Rules** in the **Actions** column.
5. On the Rules tab, click the **Inbound** or **Outbound** tab.
6. Find the rule that you want to clone and click **Clone** in the **Actions** column.
7. In the **Clone Rule** dialog box, modify the attributes of the rule.

   For more information, see Add a security group rule.
8. Click **OK**.

   After the rule is cloned, a new rule is displayed in the rule list.

10.6.3. Export security group rules

You can export security group rules as Excel files to your computer for backup purposes.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the security group from which you want to export rules. Click **Manage Rules** in the **Actions** column.
5. On the Rules tab, click the **Inbound** or **Outbound** tab.
6. Click **Export** in the upper-right corner to download and save the rules to your computer.

10.6.4. Import security group rules

Security group rules can be imported to security groups. You can import a backup file of security group rules from your computer to a security group to create or restore security group rules.

Context

You can download a template file (XLS file), configure security group rules in the file based on the template requirements, and then import the file.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the security group to which you want to import rules. Click **Manage Rules** in the **Actions** column.
5. On the **Rules** tab, click **Import**.
6. In the **Import Security Group Rule** dialog box, click **Upload File**.
7. Select a backup file of security group rules from your computer and click **Open** to upload the file.

10.6.5. Delete a security group rule

You can delete security group rules that are no longer needed from a security group.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > Security Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the security group from which you want to delete a rule. Click **Manage Rules** in the **Actions** column.
5. On the Rules tab, click the **Inbound** or **Outbound** tab.
6. Use one of the following methods to delete a security group rule:
   - To delete a single rule, find the rule and click **Delete** in the **Actions** column.
   - To delete one or more rules at a time, select the rules and click **Delete** in the lower-left corner of the tab.
7. In the message that appears, click **Delete**.

11.RAM role management

11.1. Attach an instance RAM role to an ECS instance

This topic describes how to attach an instance RAM role to an Elastic Compute Service (ECS) instance.

Prerequisites

- An instance RAM role is created. For more information, see *Create an instance RAM role* in *Apsara Uni-manager Management Console User Guide*.
- The ECS instance to which you want to attach a RAM role is located in a virtual private cloud (VPC).

> **Note** An instance RAM role can be attached to a single instance at a time.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the instance to which you want to attach a RAM role, click the ... icon in the **Actions** column, and then choose **Instance Settings > Attach/Detach RAM Role**.
5. In the Attach/Detach RAM Role dialog box, select the RAM role from the RAM Role drop-down list and click **OK**.

11.2. Replace the instance RAM role of an ECS instance

After you attach an instance RAM role to an Elastic Compute Service (ECS) instance, you can replace the role at anytime.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the ECS instance that has an instance RAM role attached, click the ... icon in the **Actions** column, and then choose **Instance Settings > Attach/Detach RAM Role**.
5. Set **Operation Type** to **Attach**. Select another instance RAM role from the **RAM Role** drop-down list and click **OK** to replace the role.

11.3. Detach an instance RAM role from an ECS instance

After you attach an instance RAM role to an Elastic Compute Service (ECS) instance, you can detach the role from the instance at anytime.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find an ECS instance that has an instance RAM role attached, click the ... icon in the **Actions** column, and then choose **Instance Settings > Attach/Detach RAM Role**.
5. Set **Operation Type** to **Detach** and click **OK** to detach the role.

12.Elastic Network Interfaces

12.1. Overview

An elastic network interface (ENI) is a virtual network interface controller (NIC) that can be bound to an Elastic Compute Service (ECS) instance of the Virtual Private Cloud (VPC) type. You can use ENIs to deploy high-availability clusters and perform low-cost failover and fine-grained network management.

Attributes

An ENI is a virtual network interface that must be bound to an instance of the VPC type before the ENI can be used. The following table describes the attributes of an ENI.

Attribute	Description
ENI type	ENIs are classified into primary and secondary ENIs. • Primary ENIs: created together with instances. The lifecycle of a primary ENI is the same as the instance to which the primary ENI is bound. You cannot unbind a primary ENI from the instance to which the primary ENI is bound. • Secondary ENIs: can be separately created. You can bind or unbind a secondary ENI to or from an instance.
VPC	Only instances of the VPC type support ENIs. An ENI must reside within the same VPC as the instance to which the ENI is bound.
Organization	An ENI must belong to the same organization as the instance to which the ENI is bound.
Resource set	An ENI must belong to the same resource set as the instance to which the ENI is bound.
Region	An ENI must reside within the same region as the instance to which the ENI is bound.
Zone	The vSwitch to which the ENI belongs must reside within the same zone as the instance to which the ENI is bound.
Security group	An ENI must be added to at least one security group. The security group controls the inbound and outbound traffic of the ENI.
EIP	An ENI can be associated with one or more elastic IP addresses (EIPs).
Primary private IP address	The primary private IP address is an IP address specified by the user or assigned by the system during ENI creation. The primary private IP address must be an idle IP address within the CIDR block of the vSwitch to which the ENI is connected.

Features

An ENI is an independent virtual NIC that can be migrated between multiple instances to support the flexible scaling and migration of services. When you create an ENI together with an instance, the ENI is automatically bound to the instance. You can also separately create a secondary ENI and bind it to an instance.

ENIs have the following features:

- In addition to the primary ENI that is created together with an instance, you can also bind multiple secondary ENIs to the instance. The ECS instance and the secondary ENIs that you want to bind to the instance must reside within the same zone and VPC, but can be connected to different vSwitches and belong to different security groups.
- Each ENI can be assigned multiple secondary private IP addresses based on the instance type of the instance to which the ENI is bound.
- When you unbind a secondary ENI from an instance and bind the ENI to another instance, the attributes of the ENI remain unchanged and the network traffic is redirected to the new instance.
- ENIs support hot-plug and can be migrated among instances. When you unbind an ENI from an instance and bind the ENI to another instance, services on the instances are not affected, and you do not need to restart the instances.

Limits

- The following limits apply to ENIs:
  - Each ENI can be assigned only a single primary private IP address.
  - Each ENI can be assigned one or more secondary private IP addresses. The number of secondary private IP addresses is determined based on the instance type of the instance to which the ENI is bound.
  - Each ENI can be assigned one or more EIPs. The number of EIPs is determined based on how the EIPs are associated with the ENI. For more information, see the "Create a VPC" and "Create a vSwitch" topics in *Apsara Stack VPC User Guide*.
  - Each ENI must be added to at least one security group and can be added to up to five security groups.
- ENIs that can be created per region in an account is limited. The ENI quota is displayed on the **Quota overview** page of the Apsara Uni-manager Management Console.
- The ENI and the instance to which the ENI is bound must reside within the same zone and VPC, but can belong to different vSwitches and security groups.
- The number of secondary ENIs that can be bound to an ECS instance is determined based on the instance type.
- The instance bandwidth is determined based on the instance type. You cannot increase the bandwidth of an ECS instance by binding multiple secondary ENIs to the instance.

Use scenarios

ENIs are suitable for the following scenarios:

- Deployment of high-availability clusters

  To implement a high-availability architecture, you can bind multiple ENIs to a single ECS instance.

- Low-cost failover

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

- Fine-grained network management

You can configure multiple ENIs for an instance. For example, you can use some ENIs for internal management and other ENIs for Internet business access to isolate management data from business data. You can also configure specific security group rules for each ENI based on the source IP addresses, protocols, and ports to achieve access control.

- Configuration of multiple private IP addresses for a single instance

  You can assign multiple secondary private IP addresses to an ENI. If multiple applications are managed on your instance, you can assign an independent IP address for each application to improve the utilization of your instance.

- Configuration of multiple public IP addresses for a single instance

  Only a single public IP address can be assigned to an ECS instance that has no ENIs bound. To assign multiple public IP addresses to an instance, you can associate EIPs with one or more ENIs of the instance. In NAT mode, each private IP address of an ENI can have EIPs associated.

12.2. Create an ENI

You can bind elastic network interfaces (ENIs) to instances to create high-availability clusters and implement fine-grained network management. You can also unbind an ENI from an instance and then bind the ENI to another instance to perform low-cost failover.

Prerequisites

- A virtual private cloud (VPC) and a vSwitch are created. For more information, see Create a VPC and Create a vSwitch in *Apsara Stack VPC User Guide*.
- A security group is available in the VPC. If no security group is available in the VPC, create a security group. For more information, see Create a security group.

Context

You can use one of the following methods to create an ENI. This topic describes how to separately create an ENI.

- Create an ENI when you create an instance.

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

- Separately create an ENI.

  ENIs created separately are secondary ENIs. You can bind secondary ENIs to or unbind them from instances.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > ENIs**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create ENI**.
5. Configure the parameters described in the following table.

Section	Parameter	Required	Description

Section	Parameter	Required	Description
Area	Organization	Yes	The organization in which to create the ENI.
	Resource Set	Yes	The resource set in which to create the ENI.
	Region	Yes	The region in which to create the ENI.
	Zone	Yes	The zone in which to create the ENI.
Basic Configurations	VPC	Yes	The VPC in which to create the ENI. The secondary ENI can be bound only to an instance in the same VPC. Select the VPC in which the instance resides. **Note** After an ENI is created, its VPC cannot be changed.
	VSwitch	Yes	The vSwitch to which the ENI is connected. The secondary ENI can be bound to only an instance that is in the same zone as the VPC. Select a vSwitch that is in the same zone as the instance to which the ENI is to be bound. The instance and the ENI can connect to different vSwitches. **Note** After an ENI is created, its vSwitch cannot be changed.
	Security Group	Yes	The security group in which to create the ENI within the specified VPC. The rules of the security group automatically apply to the ENI.
	ENI Name	Yes	The name of the ENI. The name must be 2 to 128 characters in length. It must start with a letter and cannot start with http:// or https://. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,).

Section	Parameter	Required	Description
	Description	No	The description of the ENI. We recommend that you provide enough information to facilitate easy management. The description must be 2 to 256 characters in length. It must start with a letter and cannot start with http:// or https://. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,).
	Primary Private IP Address	No	The primary private IPv4 address of the ENI. The IPv4 address must be within the CIDR block of the specified vSwitch. If you do not specify a primary private IP address, the system automatically assigns a private IP address to the ENI.

6. Click **Submit**.

Result

If the ENI is created, **Available** is displayed in the Status column in the ENI list.

What's next

After you create an ENI separately, you can bind it to an instance. For more information, see Bind a secondary ENI to an instance.

12.3. Bind a secondary ENI to an instance

You can bind a secondary elastic network interface (ENI) to an instance. After the ENI is bound, the traffic on the ENI is directed to the instance.

Prerequisites

- The secondary ENI that you want to bind is in the **Available** state.
- The instance to which you want to bind the secondary ENI is in the **Running** or **Stopped** state.
- The instance and the secondary ENI belong to the same virtual private cloud (VPC).
- The vSwitch with which the secondary ENI is associated is located within the same zone as the vSwitch to which the instance is connected. An ENI can be bound only to an instance within the same zone. The vSwitches of the ENI and of the instance can be different but must be located within the same zone.

Context

The following limits apply when you bind an ENI to an instance:

- Only secondary ENIs can be manually bound. Primary ENIs share the same lifecycle as instances and cannot be manually bound.

- An ENI can be bound only to a single instance at the same time. Each instance can have one or more bound ENIs. The maximum number of ENIs that can be bound to an instance is determined based on the instance type.

Bind an ENI to an existing instance on the instance details page

To bind multiple secondary ENIs to an instance, you can go to the details page of the instance.

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Instances page, find the instance to which you want to bind a secondary ENI and click the instance ID.
5. Click the **ENIs** tab.
6. Click **Bind ENI**.
7. In the Bind ENI dialog box, select an ENI from the **ENI** drop-down list.
8. Click **OK**.

Bind an ENI to an existing instance on the ENI page

To bind secondary ENIs to multiple instances, you can go to the ENI page.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > ENIs**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the secondary ENI that you want to bind and click **Bind to Instance** in the **Actions** column.
5. In the **Bind to Instance** dialog box, select an instance and click **OK**.
   If the ENI is bound to the instance, **Bound** is displayed in the **Status** column corresponding to the ENI.

12.4. Configure a secondary ENI

After secondary elastic network interfaces (ENIs) are bound to Elastic Compute Service (ECS) instances, some images used by these instances may not recognize the IP addresses of these ENIs or configure routes for the ENIs. If this occurs, the secondary ENIs cannot be used on the instances. This topic describes how to configure secondary ENIs from within instances to have their IP addresses recognizable and how to configure routes for the secondary ENIs.

Prerequisites

- A secondary ENI is bound to an ECS instance. For information about how to bind a secondary ENI to an ECS instance, see Bind a secondary ENI to an instance.
- You are connected to the ECS instance. For information about how to connect to an ECS instance, see Instance connecting overview.

Context

If automatic configuration tools are pre-installed on the images that are used by instances, secondary ENIs that are bound to the instances can be automatically configured by the tools. You can use these secondary ENIs without the need to manually configure them. Skip the topic if your instance uses an image of one of the following versions:

- Alibaba Cloud Linux 3.2104 64-bit
- CentOS 8.0 64-bit, CentOS 8.1 64-bit, and CentOS 8.2 64-bit
- CentOS 7.3 64-bit, CentOS 7.4 64-bit, and CentOS 7.5 64-bit
- CentOS 6.8 64-bit and CentOS 6.9 64-bit
- Debian 10.5 64-bit and Debian 10.6 64-bit
- Windows Server 2008 R2 and later

Procedure

1. Check whether the IP addresses of a secondary ENI bound to an instance are recognized.

   For more information, see Check whether the IP addresses of ENIs are recognized. If yes, skip the following steps. If not, proceed with the following steps to configure the secondary ENI.

2. Obtain the information of the secondary ENI.

   When you configure a secondary ENI, you may need to specify its information, such as the primary private IP address and media access control (MAC) address. Obtain the information required for subsequent configurations. For more information, see Obtain the information of an ENI.

   In the examples provided in this topic, the sample values listed in the following table are used. In actual scenarios, replace them with the attribute values of your secondary ENI.

Secondary ENI attribute	Sample value
ENI name	eth1
MAC address	00:16:3e:0f:**:**
Primary private IP address	192.168.**.*2
Subnet mask	255.255.255.0
Gateway address	192.168.**.253

3. Configure the secondary ENI to have its IP address recognized.

   The operations required to configure secondary ENIs vary based on the operating systems of the instances to which the secondary ENIs are bound.

Operating system	References
◦ Alibaba Cloud Linux 2 (Instances that run this operating system use the network-scripts network service) ◦ CentOS	Configure a secondary ENI for an instance that runs an Alibaba Cloud Linux 2, CentOS 6, or CentOS 7 operating system and uses the network-scripts network service

Operating system	References
Alibaba Cloud Linux 2 (Instances that run this operating system use the systemd-networkd network service)	Configure a secondary ENI for an instance that runs an Alibaba Cloud Linux 2 operating system and uses the systemd-networkd network service
○ Ubuntu ○ Debian	Configure a secondary ENI for an instance that runs a Ubuntu or Debian operating system
○ SUSE ○ openSUSE	Configure a secondary ENI for an instance that runs a SUSE or openSUSE operating system

4. Check whether routes are configured for the secondary ENI.

   You can run the `route -n` command to check route information. If no routes are configured for the secondary ENI or if the existing routes do not meet your requirements, manually configure routes for the secondary ENI. The following sections provide examples on how to configure routes for a secondary ENI that is bound to an instance that runs one of the following operating systems:

   - Configure routes for a secondary ENI of an instance that runs an Alibaba Cloud Linux 2 or CentOS 7 operating system
   - Configure routes for a secondary ENI of an instance that runs a CentOS 8 operating system

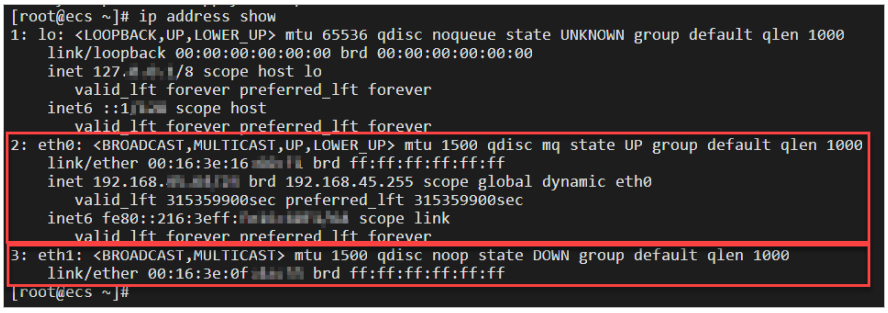
Check whether the IP addresses of ENIs are recognized

Run the following command to check whether the IP addresses of ENIs are recognized:

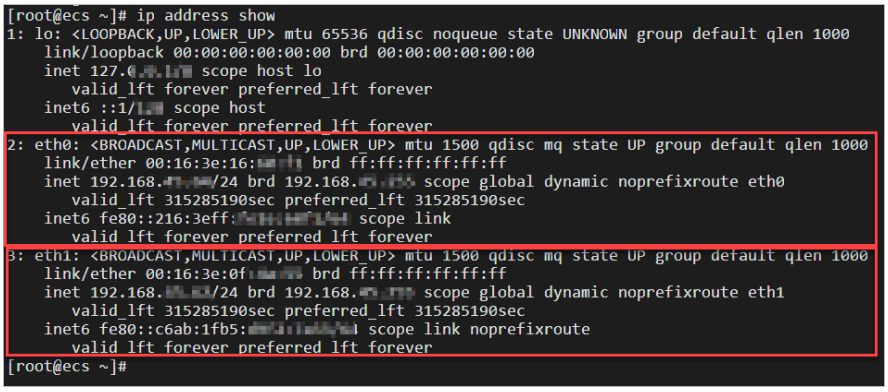
```
ip address show
```

Sample command outputs:

- The following command output shows that the IP address of the eth0 primary ENI is recognized but the IP address of the eth1 secondary ENI is not recognized. You can perform operations described in this topic to configure the secondary ENI.



- The following command output shows that the IP addresses of both the eth0 primary ENI and the eth1 secondary ENI are recognized. You do not need to configure the secondary ENI.



> **Note** In the preceding command outputs, `00:16:3e:16:**:**` is the MAC address of the primary ENI and `00:16:3e:0f:**:**` is the MAC address of the secondary ENI.

Obtain the information of an ENI

You can obtain the information of an ENI from instance metadata, by using the ECS console, or by calling an API operation.

- Obtain the information of an ENI from instance metadata.
  - Obtain the MAC addresses of the ENIs that are bound to an instance.

    ```
    curl http://100.100.100.200/latest/meta-data/network/interfaces/macs/
    ```

    > **Note** The MAC addresses of ENIs are required to obtain their primary private IP addresses, subnet masks, and gateway addresses.

  - Obtain the primary private IP address of the specified ENI.

    ```
    curl http://100.100.100.200/latest/meta-data/network/interfaces/macs/00:16:3e:19:**:**/primary-ip-address
    ```

  - Obtain the subnet mask of the specified ENI.

    ```
    curl http://100.100.100.200/latest/meta-data/network/interfaces/macs/00:16:3e:19:**:**/netmask
    ```

  - Obtain the gateway address of the specified ENI.

    ```
    curl http://100.100.100.200/latest/meta-data/network/interfaces/macs/00:16:3e:19:**:**/gateway
    ```

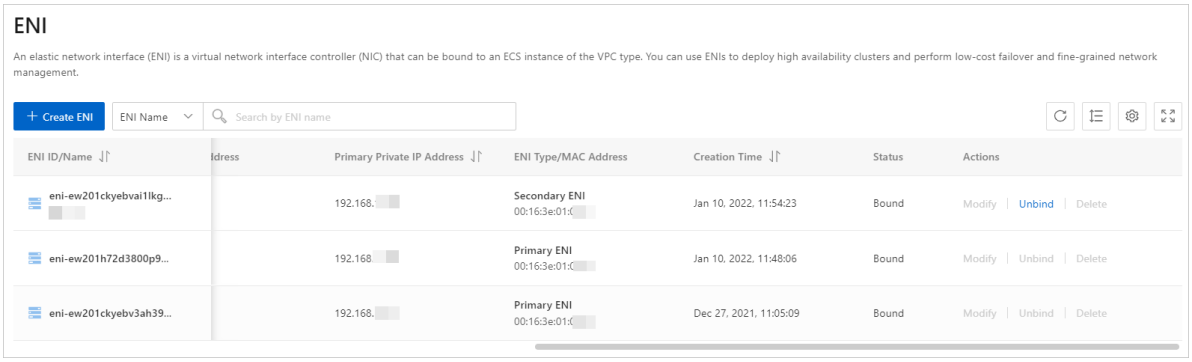
  The following figure shows sample command outputs. In the command outputs, `00:16:3e:16:**:**` is the MAC address of the primary ENI and `00:16:3e:0f:**:**` is the MAC address of the secondary ENI.

  > **Note** After you run the `ip address show` command, you can determine which is the primary ENI and which is the secondary ENI based on the order in which the MAC addresses are displayed in the command output.

```
[root@ecs ~]# curl http://100.100.100.200/latest/meta-data/network/interfaces/macs/
00:16:3e:0f
00:16:3e:16curl http://100.100.100.200/latest/meta-data/network/interfaces/macs/00:16:3e:0f/primary-ip-address
192.168.curl http://100.100.100.200/latest/meta-data/network/interfaces/macs/00:16:3e:0f/netmask
255.255.curl http://100.100.100.200/latest/meta-data/network/interfaces/macs/00:16:3e:0f/gateway
192.168.[root@ecs ~]#
```

- Obtain the information of an ENI by using the ECS console.
  i. Log on to the ECS console.
  ii. In the left-side navigation pane, choose **Network & Security > ENIs**.
  iii. On the **ENI** page, find the ENI whose information you want to query and view its primary private IP address and MAC address in the Private IP Address and ENI Type/MAC Address columns.

  The following figure shows an example of the ENI page in the ECS console.



- Obtain the information of an ENI by using an Alibaba Cloud SDK to call the DescribeNetworkInterfaces operation.

```
aliyun ecs DescribeNetworkInterfaces \
--output cols=MacAddress,PrivateIpAddress rows=NetworkInterfaceSets.NetworkInterfaceSet[]
\
--RegionId 'cn-qingdao-****-d01' \
--InstanceId 'i-bp1a5gj0bzhwz7q****'
```

The following figure shows the sample response. In the response, `00:16:3e:16:**:**` is the MAC address of the primary ENI and `00:16:3e:0f:**:**` is the MAC address of the secondary ENI.

> **Note** After you run the `ip address show` command, you can determine which is the primary ENI and which is the secondary ENI based on the order in which the MAC addresses are displayed in the command output.

```
shell@Alicloud:~$ aliyun ecs DescribeNetworkInterfaces --output cols=MacAddress,PrivateIpAddress
rows=NetworkInterfaceSets.NetworkInterfaceSet[] --RegionId 'cn-hangzhou' --InstanceId 'i-bp1a5gj0
'
MacAddress          | PrivateIpAddress
--------------------| ----------------
00:16:3e:16:        | 192.168.
00:16:3e:0f:        | 192.168.
```

Configure a secondary ENI for an instance that runs an Alibaba Cloud Linux 2, CentOS 6, or CentOS 7 operating system and uses the network-scripts network service

If your instance runs an Alibaba Cloud Linux 2, CentOS 6, or CentOS 7 operating system and uses the network-scripts network service, you can use the multi-nic-util tool to have the ENIs of the instance automatically configured or you can manually configure the ENIs by modifying their configuration files.

- Use the multi-nic-util tool to have a secondary ENI automatically configured.

  > **Note** If you want to use the multi-nic-util tool to have secondary ENIs automatically configured for CentOS instances, note that the multi-nic-util tool is supported only on some versions of CentOS images. If your instance uses a CentOS 6 image, make sure that the image version is CentOS 6.8 or later. If your instance uses a CentOS 7 image, make sure that the image version is CentOS 7.3 or later. If the multi-nic-util tool is not supported on the image version that your instance uses, manually configure secondary ENIs by modifying their configuration files.

  i. Download and install the multi-nic-util tool.

  ```
  wget https://image-offline.oss-cn-hangzhou.aliyuncs.com/multi-nic-util/multi-nic-util-0.6.tgz && \
  tar -zxvf multi-nic-util-0.6.tgz && \
  cd multi-nic-util-0.6 && \
  bash install.sh
  ```

  ii. Restart the ENI service.

  ```
  systemctl restart eni.service
  ```

- Manually configure a secondary ENI by modifying its configuration file.

  i. Open the configuration file of the secondary ENI.

  ```
  vi /etc/sysconfig/network-scripts/ifcfg-eth1
  ```

  ii. Add the information of the secondary ENI to the configuration file. Then, save and close the configuration file.

  The following section provides an example of the ENI information to add to the configuration file:

  ```
  DEVICE=eth1  # Specify the ENI that you want to configure.
  BOOTPROTO=dhcp
  ONBOOT=yes
  TYPE=Ethernet
  USERCTL=yes
  PEERDNS=no
  IPV6INIT=no
  PERSISTENT_DHCLIENT=yes
  HWADDR=00:16:3e:0f:**:**  # Enter the obtained MAC address of the ENI.
  DEFROUTE=no  # Specify not to use the ENI as the default route. To prevent the default route of the ECS instance from being changed when you run the ifup command to start the secondary ENI, do not set eth1 as the default route.
  ```

  iii. Restart the network service.

  - For versions earlier than CentOS 7, such as CentOS 6, run the following command:

    ```
    service network restart
    ```

  - For CentOS 7 or later and Alibaba Cloud Linux 2, run the following command:

    ```
    systemctl restart network
    ```

> **Note**
>
> - After you configure the secondary ENI, you can configure routes for the ENI. For more information, see Configure routes for a secondary ENI of an instance that runs an Alibaba Cloud Linux 2 or CentOS 7 operating system.
> - If you want to create custom images from the instance whose ENIs are configured, you must first run the **/etc/eni_utils/eni-cleanup** command to remove network configurations from */etc/udev/rules.d/70-persistent-net.rules* and */etc/sysconfig/network-scripts/*.

Configure a secondary ENI for an instance that runs an Alibaba Cloud Linux 2 operating system and uses the systemd-networkd network service

If your instance runs an Alibaba Cloud Linux 2 operating system and uses the systemd-networkd network service, you must manually configure a secondary ENI by modifying the configuration file of the ENI.

1. Open the configuration file of the secondary ENI.

   ```
   vi /etc/systemd/network/60-eth1.network
   ```

2. Add the information of the secondary ENI to the configuration file. Then, save and close the configuration file.

   You can use one of the following methods to assign a dynamic or static IP address to the secondary ENI based on your requirements. The following section provides examples of the ENI information to add to the configuration file when different methods are used:

   - Assign a dynamic IP address to the secondary ENI by using the Dynamic Host Configuration Protocol (DHCP).

     ```
     [Match]
     Name=eth1 # Specify the ENI that you want to configure.
     [Network]
     DHCP=yes
     [DHCP]
     UseDNS=yes
     ```

   - Assign a static IP address to the secondary ENI.

     ```
     [Match]
     Name=eth1 # Specify the ENI that you want to configure.
     [Network]
     Address=192.168.**.*2/24 # Specify the static IP address and subnet mask to be assigned.
     ```

     > **Note** In the preceding example, `192.168.**.*2` is the primary private IP address and the `/24` subnet mask is `255.255.255.0`.

3. Restart the network service.

   ```
   systemctl restart systemd-networkd
   ```

Configure a secondary ENI for an instance that runs a Ubuntu or Debian operating system

If your instance runs a Ubuntu or Debian operating system, you must configure a secondary ENI by modifying the network interface controller (NIC) configuration file based on your image version.

- Perform the following operations on an instance that runs a Ubuntu 14.04, Ubuntu 16.04, or Debian operating system:
  i. Open the NIC configuration file.

  ```
  vi /etc/network/interfaces
  ```

  ii. Add the information of the secondary ENI to the configuration file. Then, save and close the configuration file.

  The following section provides an example of the ENI information to add to the configuration file:

  ```
  auto eth0
  iface eth0 inet dhcp
  auto eth1  # Specify the ENI that you want to configure.
  iface eth1 inet dhcp
  ```

  > **Note** The eth0 primary ENI is configured in the same configuration file as the eth1 secondary ENI. You must also add the information of the primary ENI to the configuration file.

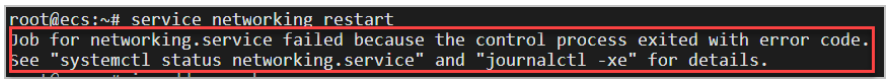
  iii. Restart the network service.
  - For versions earlier than Ubuntu 16.04, such as Ubuntu 14.04, run the following command:

    ```
    service networking restart
    ```

  - For Ubuntu 16.04 and Debian, run the following command:

    ```
    systemctl restart networking
    ```

  The configurations of the secondary ENI take effect regardless of whether the following alert notification appears. You can run the `ip address show` command to check whether the IP addresses of the secondary ENI are recognized.

- Perform the following operations on an instance that runs Ubuntu 18.04:
  i. Open the configuration file of the secondary ENI.

  ```
  vi /etc/netplan/eth1-netcfg.yaml
  ```

  ii. Add the information of the secondary ENI to the configuration file. Then, save and close the configuration file.

> **Note** When you modify the configuration file, take note of the following items:
>
> - The configuration file is in the `YAML` format. You must follow the `YAML` syntax rules when you configure the file.
> - Tabs cannot be used for indentation in `YAML` files. Use spaces instead.
> - We recommend that you copy information from the default */etc/netplan/99-netcfg.yaml* configuration file to prevent format issues.

The following section provides an example of the ENI information to add to the configuration file:

```
network:
  version: 2
  renderer: networkd
  ethernets:
    eth1:
      dhcp4: yes
      dhcp6: no
```

iii. Apply the added configurations.

```
netplan apply
```

Configure a secondary ENI for an instance that runs a SUSE or openSUSE operating system

If your instance run a SUSE or openSUSE operating system, you must manually configure a secondary ENI by modifying its configuration file.

1. Open the configuration file of the secondary ENI.

   ```
   vi /etc/sysconfig/network/ifcfg-eth1
   ```

2. Add the information of the secondary ENI to the configuration file. Then, save and close the configuration file.

   In the following example, a dynamic IP address is assigned to the secondary ENI by using DHCP.

   ```
   BOOTPROTO='dhcp4'
   STARTMODE='auto'
   USERCONTROL='no'
   ```

3. Restart the network service.
   - For versions earlier than SUSE Linux Enterprise Server 12, run the following command:

     ```
     service network restart
     ```

   - For SUSE Linux Enterprise Server 12 or later, run the following command:

     ```
     systemctl restart network
     ```

Configure routes for a secondary ENI of an instance that runs an Alibaba Cloud Linux 2 or CentOS 7 operating system

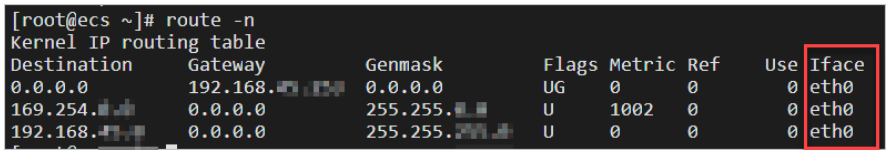
If you manually configure secondary ENIs but do not configure routes for them or if routes configured by the multi-nic-util tool do not meet your requirements, perform the following steps to configure routes.

1. View route information.

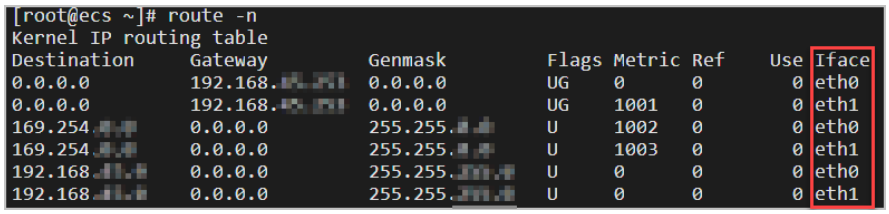
   ```
   route -n
   ```

   Sample command outputs:

   - The following command output shows only the route information of the eth0 primary ENI, which indicates that no routes are configured for the eth1 secondary ENI.

   - The following command output shows the route information of both the eth0 primary ENI and the eth1 secondary ENI. If the configured routes do not meet your requirements, you can modify the routes.

2. Plan the default route based on your requirements.

   In this example, the sample values listed in the following table are used.

Secondary ENI attribute	Sample value
ENI name	eth1
Primary private IP address	192.168.**.*2
Gateway address	192.168.**.253
metric	1001

3. Configure the default route.

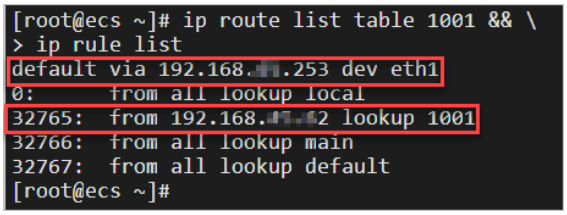
   You can run the following commands to add the default route for the eth1 secondary ENI, create a route table, and then attach a routing policy to the table. In this example, a route table named *table 1001* is created. We recommend that you keep the route table name the same as the metric value in the default route of the ENI. *192.168.**.253* is the gateway address and *192.168.*.*2* is the primary private IP address of the eth1 secondary ENI.

   ```
   ip -4 route add default via 192.168.**.253 dev eth1 metric 1001 && \
   ip -4 route add default via 192.168.**.253 dev eth1 table 1001 && \
   ip -4 rule add from 192.168.**.*2 lookup 1001
   ```

4. View the created route table and routing policy.

```
ip route list table 1001 && \
ip rule list
```

The following figure shows that the route table and routing policy are created.



5. Configure routes to automatically update on instance startup.

   After you perform the preceding steps to configure routes for the eth1 secondary ENI, you must perform the following steps to configure the routes to automatically update on instance startup. Otherwise, the routes become invalid when the instance is restarted.

   i. Open the */etc/rc.local* file.

   ```
   vim /etc/rc.local
   ```

   ii. Add the configuration information of the routes to the */etc/rc.local* file. Then, save and close the file.

   ```
   ip -4 route add default via 192.168.**.253 dev eth1 metric 1001
   ip -4 route add default via 192.168.**.253 dev eth1 table 1001
   ip -4 rule add from 192.168.**.*2 lookup 1001
   ```

   iii. Grant execution permissions on the */etc/rc.local* file.

   ```
   chmod +x /etc/rc.local
   ```

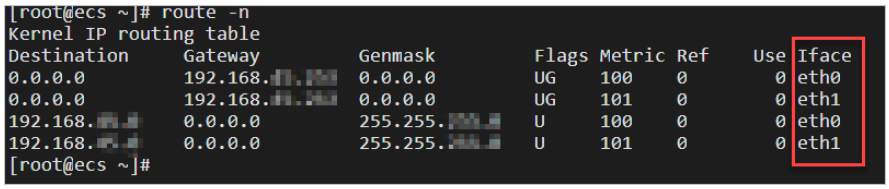
Configure routes for a secondary ENI of an instance that runs a CentOS 8 operating system

If routes configured by the system do not meet your requirements, perform the following steps to configure routes.

1. View route information.

   ```
   route -n
   ```

   The following command output shows the route information of both the eth0 primary ENI and the eth1 secondary ENI. If the configured routes do not meet your requirements, you can modify the routes.



2. Plan the default route based on your requirements.

   In this example, the sample values listed in the following table are used.

Secondary ENI attribute	Sample value
ENI name	eth1
Primary private IP address	192.168.**.*2
Gateway address	192.168.**.253
table	1001

3. Create a script for configuring routes.
   i. Create and open the */root/route.sh* file.
   ii. Add the configuration information of the routes to the */root/route.sh* file. Then, save and close the file.

   The following section shows how to create a route table and attach a routing policy to the route table for the eth1 secondary ENI. In this example, a route table named *table 1001* is created. *192.168.**.253* is the gateway address and *192.168.*.*2* is the primary private IP address of the eth1 secondary ENI.

```
#!/bin/bash
i=0
while true; do
        /usr/sbin/ip -4 route add default via 192.168.**.253 dev eth1 table 1001
        if [ $? -eq 0 ]; then
                break
    fi
        sleep 3
        let i++
        if [ $i -gt 10 ]; then
                exit -1
        fi
done
i=0
while true; do
        /usr/sbin/ip -4 rule add from 192.168.**.*2 lookup 1001
        if [ $? -eq 0 ]; then
                break
    fi
        sleep 3
        let i++
        if [ $i -gt 10 ]; then
                exit -1
        fi
done
```

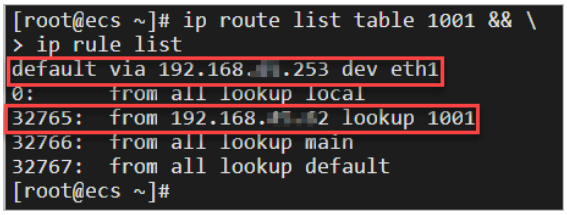
4. Configure the default route.

```
sh /root/route.sh
```

5. View the created route table and routing policy.

```
ip route list table 1001 && \
ip rule list
```

The following figure shows that the route table and routing policy are created.



6. Configure routes to automatically update on instance startup.

   After you perform the preceding steps to configure routes for the eth1 secondary ENI, you must perform the following steps to configure the routes to automatically update on instance startup. Otherwise, the routes become invalid when the instance is restarted.

   i. Open the */etc/rc.local* file.

   ```
   vim /etc/rc.local
   ```

   ii. Add the configuration information of the routes to the */etc/rc.local* file. Then, save and close the file.

   ```
   sh /root/route.sh
   ```

   iii. Grant execution permissions on the */etc/rc.local* file.

   ```
   chmod +x /etc/rc.local
   ```

12.5. View ENIs

You can view the list of created elastic network interfaces (ENIs).

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > ENIs**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Use one of the following methods to search for ENIs that meet specified conditions:
   - Select an option as the filter condition from the drop-down list, enter the corresponding information, and then click the search icon. The ENI list displays the ENIs that match the specified condition.
   - Click **Advanced Filter** on the ENI page. Then, select one or more options as filter conditions, enter the corresponding information, and then click **Search**. The ENI list displays the ENIs that match the specified conditions.

   > **Note** When you use the advanced filtering feature, you can specify multiple filter conditions to narrow down search results.

Filter option	Description
ENI Name	Enter an ENI name to search for the ENI.
ENI ID	Enter an ENI ID to search for the ENI.
vSwitch ID	Enter a vSwitch ID to search for ENIs that are associated with the vSwitch.
Security Group ID	Enter a security group ID to search for ENIs that belong to the security group.
Instance ID	Enter an instance ID to search for ENIs that are bound to the instance.
Tag	Enter the key or value of a tag to search for ENIs to which the tag is added.

12.6. Modify the attributes of a secondary ENI

You can modify the attributes of a secondary elastic network interface (ENI), including its name and security group.

Prerequisites

The secondary ENI is in the **Available** state.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > ENIs**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the secondary ENI whose attributes you want to modify and click **Modify** in the **Actions** column.
5. In the **Modify ENI** dialog box, modify the name and security group of the ENI.
6. Click **OK**.

12.7. Unbind a secondary ENI from an ECS instance

You can unbind a secondary elastic network interface (ENI) from an Elastic Compute Service (ECS) instance. After the secondary ENI is unbound from the instance, the instance no longer processes the traffic on the ENI.

Prerequisites

Before you unbind a secondary ENI, make sure that the following requirements are met:

- The secondary ENI is in the **Bound** state.
- The instance to which the secondary ENI is bound is in the **Running** or **Stopped** state.

Context

Only secondary ENIs can be manually unbound. Primary ENIs share the same lifecycle as the associated instances and cannot be manually unbound.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > ENIs**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the secondary ENI that you want to unbind and click **Unbind** in the **Actions** column.
5. In the message that appears, click **Unbind**.

Result

If the secondary ENI is unbound from the instance, **Available** is displayed in the **Status** column corresponding to the ENI.

12.8. Delete a secondary ENI

You can delete elastic network interfaces (ENIs) that are no longer needed. Secondary ENIs can be deleted, whereas primary ENIs cannot.

Prerequisites

The secondary ENI is in the **Available** state.

Context

When an ENI is deleted, the following results occur:

- The primary private IP address (PrimaryIpAddress) of the ENI is automatically released.
- The ENI is removed from all security groups to which it was added.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Networks and Security > ENIs**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the secondary ENI that you want to delete and click **Delete** in the **Actions** column.
5. In the message that appears, click **Delete**.

Result

On the ENI page, refresh the ENI list. If the ENI is deleted, it is no longer displayed.

13.Key pairs

13.1. Overview

Key pairs are a secure and convenient authentication method provided by Alibaba Cloud for Elastic Compute Service (ECS) instance logons. A key pair consists of a public key and a private key. Only Linux instances support logon based on key pairs.

Introduction

A key pair consists of a pair of public and private keys that are generated based on an encryption algorithm. By default, 2048-bit RSA key pairs are used. Before you log on to a Linux instance by using a key pair, you must first create the key pair. You can specify a key pair when you create an instance, or bind a key pair to an instance after the instance is created. Then, you can use the private key to connect to the instance.

Benefits

Key-based authentication has the following advantages over username and password-based authentication:

- Security: Key pairs provide higher security and reliability for logons.
  - Key pairs are more secure than general user passwords against brute-force attacks.
  - Private keys cannot be deduced even if the public keys are maliciously acquired.
- Ease of use:
  - If you configure a public key on a Linux instance, you can use the corresponding private key to run SSH commands or other tools for passwordless logon to the instance.
  - You can log on to a large number of Linux instances at the same time. If you want to manage multiple Linux instances, we recommend that you use this method.

Limits

Key pairs have the following limits:

- If you use a key pair to log on to a Linux instance, the password logon method is disabled for higher security.
- Key pairs apply only to Linux instances.
- Currently, only RSA 2048-bit key pairs can be created in the ECS console.
- Only a single key pair can be bound to each Linux instance in the ECS console. If you bind a key pair to an instance to which another key pair is already bound, the new key pair replaces the original one. If you want to use multiple key pairs to log on to a Linux instance, you must manually modify the *~/.ssh/authorized_keys* file from within the instance to add multiple key pairs.
- If you bind a key pair to or unbind a key pair from an instance in the **Running** (`Running`) state, you must restart the instance for the operation to take effect. This enhances data security.

Creation methods

You can use one of the following methods to create a key pair:

- Create a key pair in the ECS console. By default, the key pair is generated in the RSA 2048-bit format.

For more information, see Create a key pair.

> **Notice** If you create a key pair in the ECS console, you must download and securely lock away the private key. After the key pair is bound to an instance, you cannot log on to the instance if you do not have the private key.

- Create a key pair by using a key pair generator and then import the key pair to the ECS console. The imported key pair must support one of the following encryption methods:
  - rsa
  - dsa
  - ssh-rsa
  - ssh-dss
  - ecdsa
  - ssh-rsa-cert-v00@openssh.com
  - ssh-dss-cert-v00@openssh.com
  - ssh-rsa-cert-v01@openssh.com
  - ssh-dss-cert-v01@openssh.com
  - ecdsa-sha2-nistp256-cert-v01@openssh.com
  - ecdsa-sha2-nistp384-cert-v01@openssh.com
  - ecdsa-sha2-nistp521-cert-v01@openssh.com

13.2. Create a key pair

This topic describes how to create a key pair in the Apsara Uni-manager Management Console. After a key pair is created, its private key is automatically downloaded. You must securely store the private key and ensure its confidentiality. To log on to an Elastic Compute Service (ECS) instance to which a key pair is bound, you must provide the private key. This topic describes how to create a key pair in the Apsara Uni-manager Management Console.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Network & Security > SSH Key Pairs**.
3. Click **Create Key Pair**.
4. On the Create Key Pair page, configure the parameters described in the following table.

Section	Parameter	Required	Description
Region	Organization	Yes	Select an organization in which to create the key pair. Make sure that the key pair and the VPC belong to the same organization.
	Resource Set	Yes	Select a resource set in which to create the key pair. Make sure that the key pair and the VPC belong to the same resource set.

Section	Parameter	Required	Description
	Region	Yes	Select a region in which to create the key pair. Make sure that the key pair and VPC belong to the same region.
Basic Settings	Key Pair Name	Yes	Enter a name for the key pair. The key pair name must be unique. The name must be 2 to 128 characters in length and can contain special characteristics periods (.), underscores (_), hyphens (-), and colons (:). The name cannot start with a special character or a digit.
	Creation Mode	Yes	Select a method of creating the key pair. We recommend that you select **Auto-create**. Then, you must securely store the private key in a timely manner and ensure its confidentiality. ○ **Auto-create**: The system creates a key pair for you. The private key is automatically downloaded after the key pair is created. The private key can be downloaded only once. You must securely store the private key file and ensure its confidentiality. ○ **Import**: You can enter a Base64-encoded public key in the Public Key field.

5. Click **Create**.

Result

After the key pair is created, your browser downloads the private key file to your computer.

Notice Private key files are downloaded to your computer only when Auto-create is selected. Private key files are not saved in the ECS console and cannot be recovered if they are lost. Make sure that you securely store your private key files to ensure their confidentiality.

What's next

Before you can use a created key pair to log on to an instance, you must bind the key pair to the instance. For more information, see Bind a key pair to an instance.

13.3. Bind a key pair to an instance

You can specify a key pair when you create an Elastic Compute Service (ECS) instance, or bind a key pair to the instance after the instance is created. This topic describes how to bind a key pair to an instance after the instance is created. If your ECS instance originally uses password-based authentication, the password-based authentication is automatically disabled after the key pair is bound.

Context

Only a single key pair can be bound to each ECS instance in the ECS console. If you bind a key pair to an instance to which another key pair is already bound, the new key pair replaces the original one.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Network & Security > SSH Key Pairs**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the key pair that you want bind and click **Bind Key Pair** in the **Actions** column.
5. In the **Select Instances** section of the Bind Key Pair dialog box, click the name of the ECS instance to which you want to bind the key pair.

   If instance names in the **Select Instance** section are dimmed, the instances are Windows instances and cannot have key pairs bound.
6. Click **OK**.
7. If the selected ECS instance is in the **Running** (*Running*) state, perform the following operations to restart the instance to make the binding operation take effect:
   i. In the left-side navigation pane, choose **Instances & Images > Instances**.
   ii. Find the instance that you want to restart and choose **Instance Status > Restart Instance** in the **Actions** column.
   iii. In the message that appears, click **Restart Instance**.

13.4. Unbind SSH key pairs

This topic describes how to unbind an SSH key pair in the Elastic Compute Service (ECS) console.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Network & Security > SSH Key Pairs**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the key pair that you want to unbind and click **Unbind Key Pair** in the **Actions** column.
5. In the **Select Instances** section of the Unbind dialog box, click the ID or name of the instance from which you want to unbind the key pair. Then, the instance is moved to the right-hand column.
6. Click **OK**.
7. If the selected ECS instance is in the **Running** (*Running*) state, perform the following operations to restart the instance to make the binding operation take effect:
   i. In the left-side navigation pane, choose **Instances & Images > Instances**.
   ii. Find the instance that you want to restart and choose **Instance Status > Restart Instance** in the **Actions** column.
   iii. In the message that appears, click **Restart Instance**.

What's next

After the SSH key pair is unbound, you must reset the password of the instance before you log on to the instance as the root user. For more information, see Change the logon password of an instance.

> **Note** If you have reset the password before you unbind the key pair, you can log on by using the new password after you unbind the key pair.

14.Deployment sets

14.1. Overview

Deployment Set is a service provided by Elastic Compute Service (ECS) that allows you to view the physical topology of hosts, racks, and vSwitches. You can select deployment policies based on your business requirements to improve the reliability and performance of your business.

Benefits

When you use multiple ECS instances within the same zone, you may have the following requirements on business reliability or performance:

- **High business reliability**

  To prevent the impacts caused by the failure of physical hosts, racks, or vSwitches, you want to distribute identical application instances across different physical hosts, racks, or vSwitches.

- **High network performance**

  In scenarios that involve frequent network interactions between instances, you want to associate the instances with the same vSwitch to achieve low-latency and high-bandwidth communication between the instances.

Deployment granularities and policies

- **Deployment Granularity**
  - **Host**: The minimum granularity for scheduling is a physical server.

    > **Note** When you create a deployment set, **Deployment Granularity** is set to **Host** by default.

  - **Rack**: The minimum granularity for scheduling is a rack.
  - **Network Switch**: The minimum granularity for scheduling is a vSwitch.
- **Deployment policy**
  - **Loose Aggregation**
  - **Strict Aggregation**
  - **Loose Dispersion**
  - **Strict Dispersion**

  Loose Aggregation and **Strict Aggregation** are intended for higher performance, while **Loose Dispersion** and **Strict Dispersion** are intended for higher availability.

The following table describes the deployment policies and use scenarios corresponding to each deployment granularity.

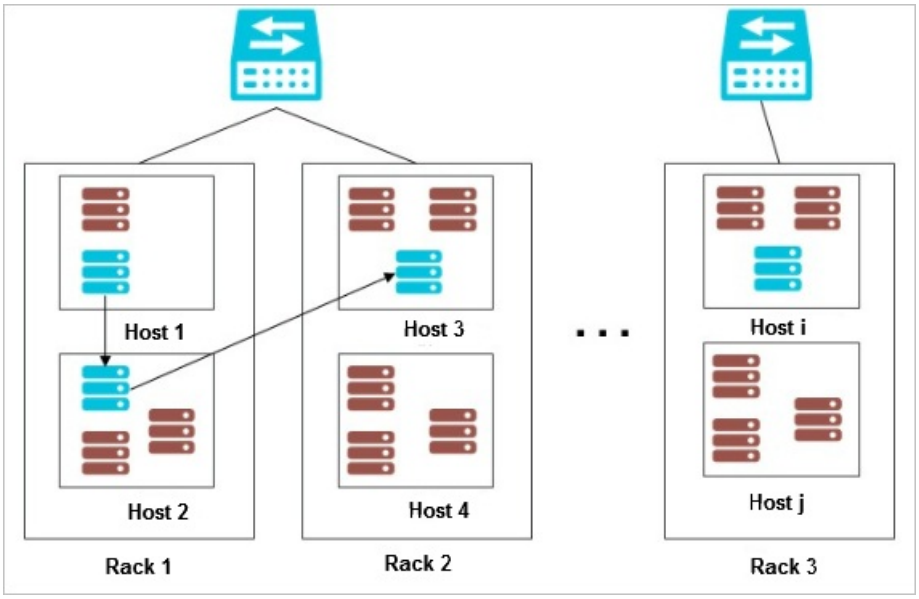
Granularities and policies

Deployment granularity	Deployment policy	Use Scenario
	Strict dispersion	

Deployment granularity	Deployment policy	Use Scenario
Host	Loose dispersion	General-purpose business scenarios.
Rack	Strict dispersion	Big data and databases
	Loose dispersion	Game customers
vSwitch	Strict dispersion	VPN
	Loose dispersion	Game customers
	Strict aggregation	Big data and databases
	Loose aggregation	Game customers

Example

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



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

14.2. Create a deployment set

You can use deployment sets to distribute or aggregate instances involved in your business. You can select hosts, racks, or switches to improve service availability or network performance.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Deployment & Elasticity > Deployment Sets**.

3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click **Create Deployment Set**.
5. Configure the parameters described in the following table to create a deployment set.

Section	Parameter	Required	Description
Area	Organization	Yes	The organization to which the deployment set belongs.
	Resource Set	Yes	The resource set to which the deployment set belongs.
	Region	Yes	The region in which the deployment set resides.
Basic Configurations	Deployment Domain	Yes	This parameter is used to determine the valid values of Deployment Granularity.
	Deployment Granularity	Yes	The basic unit in which instances can be scheduled when you deploy instances. Valid values: ○ **Host**: Instances are distributed at the host level. ○ **Rack**: Instances are distributed at the rack level. ○ **Network Switch**: Instances are distributed or aggregated at the switch level.
	Deployment Policy	No	The dispersion policies are used to improve service availability to prevent impacts on your business when a host, rack, or switch fails. The aggregation policies are used to improve network performance to minimize the access latency between instances. Valid values: ○ **Loose Dispersion** ○ **Strict Dispersion** ○ **Loose Aggregation** ○ **Strict Aggregation**
	Deployment Set Name	Yes	The name of the deployment set. The name must be 2 to 128 characters in length and can contain letters, digits, colons (:), underscores (_), and hyphens (-). It must start with a letter and cannot start with http:// or https://.

Section	Parameter	Required	Description
	Description	No	The description of the deployment set. To simplify future management operations, we recommend that you provide a suitable description. The description must be 2 to 256 characters in length and can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,). It must start with a letter and cannot start with http:// or https://.

6. Click **Submit**.

14.3. View deployment sets

You can view the list of deployment sets.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Deployment & Elasticity > Deployment Sets**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Select a filter option from the drop-down list and enter relevant information in the search box to search for a specific deployment set.

Filter option	Description
Deployment Set Name	Enter a deployment set name to search for the deployment set.
Deployment Set ID	Enter a deployment set ID to search for the deployment set.

14.4. Change the deployment set of an instance

This topic describes how to add an Elastic Compute Service (ECS) instance to a deployment set or migrate an instance from one deployment set to another.

Prerequisites

The instance is in the **Stopped** or **Running** state.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select an organization, a resource set, and a region.

4. Find the instance for which you want to change the deployment set and click the ... icon in the **Actions** column and choose **Instance Settings > Change Deployment Set**.
5. In the **Change Deployment Set** dialog box, select the destination deployment set and set the Force Change parameter.

   Valid values of Forced Change:

   - **Yes**: allows the instance to be migrated to another host, which may cause the instance to restart.
   - **No**: does not allow the instance to be migrated to another host. The instance must remain on the current host, which may cause a failure to change the deployment set of the instance.
6. Click **OK**.

14.5. Modify a deployment set

After you create a deployment set, you can modify the name and description of the deployment set based on your business needs.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Deployment & Elasticity > Deployment Sets**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the deployment set that you want to modify and click **Modify** in the **Actions** column.
5. Modify the name and description of the deployment set.
6. Click **OK**.

14.6. Delete a deployment set

You can delete deployment sets that are no longer needed.

Prerequisites

No instances exist in the deployment set.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Deployment & Elasticity > Deployment Sets**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Use one of the following methods to delete deployment sets:
   - To delete a single deployment set at a time, find the deployment set and click **Delete** in the **Actions** column.
   - To delete multiple deployment sets at a time, select the deployment sets and click **Delete** below the deployment set list.
5. In the message that appears, click **Delete**.

15.Monitoring & maintenance

15.1. Configure monitoring thresholds for organizations

You can specify a monitoring period and configure thresholds for a selection of metrics to monitor resource utilization in your organization. Then, you can improve resource utilization efficiency based on monitoring results and optimization suggestions.

Limits

Monitoring thresholds can be configured only for organizations that use an admin account.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Monitoring & Maintenance**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Organization** tab.
5. Select one of the following quota usage types and then click **Set Thresholds**:
   - **High Quota Usage**: indicates that the metric usage is greater than or equal to the threshold that you specify.
   - **Low Quota Usage**: indicates that the metric usage is less than or equal to the threshold that you specify.
   - In the dialog box that appears, specify the monitoring period and metric thresholds. The following table describes the monitoring period and metric thresholds that you configure.

Parameter	Description
Monitoring Period	Specify the monitoring period. Valid values: 7 Days, 15 Days, and 30 Days.

Parameter	Description
Average CPU Quota Usage	Specify a threshold for the average CPU quota usage. ■ Threshold for the high quota usage: The average CPU quota usage is greater than or equal to the threshold that you specify. The threshold ranges from 60% to 95%. ■ Threshold for the low quota usage: The average CPU quota usage is less than or equal to the threshold that you specify. The threshold ranges from 5% to 40%. **Note** The threshold can only be an integer multiple of 5%.
Average Memory Quota Usage	Specify a threshold for the average memory quota usage. ■ Threshold for the high quota usage: The average memory quota usage is greater than or equal to the threshold that you specify. The threshold ranges from 60% to 95%. ■ Threshold for the low quota usage: The average memory quota usage is less than or equal to the threshold that you specify. The threshold ranges from 5% to 40%. **Note** The threshold can only be an integer multiple of 5%.
Average Storage Quota Usage	Specify a threshold for the average storage quota usage. ■ Threshold for the high quota usage: The average storage quota usage is greater than or equal to the threshold that you specify. The threshold ranges from 60% to 95%. ■ Threshold for the low quota usage: The average storage quota usage is less than or equal to the threshold that you specify. The threshold ranges from 5% to 40%. **Note** The threshold can only be an integer multiple of 5%.

- Click **OK**.

Result

- The specified monitoring period and metric thresholds are displayed above the organization list.
- The organizations that match the specified monitoring period and metric thresholds are displayed in the **Monitoring & Maintenance** section of the **Overview** page in the console.

15.2. View the list of monitored organizations

You can view the organizations that meet the monitoring criteria, specified thresholds, and optimization suggestions within the monitoring period.

Procedure

1. Log on to the ECS console.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Use one of the following methods to go to the **Monitoring & Maintenance** page:
   - In the **Monitoring & Maintenance** section of the **Overview** page, click **Handle Now**.
   - In the left-side navigation pane, click **Monitoring & Maintenance**.
4. Click the **Organization** tab and click one of the following tabs as needed:
   - **High Quota Usage**: displays organizations whose quota usage reaches or exceeds the upper thresholds that you specify.
   - **Low Quota Usage**: displays organizations whose quota usage reaches or falls below the lower thresholds that you specify.

Result

- You can view the configured monitoring period and metric thresholds on the top of the organization list.
- You can view the information of organizations that meet monitoring criteria within the specified monitoring period in the organization list, such as the organization name, CPU quota usage, memory quota usage, storage quota usage, instances, and actions.

What's next

You can adjust the quota for an organization based on your business needs. For more information, see the *Quota management* chapter in *Apsara Uni-manager Management Console User Guide*.

15.3. Configure monitoring metric thresholds for instances

You can configure the monitoring period and multiple monitoring metric thresholds for instances to improve resource utilization based on the monitoring results and optimization suggestions.

Limits

You can configure monitoring metric thresholds only as as an operations administrator.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Monitoring & Maintenance**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Monitoring & Maintenance** page, click the **Instances** tab.
5. Click one of the following tabs and then click **Set Thresholds**:
   - **High Resource Usage**: displays instances whose resource usage reaches or exceeds the upper thresholds that you specify.
   - **Low Resource Usage**: displays instances whose resource usage reaches or falls below the lower thresholds that you specify.
   - In the dialog box that appears, configure the parameters described in the following table to specify the monitoring period and metric thresholds.

Parameter	Description
Monitoring Period	Specify the monitoring period. By default, the monitoring period is seven days.
Average CPU Utilization	Specify the threshold for the average CPU utilization. ■ High resource usage: The resource usage of an instance is considered high if its average CPU utilization is greater than or equal to the upper threshold that you specify. This threshold can range from 60% to 95%. ■ Low resource usage: The resource usage of an instance is considered low if its average CPU utilization is smaller than or equal to the lower threshold that you specify. This threshold can range from 5% to 40%. **Note** The threshold must be an integral multiple of 5%.

Parameter	Description
Average Memory Usage	Specify the threshold for the average memory usage. ■ High resource usage: The resource usage of an instance is considered high if its average memory usage is greater than or equal to the upper threshold that you specify. This threshold can range from 60% to 95%. ■ Low resource usage: The resource usage of an instance is considered low if its average memory usage is smaller than or equal to the lower threshold that you specify. This threshold can range from 5% to 40%. **Note** The threshold must be an integral multiple of 5%.

- Click **OK**.

Result

- The specified monitoring period and metric thresholds are displayed on the top of the instance list.
- In the **Monitoring & Maintenance** section of the **Overview** page in the ECS console, the instances that meet monitoring criteria within the specified monitoring period are displayed.

15.4. View the list of monitored instances

You can view the instances that meet the monitoring criteria, specified thresholds, and optimization suggestions within the monitoring period.

Procedure

1. Log on to the ECS console.
2. In the top navigation bar, select an organization, a resource set, and a region.
3. Use one of the following methods to go to the **Monitoring & Maintenance** page:
   - In the **Monitoring & Maintenance** section of the **Overview** page, click **Handle Now**.
   - In the left-side navigation pane, click **Monitoring & Maintenance**.
4. Click the **Instances** tab and click one of the following tabs as needed:
   - **High Resource Usage**: displays instances whose resource usage reaches or exceeds the upper thresholds that you specify.
   - **Low Resource Usage**: displays instances whose resource usage reaches or falls below the lower thresholds that you specify.

Result

- You can view the configured monitoring period and metric thresholds on the top of the instance list.
- You can view the information of instances that meet monitoring criteria within the specified monitoring period in the instance list, such as the instance ID and name and resource set.

15.5. Reclaim instance resources

You can optimize your business layout by reclaiming under-used instance resources based on the monitoring results of metrics that you set thresholds for.

Limits

- Only instance resources with low usage can be reclaimed.
- After an Elastic Compute Service (ECS) instance is moved to the recycle bin, the storage resources of the instance are retained, but its computing resources such as vCPUs and memory are released.
- After an ECS instance is moved to the recycle bin, the instance is retained here for a specified period of time. During the retention period, the instance can be recovered. After the retention period, the instance is automatically deleted.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Monitoring & Maintenance**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Instances** tab and click **Low Resource Usage**.
5. Find the instance that you want to reclaim, click **Recycle** in the **Actions** column.
6. In the message that appears, click **OK**.
   - After the instance is reclaimed, **Stopped - Pending Release** is displayed in the Status column of the instance list.
   - You can view the instance in the recycle bin.

Related operations

You can restore instances in the recycle bin based on business needs. For more information, see Restore an instance.

16.Cloud Assistant

16.1. Overview

Cloud Assistant is a native automated operations and maintenance (O&M) tool developed for Elastic Compute Service (ECS). It allows you to batch maintain ECS instances and batch run commands on and send files to ECS instances in a password-free, logon-free manner without the use of jump servers. These commands can consist of shell, PowerShell, or batch scripts. Typically, you can use Cloud Assistant to install and uninstall software, start and stop services, distribute configuration files, and run commonly used commands (scripts).

Features

After ECS instances that have the Cloud Assistant client installed enter the **Running** (`Running`) state, you can use Cloud Assistant to perform the following operations on the instances by using the ECS console or by calling API operations:

- Run batch and PowerShell scripts on Windows instances, or run shell, Python, and Perl scripts on Linux instances.
- Upload files to the instances.
- Run the same command on multiple instances. The execution state and results of one instance do not affect other instances.
- Configure custom parameters in Cloud Assistant commands to adapt to different scenarios.

> **Note** Cloud Assistant does not proactively initiate any operations. You have full control over all Cloud Assistant operations.

Use scenarios

Cloud Assistant can help you perform deployment and O&M tasks on ECS instances. The following list provides some examples:

- Uploading and running automated O&M scripts
- Running scripts that are already uploaded to instances
- Managing software lifecycle
- Deploying code or applications
- Polling processes
- Installing service-related patches or security updates
- Obtaining files from Object Storage Service (OSS) or YUM repositories to update applications on ECS instances
- Changing hostnames or user logon passwords

Limits

- Recurring Cloud Assistant tasks can be configured only by calling API operations. The interval at which a command is run cannot be less than 10 seconds.
- The size of a command cannot exceed 16 KB. This includes the total size of the Base64-encoded batch, PowerShell, or shell scripts together with the Base64-encoded custom parameters.
- The size of a file to be uploaded cannot exceed 32 KB after Base64 encoding.

- Each command can contain a maximum of 20 custom parameters.
- Cloud Assistant commands can be run only on instances that use the following operating systems:
  - Alibaba Cloud Linux
  - CentOS 7, CentOS 8, and later

    > **Note** CentOS 8 has reached its end of life (EOL) and is no longer maintained by the Linux community. We recommend that you migrate to Anolis or Alibaba Cloud Linux.

  - Debian 9
  - OpenSUSE
  - SUSE Linux Enterprise Server (SLES) 11, SLES 12, and SLES 15 and later
  - Ubuntu 14, Ubuntu 16, and Ubuntu 18 and later
  - Windows Server 2012, Windows Server 2016, and Windows Server 2019 and later

Terms

The following table describes relevant terms in Cloud Assistant.

Term	Description
Cloud Assistant	A tool provided by Alibaba Cloud that can help you perform routine maintenance tasks on multiple ECS instances and ECS bare metal instances. Cloud Assistant is available in all Alibaba Cloud regions.
Cloud Assistant client	A lightweight plug-in that can be installed on ECS instances to run Cloud Assistant commands. • On Windows instances, the process of the client program is named AliyunService. • On Linux instances, the process of the client program is named aliyun.service.
Cloud Assistant daemon process	A daemon process that is used to monitor the resource consumption of the Cloud Assistant client, report the running state of the client, and restart the client if the client fails. • Service name: `AssistDaemon` • Path: */usr/local/share/assist-daemon/assist_daemon* **Note** The Cloud Assistant daemon process is available only for Linux instances.
task execution path	A path in which Cloud Assistant saves your command as a file on an ECS instance and executes the file. The path varies based on the operating system. • Linux: */tmp* • Windows: *<Installation path of Cloud Assistant>/work/script*
command	A specific command such as a shell script or a PowerShell script that can be run on ECS instances.

Term	Description
custom parameter	A variable that is configured in the {{key}} format in a command. You can specify a custom parameter and its value in the {{"<key>":"<value>"}} format when you create a task to run the command. The number of Cloud Assistant commands that you can have within each Alibaba Cloud region is limited. To adapt Cloud Assistant commands to multiple scenarios, we recommend that you configure custom parameters.
one-time execution	An execution (`Invocation`) of a one-time task. A one-time task runs a command only once on one or more instances.
recurring execution	An execution of a recurring task. A recurring task runs a command on one or more instances based on your specified schedule.
execution status	The relationships among different types of execution states. For more information, see Execution states.

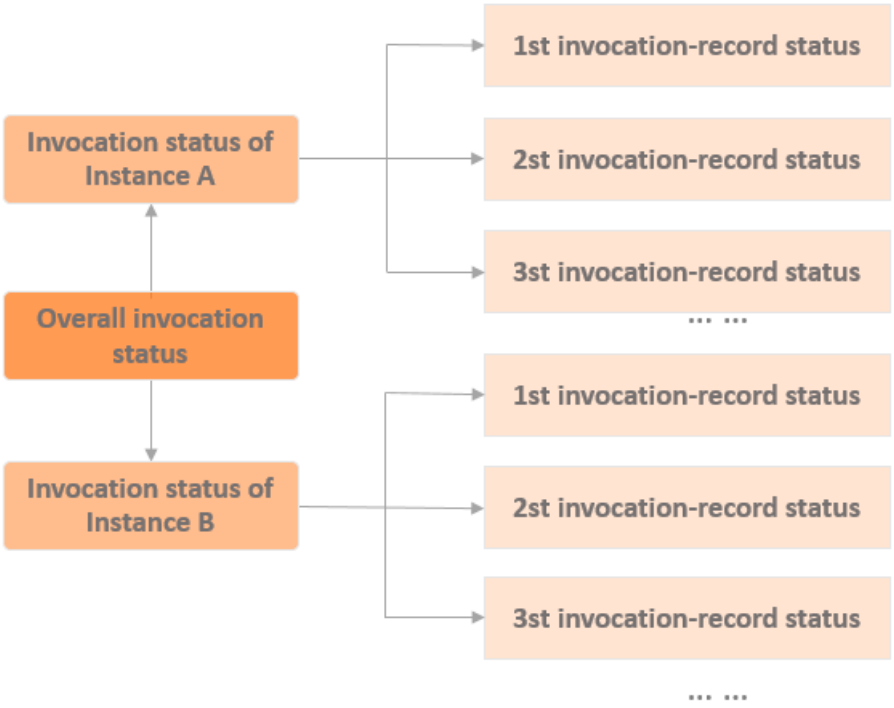
Execution states

The following table describes the instance-level execution state of a command that is run on a single instance. The InvocationStatus parameter in API indicates the execution state of a command.

Execution state in API operations	Execution state in the ECS console	Description
`Running`	Running	The command is being run.
`Stopping`	Stopping	The command is being stopped.
`Stopped`	Stopped	The command is stopped.
`Finished`	Completed	The command is run to completion. This does not indicate that the command is successful. You can check whether the command is successful based on the command output (`Output`) and the exit code (`ExitCode`).
`Failed`	Failed	The command cannot be run or the command process did not complete before the timeout period specified by `Timeout` expires.

States of batch executions and scheduled executions

A batch execution is a one-time execution in which a command is run on multiple instances. To better manage batch executions and recurring executions, you can manage the lifecycles of the executions based on the overall execution states, instance-level execution states, and execution-level states. The InvokeStatus parameter in API operations indicates the execution status of a command. The following figure shows the relationships among the three types of execution states.

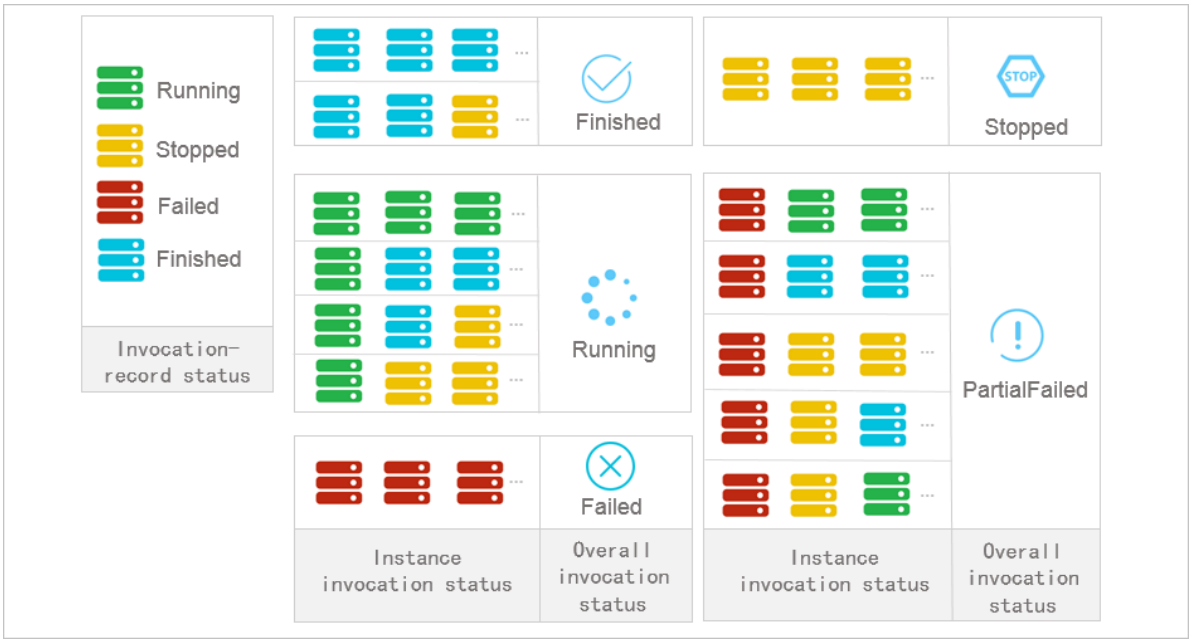


- The following table describes the overall execution states of a command that is run on multiple instances at the same time.

Execution state in API operations	Execution state in the ECS console	Description	Priority
`Running`	Running	The instance-level execution state is Running on some or all instances.	1
`Stopping`	Stopping	The instance-level execution state is Stopping on some or all instances.	2
`Stopped`	Stopped	The instance-level execution state is Stopped on all instances.	3
`Failed`	Failed	The instance-level execution state is Failed on all instances, or is Failed on some instances and is Stopped on the other instances.	4
`Finished`	Completed	The instance-level execution state is Finished on all instances, or is Finished on some instances and is Stopped on the other instances.	5

Execution state in API operations	Execution state in the ECS console	Description	Priority
`PartialFailed`	Partially Failed	The instance-level execution state is Failed on some instances and is Finished on the other instances.	6

The following figure shows the relationship between the overall execution states and the instance-level execution states of a one-time execution in which a command is run on three instances at the same time.



- The following table describes the states of recurring executions of a command.

State	Description
Overall execution state	The overall execution state is always **Running** (`Running`) unless you stop the command on all the instances.
Instance-level execution state	For each instance, the instance-level execution state is always **Running** (`Running`) unless you stop the command on the instance.
Execution-level state	For more information, see Execution states.

16.2. Configure the Cloud Assistant client

16.2.1. Install the Cloud Assistant client

The Cloud Assistant client is used to run Cloud Assistant commands on Elastic Compute Service (ECS) instances. This topic describes how to install the Cloud Assistant client.

Prerequisites

- An administrator account is used to install and use the Cloud Assistant client. The administrator username is root for Linux instances, and system for Windows instances.
- Before you install the Cloud Assistant client, make sure that your instance type and operating system support Cloud Assistant. For more information, see Limits.

Context

If ECS instances are created from public images in Apsara Stack V3.16.1, the Cloud Assistant client is pre-installed on the instances.

If ECS instances are created in Apsara Stack V3.13.0 to V3.16.0, the Cloud Assistant client is not pre-installed on the instances. You can call API operations to use Cloud Assistant on the instances. If you want to use features related to Cloud Assistant, install the Cloud Assistant client.

The following table describes how to install the Cloud Assistant client on different operating systems.

Operating system	Installation method
Windows	Install the Cloud Assistant client on Windows instances
Linux operating systems such as Alibaba Cloud Linux, CentOS, openSUSE, and SUSE Linux	• Install the client on Linux instances by using the RPM package • Install the client on Linux instances by using source code
Linux operating systems such as Debian and Ubuntu	• Install the client on Linux instances by using the Debian package • Install the client on Linux instances by using source code

Install the Cloud Assistant client on Windows instances

1. Connect to an ECS instance as the administrator.

   For more information, see Instance connecting overview.

2. Download the client installation file.

   You can download the installation file for a specific version of the Cloud Assistant client from one of the following URLs:

   - Public URL for the latest version: latest version of the Cloud Assistant client
   - Public URL for a specified version:

     ```
     https://aliyun-client-assist.oss-accelerate.aliyuncs.com/windows/aliyun_agent_{version}_setup.exe
     ```

     Note *{version}* indicates the version number of the Cloud Assistant client.

3. Double-click the installation file and install the client as instructed.
   The default installation path is *C:\ProgramData\aliyun\assist* for Windows instances.

Install the client on Linux instances by using the RPM package

This method is applicable to operating systems such as Alibaba Cloud Linux, CentOS, openSUSE, and SUSE Linux.

1. Connect to an ECS instance as the administrator.

   For more information, see Instance connecting overview.

2. Download the RPM package of the Cloud Assistant client.

   ```
   wget "https://aliyun-client-assist.oss-accelerate.aliyuncs.com/linux/aliyun_assist_latest.rpm"
   ```

   > **Note** *{version}* indicates the version number of the Cloud Assistant client.

3. Install the Cloud Assistant client.

   In this example, the latest version of the Cloud Assistant client is installed.

   ```
   rpm -ivh --force aliyun_assist_latest.rpm
   ```

   > **Note** The default installation path is /usr/local/share/aliyun-assist/ for Linux instances.

Install the client on Linux instances by using the Debian package

This method is applicable to operating systems such as Debian and Ubuntu.

1. Connect to an ECS instance as the administrator.

   For more information, see Instance connecting overview.

2. Download the Debian package for a specific version of the Cloud Assistant client from one of the following URLs:

   - Public URL for the latest version:

     ```
     wget "https://aliyun-client-assist.oss-accelerate.aliyuncs.com/linux/aliyun_assist_latest.deb"
     ```

   - Public URL for a specified version:

     ```
     wget "https://aliyun-client-assist.oss-accelerate.aliyuncs.com/linux/aliyun_assist_{version}.deb"
     ```

   > **Note** *{version}* indicates the version number of the Cloud Assistant client.

3. If an earlier version of the Cloud Assistant client is installed on the instance, uninstall the earlier version.

   ```
   dpkg -r aliyun-assist
   ```

4. Install the Cloud Assistant client.

   In this example, the latest version of the Cloud Assistant client is installed.

   ```
   dpkg -i aliyun_assist_latest.deb
   ```

   > **Note** The default installation path is /usr/local/share/aliyun-assist/ for Linux instances.

Install the client on Linux instances by using source code

1. Connect to an ECS instance as the administrator.

   For more information, see Instance connecting overview.

2. Install necessary software such as Git and Go.

   In this example, YUM is used to install Git and Go. If you use other versions of Linux, use the corresponding package manager.

   - Install Git.

     ```
     yum install git -y
     ```

   - Install Go.

     ```
     yum install go -y
     ```

3. Download the source code of the Cloud Assistant client.

   ```
   git clone https://github.com/aliyun/aliyun_assist_client
   ```

4. Access the source code directory.

   ```
   cd ./aliyun_assist_client
   ```

5. Compile the source code.

   ```
   go build
   ```

   If no error message is returned, the client is installed.

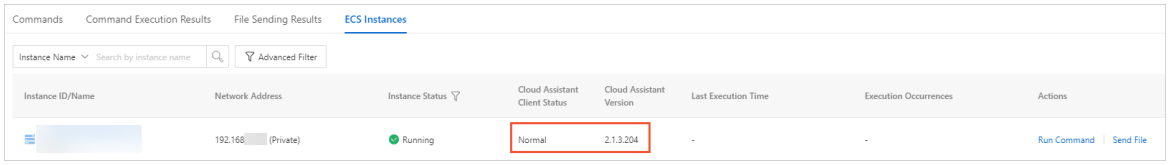
6. Run the Cloud Assistant client.

   ```
   aliyun-service -d
   ```

View information of the Cloud Assistant client on an ECS instance

After the Cloud Assistant client is installed on an instance, you can perform the following steps to query the version number and state of the client on the instance.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Maintenance & Monitoring > Cloud Assistant**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **ECS Instances** tab to view the information about the Cloud Assistant client on the ECS instances within the current region.

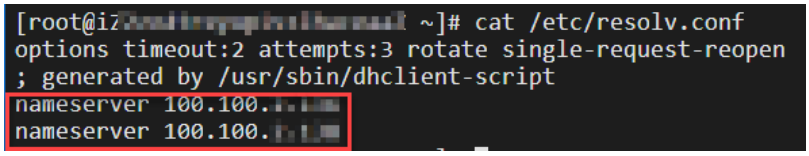


16.2.2. Configure DNS resolution for Cloud Assistant

This topic describes how to configure Domain Name System (DNS) resolution for Cloud Assistant. During the configuration procedure, you must obtain the IP addresses that correspond to the Cloud Assistant endpoints and then modify the hosts file.

Context

When you use features such as Cloud Assistant on an Elastic Compute Service (ECS) instance, the instance must have access to the endpoints required to perform actions that you specify, such as running a Cloud Assistant command. The default DNS Nameserver is installed on each ECS instance to resolve domain names. You can run the **cat /etc/resolv.conf** command to view the DNS Nameserver settings. Example command output:



If you modify the configuration file to override the default DNS Nameserver settings, domain names may fail to resolve or resolve slowly when you use features such as Cloud Assistant. This may cause the features to be unavailable. For example, Cloud Assistant commands cannot be run. In this case, you can perform the following procedure to configure DNS resolution for Cloud Assistant.

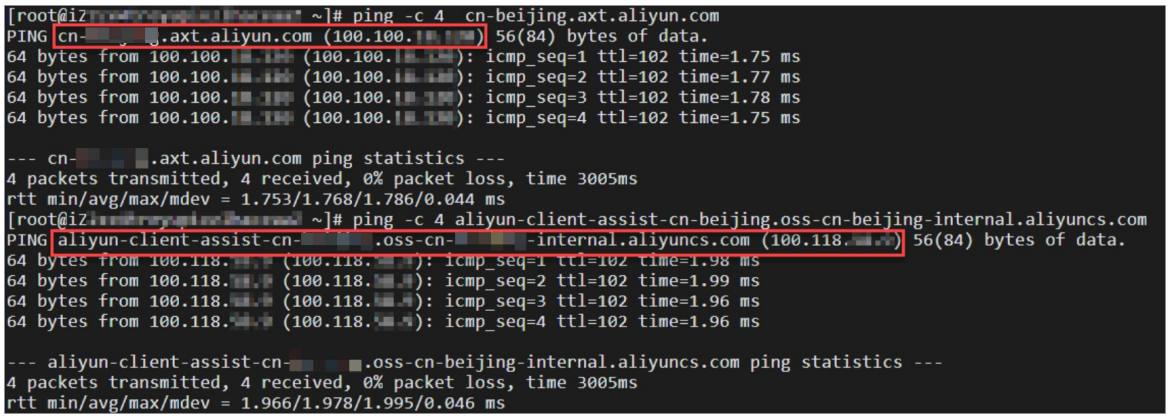
Procedure

1. Connect to an ECS instance. For more information, see Instance connecting overview.
2. Obtain the IP addresses that correspond to the Cloud Assistant endpoints.

   The following Cloud Assistant endpoints are available:

   - Endpoint used to run Cloud Assistant commands, in the format of `<region-id>.axt.aliyun.com`
   - Endpoint used to obtain the Cloud Assistant plug-in and update packages, in the format of `aliyun-client-assist-<region-id>.oss-<region-id>-internal.aliyuncs.com`

   > **Note** Replace *<region-id>* with a region ID.



3. Modify the hosts file.

```
echo "100.100.XX.XX <region-id>.axt.aliyun.com" >> /etc/hosts && \
echo "100.118.XX.XX aliyun-client-assist-<region-id>.oss-<region-id>-internal.aliyuncs.
com" >> /etc/hosts
```

> **Note** Replace `100.100.XX.XX` and `100.118.XX.XX` with the IP addresses that you obtained in the previous step.

4. Check whether the hosts file is modified.

```
cat /etc/hosts
```

If the hosts file is modified, the command output includes the Cloud Assistant endpoints and their corresponding IP addresses. Example command output:

```
[root@iZ                etc]# cat /etc/hosts
::1     localhost       localhost.localdomain   localhost6      localhost6.localdomain6
127.0.  localhost  localhost

192.168.  iZ2                    iZ2
100.100.    cn-      .axt.aliyun.com
100.118.    aliyun-client-assist-cn-      .oss-cn-      -internal.aliyuncs.com
```

After the hosts file is modified, the ECS instance can automatically obtain IP addresses from the hosts file to resolve the Cloud Assistant endpoints.

16.2.3. Start or stop the Cloud Assistant client

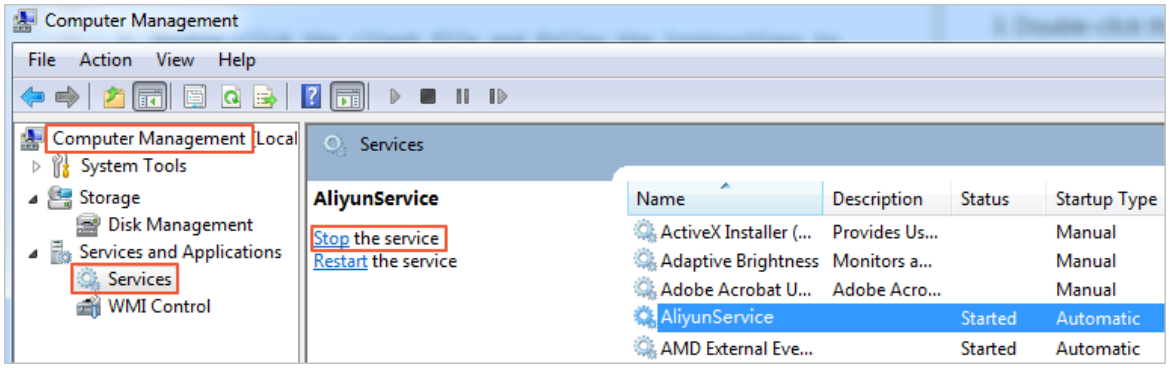
The Cloud Assistant client is an agent that runs Cloud Assistant commands on Elastic Compute Service (ECS) instances. This topic describes how to start or stop the Cloud Assistant client.

Start or stop the Cloud Assistant client on a Windows instance

To start or stop the Cloud Assistant client on a Windows instance, perform the following steps.

> **Warning** **Aliyun Assist Service** is the process of the Cloud Assistant client. If you stop **Aliyun Assist Service**, the Cloud Assistant client stops. This may cause an exception on the instance, and the instance cannot be stopped by using the ECS console. Proceed with caution when you stop the Cloud Assistant client.

1. Connect to the Windows instance. For more information, see Connect to a Windows instance by using RDC.
2. Click Start and choose **Windows Administrative Tools > Computer Management**.
3. Choose **Computer Management (Local) > Services and Applications > Services**.
4. Find **Aliyun Assist Service** and click **Stop the service** or **Restart the service**.



Uninstall the Cloud Assistant daemon process from a Linux instance

The Cloud Assistant daemon process is used to monitor the resource consumption of the Cloud Assistant client, report the running status of the client, and restart the client when the client fails. Before you stop the Cloud Assistant client, you must first uninstall the Cloud Assistant daemon process.

> **Note** The Cloud Assistant daemon process is available only for Linux instances.

1. Connect to the Linux instance. For more information, see Connect to a Linux instance by using SSH commands in Linux or Mac OS X or Connect to a Linux-based instance by using remote connection tools in Windows.
2. Stop the Cloud Assistant daemon process.

   ```
   /usr/local/share/assist-daemon/assist_daemon --stop
   ```

   > **Note** In the preceding command, */usr/local/share/assist-daemon/assist_daemon* specifies the default path of the Cloud Assistant daemon process.

3. Uninstall the Cloud Assistant daemon process.

   ```
   /usr/local/share/assist-daemon/assist_daemon --delete
   ```

4. Delete the directory of the Cloud Assistant daemon process.

   ```
   rm -rf /usr/local/share/assist-daemon
   ```

Start or stop the Cloud Assistant client on a Linux instance

> **Note** Before you stop the Cloud Assistant client, you must first uninstall the Cloud Assistant daemon process. For more information, see the Uninstall the Cloud Assistant daemon process from a Linux instance section.

To start or stop the Cloud Assistant client on a Linux instance, perform the following steps:

1. Connect to the Linux instance. For more information, see Connect to a Linux instance by using SSH commands in Linux or Mac OS X or Connect to a Linux-based instance by using remote connection tools in Windows.
2. Run the following commands based on the initialization process of the Linux instance.
   - Linux operating systems that are based on new versions of the Linux kernel typically use the **systemd** initialization process. Perform the following steps:
     - Check whether the instance uses the systemd initialization process. If the instance uses systemd, a command output is displayed.

       ```
       strings /sbin/init | grep "/lib/system"
       ```

     - Stop the Cloud Assistant client.

       ```
       systemctl stop aliyun.service
       ```

     - Start the Cloud Assistant client.

       ```
       systemctl start aliyun.service
       ```

- Restart the Cloud Assistant client.

```
systemctl restart aliyun.service
```

- Ubuntu 14 and earlier operating systems typically use the **UpStart** initialization process. Perform the following steps:
  - Check whether the instance uses the UpStart initialization process. If the instance uses UpStart, a command output is displayed.

```
strings /sbin/init | grep "upstart"
```

  - Stop the Cloud Assistant client.

```
/sbin/initctl stop aliyun-service
```

  - Start the Cloud Assistant client.

```
/sbin/initctl start aliyun-service
```

  - Restart the Cloud Assistant client.

```
/sbin/initctl restart aliyun-service
```

- Linux operating systems that are based on earlier versions of the Linux kernel typically use the **sysvinit** initialization process. Perform the following steps:
  - Check whether the instance uses the sysvinit initialization process. If the instance uses sysvinit, a command output is displayed.

```
strings /sbin/init | grep "sysvinit"
```

  - Stop the Cloud Assistant client.

```
/etc/init.d/aliyun-service stop
```

  - Start the Cloud Assistant client.

```
/etc/init.d/aliyun-service start
```

  - Restart the Cloud Assistant client.

```
/etc/init.d/aliyun-service restart
```

16.3. Use Cloud Assistant

16.3.1. Create a command

You can use Cloud Assistant commands to perform routine tasks on Elastic Compute Service (ECS) instances. These tasks include running automated O&M scripts, polling processes, resetting user passwords, installing or uninstalling software, updating applications, and installing patches. The commands can be batch or PowerShell commands for Windows instances, and shell commands for Linux instances. You can specify custom parameters as variables in Cloud Assistant commands.

Prerequisites

- The instances on which to run a command are in the **Running** (Running) state.
- You can retain up to 100 Cloud Assistant commands within an Alibaba Cloud region. This quota may increase based on your ECS usage. If you click Run when you create a command in the Create Command page, the command does not count against your command quota.

  > **Note** You can also call the DescribeAccountAttributes operation with AttributeName.N set to *max-axt-command-count* to query the maximum number of Cloud Assistant commands that you can retain within a region.

- You can run Cloud Assistant commands up to 5,000 times per day within each region. This quota may increase based on your ECS usage.

  > **Note** You can also call the DescribeAccountAttributes operation with the AttributeName.N parameter set to *max-axt-invocation-daily* to query the maximum number of times that you can run Cloud Assistant commands per day within each region.

Context

When you use the immediate execution feature, take note of the following items:

- Only operations administrators, organization administrators, and resource set administrators can create commands.
- The size of the command after Base64 encoding cannot exceed 16 KB.
- Up to 20 custom parameters can be specified in a single Cloud Assistant command.
- You can call an API operation to run a command on up to 50 instances each time.
- When you create a command, make sure that the syntax, logic, or algorithm associated with the command are correct.

  For example, assume that you have created the */backup* directory (`mkdir /backup`) on an instance. You can run the following shell commands to archive a file in this directory:

  ```
  #!/bin/bash
  OF=/backup/my-backup-$(date +%Y%m%d).tgz
  tar -cf $OF {{file}}
  ```

  > **Note** In the preceding example, `{{file}}` is a custom parameter. When you run the commands, you can set this custom parameter to the name of the file to be archived. Example: */app/usrcredential*. Custom parameters can be used in scenarios that require dynamic values and values that are shared across multiple commands. We recommend that you specify custom parameters for sensitive data or data that changes with the environment, such as AccessKey pairs, instance IDs, authorization codes, time parameters, and critical system files.

Procedure in the ECS console

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Maintenance & Monitoring > Cloud Assistant**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the ECS Cloud Assistant, click **Create /Run Command**.

5. On the **Create Command** page, configure the parameters described in the following table.

Section	Parameter	Description
Resource Owner	Organization	Select the organization to which the instances belong.
	Resource Set	Select the resource set to which the instances belong.
Command Information	Command Source	Select the command source. ○ **Enter Command Content**: creates a command. ○ **Select Saved Command**: selects an existing command.
	Command Name	Enter a name for the command.
	Execution Plan	Select a plan on how to run the command. ○ **Run Now**: The command is immediately run after you click **Run** or **Run and Save**. ○ **Run on Next System Startup**: The command is run the next time the selected instances are started after you click **Run** or **Run and Save**. ○ **Run on Each System Startup**: The command is run each time the selected instances are started after you click **Run** or **Run and Save**.
	Command Type	Select a command type. ○ For Linux instances, select **Shell**. ○ For Windows instances, select **Bat** or **PowerShell**.
	Command Content	Enter or paste the content of the command.
	Use Parameters	Specifies whether to use parameters. If you turn on **Use Parameters**, specify custom parameters in the `{{key}}` format in the **Command Content** field.
	Command Parameters	Specify the values of the custom parameters specified in the `{{key}}` format in the **Command Content** field. This parameter is available only when **Use Parameters** is turned on.
	Command Description	Enter a description for the command. We recommend that you set a description with information such as the command purpose that makes the command easy to identify, manage, and maintain.

Section	Parameter	Description
	User	Specify the username that is used to run the command on ECS instances. By default, Cloud Assistant commands are run by the root user on Linux instances and by the system user on Windows instances.
	Execution Path	Specify an execution path for the command. Different default execution paths are provided based on the operating system of instances on which the command is run. ○ For Linux instances, the default execution path is the */home* directory of the root user. ○ For Windows instances, the default execution path is the directory where the process of the Cloud Assistant client is located. Example: *C:\ProgramData\aliyun\assist\$(version)*.
	Timeout Period	Specify a **timeout period** for the command to run on instances. If a task that runs the command times out, Cloud Assistant forcefully terminates the task process. Unit: seconds. Default value: 60. Minimum value: 10. If you set **Timeout Period** to a value of less than 10, the system changes the value to 10 to ensure that the command can be run.
Instances	Instance ID/Name	Select the instances on which you want to run the command.

6. Click **Save**, **Run and Save**, or **Run**.

Procedure by using the CLI

- Sample request:

  Call the RunCommand operation to create and run a Cloud Assistant command named update to update the operating system on instances.

```
aliyun ecs RunCommand --RegionId 'cn-qingdao-****-d01' \
--Name 'update' --Username 'root' --Type 'RunShellScript' \
--CommandContent 'eXVtIC15IHVwZGF0ZQ==' \
--Timeout '60' --RepeatMode 'Once' --ContentEncoding 'Base64' \
--InstanceId.1 'i-bp12e0ib2ztibede****'
```

> **Note** Values enclosed in single quotation marks ('') are example values of the parameters. Configure the parameters based on actual conditions.

Parameter	Example	Description
RegionId	cn-qingdao-****-d01	The ID of region in which to create the command.

Parameter	Example	Description
Name	update	The name of the command.
Username	root	The username used to run the command on ECS instances.
Type	RunShellScript	The type of the command. ○ For Linux instances, set the value to RunShellScript. ○ For Windows instances, set the value to RunBatScript or RunPowershellScript.
CommandContent	eXVtIHVwZGF0ZSAteQ==	The Base64-encoded content of the command.
Timeout	60	The timeout period.
RepeatMode	Once	The execution plan.
ContentEncoding	Base64	The encoding format.
InstanceId.1	i-bp12e0ib2ztibede****	The ID of ECS instance N on which you want to run the command. In this example, the value of N is 1.

For more information, see the *CreateCommand* topic in the *ECS Developer Guide*.

- Sample success response:

```
{
        "CommandId": "c-hz018qlm868****",
        "InvokeId": "t-hz018qlm86d****",
        "RequestId": "1D24FA80-64DB-4842-AB20-2520799****"
}
```

16.3.2. Run a command

After you create a Cloud Assistant command, you can run it on one or more Elastic Compute Service (ECS) instances. The execution status and results of the command on multiple instances do not affect each other.

Prerequisites

Before you run a Cloud Assistant command on ECS instances, make sure that the instances meet the following requirements:

- The instances are in the **Running** (`Running`) state.
- The Cloud Assistant client is installed on the instances. For more information, see Install the Cloud Assistant client.

> **Notice** Only operations administrators, organization administrators, and resource set administrators can run Cloud Assistant commands.

Context

- You can call an API operation to run a command on up to 50 instances each time.
- If you select more than 50 instances to run a command in the ECS console, the system runs the command on the instances in batches.
- You can run Cloud Assistant commands up to 5,000 times per day within each region. This quota may increase based on your ECS usage.

> **Note** You can also call the DescribeAccountAttributes operation with the AttributeName.N parameter set to *max-axt-invocation-daily* to query the maximum number of times that you can run Cloud Assistant commands per day within each region.

Run Cloud Assistant commands by using the ECS console

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Maintenance & Monitoring > Cloud Assistant**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the command that you want to run and click **Run** in the **Actions** column.
5. In the **Run Command** page, configure parameters.
   i. In the **Command Information** section, check command content and configure the parameters described in the following table.

Parameter	Description
	Select a plan on how to run the command. ■ **Immediate execution**: The command is run immediately after you click **Run** or **Execute and Save**. ■ **After the next startup of the system**: The command is run the next time the selected instances are started after you click **Run** or **Execute and Save**. ■ **After each system startup**: The command is run each time the selected instances are started after you click **Run** or **Execute and Save**. ■ **Run on Schedule**: The command is run at a specified interval, at a specified time, or on a schedule after you click **Run** or **Execute and Save**. If you set Implementation Plan to Run on Schedule, the following options are available:

Parameter	Description
Execution Plan	■ **Run at Fixed Interval**: Use a rate expression to specify an interval at which to run the command. You can specify the interval in seconds, minutes, hours, or days. This option is applicable when tasks need to be executed at a fixed interval. **Note** When you set an interval, take note of the following limits: ■ The specified interval can be anywhere from 60 seconds to 7 days and must be longer than the timeout period of the scheduled task. ■ The interval is the duration between two consecutive executions. The interval is irrelevant to the amount of time required to run the command once. For example, if a command takes 2 minutes to complete and you set the interval to 5 minutes, the command is run again 3 minutes after the previous command completes. ■ Tasks are not run immediately after they are created. For example, assume that you set the interval to 5 minutes for a task. The task begins to be executed 5 minutes after it is created. ■ **Run Only Once at Specified Time**: Specify a point in time and a time zone to run the command only once. For example, if you set **Execution Time** to **2022-05-17 17:30:50** and **Time Zone** to **(GMT+8:00) Asia/Shanghai**, the command was run only once at 17:30:50 on May 17, 2022 (UTC+8). ■ **Run on Clock-based Schedule**: Use a cron expression to specify a schedule on which to run the command. Specify the schedule with second, minute, hour, day of the month, and month, and select a time zone from the Time Zone drop-down list. The system calculates the reccurence schedule based on the cron expression and time zone and runs the command as scheduled. This option provides flexibility and is applicable when tasks need to run on a schedule. For more information about cron expressions, see Cron expression. **Note** The minimum interval must be 10 seconds or more and cannot be shorter than the timeout period of scheduled executions. For example, if you set **Execution Frequency** to **0 0 12 ? * WED 2022** and set **Time Zone** to **(GMT+8:00) Asia/Shanghai**, the system runs the command at 12:00 every Wednesday in 2022 (UTC+8).
Command Content	Confirm the command content in the **Command Content** field.

Parameter	Description
Command Parameters	In the **Command Parameters** field, enter values for the custom parameters that are specified in the command. No format limits apply to the data types of values for the custom parameters. If the current task does not require values for these fields, you can leave the fields empty. **Note** If you did not select **Use Parameters** when you create the command, the **Command Parameters** fields are not displayed on the Run Command page.
User	Specify the username that is used to run the command on ECS instances. By default, Cloud Assistant commands are run by the root user on Linux instances and by the system user on Windows instances.

ii. In the **Instances** section, select one or more instances on which you want to run the command.

6. Click **Run**.

Run Cloud Assistant command by using Alibaba Cloud CLI

1. (Optional)Check the state of the instances on which you want to run a command. If the instances are not in the **Running** (`Running`) state, call the StartInstance operation to start the instances.

```
aliyun ecs StartInstance --InstanceId 'i-bp1f4f6o8lv0wqof****'
```

Note Values enclosed in single quotation marks ('') are example values of the parameters. You must configure the parameters based on actual conditions.

For more information, see or the "StartInstance" topic in *ECS Developer Guide*.

2. (Optional)Call the DescribeCloudAssistantStatus operation to check whether the Cloud Assistant client is installed on the instances.

```
aliyun ecs DescribeCloudAssistantStatus --RegionId 'cn-qingdao-****-d01' \
--InstanceId.1 'i-bp1f4f6o8lv0wqof****'
```

If the value of `CloudAssistantStatus` is true in the response, the Cloud Assistant client is installed on the instances.

3. Call the InvokeCommand operation to run a created Cloud Assistant command on the instances and obtain the value of InvokeId in the response.

```
aliyun ecs InvokeCommand --RegionId 'cn-qingdao-****-d01' \
--InstanceId.1 'i-bp1f4f6o8lv0wqof****' \
--InstanceId.2 'i-bp137qu6142s3mhm****' \
--CommandId 'c-hz018qp243j****' \
--Timed 'false' \
--output cols=InvokeId
```

Parameter	Example	Description
RegionId	cn-qingdao-****-d01	The ID of the region in which to run the command.
InstanceId.1	i-bp1f4f6o8lv0wqof****	The ID of the first instance on which to run the command.
InstanceId.2	i-bp137qu6142s3mhm****	The ID of the second instance on which to run the command.
CommandId	c-hz018qp243j****	The ID of the command.
Timed	false	Specifies whether to periodically run the command. If you want to periodically run a command, set **Timed** to true. The **Frequency** parameter specifies the execution cycle. For example, you can set Frequency to *0 */20 * * * *to run the command every 20 minutes. For more information, see Cron expression.

For more information, see the *InvokeCommand* topic in the *ECS Developer Guide*.

16.3.3. Upload files to ECS instances

This topic describes how to use the Cloud Assistant client to upload files such as configuration files and scripts to Elastic Compute Service (ECS) instances.

Prerequisites

- The ECS instances to which you want to upload a file are in the Running (Running) state.
- The Cloud Assistant client is installed on the instance. For more information, see Install the Cloud Assistant client.
- You can call an API operation to send a file to up to 50 instances at a time.
- The file that you want to upload cannot exceed 32 KB in size after it is encoded in Base64.

Context

You can use the Cloud Assistant client to upload files that cannot exceed 32 KB in size. If you want to upload files that are larger than 32 KB in size or if you want to download files from ECS instances, we recommend that you use the FileZilla tool over the SSH File Transfer Protocol (SFTP) and port 22.

> **Notice** Only operations administrators, organization administrators, and resource set administrators can upload files to ECS instances.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Maintenance & Monitoring > Cloud Assistant**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the ECS Cloud Assistant page, click **Send File**.
5. In the **Command Information** section, configure the parameters described in the following table.

Parameter	Description
Destination Operating System	Select the operating system of the ECS instances. Valid values: ○ Linux Operation System ○ Windows Operating System
File Content	Select a method to use to upload the file. Valid values: ○ Upload File: You can click Upload File to select a file or drag a file to the Upload File section. ○ Paste File Content: You can paste the file content to the field. **Note** The file that you want to upload cannot exceed 32 KB in size after it is encoded in Base64.
File Name	Specify a name for the file. **Note** If you turn off Overwrite, make sure that the file name is unique across the destination path of the ECS instances.
Destination Path	Specify the destination path to save the file. ○ Default value when Destination Operating System is set to Linux: /root ○ Default value when Destination Operating System is set to Windows: C:/Users/Administrator/Documents
File Description	Specify a description for the file.
User	Specify the user to which the file belongs. This parameter is required only for Linux instances.

Parameter	Description
User Group	Specify the user group to which the file belongs. This parameter is required only for Linux instances.
Permissions	Configure permissions on the file. Default value: `0644`. This value indicates that the file owner has read and write permissions on the file, and that other users in the same user group as the file owner and public users have read permissions on the file. This parameter is required only for Linux instances.
Overwrite	Specify whether to overwrite the file that has the same name as the uploaded file in the destination path.
Timeout Period	Set the timeout period for the file sending task. When the file sending task times out, Cloud Assistant forcibly stops the task process. Unit: seconds. Valid values: 10 to 86400. Default value: 60.

6. In the **Instances** section, select the instances to which to send the file and configure the parameters described in the following table.

Parameter	Required	Description
Organization	Yes	Select the organization to which the file belongs.
Resource Set	Yes	Select the resource set to which the file belongs.
Instance	Yes	Select the instances to which you want to send the file.

7. Click **Run**.

View the execution results of the file sending task

View the execution result of this file sending task.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Maintenance & Monitoring > Cloud Assistant**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the Cloud Assistant page, click the File Sending Results tab.
5. In the task list, view the execution states, execution IDs, and destination paths of file sending tasks.

   You can perform the following operations in the Actions column corresponding to a file sending task:

   - Click **View** to view the execution results of the task on each instance.
   - Click **Export** to export the task execution results.

- Click **Resend** to execute the task again.

16.3.4. Clone a command

You can clone an existing Cloud Assistant command to create another command. You can retain all the information of the original command (cloned command), or you can modify information such as the name, description, type, content, execution path, or timeout period in the new command (command clone).

Procedure

Notice Only operations administrators, organization administrators, and resource set administrators can clone commands.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Maintenance & Monitoring > Cloud Assistant**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the Cloud Assistant command that you want to clone and click **Clone** in the **Actions** column.
5. On the **Clone Command** page, configure the parameters described in the following table.

Section	Parameter	Description
Resource Owner	Organization	Select the organization to which the instances belong.
	Resource Set	Select the resource set to which the instances belong.
Command	Command Name	Specify a name for the new command.
	Execution Plan	Select a plan on how to run the new command. ○ **Run Now**: The command is immediately run after you click **Run** or **Run and Save**. ○ **Run on Next System Startup**: The command is run the next time the selected instances are started after you click **Run** or **Run and Save**. ○ **Run on Each System Startup**: The command is run each time the selected instances are started after you click **Run** or **Run and Save**.
	Command type	Select a command type. ○ For Linux instances, select **Shell**. ○ For Windows instances, select **Bat** or **PowerShell**.
	Command Content	Enter or paste the content of the command.

Information Section	Parameter	Description
	Command Description	Specify a description for the new command. We recommend that you set a description with information such as the command purpose that makes the command easy to identify, manage, and maintain.
	Execution Path	Specify an execution path for the new command. Different default execution paths are provided based on the operating system of instances on which the command is run. ◦ For Linux instances, the default execution path is the */home* directory of the root user. ◦ For Windows instances, the default execution path is the directory where the process of the Cloud Assistant client is located. Example: *C:\ProgramData\aliyun\assist\$(version)*.
	Timeout Period	Specify a **timeout period** for the new command to run on instances. If a task that runs the command times out, Cloud Assistant forcefully terminates the task process. Unit: seconds. Default value: 60. Minimum value: 10. If you set **Timeout Period** to a value of less than 10, the system changes the value to 10 to ensure that the command can be run.

6. Click **Save**.

16.3.5. Delete a command

The number of Cloud Assistant commands that you can have within each Alibaba Cloud region is limited. To ensure a sufficient command quota, we recommend that you regularly delete commands that are no longer needed.

Procedure

Notice Only operations administrators, organization administrators, and resource set administrators can delete commands.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Maintenance & Monitoring > Cloud Assistant**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the Cloud Assistant command that you want to delete and click **Delete** in the **Actions** column.
5. In the **message** that appears, click **OK**.

16.4. Cron expression

When you run a Cloud Assistant command, you can call an API operation and use the Timed and Frequency parameters to set when to run the Cloud Assistant command. The value of the Frequency parameter is a cron expression. This parameter specifies the frequency of scheduled tasks, frequency of routine maintenance, and the point in time at which to complete a one-time task.

Introduction

A cron expression is a string that represents time. The string consists of five spaces and six fields, which is in the `X X X X X X X` format. `X` is a placeholder of a field. If a field contains multiple values, the values are separated with commas (,). Each field can be a specific value or special characters that have logical representations.

Field values

The following table describes valid values and supported special characters for each field in cron expressions.

Field	Required	Valid value range	Special character
Second	Yes	[0, 59]	* , - /
Minute	Yes	[0, 59]	* , - /
Hour	Yes	[0, 23]	* , - /
Day	Yes	[1, 31]	* , - / ? L W
Month	Yes	[1, 12] or [JAN, DEC]	* , - /
Week	Yes	[1, 7] or [MON, SUN]. If you use the [1, 7] format, `1` indicates Monday and `7` indicates Sunday.	* , - / ? L #

Special characters

Each field in a cron expression can contain a specific number of special characters. Each special character represents a logical argument.

Special character	Description	Example
`*`	Indicates all valid values.	In the Month field, an asterisk (`*`) indicates every month. In the Week field, an asterisk (`*`) indicates every day of the week.
`,`	Lists enumerated values.	In the Minute field, `5,20` indicates that the task is triggered once at both the 5th and 20th minutes.
`-`	Indicates a range.	In the Minute field, `5-20` indicates that the task is triggered once every minute from the 5th to 20th minute.

Special character	Description	Example
`/`	Indicates increments.	In the Minute field, `0/15` indicates that the task is triggered once every 15 minutes from the beginning of an hour. In the Minute field, `3/20` indicates that the task is triggered once every 20 minutes from the 3rd minute of an hour.
`?`	Indicates an unspecified value. Only the Day and Week fields support this character.	If the Day or Week field is specified, the other field must be set to a question mark (`?`) to prevent conflicts.
`L`	Indicates the last day of a specific period. Only the Day and Week fields support this character. **Note** To prevent logic errors, do not specify a list or range when you use the `L` character.	• In the Day field, `L` indicates the last day of a month. In the Day of a week field, `L` indicates the last day of a week, which is Sunday (`SUN`). • `L` can be preceded by a value. For example, `6L` in the Week field indicates the last Saturday of a month.
`W`	The weekday that is nearest to the specified day of the month. The weekday that the `W` character indicates is in the same month as the specified day of the month. `LW` indicates the last weekday of the specified month.	If `5W` is specified in the Day field and the 5th day of the month falls on Saturday, the task is triggered on Friday, which is the 4th day of the month. If the 5th day of the month falls on Sunday, the scheduled task is triggered on Monday, which is the 6th day of the month. If the 5th day of the month falls on a weekday, the scheduled task is triggered on the 5th day of the month.
`#`	A specific day of a specific week in every month. Only the Day of a week field supports this character.	In the Week field, `4#2` indicates the second Thursday of a month.

Examples

The following table describes some example values of cron expressions.

Example	Limit
`0 15 10 ? * *`	Executes the task at 10:15 every day.
`0 15 10 * * ?`	Executes the task at 10:15 every day.
`0 0 12 * * ?`	Executes the task at 12:00 every day.
`0 0 10,14,16 * * ?`	Executes the task at 10:00, 14:00, and 16:00 every day.
`0 0/30 9-17 * * ?`	Executes the task every half an hour between 09:00 and 17:00 every day.

Example	Limit
`0 * 14 * * ?`	Executes the task every minute between 14:00 and 14:59 every day.
`0 0-5 14 * * ?`	Executes the task every minute between 14:00 and 14:05 every day.
`0 0/5 14 * * ?`	Executes the task every 5 minutes between 14:00 and 14:55 every day.
`0 0/5 14,18 * * ?`	Executes the task every 5 minutes between 14:00 and 14:55 and between 18:00 and 18:55 every day.
`0 0 12 ? * WED`	Executes the task at 12:00 every Wednesday.
`0 15 10 15 * ?`	Executes the task at 10:15 on the 15th day of every month.
`0 15 10 L * ?`	Executes the task at 10:15 on the last day of every month.
`0 15 10 ? * 6L`	Executes the task at 10:15 on the last Saturday of every month.
`0 15 10 ? * 6#3`	Executes the task at 10:15 on the third Saturday of every month.
`0 10,44 14 ? 3 WED`	Executes the task at 14:10 and 14:44 every Wednesday in March every year.

17.Dedicated hosts

17.1. Create a dedicated host

A dedicated host is a cloud host whose physical resources are reserved for the exclusive use of a single tenant. Elastic Compute Service (ECS) instances created on a dedicated host are physically isolated from those created on other hosts. This topic describes how to create a dedicated host.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Dedicated Hosts**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Hosts** tab, click **Create Host**.
5. On the **Create Host** page, configure the parameters described in the following table.

Section	Parameter	Description
Basic Configurations	Organization	The organization in which to create the dedicated host.
	Resource Set	The resource set in which to create the dedicated host.
Region and Zone	Region	The region in which to create the dedicated host.
	Zone	The zone in which to create the dedicated host.
Instance	Dedicated Host Type	The type of the dedicated host. The dedicated host type determines the instance family and the maximum number of ECS instances that you can deploy on the dedicated host.
	DDH Name	The name of the dedicated host. The name must be 2 to 128 characters in length and can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,). It must start with a letter and cannot start with http:// or https://.
	Quantity	The number of dedicated hosts that you want to create.
DDH Settings	Allow Automatic Deployment	Specify whether to add the dedicated host to the resource pool for automatic deployment. Valid values: ○ Allow ○ Forbid

Section	Parameter	Description
	Automatic Instance Migration upon DDH Failure	Specify whether to fail over the instances deployed on the dedicated host when it fails. Valid values: ○ Enable ○ Disable

6. Click **Submit**.

Result

After the dedicated host is created, you can view it in the Dedicated Host list and create instances on it. For more information about the parameters used to create an ECS instance, see Create an instance.

17.2. Create a host group

You can group dedicated hosts into host groups for easy management. This topic describes how to create a host group.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Dedicated Hosts**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Host Groups** tab.
5. Click **Create Host Group**.
6. On the **Create Host Group** page, configure the parameters described in the following table.

Section	Parameter	Required	Description
Basic Settings	Organization	Yes	The organization in which to create the host group.
	Resource Set	Yes	The resource set in which to create the host group.
Region and Zone	Region	Yes	The region in which to create the host group.
	Zone	Yes	The zone in which to create the host group.
Instance	Host Group Name	No	The name of the host group. The name must be 2 to 128 characters in length. It must start with a letter and cannot start with http:// or https://. It can contain letters, digits, periods (.), underscores (_), hyphens (-), colons (:), and commas (,).

7. Click **Submit**.

Result

After the host group is created, you can view it in the host group list.

17.3. Add dedicated hosts to a host group

After you create a host group, you can add dedicated hosts to the host group for easy management. This topic describes how to add dedicated hosts to a host group.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Dedicated Hosts**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the **Host Groups** tab.
5. Find the host group to which you want to add dedicated hosts and click **Add Host** in the **Actions** column.
6. In the **Add Host** panel, select a dedicated host and click **Add Host**.

   To add a new dedicated host, you can click **Create Host** in the **Add Host** panel. For information about the parameters used to create a host, see Create a dedicated host. After the host is created, add it to the host group.

Result

After the dedicated host is added to the host group, you can click the host group name to view the dedicated host in the **Hosts** list.

18.Install FTP software

18.1. Overview

File Transfer Protocol (FTP) transfers files between a client and a server by establishing two TCP connections. One is the command link for transferring commands between a client and a server. The other is the data link used to upload or download data. Before uploading files to an instance, you must build an FTP site for the instance.

18.2. Install and configure vsftp in CentOS

This topic describes how to install and configure vsftp in CentOS to transfer files.

Procedure

1. Install vsftp.

   ```
   yum install vsftpd -y
   ```

2. Add an FTP account and a directory.

   i. Check the location of the *nologin* file,

   which is usually under the */usr/sbin* or */sbin* directory.

   ii. Create an FTP account.

   Run the following commands to create the */alidata/www/wwwroot* directory and specify this directory as the home directory of the account pwftp. You can also customize the account name and directory.

   ```
   mkdir -p /alidata/www/wwwroot
   useradd -d /alidata/www/wwwroot -s /sbin/nologin pwftp
   ```

   iii. Modify the account password.

   ```
   passwd pwftp
   ```

   iv. Modify the permissions on the specified directory.

   ```
   chown -R pwftp.pwftp /alidata/www/wwwroot
   ```

3. Configure vsftp.

   i. Open the vsftp configuration file.

   ```
   vi /etc/vsftpd/vsftpd.conf
   ```

   ii. Change the value of anonymous_enable from `YES` to `NO`.

   iii. Delete the comment delimiter (`#`) from the following configuration lines:

   ```
   local_enable=YES
           write_enable=YES
           chroot_local_user=YES
   ```

iv. Press the Esc key to exit the edit mode, and enter `:wq` to save the modifications and exit.

4. Modify the shell configuration.
   i. Open the shell configuration file.

   ```
   vi /etc/shells
   ```

   ii. If the file does not contain */usr/sbin/nologin* or */sbin/nologin*, add it to the file.
5. Start vsftp and perform a logon test.
   i. Start vsftp.

   ```
   service vsftpd start
   ```

   ii. Use the account pwftp to perform an FTP logon test.

   This example uses the directory */alidata/www/wwwroot*.

18.3. Install vsftp in Ubuntu or Debian

This topic describes how to install and configure vsftp in an instance running Ubuntu or Debian to transfer files.

Procedure

1. Update the software source.

   ```
   apt-get update
   ```

2. Install vsftp.

   ```
   apt-get install vsftpd -y
   ```

3. Add an FTP account and a directory.
   i. Check the location of the *nologin* file,

   which is typically under the */usr/sbin* or */sbin* directory.

   ii. Create an FTP account.

   Run the following commands to create the */alidata/www/wwwroot* directory and specify this directory as the home directory of the account pwftp. You can also customize the account name and directory.

   ```
   mkdir -p /alidata/www/wwwroot
   useradd -d /alidata/www/wwwroot -s /sbin/nologin pwftp
   ```

   iii. Modify the account password.

   ```
   passwd pwftp
   ```

   iv. Modify the permissions on the specified directory.

   ```
   chown -R pwftp.pwftp /alidata/www/wwwroot
   ```

4. Configure vsftp.
   i. Open the vsftp configuration file.

```
vi /etc/vsftpd.conf
```

ii. Change the value of anonymous_enable from `YES` to `NO`.

iii. Delete the comment delimiter (`#`) from the following configuration lines:

```
local_enable=YES
        write_enable=YES
        chroot_local_user=YES
        chroot_list_enable=YES
        chroot_list_file=/etc/vsftpd.chroot_list
```

iv. Press the Esc key to exit the edit mode, and enter `:wq` to save the modifications and exit.

v. Open the */etc/vsftpd.chroot_list* file and add the FTP account name to the file. Save the modifications and exit.

You can follow steps a to d to open and save the file.

5. Modify shell configurations.

i. Open the shell configuration file.

```
vi /etc/shells
```

ii. If the file does not contain */usr/sbin/nologin* or */sbin/nologin*, add it to the file.

6. Start vsftp and perform a logon test.

i. Start vsftp.

```
service vsftpd restart
```

ii. Use the account pwftp to perform an FTP logon test.

This example uses the directory */alidata/www/wwwroot*.

18.4. Build an FTP site in Windows Server 2008

This topic describes how to build an FTP site on an Elastic Compute Service (ECS) instance that runs Windows Server 2008.

Prerequisites

The Web Server (IIS) role is added and FTP is installed on the ECS instance.

Procedure

1. Connect to the instance. For more information, see Instance connecting overview.
2. Choose **Start > Administrative Tools > Internet Information Services (IIS) Manager**.
3. Right-click the server name and select **Add FTP Site**.
4. Enter an FTP site name and a physical path, and then click **Next**.
5. Set **IP Address** to **All Unassigned** and **SSL** to **No SSL**, and then click **Next**.
6. Set **Authentication** to **Basic**, **Authorization** to **All Users**, and **Permissions** to **Read** and **Write**, and click **Finish**.

Result

Then you can use the administrator account and its password to upload and download files through FTP. Make sure that the following conditions are met:

- The FTP port is not in use by other applications, and the Windows firewall is not blocking the port.
- The security group of the instance contains a security group rule that allows inbound traffic on the FTP port.

18.5. Build an FTP site in Windows Server 2012

This topic describes how to build an FTP site on an Elastic Compute Service (ECS) instance that runs Windows Server 2012.

Prerequisites

The Web Server (IIS) role is added and FTP is installed on the ECS instance.

Procedure

1. Connect to the instance. For more information, see Instance connecting overview.
2. In the left-side navigation pane, click **IIS**.
3. In the left-side navigation pane, click **IIS**.
4. In the **SERVERS** section, right-click the server name and select **Internet Information Services (IIS) Manager**.
5. Right-click the server name and select **Add FTP Site**
6. Enter an FTP site name and a physical path, and then click **Next**.
7. Set **IP Address** to **All Unassigned** and **SSL** to **No SSL**, and then click **Next**.
8. Set **Authentication** to **Basic**, **Authorization** to **All Users**, and **Permissions** to **Read** and **Write**, and click **Finish**.

Result

Then you can use the administrator account and its password to upload and download files through FTP. Make sure that the following conditions are met:

- The FTP port is not in use by other applications, and the Windows firewall is not blocking the port.
- The security group of the instance contains a security group rule that allows inbound traffic on the FTP port.