

Elastic Compute Service

User Guide

User Guide

Quick reference

This article provides a quick reference for you on the use of ECS on the ECS Management Console.

Must read: ECS operation instructions

- ECS operation instructions

Log on to an instance

How to use a username and password to log on to a Linux instance?

How to use an SSH key pair to log on to a Linux instance?

How to log on to a Windows instance?

If you forget your instance logon password (not the Management Terminal password), you can reset the password.

Operate disks

- How to attach a data disk after you purchase a cloud disk?

Change the operating system

You can change the operating system, such as:

From Windows to Linux, or from Linux to Windows.

From one version to another, for example, from Windows Server 2008 to Windows Server

2012.

Change the image, for example, change to custom images, or shared images.

Use images and snapshots

How to copy images across different regions?

How to define automatic snapshot policies when you want an automatic update policy for configurations or applications?

Enable Intranet communication

- How to use security groups to enable intranet communication?

For the answers to the above questions, click the corresponding node in the left navigation bar.

To ensure proper operation of your ECS instance, you must take the considerations outlined in this section into account before use.

Prohibitions

Alibaba Cloud prohibits you from:

- Using your instance for flow-through services. Any violation will result in punishment up to shutdown and lockout of instance, and termination of services.
- Upgrading the ECS kernel or OS without prior authorization.
- Activating SELinux.
- Uninstalling PVDriver.
- Arbitrarily modifying the MAC address of the network adapter.

General operating system considerations

- For an ECS with more than 4 GB RAM, use of a 64-bit OS is recommended, because a 32-bit OS supports a maximum of 4 GB RAM. Currently available 64-bit systems include:
 - Aliyun Linux
 - CoreOS
 - CentOS
 - Debian

- FreeBSD
 - OpenSUSE
 - SUSE Linux
 - Ubuntu
 - Windows 2008/2012
- 32-bit Windows OS supports CPUs with up to 4 cores.
 - To ensure service continuity and avoid service downtime, enable auto-start of service applications upon OS boot.
 - For I/O-optimized instances, do not stop the **aliyun-service** process.

Linux restrictions

To ensure stable system operation, **DO NOT**:

- Modify the content of the default `/etc/issue` file. Modifying this file will render management console buttons unusable.
- Modify directory permissions in partitions, particularly permissions for directories such as `/etc`, `/sbin`, `/bin`, `/boot`, `/dev`, `/usr`, and `/lib`. Improper modification of permissions may cause errors.
- Rename, delete, or disable the Linux **root** account.
- Compile or perform any other operations on the Linux kernel.
- Enable the **NetWorkManager** service. This service conflicts with the internal network service of the system and will cause network errors.

Windows restrictions

- Do not close the built-in **shutdownmon.exe** process. This may delay the restart of your Windows server.
- Do not rename, delete, or disable the **Administrator** account.
- The use of virtual memory is not recommended.

Pay-As-You-Go restrictions

The Pay-As-You-Go service is subject to the following restrictions:

- An ECS server can use only one billing method, meaning that you cannot switch between Subscription and Pay-As-You-Go services. You cannot modify the configuration of the Pay-As-You-Go service, including upgrades of bandwidth, CPU/memory, and data disks.
- If a fixed 0 Mbps bandwidth is selected, you will not be allocated an external IP address, neither can you upgrade the bandwidth.
- A record-filing service is not offered for Pay-As-You-Go services.
- The fee for Pay-As-You-Go services is calculated per hour as follows: Total fee = CPU fee + memory fee + data disk fee + public bandwidth fee.

ECS API product and business restrictions

Unless otherwise specified, all the resource limits listed in the table are for one region.

Restricted item	Parameter	To apply for an exception or unlock configuration rights
User creation of ECS resources	N/A	Users must undergo real-name authentication
Zones in which users may create instances	1 online zone	Request increase via ticket
Zones in which users may create disks	The zone combinations in which users may create instances and the zones where instances will remain after overlap removal	N/A
Default Pay-As-You-Go instance types	ecs.t1.small (single-core 1 GB)	Request change via ticket
	ecs.s1.small (single-core 2 GB)	
	ecs.s1.medium (single-core 4 GB)	
	ecs.s2.small (dual-core 2 GB)	
	ecs.s2.large (dual-core 4 GB)	
	ecs.s2.xlarge (dual-core 8 GB)	
	ecs.s3.medium (quad-core 4 GB)	
	ecs.s3.large (quad-core 8 GB)	
	ecs.m1.medium (quad-core 16 GB)	
Default Pay-As-You-Go instance quota in all regions	30	Request increase via ticket
Disk quota within a single instance	17, including one system disk and 16 data disks	No higher configurations exist
Support for Pay-As-You-Go ephemeral disks	Supported	For APIs, open a ticket. Other types do not support this disk type.
Capacity of a single ephemeral disk	20 to 1,024 GB	No higher configurations exist
Total ephemeral disk size for a single instance	2,048 GB	No higher configurations exist

Number of snapshots	(Number of disks)*64	No higher configurations exist
Capacity of a single basic cloud disk	5 to 2,000 GB	No higher configurations exist
Number of accounts to share a single custom image with	50	No higher configurations exist
List of available public images	List of images for sale on the official website	Common users cannot change; for others, open a ticket to add other images
Available inbound bandwidth for Internet access	Up to 200 Mbps	No higher configurations exist
Available outbound bandwidth for Internet access	Up to 200 Mbps	No higher configurations exist
Number of instances allowed in a single security group	1,000	No higher configurations exist
Number of authorization rules for a single security group	100	No higher configurations exist
Security group quota	100	Request increase via ticket
Maximum number of security groups that a single instance can belong to	5	No higher configurations exist
Restrictions on image and instance types	Instances with 4 GB or more of memory cannot use 32-bit images	N/A
Adding new disks for ephemeral disk instances	Not supported	N/A
Changing the configuration of an instance with ephemeral disks	Bandwidth is adjustable	No exceptions (cloud disks can be used for attachment)
Relationship between the system disk and data disk	If the system disk is a cloud disk, all data disks must be cloud disks	N/A
Total number of Pay-As-You-Go cloud disks that can be purchased	ECS Instance Quota * 5	Open a ticket
Basic cloud disk capacity	5 to 2,000 GB	No higher configurations exist
User restrictions for the creation of Pay-As-You-Go cloud disks	Users must pass real-name authentication (for buy only)	N/A
Range of system disk attaching points	/dev/xvda	N/A

Range of data disk attaching points	/dev/xvd[b-z]	N/A
Capacity of a single SSD Cloud Disk	20 to 2,048 GB	No higher configurations exist
Capacity of a single ultra Cloud Disk	20 to 2,048 GB	No higher configurations exist

Besides all listed in the preceding table, neither does ECS support:

- Sound card applications.
- The installation of external hardware devices such as hardware dongles, USB drives, external hard drives, and the USB security keys issued by banks.
- SNAT and other IP packet address translation services. Achieve this using an external VPN or proxy.
- Multicast protocol. If multicasting services are required, unicast point-to-point method is recommended.
- Virtual application installation or subsequent virtualization such as when using VMware.

For limitations of VPC, see **Limitations** in *VPC Product Introduction* for details.

Connect

You can use the Management Terminal, also called VNC, to connect to an ECS instance, especially when the remote access software program that you are using, such as PuTTY, Xshell, or SecureCRT, cannot work.

Usage scenarios

The Management Terminal is used to:

- Check the status of an ECS instance if boot speed is slow.
- Reconfigure the firewall if a remote connection fails due to a software error within the ECS instance.
- End abnormal processes consuming excessive CPU usage or bandwidth.

Note:The Management Terminal can be used to connect to an instance even if purchased bandwidth is insufficient.

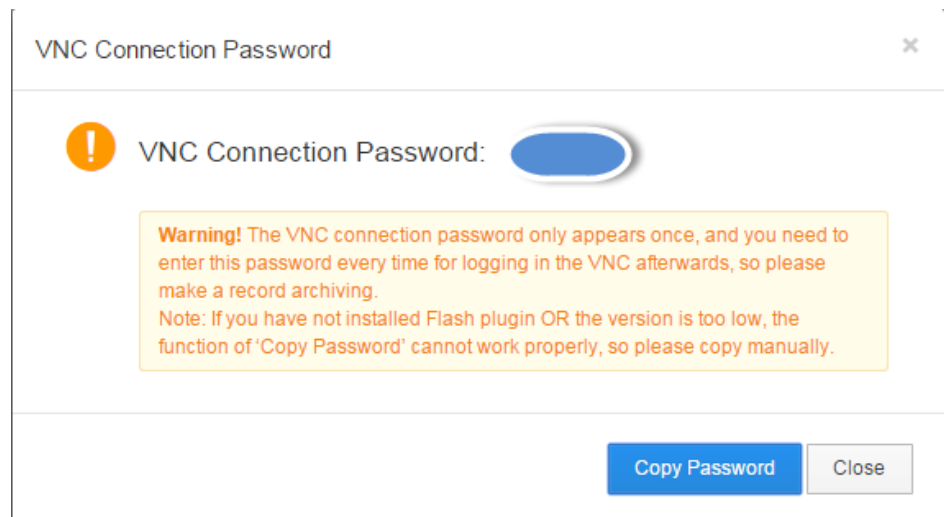
Procedure

Log on to the ECS console.

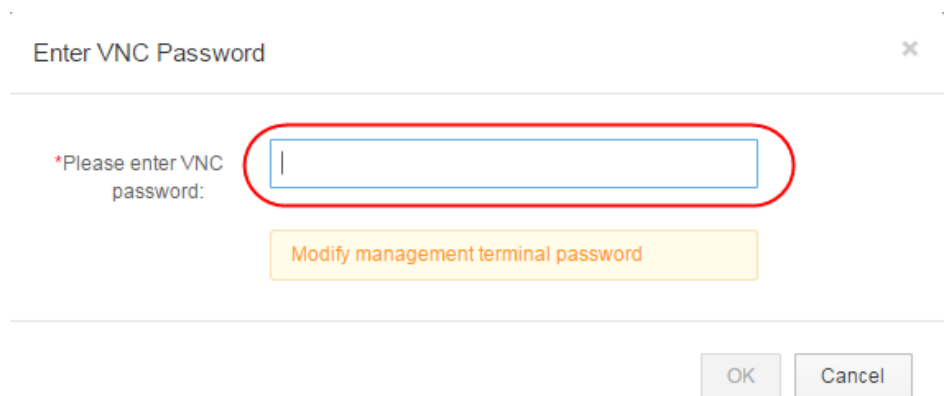
Go to the ECS instance to connect to, and click **VNC** on the right.

Follow the tips below to connect to the **Management Terminal**:

- If you connect the **Management Terminal** for the first time, follow the steps below:
 - a. On the **VNC Connection Password** dialog, copy the password. This dialog appears only once, but you need to enter the connection password each time you want to connect to the **Management Terminal**, so **write down the password**.



- b. Click the **Close** button to close the **VNC Connection Password** dialog.
- c. On the **Enter VNC Password** dialog, paste the connection password that you copied, and then click the **OK** button to connect to the **Management Terminal**.



- If this is not your initial connection to the **Management Terminal**, the **Enter VNC**

Password dialog appears, and you need to enter the connection password and click the **OK** button to connect to the **Management Terminal**.

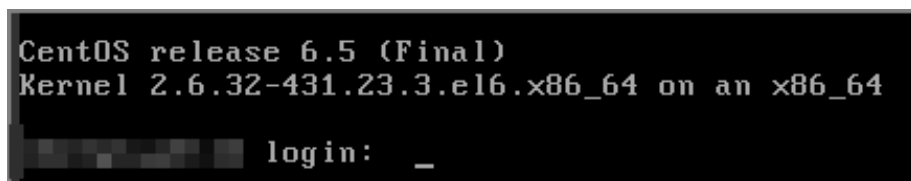
- If you forget the password, you can follow the steps below to connect to the **Management Terminal**:

- a. Change password.
- b. On the upper left corner of the **Management Terminal** interface, click **Send remote command** > **Connect to management terminal**.
- c. On the **Enter VNC Password** dialog, enter the new password to finish connection.

Follow the steps below to connect to an instance:

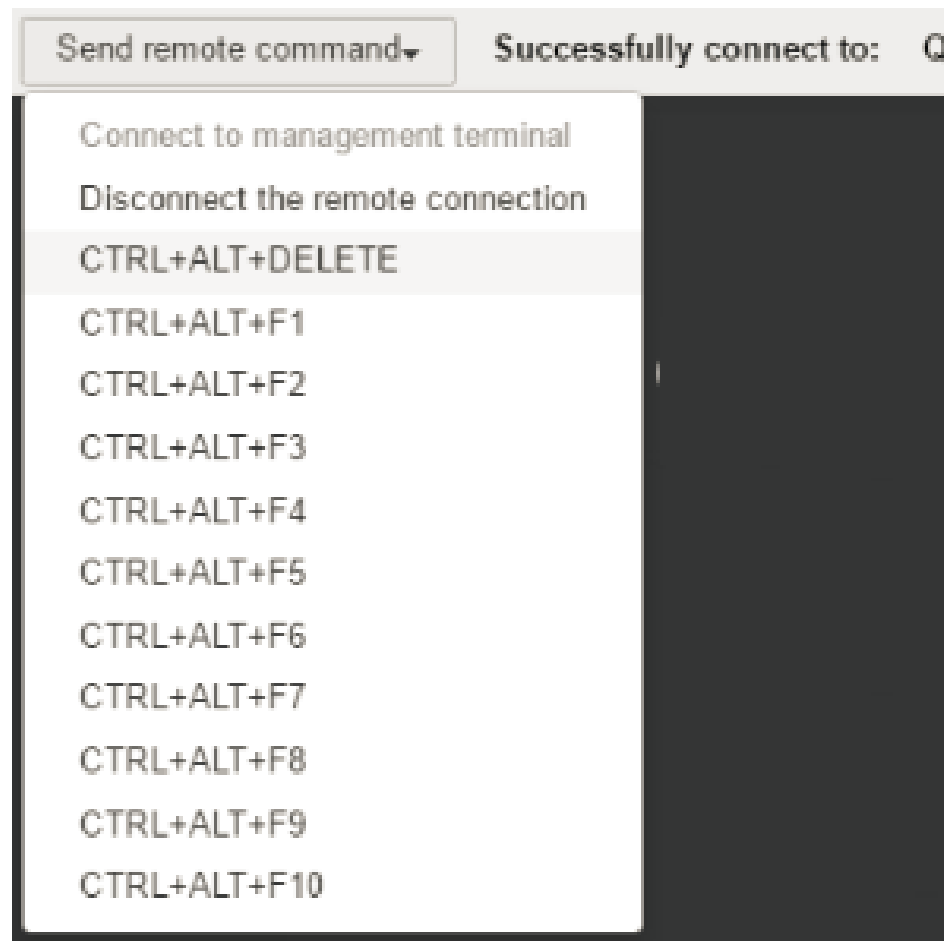
For a Linux instance, enter the user name ("root") and the password to connect to it. Your screen may go black constantly, which occurs when the Linux instance is in sleep mode. Click the mouse or press any key to wake it up.

If you are operating several Linux instances, you can click **Send remote command** > **CTRL+ALT+Fx**, of which **Fx** can be any one from **F1** to **F10**, to switch management terminals.

A screenshot of a terminal window with a black background and white text. The text displays the system information: 'CentOS release 6.5 (Final)' and 'Kernel 2.6.32-431.23.3.el6.x86_64 on an x86_64'. Below this, the 'login:' prompt is shown with a single underscore character as input.

```
CentOS release 6.5 (Final)
Kernel 2.6.32-431.23.3.el6.x86_64 on an x86_64
login: _
```

For a Windows instance, on the upper left corner of the **Management Terminal** interface, click **Send remote command** > **CTRL+ALT+DELETE** to get to the logon screen. Enter the user name, Administrator, and password to log on.



Change password

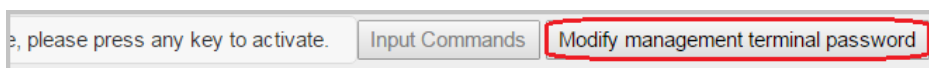
If you prefer a password that you are familiar with rather than the password displayed on the **VNC Connection Password** dialog, or if you forget your password, you can change the connection password.

Note: If the instance that you are connecting to is not I/O optimized, you need restart your instance to make the new VNC connection password take effect after you change it. The restart operation will stop the work of your instance and interrupt your business. So you must be cautious to change the password.

Log on to the ECS console.

Go to the ECS instance to connect to, and click **VNC** on the right.

Close the **VNC Connection Password** dialog or **Enter VNC Password** dialog, and on the upper right corner of the **Management Terminal** interface, click the **Modify management terminal password** button to start changing the password.



Enter a new password, which must be of 6-character length, composed of uppercase letters, lowercase letters, digits, or a combination of them, but not special characters.

Make the new password take effect:

- If the instance that you are connecting to is I/O optimized, the new password takes effect immediately.
- If the instance that you are connecting to is not I/O optimized, restart the instance through the Management Console for the new password to take effect. Restarting within the instance does not work.

FAQ

Q: Can multiple users be simultaneously connected to the Management Terminal?

A: No. Only one user can be connected to the Management Terminal at any given time.

Q: Why can I not connect to an instance using the Management Terminal after changing the password?

A: Ensure you are entering the right **VNC connection password**. If the instance that you are connecting to is not I/O optimized, you must restart the instance through the Management Console for the new **VNC connection password** to take effect. Restarting directly within the instance will not make the password take effect.

Q: Why do I see a black screen after logging on to my instance?

A: A black screen indicates that the instance is in sleep mode.

- For a Linux instance, press any key to wake it up.
- For a Windows instance, click **Send remote command** > **Ctrl+ALT+DELETE** to bring back the logon interface.

Q: Why can I not access the Management Terminal?

A: To address login issues, open your browser and connect to the Management Terminal. Press the F12 key to open the developer tool. The Management Terminal information can then be analyzed to locate faults under the Console tab.

Q: I cannot use IE8.0 or Firefox to open the Management Terminal. How can I resolve this?

A: Only IE10 or higher is supported, and only certain versions of Firefox are supported.

To resolve this issue, update or change your browser to a compatible version.

Note: Google Chrome offers the best support for the Management Terminal function, and is recommended for use when you connect to the Management Terminal.

Log on to an instance using an SSH key pair

How to use a key pair to log on to a Linux instance depends on the local operating system.

- Using Windows OS
- Using Linux OS or other systems supporting SSH commands

Using Windows OS

In this section, it is demonstrated how to use a key pair to log on to a Linux instance on a Windows system, using the popular SSH tools PuTTY and PuTTYgen as an example.

Prerequisites

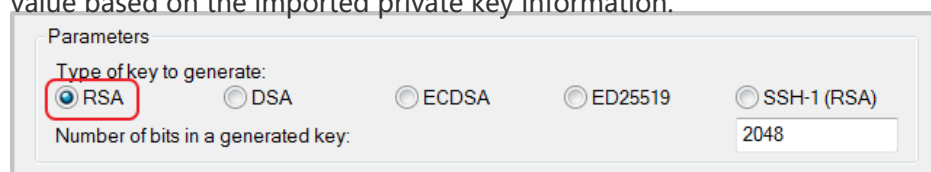
- PuTTY and PuTTYgen must have been installed. You can download them at:
 - For PuTTY: <https://the.earth.li/~sgtatham/putty/latest/w64/putty.exe>.
 - For PuTTYgen: <https://the.earth.li/~sgtatham/putty/latest/w64/puttygen.exe>.
- You must have a Linux instance that has been bound to an instance. You can allocate an SSH key pair when creating an instance or bind an SSH key pair to an instance.

Operating procedure

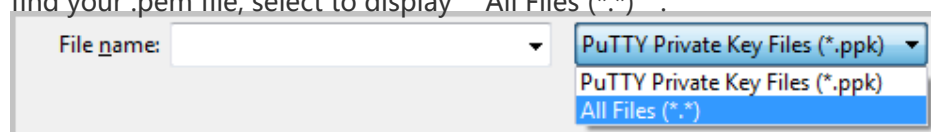
(Optional) If you are using a key pair generated by Alibaba Cloud, of which the private key is a .pem file, you must convert it to a .ppk file. If your private key is a .ppk file, you can skip this step.

- Start PuTTYgen. In this example, we use PuTTYgen version 0.68.

Under the **Type of key to generate** option, select **RSA**. The value of **Number of bits in a generated key** can be left as is. The software will automatically update the value based on the imported private key information.

The image shows the 'Parameters' dialog box in PuTTYgen. It has a title bar 'Parameters'. Inside, there's a section 'Type of key to generate:' with five radio buttons: 'RSA' (selected and circled in red), 'DSA', 'ECDSA', 'ED25519', and 'SSH-1 (RSA)'. Below this is a label 'Number of bits in a generated key:' followed by a text box containing the value '2048'.

Click **Load**. By default, PuTTYgen only displays files with an extension of .ppk. To find your .pem file, select to display "All Files (*.*)" .

The image shows the 'File name' dialog box in PuTTYgen. It has a label 'File name:' followed by a text box. To the right of the text box is a dropdown menu. The dropdown menu is open, showing three options: 'PuTTY Private Key Files (*.ppk)' (selected), 'PuTTY Private Key Files (*.ppk)', and 'All Files (*.*)'.

Select the downloaded private key file from Alibaba Cloud, or the ready private key file, and click **Open**.

Click **OK** to close the confirmation dialog box.

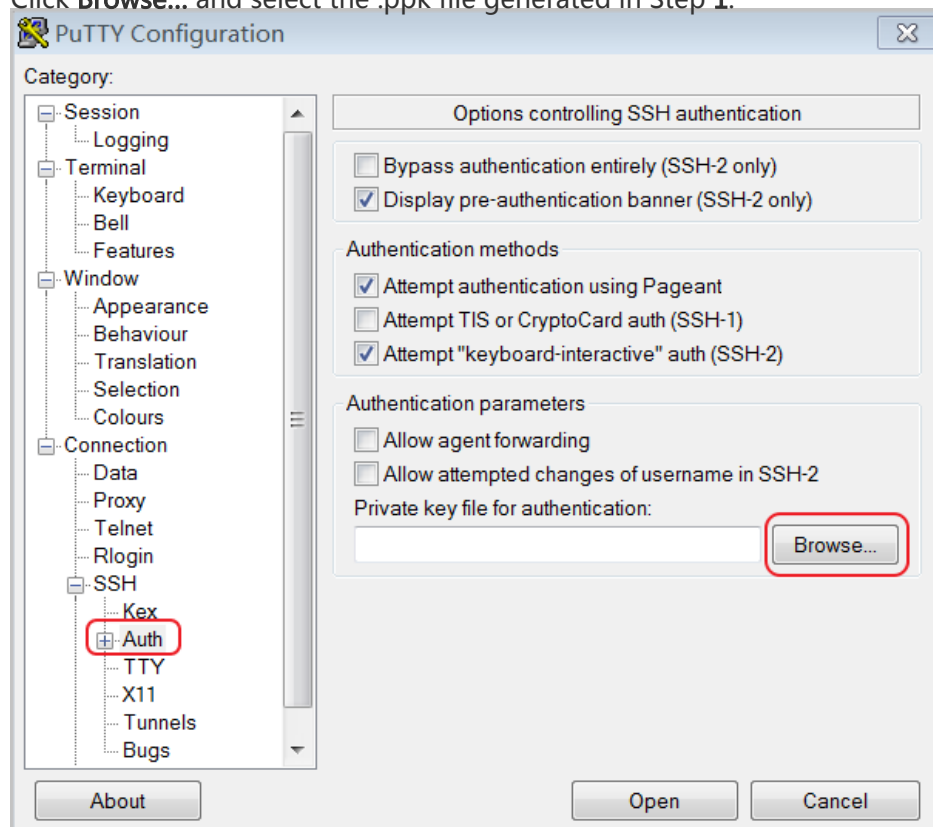
Click **Save private key**. PuTTYgen will display a warning about saving the key without a password. Click **Yes**.

Specify the same name for the private key with the key pair, and save the settings. PuTTY automatically adds the .ppk file.

Start PuTTY.

Click **Connection > SSH > Auth**.

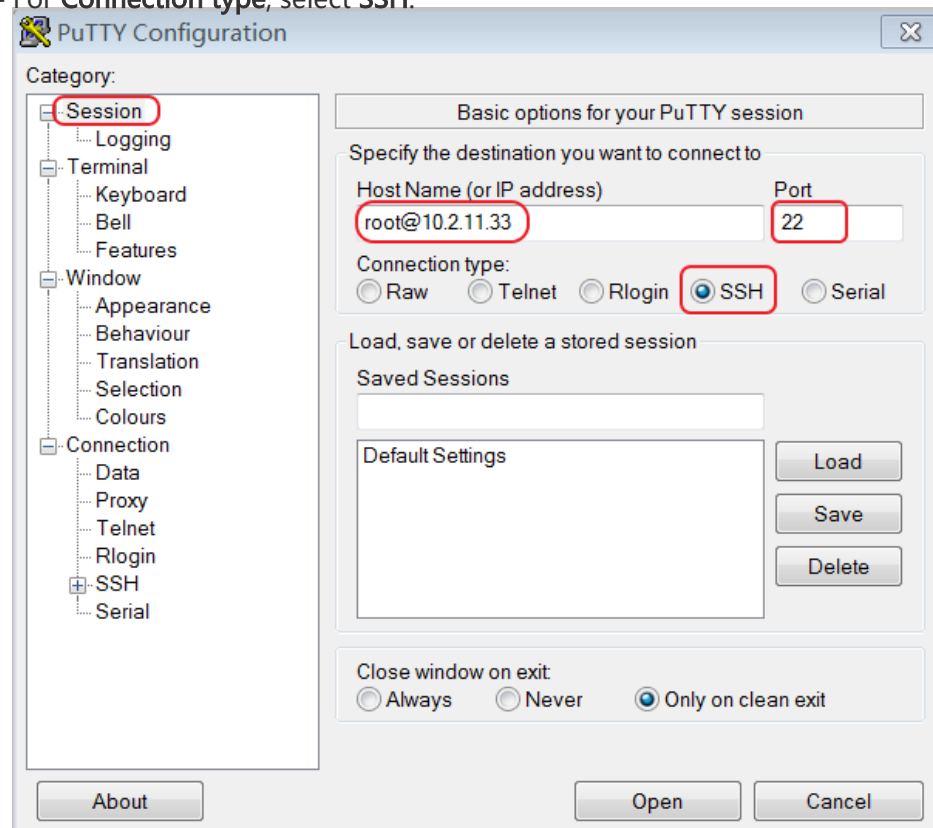
Click **Browse...** and select the .ppk file generated in Step 1.



Click **Session**.

- In **Host Name (or IP address)**, enter your account and the public IP address of the instance to be connected to. The format is "root@IP address" .
- In **Port**, enter the port number **22**.

- For **Connection type**, select **SSH**.



Click **Open** to start accessing your Linux instance.

When the window shows *Connection established*., it indicates you have successfully logged on to the instance using the key pair.

Using Linux OS or other systems supporting SSH commands

In this section, it is demonstrated how to use a key pair to log on to a Linux instance on a Linux system or a system supporting SSH commands, such as MobaXterm for Windows.

Prerequisites

You must have a Linux instance that has been bound to an instance. You can allocate an SSH key pair when creating an instance or bind an SSH key pair to an instance.

Operating procedure

Locate directory of your private key, for example, `/root/xxx.pem`.

xxx.pem can be replaced with your private key.

Use the command to modify the attributes of the private key: `chmod 400 xxx.pem`.

Use the command to connect to the instance: `ssh root@10.10.10.100 -i /root/xxx.pem`.

Replace the Internet IP address in the example, 10.10.10.100, with that of your own instance.

Connect to a Linux instance

The utilities used to remotely connect to Linux ECS instances vary based on the local OS as follows:

- For a Linux OS, use Secure Shell (SSH) Command Line.
- For a Windows OS, use either Management Terminal or use SSH Command Line through PuTTY or other SSH clients.
- For a Mac OS, use Management Terminal or SSH Command Line.
- For iPhone, use SSH Control Lite.
- For Android, use SSH Control Lite.

Connect to a Linux instance using Windows OS

On a Windows system, you can connect to a Linux instance using either of the following methods:

Remote access software

This method is available only if you purchase bandwidth when creating your instance. Prior to using this method, ensure the instance can be accessed through the Internet.

Management Terminal (VNC)

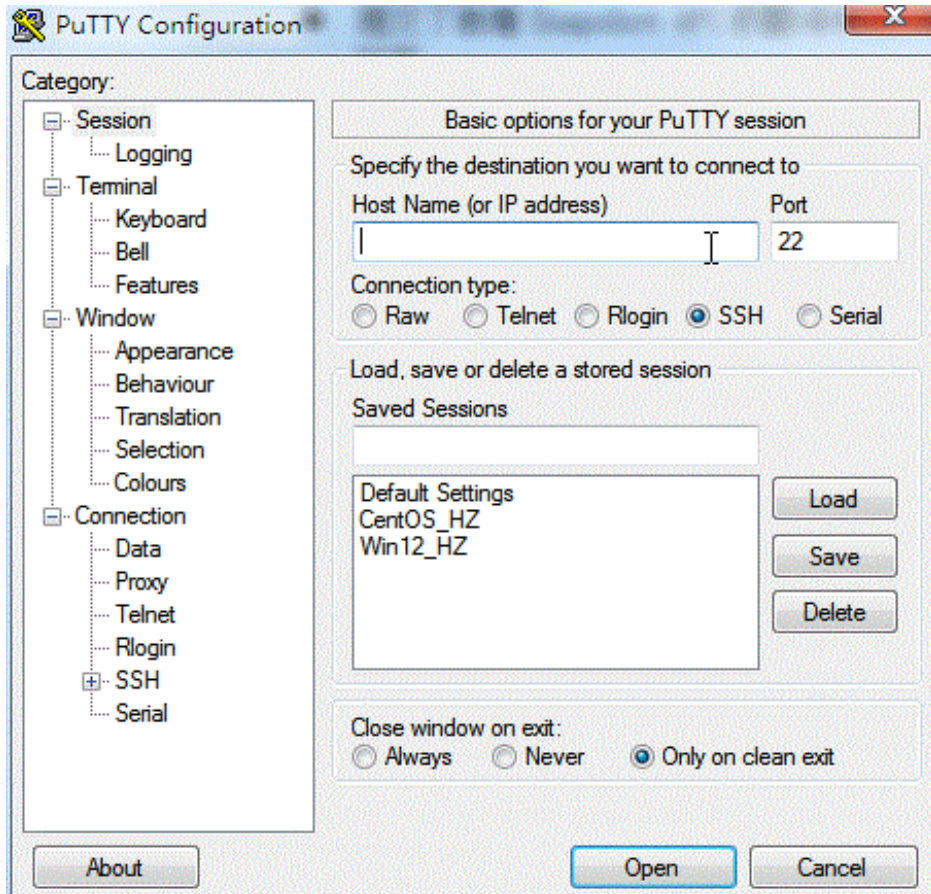
Connection through the Management Terminal (VNC) can be completed disregarding whether bandwidth has been purchased.

Use a remote connection Application

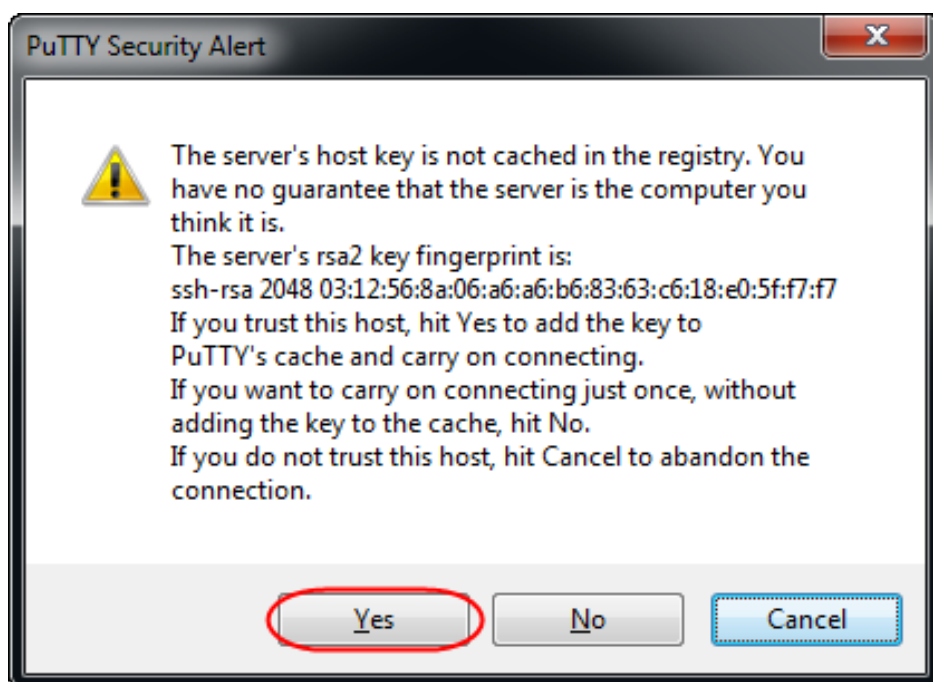
This section uses PuTTY as an example. PuTTY can be downloaded at <http://www.chiark.greenend.org.uk/~sgtatham/putty/>.

You can connect to a Linux instance via PuTTY as follows:

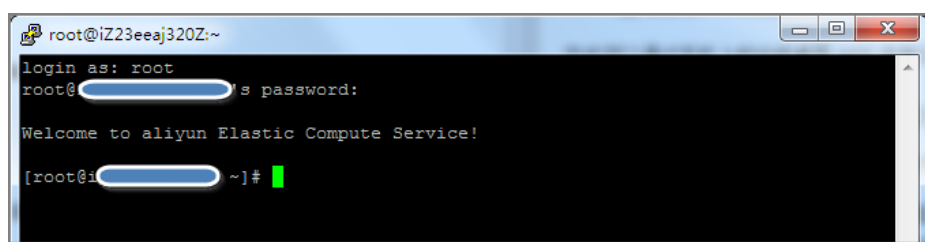
1. Start Putty.exe.
2. Enter the public IP address of the instance in **Host Name (or IP address)**.
3. Use the default port **22**.
4. Select **SSH** as **Connection Type**.
5. Type a session name in **Saved Sessions**, and then click **Save**. In later logins, you may directly load the session without re-entering the IP address.
6. Click **Open** to connect.



7. Upon first connection, the following dialog box will be displayed. Click **Yes**.



8. As prompted, enter the username and password for the Linux ECS instance. The password will not be displayed on-screen. Press the **Enter** key to complete connection to the instance.



When you connect your computer to the Linux instance successfully, you can operate the instance from your computer.

Use Management Terminal to connect to an ECS instance

Refer to [Use Management Terminal \(VNC\) to connect to an ECS instance](#).

Connect to a Linux instance using Linux OS or Mac OS X

1. Connect to the instance using SSH commands. For example, `ssh root@Instance'` s public IP address.
2. Enter the root user password.

Connect to a Linux instance using an mobile app

You can connect to an instance via a remote desktop application installed on your smart phone. For

example, iPhone users can download **SSH Control Lite** from the App Store and use it to connect to Linux instances.

What if I forget my logon password?

If you forget your instance logon password (not the Management Terminal password), reset the logon password. For details, refer to [Reset an instance password](#).

Connect to a Windows instance

The utilities used to remotely connect to Linux ECS instances vary based on local OS as follows:

- For a Linux OS, use rdesktop.
- For a Windows OS, use Management Terminal or Microsoft Terminal Services Client (MSTSC).
- For a Mac OS, use Management Terminal or MSTSC.
- For iPhone, use the Microsoft Remote Desktop app.
- For Android, use the Microsoft Remote Desktop app.

Connect to a Windows instance using Windows OS

Using a local Windows OS, connect to a Windows instance using one of the following:

Microsoft Terminal Services Client (MSTSC)

This method is available only if you purchased bandwidth when creating your instance. Prior to using this method, ensure the instance can be accessed through the Internet.

Management Terminal (VNC)

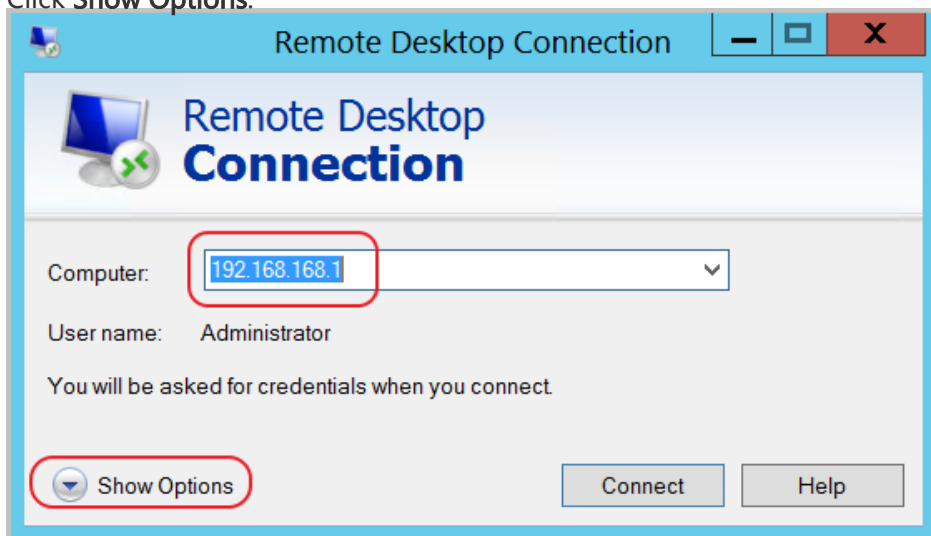
Connection through the Management Terminal can be completed disregarding whether bandwidth has been purchased.

Use MSTSC

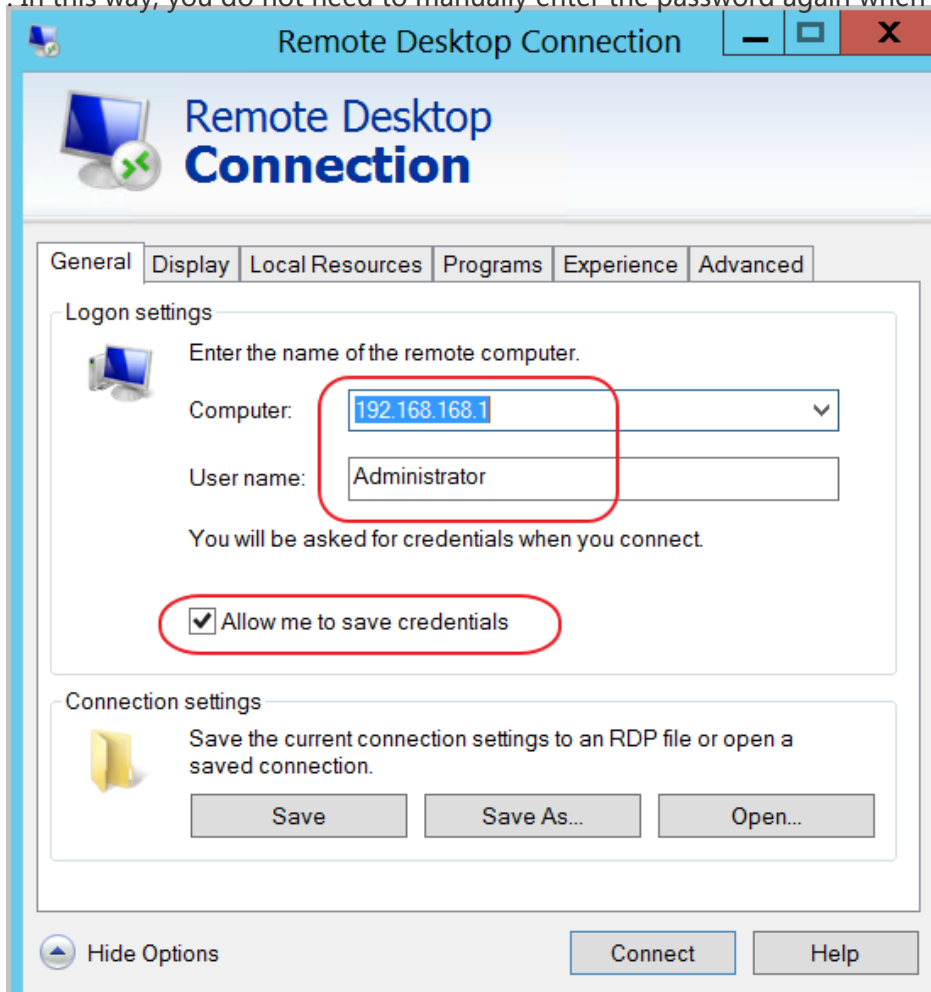
Perform one of the following to start Remote Desktop Connection:

- Click **Start > Remote Desktop Connection**.
- Click the **Start** icon and enter **mstsc** in the search box.
- Press the shortcut key **Windows Logo + R** to open the **Run** window, enter **mstsc**, and then press the **Enter** key.

In the **Remote Desktop Connection** dialog box, enter the public IP address of the instance. Click **Show Options**.

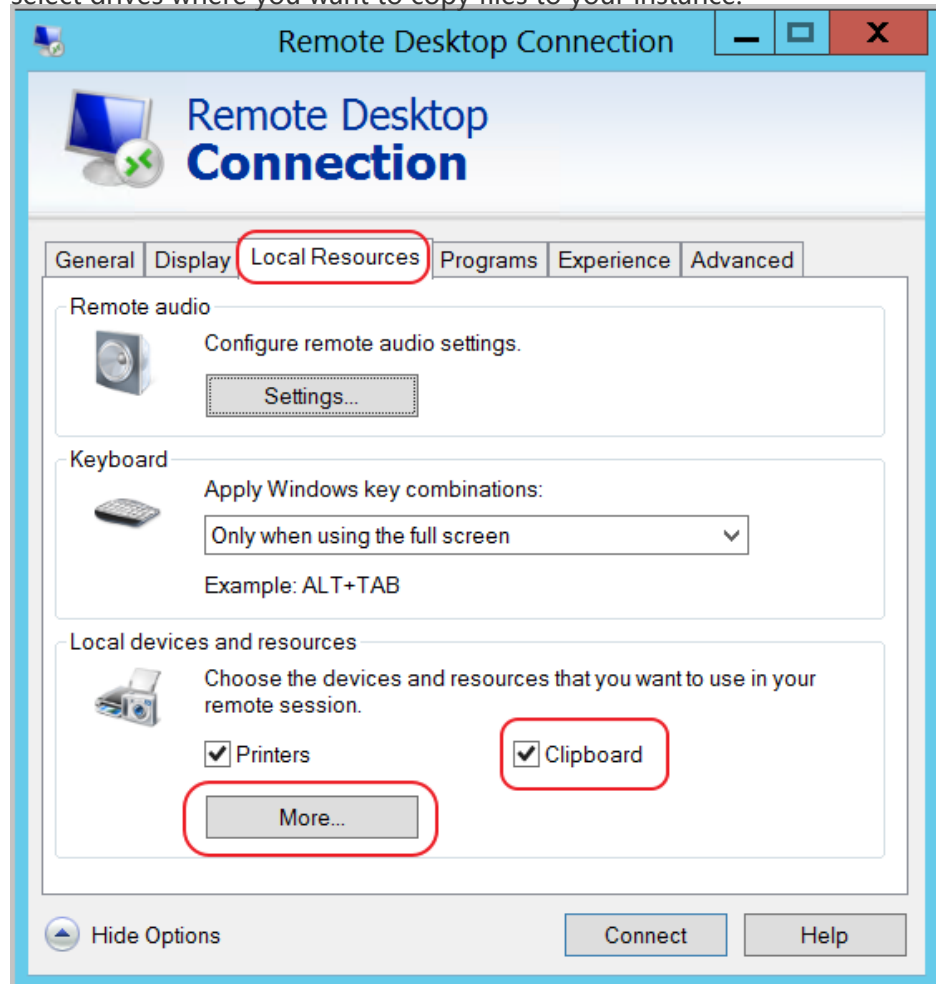


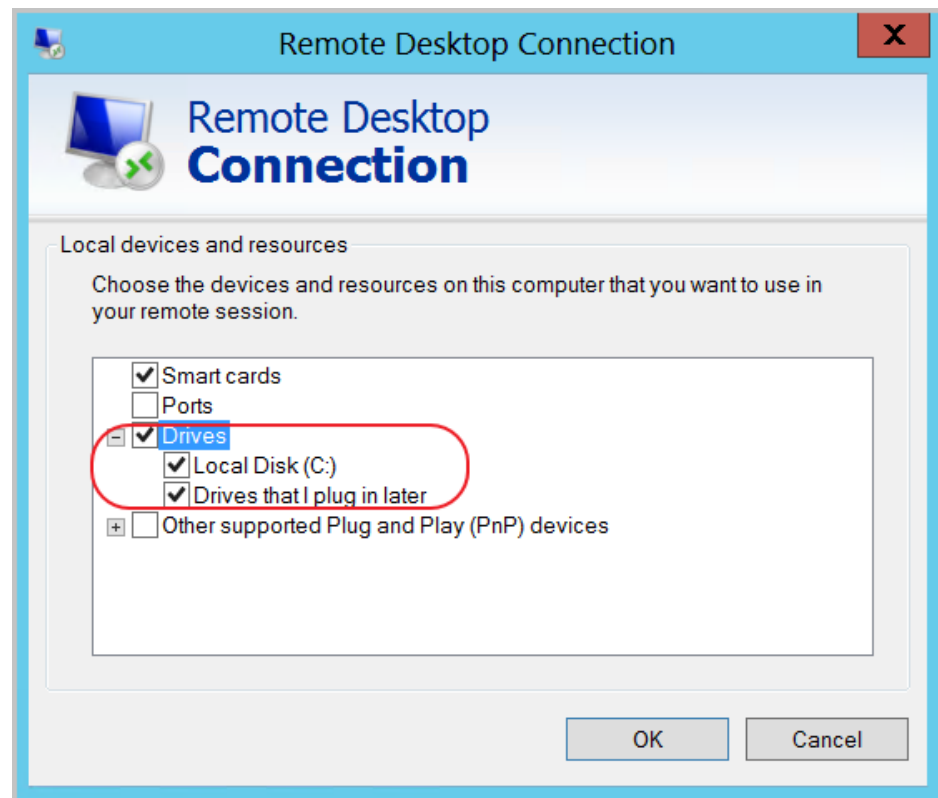
Enter the user name. The default value is **Administrator**. Check **Allow me to save credentials**. In this way, you do not need to manually enter the password again when you log on later.



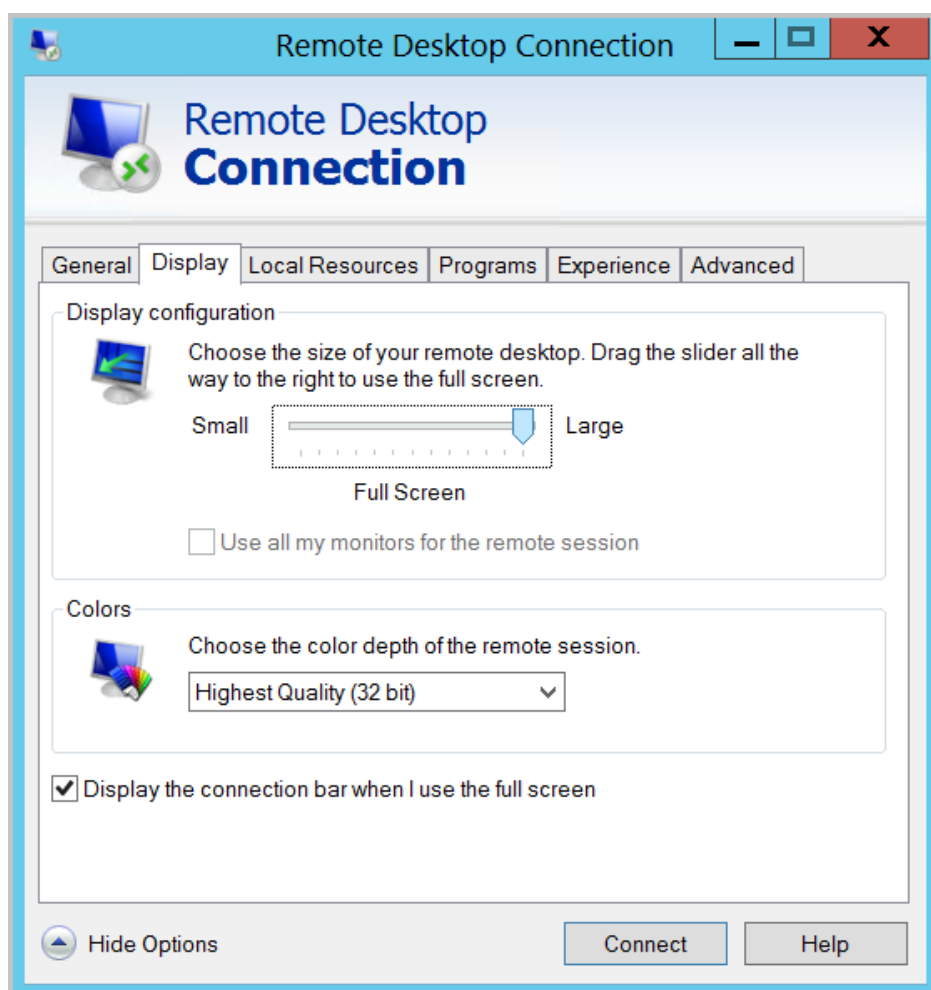
(Optional) If you want to copy text or files from your computer to the instance, click the **Local Resources** tab to see options for sharing local computer resources.

- If you need to copy text only, check **Clipboard**.
- If you need to copy files from your computer as well, click **More > Drives**, and select drives where you want to copy files to your instance.





(Optional) Open the **Display** tab to resize the remote desktop window. **Full Screen** is recommended.



Click the **Connect** button to complete connection to the instance.

When you connect your computer to the Windows instance successfully, you can operate the instance from your computer.

Use Management Terminal (VNC) to connect to an ECS instance

Refer to Use Management Terminal (VNC) to connect to an ECS instance.

Connect to a Windows instance on a Linux OS

If you are connecting to a Windows Instance on a Linux OS, you can use either remote access software or the Management Terminal.

If no bandwidth has been purchased, you must log on to the Alibaba Cloud Console and use the Management Terminal to connect to the instance.

Use a remote connection application

This section uses the rdesktop application as an example. rdesktop can be downloaded at <http://www.rdesktop.org/>. To connect to a Windows instance using rdesktop, perform the following:

1. Start rdesktop.
2. Enter the following command (using your parameters):

```
rdesktop -u administrator -p password -f -g 1024*720 192.168.1.1 -r clipboard:PRIMARYCLIPBOARD -r disk:sunray=/home/yz16184
```

The following table defines the parameters used in the rdesktop command.

Parameter	Description
-u	Indicates username. For a Windows instance, the default username is Administrator.
-p	Indicates the Windows instance login password.
-f	Indicates full screen is the default view. Use the key combination Ctrl+Alt+Enter to exit full screen mode.
-g	Indicates the resolution. If the connector "*" is omitted, the default resolution will be the native resolution in full screen.
IPADDRESS	Enter the IP address of your Windows instance
-d	Indicates the domain name. For example, INC domains will use -d INC.
-r	Indicates a multimedia redirection. For example: To enable sound, use -r sound. To use the local sound card, use -r sound : local. To enable a Udisk, use -r disk:usb=/mnt/usbdevice
-r clipboard:PRIMARYCLIPBOARD	This parameter can be used to directly copy and paste text between the local Linux system and the remote Windows instance. Chinese characters are also supported.
-r disk:sunray=/home/yz16184	This indicates that a directory on the local Linux system is mapped onto the Windows hard disk. This allows you to transfer files without relying on Samba or FTP.

Use the Management Terminal (VNC) to connect to an ECS instance

The operation procedure is the same as that from a local Windows OS.

Connect to a Windows instance using Mac OS X

Download and install the remote desktop client for Mac OS X at <https://itunes.apple.com/us/app/microsoft-remote-desktop/id715768417>.

Follow the in-app directions to complete login.

Connect to a Windows instance using a mobile app

Download and install **Microsoft Remote Desktop** from the iTunes App Store. Follow the in-app directions to complete login.

What if I forget my logon password?

If you forget your instance logon password (not the VNC connection password), see [Reset the password](#).

Instances

Create an instance

Create an instance

You can create instances running Linux, Windows, or a custom system (based on a public image). Detailed information for each is provided as follows:

- To create an instance, refer to [Quick Start](#).
- If you want to clone the operating system, installed applications, and data of an existing instance, refer to [Create an instance using a custom image](#).

Introduction to gn4 type family

For detailed information about various types in gn4 type family, see *Instance generations and type families* in *Product Introduction* of Elastic Compute Service.

Create an instance of gn4 type family

Create an instance of gn4 type family by following the steps described in *Create an instance* in *Quick Start of Elastic Compute Service (ECS)*. When creating the instance, note the following items:

- **Network Type:** Select **VPC** because currently gn4 is available for virtual private cloud network.
- **Instance Type:** Select **GPU computing type gn4** under **Generation III**.
- **Network Bandwidth:** Select the peak bandwidth as needed.

If the Windows 2008 R2 image is used and you want to connect to the instance of gn5 type family, you must enable Internet access of the instance because you cannot connect to an instance of gn5 type family by using the **Management Terminal** on the ECS management console. If you want to allocate an Internet IP address to the instance, you must not set the peak bandwidth to 0 Mbps.

Download and install the GPU driver

Before using an instance of gn4 type family, you need to install the GPU driver for the instance. Follow the steps to download and install the GPU driver.

Go to NVIDIA official website to download the corresponding driver for the operating system and P100 GPU. The download URL:

<http://www.nvidia.com/Download/index.aspx?lang=en-us>.

Manually find the driver for the instance. Set the parameters as follows:

- **Product Type:** Tesla
- **Product Series:** M-Class
- **Product:** M40
- **Operating System:** The corresponding version according to the instance image

If the operating system is not displayed in the drop-down list, click **Show all Operating Systems** at the bottom of the drop-down list.

NVIDIA Driver Downloads

Option 1: Manually find drivers for my NVIDIA products. Help

Product Type: Tesla

Product Series: M-Class

Product: M40

Operating System:
 Show less Product Series
 Windows 10 64-bit
 Windows 7 64-bit
 Windows Vista 64-bit
 Windows XP
 Windows XP 64-bit
 Windows Server 2008
 Windows Server 2008 R2 64
 Windows Server 2012 R2 64
 Windows Server 2016
 Windows Server 2003
 Windows Vista 32-bit
 Windows Server 2003 x64
 VMware vSphere ESXi 5.1
 VMware vSphere ESXi 5.5
 Linux 64-bit
 Linux 64-bit RHEL6
 Linux 64-bit RHEL7
 Linux POWER8 RHEL
 Linux 64-bit Ubuntu 16.04
 Linux POWER8 Ubuntu
 Linux 64-bit Ubuntu 14.04
 Linux 64-bit Fedora 23
 Linux 64-bit SLES 12
 Linux 64-bit Opensuse 13.2
 Show less Operating Systems

CUDA Toolkit: 8.0

Language: English (US)

SEARCH

Click **SEARCH**.

After confirming the information, click **DOWNLOAD**.

Install the GPU driver by following the **ADDITIONAL INFORMATION** on the download page.

Take Linux 64-bit Opensuse 13.2 as an example:

TESLA DRIVER FOR LINUX OPENSUSE 13.2

Version:	375.66
Release Date:	2017.5.9
Operating System:	Linux 64-bit Opensuse 13.2
Language:	English (US)
File Size:	133.05 MB

DOWNLOAD

RELEASE HIGHLIGHTS

ADDITIONAL INFORMATION

Once you accept the download please follow the steps listed below

- i) ``rpm -i nvidia-diag-driver-local-repo-opensuse132-375.66-1.x86_64.rpm``
- ii) ``zypper refresh``
- iii) ``zypper install cuda-drivers``
- iv) ``reboot``

SUPPORTED PRODUCTS

Notes

For Windows 2008 R2 or earlier version, if you go to the **Management Terminal** by clicking **Connect** in the ECS console after the GPU driver is installed, the **Management Terminal** will be stuck at either a black screen or the startup interface. If the instance can access the Internet, you need to connect to the ECS instance remotely using other protocols, such as the Remote Desktop Protocol developed by Microsoft.

The RDP does not support DirectX, OpenGL, and other related applications. Therefore, you need to install VNC services and the client, or other protocols that supports these applications, such as PCOIP or XenDesktop HDX 3D.

Introduction to gn5 type family

For detailed information about various types in gn5 type family, see *Instance generations and type families* in *Product Introduction* of Elastic Compute Service.

Create an instance of gn5 type family

Create an instance of gn5 type family by following the steps described in *Create an instance* in *Quick Start* of Elastic Compute Service (ECS). When creating the instance, note the following items:

- **Region:** Currently, gn5 is available only in the **US East 1 (Virginia)** region.
- **Network Type:** Select **VPC** because gn5 is available for virtual private cloud network.
- **Instance Type:** Select **GPU Compute Type gn5** under **Generation III**.
- **Network Bandwidth:** Select the peak bandwidth as needed.

If the Windows 2008 R2 image is used and you want to connect to the instance of gn5 type family, you must enable Internet access of the instance because you cannot connect to an instance of gn5 type family by using the **Management Terminal** on the ECS management console. If you want to allocate an Internet IP address to the instance, you must not set the peak bandwidth to 0 Mbps.

- **Image:** Select the image as needed.

Download and install the GPU driver

Before using an instance of gn5 type family, you need to install the GPU driver for the instance. Follow the steps to download and install the GPU driver.

Go to NVIDIA official website to download the corresponding driver for the operating

system and P100 GPU. The download URL:
<http://www.nvidia.com/Download/index.aspx?lang=en-us>.

Manually find the driver for the instance. Set the parameters as follows:

- **Product Type:** Tesla
- **Product Series:** P-Series
- **Product:** Tesla P100
- **Operating System:** The corresponding version according to the instance image
 - If the operating system is not displayed in the drop-down list, click **Show all Operating Systems** at the bottom of the drop-down list.
 - If the instance uses a Linux image that is not in the list, select **Linux 64-bit**.

The screenshot shows the 'NVIDIA Driver Downloads' page. Under 'Option 1: Manually find drivers for my NVIDIA products.', the following settings are shown: Product Type: Tesla, Product Series: P-Series, Product: Tesla P100. The Operating System dropdown is open, showing a list of operating systems including Windows 10 64-bit, Windows 7 64-bit, Windows 8.1 64-bit, Windows Server 2008 R2 64, Windows Server 2012 R2 64, Windows Server 2016, Linux 64-bit, Linux 64-bit RHEL6, Linux 64-bit RHEL7, Linux POWER8 RHEL, Linux 64-bit Ubuntu 16.04, Linux POWER8 Ubuntu, Linux 64-bit Ubuntu 14.04, Linux 64-bit Fedora 23, Linux 64-bit SLES 12, and Linux 64-bit Opensuse 13.2. The 'Show less Operating Systems' link is highlighted at the bottom of the list. The CUDA Toolkit is set to 8.0 and the Language is set to English (US). A green 'SEARCH' button is located at the bottom right.

Click **SEARCH**.

After confirming the information, click **DOWNLOAD**.

Install the GPU driver by following the **ADDITIONAL INFORMATION** on the download page.

Take Linux 64-bit Opensuse 13.2 as an example:

TESLA DRIVER FOR LINUX OPENSUSE 13.2

Version:	375.66
Release Date:	2017.5.9
Operating System:	Linux 64-bit Opensuse 13.2
Language:	English (US)
File Size:	133.05 MB

DOWNLOAD

RELEASE HIGHLIGHTS

SUPPORTED PRODUCTS

ADDITIONAL INFORMATION

Once you accept the download please follow the steps listed below

- `'rpm -i nvidia-diag-driver-local-repo-opensuse132-375.66-1.x86_64.rpm'`
- `'zypper refresh'`
- `'zypper install cuda-drivers'`
- `'reboot'`

Notes

For Windows 2008 R2 or earlier version, if you enter the **Management Terminal** by clicking **Connect** in the ECS console after the GPU driver is installed, the **Management Terminal** will be stuck at either a black screen or the startup interface. If the instance can access the Internet, you need to connect to the ECS instance remotely using other protocols, such as the Remote Desktop Protocol developed by Microsoft.

To create an instance that has the same operating system, software applications, and data with those of your existing ECS instance or server, create a copy of the existing ECS instance or server as a custom image, and then use it to create an instance. This method improves the deployment efficiency.

Prerequisites

If the image and the instance are in the same region, you have created a custom image by using one of the following methods:

- Importing a physical image
- Creating a custom image by using an ECS instance
- Creating a custom image by using a snapshot of a system disk

If the custom image and instance are in different regions, you have copied the custom image to the target region.

Operating procedure

Log on to the ECS console.

In the left-side navigation pane, click **Instances**.

In the top-right corner of the page, click **Create Instance**.

On the purchase page,

- Select the desired billing method, target region, instance type, network type, and other parameters. For details, see the instance creation process in the **Quick Start**.
- Select a custom image.

Note: If the selected custom image contains more than one data disk snapshot, an equal number of data disks are automatically created. By default, the size of each data disk is equal to that of the source snapshot. You are only allowed to increase, but not decrease, the size of a data disk.

Click **Buy Now**.

Change the operating system

Use the management console to convert the instance OS to your preferred OS. For details, see **Change the system disk to your custom image** or **Change the system disk to a public image**.

Note: Regions outside of mainland China do not currently support transition between Linux and Windows OSs. If your instance is hosted in one of these regions, you are not allowed to change the operating system between Windows and Linux. You can only change the version of Windows OS, or replace one Linux OS with another Linux OS.

In this article, we describe how to upgrade the CPU and memory specifications and peak bandwidth of your **Subscription** ECS instance.

If your **Pay-As-You-Go** ECS instance is bound to an elastic IP (EIP) address, you can use the **Change Configuration** function in the ECS console to modify the peak bandwidth on the EIP address. For more information about EIP address, see the **EIP addresses documents**.

Scenarios

Choose to upgrade configurations if you have the following requirements:

Upgrading the specifications of CPU and memory of your instance.

If you want to upgrade the specifications of an instance, only the higher specifications in the same generation are available. After you have applied the changes, restart your instance in the ECS console or using the **RebootInstance** interface.

Increasing the bandwidth for Internet access in either of the following scenarios:

You did not buy bandwidth (0 Mbps) when you were creating an instance of the Classic network type, and now you want to increase the bandwidth to provide Internet access. After you have applied the changes, restart your instance in the ECS console or using the **RebootInstance** interface.

You bought bandwidth for Internet access, and now you want to increase the bandwidth to meet your growing business requirements. After you have applied the changes, the new bandwidth takes effect immediately.

After you upgrade your configurations, the Internet and the intranet IP addresses remain unchanged.

Procedure

This section introduces how to upgrade configurations in the ECS console. If you want to upgrade configurations by using the API, [open a ticket](#) to apply for permission in the whitelist.

To upgrade the configurations, follow these steps:

Log on to the ECS Console.

In the left-side navigation pane, click **Instances**.

Select a region.

In the instance list, find your instance, and in the **Action** column, click **Change Configuration**.

On the **Change Configuration Guide** dialog box, choose **Upgrade configuration**.

Click **Continue** to go to the **Upgrade Configuration** page.

Select a new instance type, or set a new value for the bandwidth. Click **Pay** to complete the order.

The new configuration takes effect immediately in most cases. However, if you changed the instance type, or if you increased the bandwidth from 0 Mbps for your instance of the Classic network type, restart the instance in the console or using the `RebootInstance` interface.

- For Linux instances, the default username is **root**.
- For Windows instances, the default username is **Administrator**.

To reset the instance password, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Instances**.

Select a region.

Select your desired instance. You may select multiple instances with identical operating statuses.

Click **More > Reset Password**.

Enter a new password in the displayed dialog box. Click **Submit**.

Click **OK**.

Select the instance on which the password was changed and click **Restart**.

Note: The new password will only take effect after the instance is **restarted** through the console. Restarting directly within the instance will not apply the new password.

Click **OK** in the displayed dialog box to restart the instance.

Start, view, or stop an instance

This section describes how to start, view, and stop an instance.

Start an instance

Note: Only instances in **Stopped** status can be started.

To start an instance, perform the following:

1. Log on to the **ECS console**.
2. In the left-side navigation pane, click **Instances**.
3. Select a region.
4. Select the desired instance. You can select multiple instances, as long as they are all in **Stopped** status.
5. Click **Start** at the bottom of the page.

View an instance

Use the console to view the following instance information:

- Quantity and operating statuses of instances in each region.
- Basic configuration, payment, and monitoring information.
- Disks.
- Snapshots.
- Related security groups.

To view instances, perform the following:

1. Log on to the **ECS console**.
2. On the overview page, you can view the operating statuses of ECS instances in all regions.
3. In the left-side navigation pane, click **Instances**.
4. Select a region, and click the name of the instance you want to view.
(Optional) You can also search for the name of the instance you want to view.

You can then view instance details, including its:

- Region.
- Zone.
- Configuration specification.
- Payment status.
- CPU.
- Network usage.

Note: You can also view and manage the instance's disks, snapshots, and security group

information.

Stop an instance

On the console, you can stop an instance in a similar manner to an actual server.

Note:

- Stopping an instance may disrupt your business traffic. Proceed with caution.
- Only instances in **Running** status can be stopped.

To stop an instance, perform the following:

1. Log on to the ECS console.
2. In the left-side navigation pane, click **Instances**.
3. Select a region.
4. Select the desired instance. You can select multiple instances, as long as they are all in **Running** status.
5. Click **Stop**.
6. Select **Stop** in the displayed dialog box. Click **OK**.

Restart an instance

Instances can be restarted from within or through the management console.

Note: Restarting an instance may disrupt your business traffic. Proceed with caution.

Only instances in the **Running** status can be restarted.

To restart an instance, perform the following:

Log on to the ECS console.

Click **Instances** in the left side navigation bar.

Select your desired region.

Select the desired instance. You can select multiple instances, as long as they are all in the **Running** status.

Click **Restart**.

In the displayed dialog box, click **Restart**, and then click **OK**.

Release an instance

If you no longer require a Pay-As-You-Go instance, we recommend that you release it immediately. Charges will continue when the instance is stopped but not released.

Release instances by performing either of the following:

- **Release Now:** Immediately releases the Pay-As-You-Go instance.
- **Timed Release:** Arranges a release plan for your Pay-As-You-Go instances by defining future release times. These times are definable to the hour. Applying new time settings will overwrite previous ones.

Note: Instances can be released every half hour or every hour, however, the system will stop billing according to your specified release time.

To perform either release type, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Instances**.

Select a region.

Select the instance to be released, and then click **Release Instance**.

In the displayed dialog box, select your preferred release action.

Note: If you select **Timed Release**, first specify whether it is to be auto released, and then specify the release date and time.

Click **Next > OK**.

Disable auto release

If wish to cancel the automatic release of Pay-As-You-Go instances, you can disable the set auto release function as follows:

Log on to the ECS console.

In the left-side navigation pane, click **Instances**.

Select a region.

Select the required instance.

Select **More > Release Instance**.

In the displayed dialog box, select **Timed Release**.

Disable the **Auto-release Setting** option.

Click **Next > OK**.

Add an instance to a security group

You can add an instance to a security group using the ECS Management Console. One ECS instance can be added to up to five security groups. After adding the instance to a security group, the security group rules will automatically be applied to the instance.

To add an instance to a security group, perform the following:

Log on to the ECS console.

Click **Instances** in the left navigation bar.

Select your desired region.

Select the desired instance. Click the instance name or corresponding **Manage** button.

Click **Security Groups** in the left navigation bar.

Click **Add Security Group**. In the displayed dialog box, select the appropriate security group.

Click **OK**.

Remove an instance from a security group

You can remove instances from security groups. Note that an instance must be in at least two security group for this action to be performed, and you have done enough test before this operation to avoid any intranet communication error between instances.

To remove an instance from a security group, perform the following:

Log on to the ECS console.

Click **Instances** in the left navigation bar.

Select your desired region.

Select the desired instance. Click the instance name or corresponding **Manage** button.

Click **Security Groups** in the left navigation bar.

Select the security group to remove from and click **Remove**.

Click **OK**.

Currently you are only allowed to use the RAM role of ECS instances using OpenAPI. This feature will be available soon in the console.

What is the RAM role of an ECS instance?

The RAM role of an ECS instance, also known as the instance RAM role, is designed to enable ECS instances to adopt a role with certain permissions to grant some access permissions to instances.

The instance RAM role allows you to associate a RAM role to an ECS instance to access other cloud products using the STS temporary credential, which is updated regularly, within the instance. Using RAM roles keeps Access Key secure and achieves precise control and management over permissions.

Why do we need instance RAM roles?

Applications deployed on ECS instances usually use the Access Key (AccessKeyId/AccessKeySecret) of the cloud account or RAM user account to access APIs of various Alibaba Cloud products. To respond

to calls, the Access Key is generally fixed in the instance, such as written in the configuration file. However, this approach has the following shortcomings:

The Access Key of a cloud account has high permission, so we do not recommend that you do not use it.

The method of writing the Access Key to disk is not a good practice to keep the Access Key secure and carries a high risk of exposure.

All the instances and related images must be updated when you want to change the Access Key.

With the instance RAM role, you can:

Use an STS temporary credential to access other Alibaba Cloud services safely on the ECS instance.

Grant roles with different authorization policies to different instances to give them different access permissions to different cloud resources.

Save time and effort by changing the role's authorization policies rather than manually changing the Access Key for the instance.

Use instance RAM roles

The following are some considerations when you want to use instance RAM roles:

Only instances of VPC network type are eligible to use the instance RAM role.

Only one RAM role can be attached to an ECS instance.

To support RAM sub-account using the instance RAM role, grant the PassRole permission when creating the authorization policies. The following is an example of a authorization policy.

```
{
  "Version": "2016-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
```

```
"ecs:CreateInstance",
"ecs: AttachInstanceRamRole",
"ecs: DetachInstanceRAMRole"
],
"Resource": "*"
},
{
  "Effect": "Allow",
  "Action": "ram:PassRole",
  "Resource": "*"
}
]
```

Note: The sub-account must be authorized with the PassRole permission for a role to prevent permissions of instance roles from being abused by unauthorized sub-accounts.

You need to get the temporary authorization credential of the instance role using Meta Data.

Get the STS temporary credential for applications deployed on the instance by running the command: `http://100.100.100.200/latest/meta-data/ram/security-credentials/role-name`. The temporary credential enables you to perform the actions on the resources authorized by the role. The temporary credential is automatically updated regularly.

Search for the security certificate for the instance RAM role named **ossaccess** by running the command `curl http://100.100.100.200/latest/meta-data/Ram/security-credentials/ossaccess`. Here is the example output.

```
{
  "Code" : "Success",
  "LastUpdated" : "2012-04-26T16:39:16Z",
  "AccessKeyId" : "AKIAIOSFODNN7EXAMPLE",
  "AccessKeySecret" : "wJalrXUtnFEMI/K7MDENG/bPxrFiCYEXAMPLEKEY",
  "SecurityToken" : "token",
  "Expiration" : "2017-05-27T22:39:16Z"
}
```

Instance RAM role related APIs

The APIs related to instance RAM roles are as follows:

To query instance RAM role information: **DescribeInstanceRamRole**.

To attach an instance RAM role to an ECS instance: **AttachInstanceRamRole**.

To detach an instance RAM role from an ECS instance: `DetachInstanceRamRole`.

Instance custom data and instance meta data

Disks

Create a cloud disk

Cloud disks, also known as data disks, can be purchased from the ECS Management Console. Each user account can own up to 250 cloud disks simultaneously. Up to 16 data disks can be attached to any single ECS instance, with a maximum capacity of 32768 GB per data disk.

To purchase a cloud disk, perform the following:

Log on to the ECS console.

Click **Disks** in the left navigation bar.

Select your desired region, then click **Create Cloud Disk** in the top-right corner of the page.

Select a region and zone.

Note: A cloud disk can be attached to only a server in the same zone of the same region. Cloud disks do not support cross-regional functionality.

Select the cloud disk type, size, and quantity. Click **Buy Now** on the right side of the page.

Next step for Linux

For Linux instances, cloud disks must be attached, partitioned, and formatted before they can be displayed and used in the system.

- For details on attaching a data disk, see [Attach a data disk](#).
- For details on formatting partitions and mounting new partitions to an attached data disk, see [Format and mount a data disk](#).

Next step for Windows

For Windows instances, you must attach and format a cloud disk before using it.

- For details on attaching a data disk, see [Attach a data disk](#).
- For details on formatting an attached data disk, see [Format a data disk](#).

Create a cloud disk from a snapshot

If you need to access data from a snapshot, but do not want to roll back your disk to the snapshot, you can create a cloud disk from the snapshot and access the data from the disk.

For example, if you cannot start your instance because of a system disk failure, you can take a snapshot of the system disk, and use it to create a cloud disk. Then, you can copy and analyze its data.

Notice: Unlike a new ultra cloud disk or SSD cloud disk, which can achieve its maximum performance once it is created, a cloud disk created from a snapshot may provide reduced performance during the first access to it because during the access some time is needed to get the data from OSS and write to the disk.

We recommend that you read all data from the cloud disk created from a snapshot to avoid the reduced performance during the first access after creation.

Operating procedure

To create a cloud disk from a snapshot, perform the following:

Log on to the ECS console.

Click **Disks** in the left navigation bar.

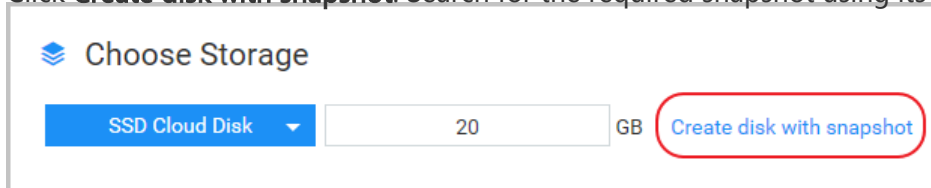
Click **Create Cloud Disk** in the top-right corner of the page.

Select a region and zone.

Note: A cloud disk can be attached to only a server in the same zone of the same region.

Cloud disks do not support cross-regional functionality.

Click **Create disk with snapshot**. Search for the required snapshot using its snapshot ID.



Choose Storage

SSD Cloud Disk ▼ 20 GB Create disk with snapshot

Confirm the order. The created disk will be displayed in the disk list.

Click **Attach Disk**. Enter the instance ID to which the disk is to be attached. Select the device name. You can now log on to the instance to view the disk data.

Note:

- For Windows instances, the disk will be displayed after login.
- For Linux instances, the disk must be mounted to be displayed.

Attach a data disk

The following disk types can be attached:

- Basic Cloud Disks
- Ultra Cloud Disks
- SSD Cloud Disks that serve as data disks.

Before attaching a cloud disk, an instance must meet the following requirements:

- The instance is in **Stopped** status.
- The security control marker is not Locked.
- The instance is not in payment arrears.

To attach a cloud disk, the following conditions must be adhered to:

- The cloud disk must be in available status.
- A single instance can have one system and up to 16 data disks attached. This includes disks of all types.
- A cloud disk can be attached to an instance of the same zone only.
- A cloud disk can be attached to only one instance at a time. Attachment to multiple instances is not supported.
- A cloud disk can be attached to any instance of the same region and zone. Both Subscription and Pay-As-You-Go instances are supported.
- When a cloud disk is acting as the system disk of an instance, it cannot be separately

attached.

You can attach a disk through either of the following:

Instances Menu

Recommended if you require multiple disks to a single instance.

Disks Menu

Recommended if you require disks be attached to different instances.

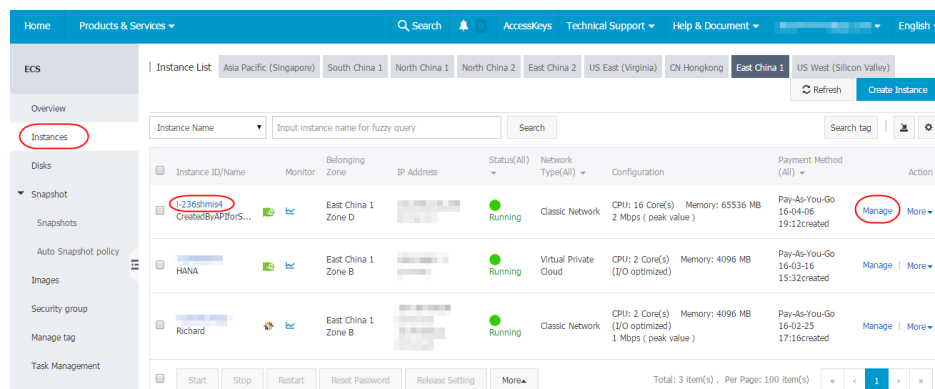
From the Instances menu

Log on to the ECS console.

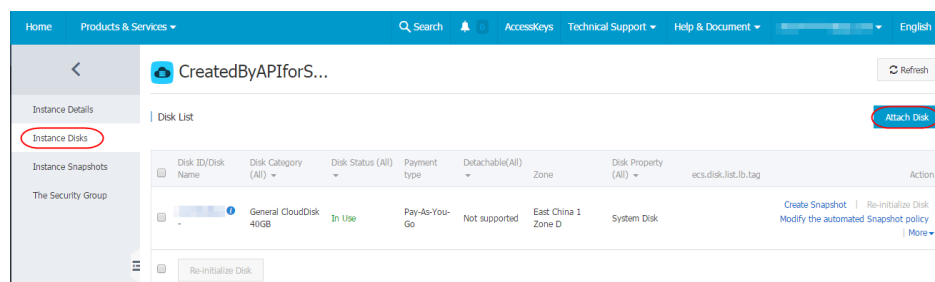
Click **Instances** in the left navigation bar.

Select your desired region.

Click the name of the instance for attachment or click **Manage**.



Click **Instance Disks** in the left navigation bar. The disks already attached to the instance is displayed.



Click **Attach Disk** on the right side of the page. Select **Available Devices** and **Target Disk** to attach the disk.

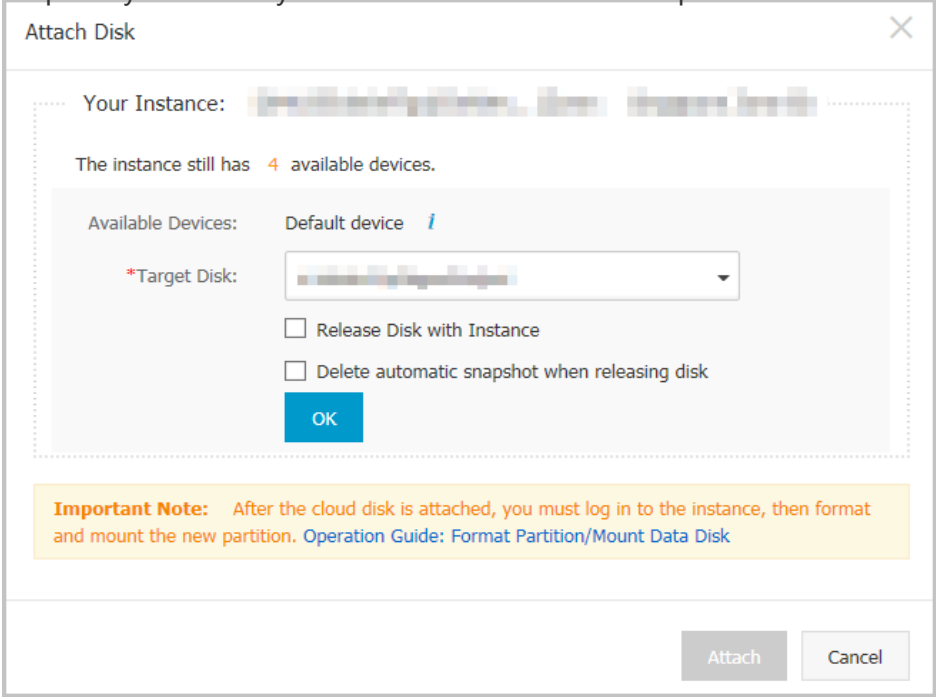
(Optional) Set whether disks are to be released with instances and whether snapshots are to be deleted with disks.

- **Release disk with instance**

When you release the instance, the disk will be released together.

- **Delete automatic snapshots when releasing disk**

When you release the disk, all auto snapshots will be deleted. However, the snapshot you manually created will be retained. This option is not recommended.



The screenshot shows the 'Attach Disk' dialog box. At the top, it says 'Your Instance:' followed by a blurred instance ID. Below that, it states 'The instance still has 4 available devices.' The 'Available Devices:' section shows a 'Default device' with a blue 'i' icon. The '*Target Disk:' section has a dropdown menu with a blurred disk ID. Below the dropdown are two checkboxes: 'Release Disk with Instance' and 'Delete automatic snapshot when releasing disk', both of which are currently unchecked. An 'OK' button is located below the checkboxes. At the bottom of the dialog, there is an 'Important Note' box with orange text: 'After the cloud disk is attached, you must log in to the instance, then format and mount the new partition. Operation Guide: [Format Partition/Mount Data Disk](#)'. At the very bottom of the dialog are 'Attach' and 'Cancel' buttons.

After attaching a disk, you need to log on to the instance to format disk partitions and attach new partitions. For details, refer to **Next step** at the bottom of this section.

From the Disks menu

Log on to the ECS console.

Click **Disks** in the left navigation bar.

Select your desired region.

Find the disk to attach. The disk status must be **Available**.

Click **More > Attach**. Select the target instance and release action:

- **Release disk with instance**
When you release the instance, the disk will be released together.
- **Delete automatic snapshots when releasing disk**
When you release the disk, all auto snapshots will be deleted. However, the snapshot you manually created will be retained. This option is not recommended.

After attaching a disk, you need to log on to the instance to format disk partitions and attach new partitions. For details, refer to **Next step** at the bottom of this section.

Next step for Linux

After attaching a disk, you must log on to the instance to format disk partitions and mount new partitions. For detailed instructions, see [Format and mount a data disk](#).

Next step for Windows

After attaching a disk, you must log on to the instance to format disk partitions. For detailed instructions, see [Format a data disk](#).

ECS supports the detachment of Basic Cloud Disks, Ultra Cloud Disks, and SSD Cloud Disks serving as data disks. System disks cannot be detached. Detach disks on the **Instances** page or the **Disks** page.

Notes

Only the data disks in the **In Use** status can be detached.

Ensure the following based on the operating system of the instance:

- For a Linux instance, log on to the instance and run the `umount` command to unmount the data disk. After the command runs, log on to the ECS console and detach the disk.
- For a Windows instance, stop read and write operations on all file systems of the disk to ensure data integrity. Otherwise, data being read and written will be lost.

Procedures

On the Instances page

Log on to the ECS console.

In the left-side navigation pane, click **Instances**.

Select a region.

Click the instance ID go to the **Instance Details** page.

In the left-side navigation pane, click **Instance Disks**, and you can see all the disks, both system disks and data disks, associated with the instance.

Find the disk to be detached and select **More > Detach**.

In the **Detach Disk** dialog box, read the note and click **Confirm Detaching**.

You have successfully detached the data disk from the instance.

On the Disks page

Log on to the ECS console.

In the left-side navigation pane, click **Disks**.

Select a region.

Find the disk to be detached and select **More > Detach**.

In the **Detach Disk** dialog box, read the note and click **Confirm Detaching**.

You have successfully detached the data disk from the instance.

After the procedures

You have to pay for the data disk even though it is not attached to an instance. So, if you do not need the data disk any more, release it.

By changing the system disk, you can change the operating system to your custom image. For instructions about how to change the OS to a public image, refer to [Change the system disk \(public image\)](#).

Changing the system disk will not change your instance IP address.

Note: Regions that are not in mainland China do not support replacement between Linux and Windows. A Linux or Windows can be only replaced by a different version of the same operating system type.

Warning:

- Stopping an instance may disrupt traffic.
- Redeploying the runtime environment on the new system disk is required once it is stopped and this may disrupt traffic.
- Automatic snapshots and data from your original system disk will be lost once the system disk is replaced. Ensure that all necessary data has been backed up in advance. If you want to retain auto snapshots, see [disable releasing auto snapshots with disk](#).
- Manually created snapshots from the original system disk are retained but cannot be used to roll back the new system disk because the disk ID is changed. You can use the retained snapshots to create custom images.
- The original system disk will be deleted once the system disk is replaced.
- Ensure that the system disk has at least 1 GB of free space. Otherwise, the instance may fail to start after you change the system disk.

Procedure of changing the system disk

A complete procedure of changing the system disk includes the following steps:

- Step 1. Back up the current system disk by creating a snapshot.
- Step 2. Create an image from the snapshot.
- Step 3. Change the system disk and selecting a new OS.
- Step 4. Set auto snapshot policies for the new system disk.

To retain enough snapshot quota for the auto snapshot policy of the new disk, please delete unwanted snapshots before proceeding. If you wish to change the OS and do not need to retain the data from the current system disk, you can proceed directly to Step 3 of this section.

Step 1: Back up the current system disk by creating a snapshot

Note:

- Skip this step if you do not want to retain the data in the system disk.
- Do not create the snapshot during busy hours.
- It takes about 40 minutes to create a snapshot of 40 GB. Ensure you reserve sufficient time for it.
- Make sure the system disk has at least 1 GB free space. Otherwise, the instance may fail to start after you change the system disk.

Log on to the ECS console.

Click **Instances** in the left navigation bar.

Select your desired region.

Select the instance containing the system disk with the OS to be changed.

Click **Instance Disks** on the left navigation bar.

Select the system disk with the OS to be changed, and then click **Create Snapshot**.

Enter a name for the snapshot.

Click **Instance Snapshots** on the left navigation bar to check the progress and status of the snapshot.

Step 2: Create an image from the snapshot

Note:

- Skip this step if you do not want to retain the data in the system disk.
- If you do want to retain the data in the current system disk, you need to create an image to replicate the data in the system disk.
- Make sure the system disk has at least 1 GB free space. Otherwise, the instance may fail to start after you change the system disk.

Select the snapshot created in Step 1 and click **Create Custom Image**.

Enter a name and description for the image.

Return to the navigation bar, and then click **Images**.

You can now check the process and status of the new image.

Step 3: Change the system disk

To change a system disk:

Log on to the ECS console.

Click **Instances** in the left navigation bar.

Select your desired region.

Stop the instance before changing the system disk. To do this, select the instance for which you wish to replace the system disk in the instance list and click **Stop**.

Once the instance is stopped, click **More > Change System Disk**.

Click **Custom Image** and select the image created in Step 2.

Click **Confirm Change**. Any expenditure that may have occurred will need to be paid at this time.

Step 4: Set auto snapshot policies for the new system disk

After changing the system disk, any auto snapshot policies you have set will no longer work for the new system disk, because the disk ID has changed. In this scenario, you must configure auto snapshot policies for the new system disk. For more information, see [Set auto snapshot policies for disks](#).

Step 5: Attach data disks (for Linux instances only)

For Linux instances, you must re-attach the data disks after changing the system disk, but you do not need to partition them. For more information, see [Attach a data disk](#).

Change the system disk to a public image

By changing the system disk, you can change the operating system to a public image, for example, from Windows Server 2003 to Windows 2012.

Note: Regions that are not in mainland China do not support replacement between Linux and Windows. A Linux or Windows can be only replaced by a different version of the same operating system type.

Considerations for changing the system disk

Risks

- This operation requires you to stop your instance, which means interruption of your

business. Therefore, perform this operation with caution.

- After replacement, you must redeploy the runtime environment on the new system disk. There is a possibility of a long interruption of your business. Therefore, perform this operation with caution.
- Replacing the system disk will result in the loss of the automatic snapshots and data on your original system disk. Make necessary backup in advance.

Note:

- To retain enough snapshot quota for the auto snapshot policy of the new disk, you can delete unwanted snapshots.
- Changing the system disk will not change your instance IP address.
- Manually created snapshots are retained after the replacement. However, since the disk ID is changed, you can no longer use the manually created snapshots on the original system disk to roll back the new system disk. The retained snapshots can be used to create custom images.
- After the system disk is replaced, the original system disk will be deleted.

Retain automatic snapshots

By default, the automatic snapshots will be released along with the disk. If you want to keep the automatic snapshots, see [Configure releasing auto snapshots together with disk](#).

Procedure of changing the system disk

A complete procedure of changing the system disk includes the following steps:

1. Create a snapshot for the current system disk.
2. Change the system disk.
3. Set automatic snapshot policies for the new system disk.
4. Attach data disks (for Linux instances only).

Step 1: Create a snapshot based on the current system disk

Skip this step if you do not want to retain the data on the system disk.

Do not create the snapshot during busy hours. It takes about 40 minutes to create a snapshot of 40GB. Ensure you reserve sufficient time for it.

Note: Make sure the system disk has at least 1GB free space; otherwise, the instance may fail to start after you change the system disk.

Log on to the ECS console.

Click **Instances** on the left navigation bar. Then click the region.

Click the instance for which you want to change the system disk.

Click **Instance Disks** on the left navigation bar.

Find the system disk you want to change, and then click **Create Snapshot**.

Enter a name for the snapshot.

Click **Snapshots** on the left navigation bar, you can check the progress and status of the snapshot.

Step 2: Change the system disk

To change a system disk:

Log on to the ECS console.

Click **Instances** on the left navigation bar. Then, select a region at the top of the page.

Stop the instance before changing the system disk. In the instance list, select the instance for system disk replacement and click **Stop** at the bottom.

After the instance is stopped, on the right end of the instance table, click **More > Change System Disk**.

A dialog box showing the considerations is displayed. Read the considerations carefully, and then confirm the operation.

Select a public image.

Set the password for Administrator or root.

Click **Pay Now**. Pay for the expenditure that incurred, if any.

An important message is prompted. Read it carefully. After confirming everything is correct, click **OK**.

Step 3: Configure the automatic snapshot policies

After changing the system disk, the automatic snapshot policies you have set will no longer work for the new system disk, because the disk ID has changed. In this case, you need to configure automatic snapshot policies for the new system disk. For more information, see [Set automatic snapshot policies for disks](#).

Step 4: Attach data disks (for Linux instances only)

For Linux instances, after you have changed the system disk, you need to attach the data disks again, but you don't need to partition them. For more information, see [Attach a data disk](#).

Resize cloud disks

If you want to resize a data disk attached to a Windows instance, refer to [Windows _ Resize a data disk](#).

As your business grows, the capacity of your data disks may not be able to meet your need for data storage, and then you can use the **Resize Disk** function to resize your data disk.

Notice:

- We recommend that you manually **create snapshots** to back up your data before resizing your data disk.
- You can resize a data disk when the data disk is either in the **Available** status or in the **In Use** status.
- If a snapshot is being created for a data disk, resizing the data disk is not allowed.
- You can resize data disks, but not system disks or ephemeral disks.
- You can resize data disks that are attached to instances only when the instance is in the **Running** or **Stopped** status. You must restart the instance to make the disk resizing take effect, which will cause your instance to stop working and your business to be interrupted, so please proceed with caution.

This document takes a data disk of the ultra cloud disk type and an ECS instance running 64-bit CentOS 7.3 as an example to describe how to resize a data disk and make the extended capacity available.

To resize a data disk, follow these steps:

Step 1. Resize a data disk in the ECS console

Step 2. Log on to the instance to enable the extended storage space

Step 1. Resize a data disk in the ECS console

To resize a data disk in the ECS console, follow these steps:

Log on to the ECS console.

In the left-side navigation pane, select **Block Storage** > **Cloud Disks**, and select a region.

If the data disk you want to resize has been attached to an instance, in the left-side navigation pane, click the **Instances**, find the corresponding instance, go to the instance details page, and click **Instance Disks**.

Find the disk to be resized, and in the **Action** column, select **More** > **Resize Disk**.

On the **Resize Disk** page, set **Capacity after resizing** (In this example, 30 GB). The capacity after resizing must be larger than the current capacity.

When the cost is calculated, click **Confirm to resize**.

(Optional) If your data disk is attached to an instance, restart the instance to make the resizing take effect.

After the disk resizing is completed in the console,

If the data disk is attached to an instance, log on to the instance to enable the extended storage space.

If the data disk is not attached to an instance, attach the disk to an instance in the console first, and then proceed depending on the data disk:

- If it is a brand new data disk, format and mount the data disk.
- If it has been formatted and partitioned, log on to the instance to enable the extended storage space.

Step 2. Log on to the instance to enable the extended storage space

In this example, assume that the data disk is attached to a Linux instance running the 64-bit CentOS 7.3. The data disk before resizing has only one primary partition (/dev/vdb1, ext4 file system), the mount point of the file system is /resizetest, and after the resizing is completed, the data disk still has only one primary partition.

Log on to the instance.

Run the `umount [file system name]` command to unmount the primary partition.

```
umount /dev/vdb1
```

Run the `df -h` command to see whether the unmounting is successful, and if you do not see the /dev/vdb1 information, it means that the unmounting is successful. The following is the sample output.

```
[root@iXXXXXX ~]# df -h
Filesystem Size Used Avail Use% Mounted on
/dev/vda1 40G 1.5G 36G 4% /
devtmpfs 487M 0 487M 0% /dev
tmpfs 497M 0 497M 0% /dev/shm
tmpfs 497M 312K 496M 1% /run
tmpfs 497M 0 497M 0% /sys/fs/cgroup
tmpfs 100M 0 100M 0% /run/user/0
```

Run the `fdisk` command to delete the original partition and create a new partition:

If you used the `parted` tool to manipulate partitions, you cannot use it in alternation with `fdisk`. Otherwise, this will result in inconsistent first sector of the partition. Instructions on how to use the `parted` tool can be found [here](#).

- i. Run the `fdisk -l` command to list the partition details and record the final size of the data disk and its first sector before resizing.
- ii. Run the `fdisk [device name of data disk]` command to use the `fdisk` tool. In this example, the device name is /dev/vdb.
- iii. Type `d` and press the Enter key to delete the original partition.

Deleting a partition will not cause loss of data in the data disk.

- iv. Type `n` and press the Enter key to start creating a new partition.
- v. Type `p` and press the Enter key to create a primary partition. In this example, you are creating a single-partition data disk, so it is sufficient to create one primary partition.

If you want to create more than four partitions, create at least one extended partition, that is, type e.

- vi. Type the partition number and press the Enter key. In this example, only one partition is to be created, so type 1.
- vii. Type a number for the First sector: For data consistency, the number for the First sector must be identical with that of the original partition. In this example, press the Enter key to use the default value of 1.

If you find that the First sector is not identical with the recorded one, you may have used the parted tool for partitioning. In that case, stop the current fdisk operation and use **parted** to start over again.

- viii. Type a number for the last sector: Because only one partition is to be created in this example, press the Enter key to use the default value.
- ix. Type wq and press the Enter key to start partitioning.

```
[root@iXXXXXX ~]# fdisk /dev/vdb
Welcome to fdisk (util-linux 2.23.2).
Changes will remain in memory only, until you decide to write them.
Be careful before using the write command.
Command (m for help): d
Selected partition 1
Partition 1 is deleted
Command (m for help): n
Partition type:
 p primary (0 primary, 0 extended, 4 free)
 e extended
Select (default p):
Using default response p
Partition number (1-4, default 1):
First sector (2048-62914559, default 2048):
Using default value 2048
Last sector, +sectors or +size{K,M,G} (2048-62914559, default 62914559):
Using default value 62914559
Partition 1 of type Linux and of size 30 GiB is set
Command (m for help): wq
The partition table has been altered!
Calling ioctl() to re-read partition table.
Syncing disks.
```

If you are using the parted tool, type p to list the current partition details. If any partition is displayed, use rm + serial number to delete the original partition table, then run the unit s command to specify the start unit, calculated by the number of sectors, and finally run the mkpart command to create it, as shown in the following figure.

```

[root@: ~]# parted /dev/xvdb
GNU Parted 3.1
Using /dev/xvdb
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) p
Model: Xen Virtual Block Device (xvd)
Disk /dev/xvdb: 5369MB
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Disk Flags:

Number  Start  End  Size  File system  Name  Flags
(parted) unit s
(parted) mkpart primary ext3 56 5369MB
Warning: The resulting partition is not properly aligned for best performance.
Ignore/Cancel? i
(parted) p
Model: Xen Virtual Block Device (xvd)
Disk /dev/xvdb: 10485760s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Disk Flags:

Number  Start  End  Size  File system  Name  Flags
1       56s   10485726s  10485671s  ext3        primary

```

(Optional) For some operating systems, the file system may be automatically mounted to the mount point again after partitioning. We recommend that you run the `df -h` command to check the disk storage space and usage. If the file system is mounted again, run the `umount [file system name]` to unmount the file system once again.

Check the file system and resize the file system.

```

e2fsck -f /dev/vdb1 # check the file system
resize2fs /dev/vdb1 # resize the file system

```

Note:

- Running the `e2fsck` command is time-consuming because the system needs to check and revise the file system metadata during that process, so be patient.
- Properly running the `e2fsck` command and the `resize2fs` command will not cause any loss of the original data.

The following is the sample output.

```

[root@iXXXXXX ~]# e2fsck -f /dev/vdb1
e2fsck 1.42.9 (28-Dec-2013)
Pass 1: Checking inodes, blocks, and sizes
Pass 2: Checking directory structure
Pass 3: Checking directory connectivity
Pass 4: Checking reference counts
Pass 5: Checking group summary information
/dev/vdb1: 11/1835008 files (0.0% non-contiguous), 159218/7339776 blocks
[root@iXXXXXX ~]# resize2fs /dev/vdb1

```



```
resize2fs 1.42.9 (28-Dec-2013)
Resizing the filesystem on /dev/vdb1 to 7864064 (4k) blocks.
The filesystem on /dev/vdb1 is now 7864064 blocks long.
```

Mount the resized file system to the original mount point (in this example, /resizetest).

```
mount /dev/vdb1 /resizetest
```

Run the `df -h` command to check disk storage space and usage. If the information about the resized file system is displayed, the mounting is successful and the resized file system is ready for use.

After the mounting is completed, you can use the resized file system without restarting the instance.

The following is the sample output.

```
[root@iXXXXXX ~]# df -h
Filesystem Size Used Avail Use% Mounted on
/dev/vda1 40G 1.5G 36G 4% /
devtmpfs 487M 0 487M 0% /dev
tmpfs 497M 0 497M 0% /dev/shm
tmpfs 497M 312K 496M 1% /run
tmpfs 497M 0 497M 0% /sys/fs/cgroup
tmpfs 100M 0 100M 0% /run/user/0
/dev/vdb1 30G 44M 28G 1% /resizetest
```

If you want to resize a data disk attached to a Linux instance, refer to [Linux _ Resize a data disk](#).

As your business grows, the capacity of your data disks may not be able to meet your need for data storage, and then you can use the **Resize Disk** function to resize your data disks.

Notice:

- We recommend that you manually **create snapshots** to back up your data before resizing your data disk.
- You can resize a data disk when the data disk is either in the **Available** status or in the **In Use** status.
- If a snapshot is being created for a data disk, resizing the data disk is not allowed.
- You can resize data disks, but not system disks or ephemeral disks.
- You can resize data disks that are attached to instances only when the instance is in the **Running** or **Stopped** status. You must restart the instance to make the disk resizing take

effect, which will cause your instance to stop working and your business to be interrupted, so please proceed with caution.

This document takes a data disk of the ultra cloud disk type and an ECS instance running 64-bit Windows Server 2008 R2 Enterprise Edition as an example to describe how to resize a data disk and make the extended capacity available. In the example, the current disk capacity is 20 GB, and we will resize it to 24 GB.

To resize a data disk, follow these steps:

Step 1. Resize a data disk in the ECS console

Step 2. Log on to the instance to enable the extended storage space

Step 1. Resize a data disk in the ECS console

To resize a data disk in the ECS console, follow these steps:

Log on to the ECS console.

In the left-side navigation pane, select **Block Storage** > **Cloud Disks**, and select a region.

If the data disk you want to resize has been attached to an instance, in the left-side navigation pane, click **Instances**, find the corresponding instance, go to the instance details page, and click **Instance Disks**.

Find the disk to be resized, and in the **Action** column, select **More** > **Resize Disk**.

On the **Resize Disk** page, set **Capacity after resizing** (In this example, 24 GB). The capacity after resizing must be larger than the current capacity.

When the cost is calculated, click **Confirm to resize**.

(Optional) If your data disk is attached to an instance, restart the instance to make the resizing take effect.

After the disk resizing is completed in the console,

If the data disk is attached to an instance, log on to the instance to enable the extended storage space.

If the data disk is not attached to an instance, attach the disk to an instance in the console first, and then proceed depending on the data disk:

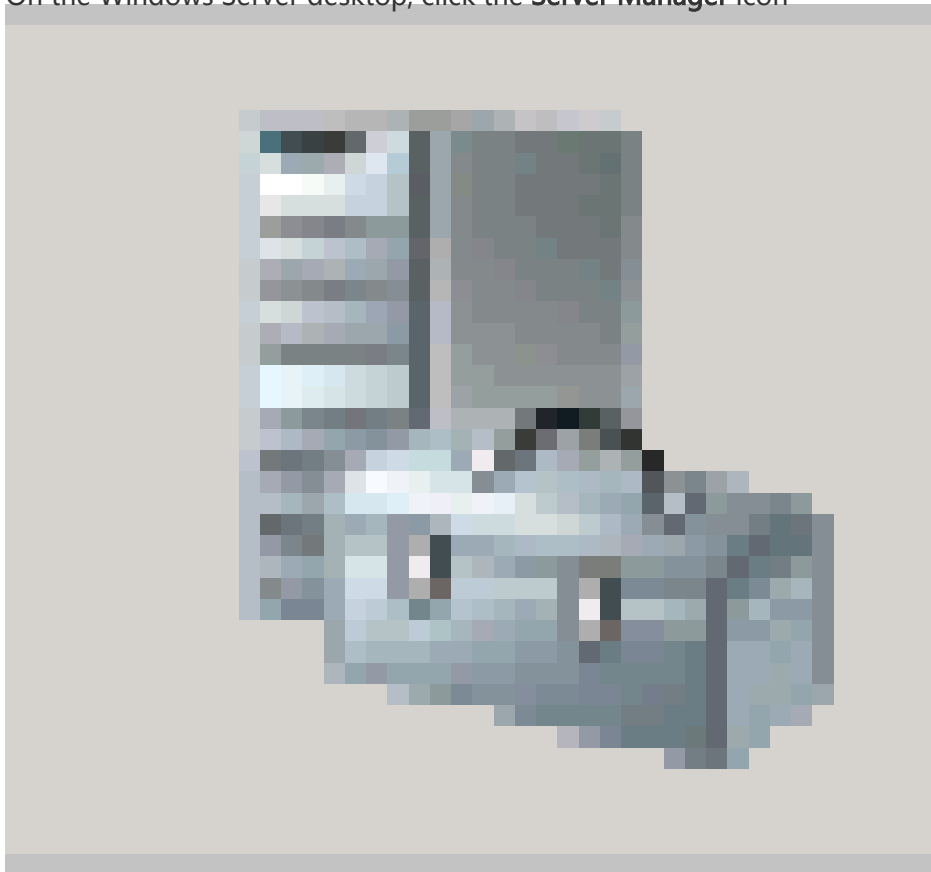
- If it is a brand new data disk, format and mount the data disk.
- If it has been formatted and partitioned, log on to the instance to enable the extended storage space.

Step 2. Log on to the instance to enable the extended storage space

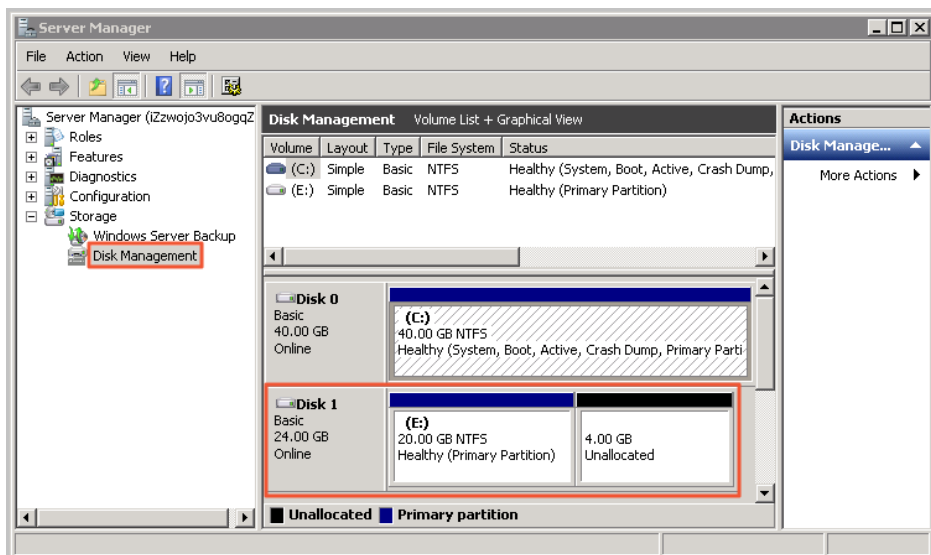
To resize a data disk within the instance, follow these steps:

Log on to the instance.

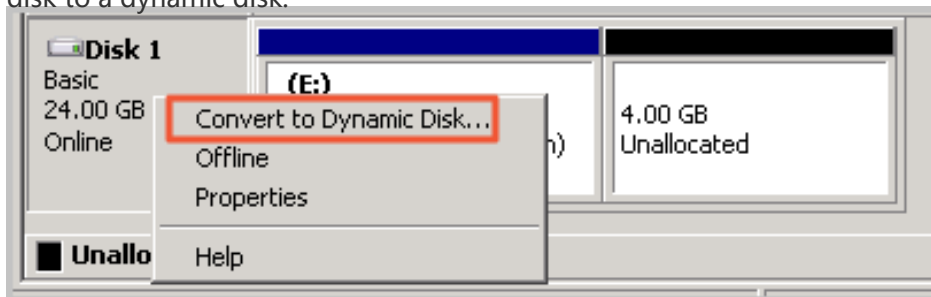
On the Windows Server desktop, click the **Server Manager** icon



In the left-side navigation pane of **Server Manager**, select **Storage > Disk Management**. In the disk management area, you can see the relationship between the new and the original data disk spaces. In this example, **Disk 1** is the resized data disk.

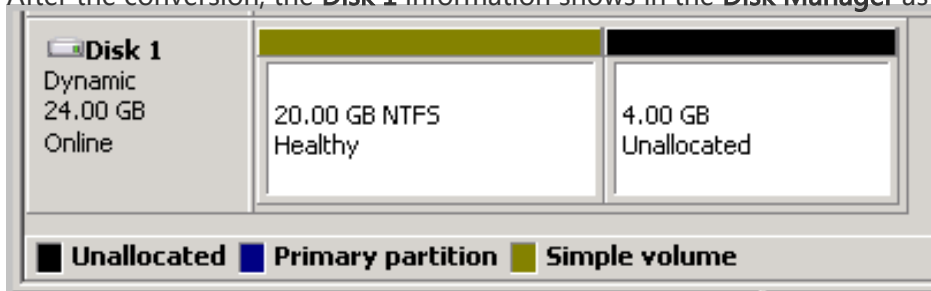


Right click **Disk 1**, select **Convert to Dynamic Disk**, and follow the wizard to convert a basic disk to a dynamic disk.

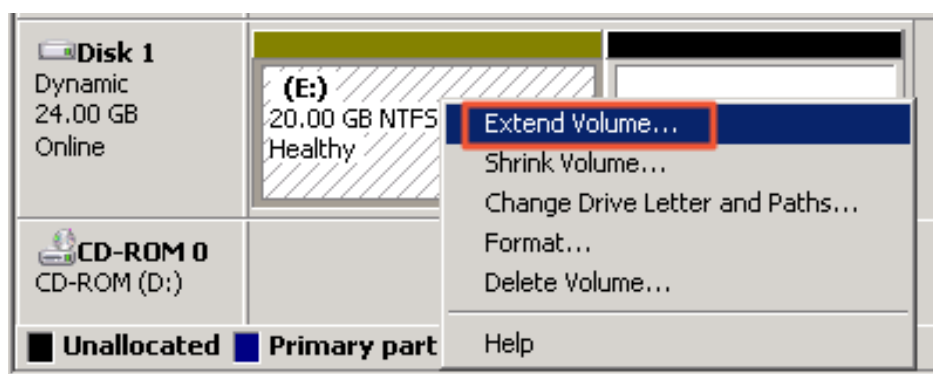


Note: Converting a basic disk into a dynamic disk will unmount the disk from the system. Applications installed in the data disk, if any, are temporarily unavailable during the conversion process. The conversion process will not cause any data loss.

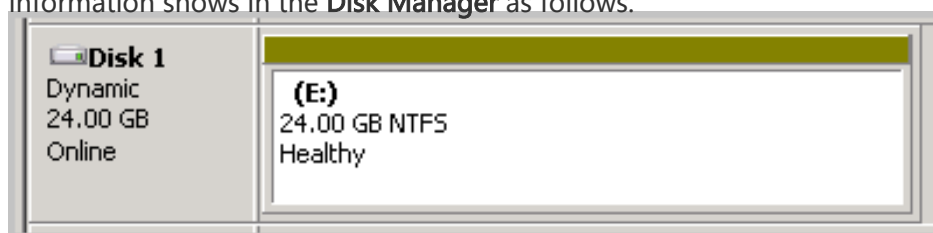
After the conversion, the **Disk 1** information shows in the **Disk Manager** as follows.



Right click an empty area of the simple volume of **Disk 1**, and select **Extend Volume**.



Follow the **Extend Volume Wizard** to extend the volume. When the wizard completes, the new data disk space is automatically merged into the original volume and the **Disk 1** information shows in the **Disk Manager** as follows.



Note: On Windows Server 2003, the extended storage space is added to the data disk but it is displayed as a separate volume in **Disk Manager**. On Windows Server 2003, one separate volume is created for each expansion and is not merged into the original volume, which does not affected the availability of the extended storage space.

Now, you have resized a data disk successfully and have the extended storage space ready for use.

Re-initialize a disk

Disk re-initialization restores a disk to its initial state and settings.

Warning: Re-initializing a disk will erase all data on that disk. Ensure that you have backed up all necessary data before proceeding.

Note:

- The operating system and version of the instance is retained, and will be restored to its initial state and settings.
- The IP address of the instance will not change. The data on the original system disk will be cleared, but the automatic backup of snapshots on the instance will be retained, and can be used to roll back the applications on the instance.
- If you are re-initializing a data disk, you do not need to attach it after re-initialization.

Perform the following steps to re-initialize a data disk:

Log on to the ECS console.

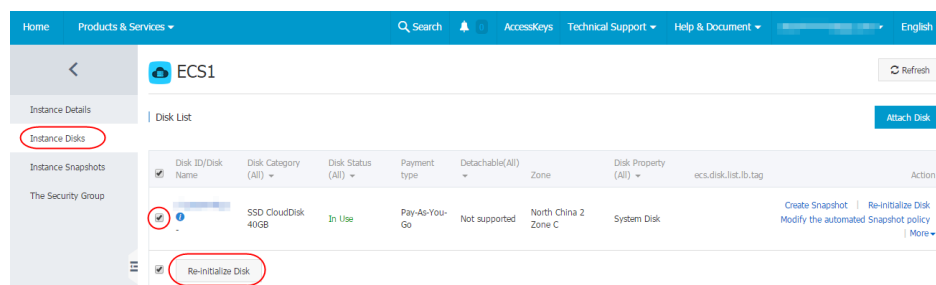
Click **Instances** in the left navigation bar.

Select your desired region.

Select the instance for disk re-initialization, and click **Stop**.

Select the instance name and then click **Instance Disks** in the left navigation bar.

Select one or more disks to re-initialize and click **Re-initialize Disk**.



Enter a new login password once the re-initialization is finished, and then click **Confirm Re-initializing Disk**.

Roll back a disk

Disk rollback restores a disk to a state and setting from previous point in time.

Note:

- Snapshot rollback is a permanent action and cannot be reversed. Once rollback is completed, the original data cannot be restored. You are recommended to proceed with caution.
- Disk rollback can only be performed when the instance is completely stopped.

To rollback a disk, perform the following:

Log on to the ECS console.

Click **Instances** in the left navigation bar.

Select your desired region.

Select the instance for disk rollback, and click **Stop**.

Click the instance name. Then, click **Snapshots** > **Snapshots** in the left navigation bar.

Select the snapshot for rollback. You can select only one snapshot at a time.

Click **Disk Rollback**.

In the displayed dialog box, click **OK**.

View monitoring information of a disk

To view the monitoring information of a disk, such as IOPS and BPS, perform the following:

Log on to the ECS console.

Select a disk to view monitoring information by using one of the following methods:

- Click the instance that the disk is attached to from the instance list page and click **Instance Disks**.
- Locate the disk from the **Disks** list.

Click **Disk Monitoring** to view the monitoring information of the selected disk.

Note: You can select pre-set time segments to initiate regular monitoring periods, such as 1 hour, 6 hours, 1 day, and 7 days. You can also set custom monitoring start and end times.

If you no longer need a data disk, you can release it. Otherwise, we continue charging you for it.

Warning: Releasing a data disk is a permanent action and cannot be reversed. Once deleted, the original data on the data disk cannot be restored. We recommend that you proceed with caution.

Prerequisites

- Ensure that the data disk is in the **Available** status. If the data disk is in the **In Use** status, detach it from the instance in the ECS console.
- Ensure that you backed up all the data you need in the data disk. You can create snapshots for backup.

Procedure

To release a data disk, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Disks**.

Select a region.

Find the disk that you want to release, and in the **Action** column, select **More > Release**.

In the **Release** dialog box, read the note and click **Confirm Release**.

Now you have successfully released the data disk.

Snapshots

Create a snapshot

You can create instance snapshots to save the system state from a certain point in time for data backup or to create images.

Note:

- Creation of the first snapshot will take relatively longer than subsequent snapshots due to the first snapshot being a full snapshot. However, depending on the amount of changed data since previous snapshots, the length of time for each snapshot creation may vary.
- Creating snapshots of a disk may reduce disk performance.

- It is recommended that you not create snapshots during peak traffic hours.
- Manually created snapshots, unlike automatic snapshots, will be retained until they are manually deleted.

To create a snapshot, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Disks**.

Select a region.

Select a system or data disk for which you want to create a snapshot. You can only select one disk at a time.

Click **Create Snapshot**.

Enter a name for the snapshot and click **OK**.

To view all snapshots, go to the left navigation bar and click **Snapshots > Snapshots**.

To view snapshots of a particular instance, navigate to **Instances**, select the particular instance, and then click **Instance Snapshots** in the left navigation bar.

An automatic snapshot policy is a set of defined parameters for automatically creating snapshots.

You can create a maximum of 100 automatic snapshot policies in each region.

Create an automatic snapshot policy

To create an automatic snapshot policy, perform the following:

Log on to the ECS console.

In the left-side navigation pane, select **Snapshots > Automatic Snapshot Policy**.

Click **Create Automatic Snapshot Policy**.

Define automatic snapshot policy parameters:

- **Automatic Snapshot Policy**

This parameter is the name of the automatic snapshot policy. It must contain 2 ~ 128 characters and begin with English letters or Chinese characters. It can include digits and periods (.), underscores (_), and hyphens (-).

- **Time**

Defines the time of day for automatically creating snapshots. There are 24 snapshot creation points available between 00:00 and 23:00.

- **Repeated day**

There are seven available repetition day configurations.

- **Retention period**

Defines the number of days a snapshot can be retained. This parameter can be set between 1–65536 days, or permanently. By default, it is set to 30 days.

Create Automatic Snapshot Policy

×

- ECS Snapshot 2.0 data service provides a quota of 64 snapshots for each disk. When the maximum number of snapshots for a disk has been reached, the oldest automatic snapshot generated by the automatic snapshot policy will be deleted when you create a new snapshot.
- If the time for creating a new snapshot exceeds the interval between two automatic snapshot points in time (due to reasons such as huge disk data volume) the next point in time will be skipped automatically without a snapshot being created. For example, if a user sets 9:00, 10:00, and 11:00 as the automatic snapshot points in time, and it takes 70 minutes to create the snapshot scheduled for 9:00, or it is to be completed at 10:10, the scheduled snapshot for 10:00 will be skipped and the next snapshot will be created at 11:00.
- The current snapshot policy execution time defaults to the Eastern Time Zone (UTC + 8). You should adjust this time policy according to the actual business requirements.

*Automatic Snapshot Schedule:

It must contain 2-128 characters and begin with English letters or Chinese characters. It can include numbers and the characters ".", "_", and "-".

*Time:

☐ 00:00 ☐ 01:00 ☐ 02:00 ☐ 03:00 ☐ 04:00 ☐ 05:00
☐ 06:00 ☐ 07:00 ☐ 08:00 ☐ 09:00 ☐ 10:00 ☐ 11:00
☐ 12:00 ☐ 13:00 ☐ 14:00 ☐ 15:00 ☐ 16:00 ☐ 17:00
☐ 18:00 ☐ 19:00 ☐ 20:00 ☐ 21:00 ☐ 22:00 ☐ 23:00

*Repeated Day:

☐ Monday ☐ Tuesday ☐ Wednesday ☐ Thursday ☐ Friday
☐ Saturday ☐ Sunday

Retention Period:

☒ Custom Duration day(s) The retention period
☐ Permanent

OK

Cancel

Click OK.

Delete an automatic snapshot policy

If you no longer need an automatic snapshot policy, navigate to the policy you want to delete and then click **Delete Automatic Snapshot Policy**.

Apply automatic snapshot policies to disks

You can apply an auto snapshot policy to disks according to your business needs. All the automatic snapshots are named in the format of `auto_yyyyMMdd_1`, for example, `auto_20140418_1`.

Note:

- Creating snapshots may disturb read and write operations on your disk. It is recommended that you perform auto snapshots during periods when service load is low to reduce effects on your service.
- Auto snapshot policies cannot be applied to basic cloud disks when they are not in use.
- Snapshots that are manually created do not conflict with auto snapshots. However, if you are performing an auto snapshot on a disk, you must wait for it to finish before manually creating a snapshot.

You can apply an auto snapshot policy to a disk through either of the following:

Disks menu

For applying an auto snapshot policy to a specific disk.

Snapshots menu

For applying a unified auto snapshot policy to several or all disks.

From the Disks menu

To apply an auto snapshot policy through the **Disks** menu, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Disks**.

Select a region.

Select the disk for which you want to execute the policy and click **Automatic Snapshot**

Policy.

Enable the auto snapshot function and select the desired snapshot policy.

Click **OK**.

From the Disks menu

To apply or disable an auto snapshot policy through Snapshot, perform the following:

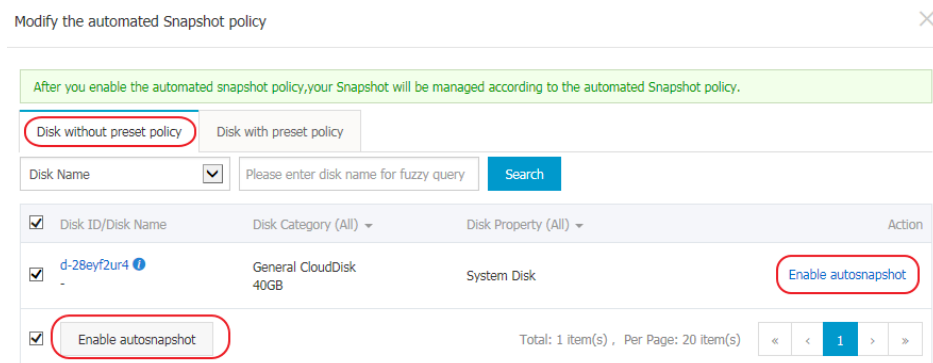
Log on to the ECS console.

In the left-side navigation pane, select **Snapshots > Automatic Snapshot Policy**.

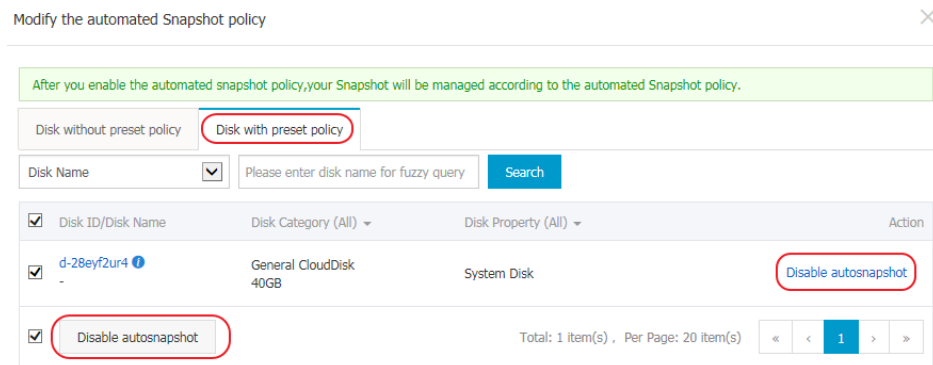
Select a region.

Select the auto snapshot policy you want to apply and click **Set Disk**.

- To enable the auto snapshot policy, select the **Disk without Preset Policy** tab to view the disks. Select the disks in which you want to enable the policy, and then click **Enable the Automatic Snapshot**.



- To disable the auto snapshot policy, select the **Disk with Preset Policy** tab to view the disks. Select the disks in which you want to disable the policy, and then click **Disable the Automatic Snapshot**.



Retain automatic snapshots when releasing disk

To retain auto snapshots when releasing the disk, perform the following:

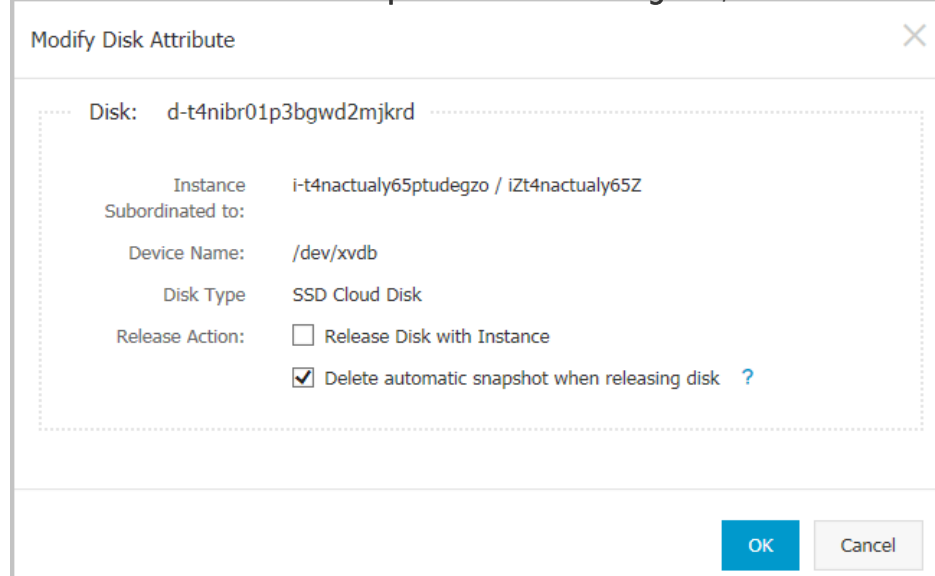
Log on to the ECS console.

In the left-side navigation pane, click **Disks**.

Select a region.

Select the disk that you want to configure and click **More > Modify Attributes**.

Deselect **Delete automatic snapshots when releasing disk**, and then click **OK**.



You may want to delete a snapshot or an automatic snapshot policy.

Delete a snapshot

When you no longer need a snapshot, or you have reached your snapshot quota, you can delete snapshots to free up space.

Note:

- Deleting a snapshot is a permanent action and cannot be reversed. Once deletion is completed, the original snapshot cannot be restored. Proceed with caution.
- If a snapshot has been used to create a custom image, you must delete the associated image before you can delete the snapshot.

To delete a snapshot, perform the following:

Log on to the ECS console.

In the left-side navigation pane, select **Snapshots > Snapshots**.

Select a region.

Select the snapshots you want to delete.

Click **Delete** at the bottom of the window.

Click **OK**.

Delete an automatic snapshot policy

Log on to the ECS console.

In the left-side navigation pane, select **Snapshots & Images > Automatic Snapshot Policy**.

Select a region.

Find the target automatic snapshot policy, and in the **Action** column, click **Delete Automatic Snapshot Policy**.

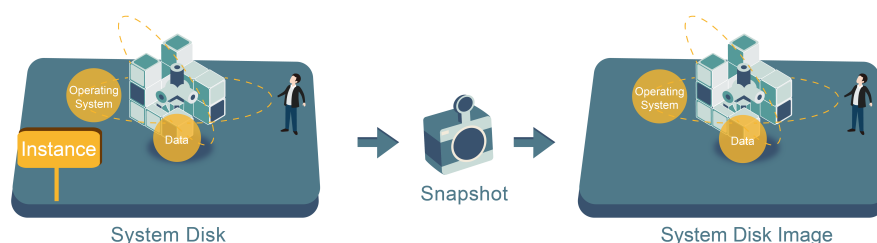
In the dialog box, confirm information and click **OK**.

Images

Create a custom image using a snapshot

Custom images help you run ECS instances effectively by allowing you to create multiple ECS instances with identical OS and environment data to meet scaling requirements.

Custom images are based on ECS disk snapshots. You can set up identical or different configurations for ECS instances that are created from images.



Considerations

Custom images are subject to the following restrictions:

- One account can have up to 100 custom images in a region.
- If the ECS used for creating a custom image expires, or the data is erased (that is, the system disk used for the snapshot expires or is released), the custom image and the ECS instances created from the custom image will not be affected. However, auto snapshots will be cleared when an ECS instance is released.
- You can upgrade the CPU, memory, bandwidth, hard drive, and, other configurations of ECS instances activated using a custom image.
- Custom images cannot be used across regions.
- A custom image applies to any ECS payment method, either yearly/monthly subscription or Pay-As-You-Go. Custom images for ECS instances under yearly/monthly subscription plans can be used to create Pay-As-You-Go instances, and vice versa.

When creating custom images on a Linux instance, adhere to the following:

- Do not load data disk information in the `/etc/fstab` file. Otherwise, instances created using this image will not start.

- It is recommended that you unmount all data disks before taking a snapshot and creating an image. Otherwise, ECS instances that are created based on this custom image may not start.
- Modifying the default logon user name **root** is not allowed (Linux OS only).

Operating procedure

To create a custom image from a snapshot, perform the following:

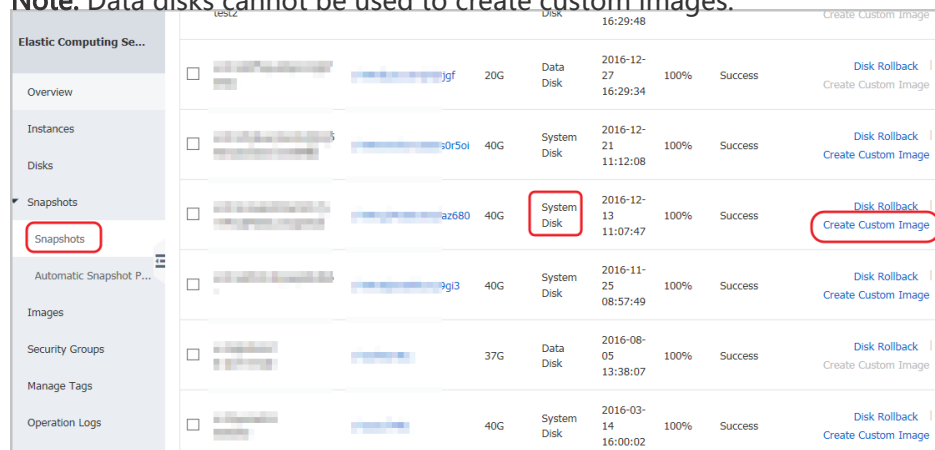
Log on to the ECS console.

Click **Snapshots** > **Snapshots** in the left navigation bar.

Select your desired region.

Select a snapshot with the disk attribute of **System Disk** and click **Create Custom Image**.

Note: Data disks cannot be used to create custom images.



	Snapshot ID	Size	Disk Type	Creation Time	Progress	Status	Actions
☐	test2-jgf	20G	Data Disk	2016-12-27 16:29:34	100%	Success	Disk Rollback Create Custom Image
☐	z0r50l	40G	System Disk	2016-12-21 11:12:08	100%	Success	Disk Rollback Create Custom Image
☐	az680	40G	System Disk	2016-12-13 11:07:47	100%	Success	Disk Rollback Create Custom Image
☐	2gl3	40G	System Disk	2016-11-25 08:57:49	100%	Success	Disk Rollback Create Custom Image
☐		37G	Data Disk	2016-08-05 13:38:07	100%	Success	Disk Rollback Create Custom Image
☐		40G	System Disk	2016-03-14 16:00:02	100%	Success	Disk Rollback Create Custom Image

In the displayed dialog box, you can view the snapshot ID. Enter a name and description for the custom image.

(Optional) Click **Add Data Disk Snapshot** to select multiple snapshots of data disks for the image.

Note: If the snapshot disk capacity is left blank, an empty disk will be created with the default capacity of 5 GB. If you select available snapshots, the disk size is the same as the size of these snapshots.

Create Custom Image

When creating a custom image with Linux system, please do not load data disk information in the `/etc/fstab` file. Otherwise, you cannot launch the instance created through the image.

System Snapshot ID: `s-bp1dbycplpp44p7wvdc9/`

* Image Name:

It must contain 2-128 characters and begin with English letters or Chinese characters. It can include numbers and the characters ".", "_", and "-".

* Image Description:

It must contain 2-256 characters and it cannot begin with `http://` or `https://`.

☒ Add Data Disk Snapshot

Snapshot Details:

Snapshot ID	Device Name:	Disk Capacity:	Action
<code>s-bp1dbycplpp44p7wvdc9(System Disk)</code>	<code>/dev/xvda</code>	GB	Delete

Add

1. Leaving the snapshot ID blank will create an empty disk. Default capacity: 5 GB, with up to 2,000 GB supported.
2. If a snapshot ID is selected, the default disk capacity will be the snapshot capacity.
3. If the device name is blank, it will be randomly allocated.

Create

Cancel

Click **Create**. The custom image is successfully created.

(Optional) To view images you have created, select **Images** in the left navigation bar.

FAQ for images of Linux instances

How to unmount a disk and delete disk table data?

If `/dev/hda5` is attached to `/mnt/hda5`, run any of the following three commands to detach the file system:

```
umount /dev/hda5
umount /mnt/hda5
umount /dev/hda5 /mnt/hda5
```

`/etc/fstab` is an important configuration file in Linux. It contains file system details and storage devices attached at startup.

If you do not want to mount a specified partition when starting the VM, delete the corresponding lines from `/etc/fstab`. For example, you can delete the following statement to disconnect `xvdb1` at startup: `/dev/xvdb1 /leejd ext4 defaults 0 0`.

How to determine whether a data disk is detached and a custom image can be

created?

You must ensure that the auto attach data disk statement line has been deleted from the **fstab** file.

Use the mount command to view information on all mounted devices. Ensure that the execution results do not contain the information of the data disk partition.

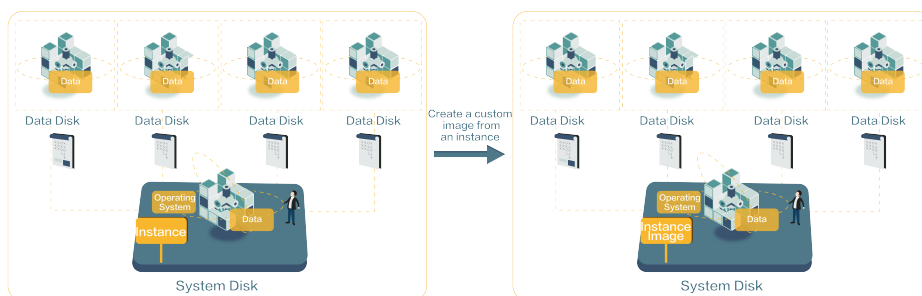
Relevant configuration files

Before creating an image, ensure the key configuration files from the following table have not been modified; otherwise, the new instance will not be able to start.

Configuration File	Purpose	Risks if changed
/etc/issue*, /etc/*-release, /etc/*_version	For system release and version	Modifying /etc/issue* will make the system release version unidentifiable, and cause instance creation failure.
/boot/grub/menu.lst, /boot/grub/grub.conf	For system boot	Modifying /boot/grub/menu.lst will result in kernel loading failure, and the system will not be able to start.
/etc/fstab	For mounting partitions during boot.	Modifying it will cause partition mounting failure, and the system will not be able to start.
/etc/shadow	For storing system passwords.	If this file is set to read-only, the password file cannot be edited, and instance creation will fail.
/etc/selinux/config	For system security policies	Modifying /etc/selinux/config and enabling SELinux will result in start failure.

Create a custom image using an instance

You can create a custom image using an instance by selecting its disks, including the system disk and data disks, and pack them into an image. During this process, snapshots are automatically created for all disks of the instance. It is these snapshots that are then used to compose the new image.

**Note:**

- Before creating a custom image, ensure no confidential data is accessible on the disks being used.
- To create a custom image from snapshots, see [Create a custom image from snapshots](#).
- During creation, do not change the status of the instance. Do not stop, start, or restart the instance.

To create a custom image, follow these steps:

Log on to the ECS console.

Click **Instances** on the left navigation bar.

Select your desired region.

Select the target instance, and then click **More > Create Custom Image**.

Enter a name and description for the image, and then click **Create**.

When all snapshots have been successfully created, the custom image will be ready for use.

Copying an image is the process in which a custom image is copied from one region to another region. When a request of copying a custom image is initiated, Alibaba Cloud copies the snapshot that the custom image is created from the source region to the target region, and then creates a custom image from the copied snapshot in the target region.

The speed of the process of copying the snapshot between regions depends on the network status, and Alibaba Cloud supports processing concurrent requests of copying images and your request maybe in a long queue. Therefore, it may take long time to complete copying.

Copying images across regions allows you to deploy a backup image system, or an identical application environment, in different regions.

To copy a custom image, follow these steps:

Log on to the ECS console.

In the left-side navigation pane, choose **Snapshots & Images > Images**.

Select a region.

Select the custom image you want to copy, and in the **Action** column, click **Copy Image**.

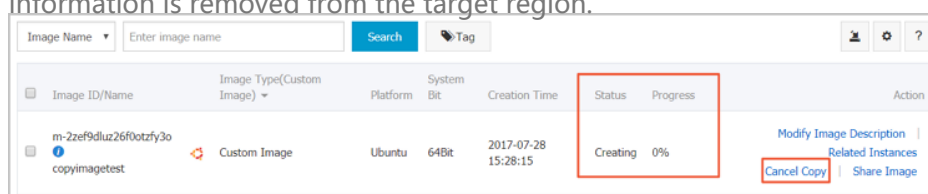
If your custom image is larger than 100 GB, when you click **Copy Image**, you are directed to open a ticket to complete copying the image.

In the **Copy Image** dialog box, the ID of the selected image is displayed, and you have to complete the configurations:

- i. Select the target region. Currently, copying images is only allowed between regions in mainland China.
- ii. Specify a name for the image to be displayed in the target region, and give a short description of the image to ease future management.
- iii. Click **OK**.

Click the target region and check the progress. When 100% progress is displayed, the image is copied successfully.

When the progress is not 100% and the status of the image is **Creating**, you can click **Cancel Copy** to cancel the copying process. After the process is canceled, the image information is removed from the target region.



When the status of the image is **Available**, you can use the custom image to create an ECS instance or change a system disk. You can check the snapshot for creating the custom image in the **Snapshot List** in the target region.

For more information about the process of copying images, refer to [Image copy FAQ](#).

You can share your custom images with other users. Through the ECS Management Console or ECS Open API, you can query images shared by other accounts with your own account, and select images shared by other accounts to create ECS instances.

Note:

- The integrity or security of images is not guaranteed. Ensure that you use only images shared by trusted accounts.
- Before using shared images to create ECS instances, log on to the ECS instances to which the shared images belong and verify that the images are secure and complete.
- Before sharing an image, ensure no confidential data is accessible on the disks to be shared.

Considerations

Restrictions

- One image can be shared with a maximum of **50** accounts.
- Shared images do not count towards your image quota.
- Shared images can only be used to create instances in the same region as the source image.
- Only image owners can share images with other accounts.

Impact of deleting shared images

- You can delete a custom image even you have shared it with other accounts. Before deleting the shared image, however, you need to unassociate it from other accounts.
- If you delete an account that has shared a custom image, the users who are using the shared image can no longer find the image through the ECS Management Console or ECS Open API, or use the image to create ECS instances.
- Deleting shared custom images may cause system disk re-initialization to fail for ECS instances created from these images.

Procedure

Log on to the ECS console.

In the left-side navigation pane, click **Images**.

Select a region.

Select the image you want to share.

Note: The image type must be **Custom Image**.

Click **Share Image**.

In the displayed dialog box, select the **Account Type** and enter the account ID you want to share the image with.

Note: The **Account ID** can be obtained from the **Account Management > Security Settings** on the Alibaba Cloud website by logging on to <https://account-intl.console.aliyun.com/#/secure>.

Click **Share Image**.

View accounts using your shared images

You can view which accounts are using your shared images.

To view accounts using your shared images, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Images**.

Select a region.

Select the image you want to check.

Click **Share Image**. A list of the accounts using the selected image is displayed.

Cancel the sharing of an image

You can cancel the sharing of an image to specific accounts at any time.

Note: When the sharing of an image is cancelled:

- Any accounts currently using the image will no longer be able to use the image. Therefore, you must disassociate the image from other accounts before cancelling it being shared.
- Any instances using the image, including instances of other accounts using the shared image, will not be able to reinitialize the system disk.

Procedure

To cancel the sharing of an image, perform the following:

Log on to the ECS console.

In the left-navigation pane, click **Images**.

Select a region.

Select the image you want to cancel sharing.

Click **Share Image**.

Click **Unshare** next to the account with which you want to stop sharing the image.

View the shared images you are using

You can view a list of the shared images from other accounts that you are using.

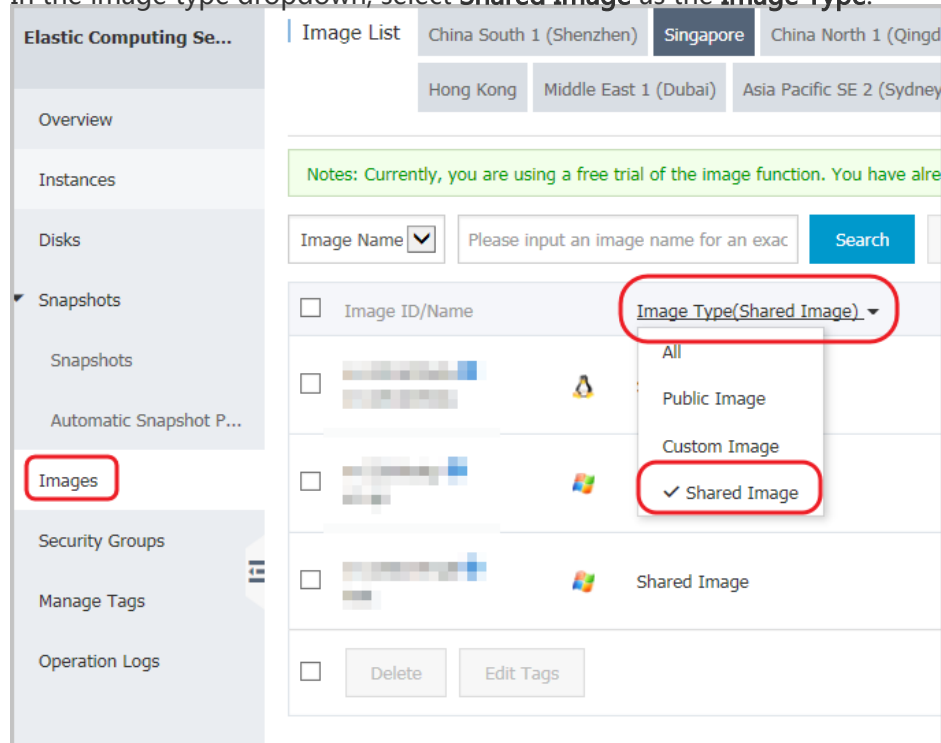
To view a list of the shared images you are using, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Images**.

Select a region.

In the image type dropdown, select **Shared Image** as the **Image Type**.



A list of the shared images you are using will be displayed.

Import images

To ensure the usability of an imported image and to improve the efficiency of importing an image, pay attention to the followings before importing an image.

Images of standard operating systems

If you want to import an image of a standard operating system, such as Windows, CentOS, Ubuntu, Debian, CoreOS, SUSE, OpenSUSE, or FreeBSD, pay attention to the followings.

Critical system configuration files

- Do not modify `/etc/issue*`. If modified, the distribution of the system cannot be properly recognized and the system creation will fail.
- Do not modify `/boot/grub/menu.lst`. If modified, the system may fail to start up.
- Do not modify `/etc/fstab`. If modified, an exception may occur preventing partitions from being loaded, leading to system startup failure.
- Do not modify `/etc/shadow` to **read-only**. If modified, the password file cannot be modified and the system startup will fail.
- Do not enable SELinux by modifying `/etc/selinux/config`. If modified, the system may fail to start up.

Requirements for system basic environments

- Do not adjust the partition of the system disk. Currently only a single root partition is supported.
- Ensure that there is sufficient free space in the system disk.
- Do not modify critical system files such as `/sbin`, `/bin`, or `/lib*`.
- Before importing an image, confirm the integrity of the file system.
- File system: Linux images only support ext3 and ext4 file systems.

Applications

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

Size and format

- Currently, the images in only RAW and VHD formats are supported. If you want to import

images in other formats, use a tool to convert the format before importing the image. We recommend that you import images in a VHD format, which has a smaller transmission capacity.

- Setting the disk size when importing an image: We recommend that you configure the disk size for importing based on the virtual file size (not the used capacity) of the image. The size of the disk for importing must be greater than or equal to 40 GB.

Images of non-standard operating system

The *non-standard operating system* in this section refers to either of the followings:

- The operating system that are not included in the list of operating systems currently supported by Alibaba Cloud
- A standard operating system that fails to comply with the requirements for a standard operating system in terms of critical system configuration files, system basic environment, and applications.

If you want to use an image of a non-standard operating system, you are only allowed to choose:

- Customized Linux: Customized image. If you import an image of this type of operating system, Alibaba Cloud will conduct necessary network or password configuration according to the pre-defined configuration norms. For detailed configuration information, see [Configuration of Customized Linux](#).
- Others Linux: Alibaba Cloud identifies all of these images as other system types. If you import an image of such operating system, Alibaba Cloud will not perform any processing on the created instance. After completing the instance creation, you need to connect to the instance by using the **Connect** function on the console and then manually configure the IP address, the router, and the password.

Note: The two types of operating system mentioned in this section must meet the requirements in **size and format** for images of standard operating systems.

Import an image

You can import image files to the ECS environment to create custom images. You can then use these image to create ECS instances.

Requirements

Supported Windows (32-bit and 64-bit) operating systems

- Microsoft Windows Server 2012 R2 (Standard Edition)
- Microsoft Windows Server 2012 (Standard Edition, Data Center Edition)
- Microsoft Windows Server 2008 R2 (Standard Edition, Data Center Edition, Enterprise Edition)
- Microsoft Windows Server 2008 (Standard Edition, Data Center Edition, Enterprise Edition)
- Microsoft Windows Server 2003 R2 (Standard Edition, Data Center Edition, Enterprise Edition)
- Microsoft Windows Server 2003 with Service Pack 1 (SP1) (Standard Edition, Data Center Edition, Enterprise Edition)

Note: Windows XP, Windows 7 (both Professional Edition and Enterprise Edition), Windows 8, and Windows 10 are not supported.

Supported Linux (32-bit and 64-bit) operating systems

- CentOS 5, 6, 7
- Ubuntu 10, 12, 13, 14
- Debian 6, 7
- OpenSUSE 13.1
- SUSE Linux 10, 11, 12
- CoreOS 681.2.0+

Supported image formats

- RAW
- Virtual hard disk (VHD)

Supported file systems

Windows (32-bit and 64-bit)
Supports NTFS format and MBR partition.

Linux/Unix (32-bit and 64-bit)
Supports ext3 and ext4 file system format and MBR partition.

Windows-specific requirements

- Recommended system disk size: 40 GB ~ 500 GB.
- The firewall must be disabled. Otherwise, remote login is not possible. Port 3389 must be enabled.
- Disable user account control (UAC) to enable changes to be made to the system.

Linux-specific requirements

- Recommended system disk size: 40 GB ~ 500 GB.
- SELinux is not activated.
- The firewall is disabled; port 22 is enabled by default.

Restrictions

- Images can only be imported from the system disk. You cannot import images from data disks.
- Image import does not support the use of multiple network interfaces or IPv6 addresses.
- Passwords must be 8 ~ 30 characters in length and contain three types of characters (uppercase or lowercase letters, digits, and special characters).
- It is recommended that you install the XEN and KVM virtualization platform drivers.

Import images using the ECS Management Console

Prerequisites

- OSS must be activated before images can be imported.
- OSS access permissions must be manually granted to the official ECS service account before images are imported.
- Use an OSS third-party tool client, OSS API or OSS SDK, to upload the file to a bucket in the same region as the ECS custom image to import.
- Ensure images are being created in accordance with image restrictions and requirements.

Operating procedure

To import images using the management console, perform the following:

Log on to the ECS console.

In the left-side navigation pane, select **Snapshots and Images > Images**.

Select the image you want to import, and click **Import Image** next to that image. Ensure the image you want to import meets the preceding requirements. If you have not granted the official ECS service account access to your OSS, image import may fail.

Complete the image import form using the following information:

Region

Select the region where you want to deploy the application.

Image File OSS Address

Copy the object address taken from the OSS console.

Image Name

The length should be 2 ~ 128 characters. It can contain uppercase letters, lowercase letters or Chinese characters. It cannot contain numbers, underscores (_), or hyphens (-).

System Disk Size

Windows system disk size: 40-500 GB; Linux system disk size: 20- 500 GB.

System Architecture

64-bit OS: x86_64. 32-bit OS: i386.

Operating System Type

Currently supported OS releases are:

- Windows: Windows Server 2003, Windows Server 2008, and Windows Server 2012.
- Linux: CentOS, SUSE, Ubuntu, Debian, FreeBSD, and CoreOS.

Note:

- **(Linux only)** Open a ticket to Alibaba Cloud to confirm the selected edition is supported.
- If your image OS is a custom edition developed on a Linux core, open a ticket to Alibaba Cloud.

Image Format

Supports RAW and VHD format. RAW format is recommended.

Note: You cannot use qemu-image to create VHD images.

Image Description

Enter a description of the image.

Click **Submit**. You can view the task progress in the image list of the import region.

Image importing takes time, usually several hours. The duration of the task depends on the size of your image file and how many other import tasks are running. You can view the task progress in the image list of the import region.

You can also find and cancel the image import task in the task manager.

Modify custom image names and description

You can modify the names and descriptions of custom images at any time.

To modify the names and descriptions of custom images, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Images**.

Select a region.

Select the image to edit.

Note: The image type must be **Custom Image**.

Modify the image name by hovering the cursor over the image name, and then clicking the **pen icon** that appears.

Modify the description of an image by clicking **Modify Image Description**, and then entering a description.

Image ID/Name	Image Type(Custom Image)	Platform	System Bit	Creation Time	Status	Progress	Action
       	 Custom Image	CentOS	64Bit	2017-03-07 11:41:14	Available	100%	Modify Image Description Related Instances Copy Image Share Image
       	 Custom Image	CentOS	64Bit	2016-11-25 10:19:00	Available	100%	Modify Image Description Related Instances Copy Image Share Image
       	 Custom Image	CENTOS5	64Bit	2016-11-24 12:22:09	Available	100%	Modify Image Description Related Instances Copy Image Share Image

Click **OK**.

Delete a custom image

You can delete custom images that you no longer require. To ensure successful deletion, check that you do not currently have any ECS instances created from this custom image.

To delete a custom image, perform the following:

Log on to the ECS console.

Click **Images** in the left-side navigation bar.

Select your desired region.

Select the image you want to delete.

Note: The image type must be **Custom Image**.

Click **Delete**.

In the dialog box, click **OK**.

Security groups

Usage scenarios

The following scenarios apply to **Classic Networks**. If you want to know how to use security groups together with VPC, refer to [Use security groups to control ECS instances' access to public cloud for a VPC](#).

Security groups provide security assurance to networks. They can be used to:

- Provide secure intranet communication.
- Block access to instances from specified IP addresses.
- Allow remote logins using only a specified IP address.
- Permit an instance to access only a specified IP address.

Scenario 1: Provide secure intranet communication

In a classic network, you can use security groups for Intranet communication between:

- ECS instances belonging to the same account in the same region.
- ECS instances belonging to different accounts in the same region.

ECS instances belonging to the same account in the same region

By setting security group rules, you can allow classic network instances belonging to the same account in the same region to communicate via intranet.

By default, ECS instances in the same security group can communicate through the intranet, but instances in different security groups cannot communicate through the intranet.

There are two ways to allow communication between ECS instances in different security groups. You can:

- Place the instances in the same security group to allow intranet communication.
- Authorize intranet communication between the two security groups by setting access-type security group rules. In **Authorization Type**, select **Security Group Access**, and then select the security group of each other for the **Authorization Object**.

Instances belonging to different accounts in the same region

By setting security group rules, you can allow classic network ECS instances belonging to different accounts in the same region to communicate via intranet.

To achieve intranet communication between instances belonging to different accounts in the same region, each user must perform the following:

- Add the other user's security group to its inbound intranet.
- Authorize the ECS instances of the other user's security group to access all instances in the their account.

Notice: To ensure the security of your instances, when you are configuring an intranet inbound rule for a security group of classic network type, **Security Group Access** is the top priority for **Authorization Type**. If you want to select **Address Field Access**, you must enter an IP address with CIDR prefix, `"/32"`, in the format of `a.b.c.d/32`. Only IPv4 is supported.

Block access to an ECS instance from an IP address

You can use security groups to block access to an ECS instance or a port of an ECS instance from specified IP addresses.

To block access to an instance from a specific IP address, perform the following:

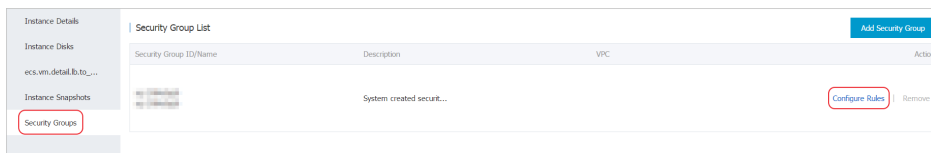
Log on to the ECS console.

Click **Instances** in the left-side navigation bar.

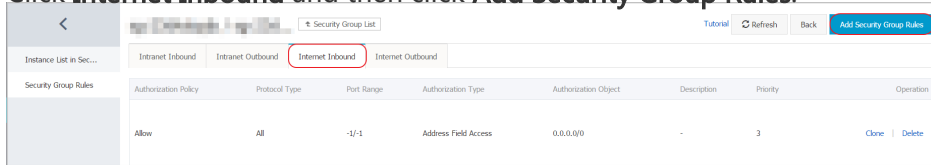
Select your desired region.

Select your desired instance, and click **Manage**.

Select **Security Groups** in the left-side navigation bar, and click **Configure Rules** of one security group in the list.



Click **Internet Inbound** and then click **Add Security Group Rules**.



If you want to drop access from an IP address, on the **Add Security Group Rules** dialog box:

- Select **Drop** for the **Authorization Policy**.
- Select **All** for the **Protocol Type**.
- Select **Address Field Access** for the **Authorization Type** and enter an IP address with or without a CIDR prefix, for example 1.2.3.4 or 1.2.3.4/24, as the **Authorization Object**.
- Enter **1** for **Priority**.
- Click **OK**.

Add Security Group Rules

NIC:
Internet

Rule Direction:
Inbound

Authorization Policy:
Drop

Protocol Type:
All

* Port Range:
-1/-1

The value range is 1-65535. For example, "1/200", "80/80".
Tutorial

Authorization Type:
Address Field Access

* Authorization Object:
e.g. 10.0.0.0/32

Please be cautious when setting authorization objects. Based on different authorization policies, 0.0.0.0/0 indicates that access by all IPs is either allowed or rejected.
Tutorial

Priority:
1

The priority value range is 1 - 100. The default value is 1 which is the highest priority.

OK
Cancel

If you want to drop an IP address to access to Port 22 of your ECS instance, on the **Add Security Group Rules** dialog:

- Select **Drop** for the **Authorization Policy**.
- Select **Custom TCP** for the **Protocol Type**.
- Enter **22/22** as the **Port Range**.
- Select **Address Field Access** for the **Authorization Type** and enter an IP address with or without a CIDR prefix, for example 1.2.3.4 or 1.2.3.4/24, as the **Authorization Object**.
- Enter **1** for **Priority**.
- Click **OK**.

Add Security Group Rules

NIC: Internet

Rule Direction: Inbound

Authorization Policy: Drop

Protocol Type: Custom TCP

* Port Range: 22/22

Authorization Type: Address Field Access

* Authorization Object: e.g. 10.0.0.0/32

Priority: 1

The value range is 1-65535. For example, "1/200", "80/80". [Tutorial](#)

Please be cautious when setting authorization objects. Based on different authorization policies, 0.0.0.0/0 indicates that access by all IPs is either allowed or rejected. [Tutorial](#)

The priority value range is 1 - 100. The default value is 1 which is the highest priority.

OK Cancel

Allow remote login from a specific IP address

Take a Linux instance as an example. Configure to allow a specific IP address to access port 22.

A Linux instance is used in the following example. Allow a specific IP address to SSH the instance.

To allow remote login from a specific IP address, perform the following:

Log on to the ECS console.

Click **Instances** in the left-side navigation bar.

Select your desired region.

Select your desired instance, and click **Manage**.

Select **Security Groups** in the left-side navigation bar, and click **Configure Rules** of one security group in the list.

Click **Internet Inbound** and then click **Add Security Group Rules**.

On the **Add Security Group Rules** dialog box:

- Select **Allow** for the **Authorization Policy**.
- Select **Custom TCP** for the **Protocol Type**.
- Enter **22/22** as the **Port Range**.
- Select **Address Field Access** for the **Authorization Type** and enter an IP address with or without a CIDR prefix, for example 1.2.3.4 or 1.2.3.4/24, as the **Authorization Object**.
- Enter **1** for **Priority**.
- Click **OK**.

Add Security Group Rules

NIC:

Internet

Rule Direction:

Inbound

Authorization Policy:

Allow

Protocol Type:

Custom TCP

* Port Range:

22/22

The value range is 1-65535. For example, "1/200", "80/80". [Tutorial](#)

Authorization Type:

Address Field Access

* Authorization Object:

1.2.3.4

Please be cautious when setting authorization objects. Based on different authorization policies, 0.0.0.0/0 indicates that access by all IPs is either allowed or rejected. [Tutorial](#)

Priority:

1

The priority value range is 1 - 100. The default value is 1 which is the highest priority.

OK

Cancel

Add another security group rule:

- Select **Drop** for the **Authorization Policy**.
- Select **Custom TCP** for the **Protocol Type**.
- Enter **22/22** as the **Port Range**.
- Select **Address Field Access** for the **Authorization Type** and enter **0.0.0.0/0** as the **Authorization Object**.
- Enter **2** for **Priority**.
- Click **OK**.

Add Security Group Rules

NIC: Internet

Rule Direction: Inbound

Authorization Policy: Drop

Protocol Type: Custom TCP

* Port Range: 22/22 The value range is 1-65535. For example, "1/200", "80/80". [Tutorial](#)

Authorization Type: Address Field Access

* Authorization Object: 0.0.0.0/0 Please be cautious when setting authorization objects. Based on different authorization policies, 0.0.0.0/0 indicates that access by all IPs is either allowed or rejected. [Tutorial](#)

Priority: 2 The priority value range is 1 - 100. The default value is 1 which is the highest priority.

OK Cancel

When both security rules are successfully configured, the following results are displayed:

- The request to access port 22 from 1.2.3.4 is allowed, according to the rule configured as priority 1.
- The request to access port 22 from other IP addresses is dropped, according to the rule configured as priority 2.

Allow instance access to specific IP addresses

You can allow instance access to specified IP addresses.

To allow access to an instance for only specific IP addresses, perform the following:

Log on to the ECS console.

Click **Instances** in the left-side navigation bar.

Select your desired region.

Select your desired instance, and click **Manage**.

Select **Security Groups** in the left-side navigation bar, and click **Configure Rules**.

Click **Internet Outbound** and then click **Add Security Group Rules**.

On the **Add Security Group Rules** dialog box:

- Select **Drop** for the **Authorization Policy**.
- Select **All** for the **Protocol Type**.
- Select **Address Field Access** for the **Authorization Type** and enter **0.0.0.0/0** as the **Authorization Object**.
- Enter **2** for **Priority**.
- Click **OK**.

Add Security Group Rules

NIC: Internet

Rule Direction: Outbound

Authorization Policy: Drop

Protocol Type: All

* Port Range: -1/-1

The value range is 1-65535. For example, "1/200", "80/80". [Tutorial](#)

Authorization Type: Address Field Access

* Authorization Object: 0.0.0.0/0

Please be cautious when setting authorization objects. Based on different authorization policies, 0.0.0.0/0 indicates that access by all IPs is either allowed or rejected. [Tutorial](#)

Priority: 2

The priority value range is 1 - 100. The default value is 1 which is the highest priority.

OK Cancel

Add another security group rule:

- Select **Allow** for the **Authorization Policy**.
- Select **All** for the **Protocol Type**.

- Select **Address Field Access** for the **Authorization Type** and enter an IP address with or without a CIDR prefix, for example 1.2.3.4 or 1.2.3.4/24, as the **Authorization Object**.
- Enter **1** for **Priority**.
- Click **OK**.

Add Security Group Rules

NIC: Internet

Rule Direction: Outbound

Authorization Policy: Allow

Protocol Type: All

* Port Range: -1/-1
The value range is 1-65535. For example, "1/200", "80/80". [Tutorial](#)

Authorization Type: Address Field Access

* Authorization Object: 1.2.3.4
Please be cautious when setting authorization objects. Based on different authorization policies, 0.0.0.0/0 indicates that access by all IPs is either allowed or rejected. [Tutorial](#)

Priority: 1
The priority value range is 1 - 100. The default value is 1 which is the highest priority.

OK Cancel

To check that the rules were successfully configured, log on to the instance and run ping or telnet. If you do not have access to IP addresses except for the IP address previously specified for access authorization, the configuration is successful.

This document introduces the default rules of security groups that are created by the system and by yourself.

The security group created by the system

The security group created by the system has only rules for access over all ICMP ports, TCP Port 22, and TCP Port 3389, of which,

- All ICMP ports are used by network devices, including routers, to send error messages and

operational information.

- TCP Port 22 is used to connect to a Linux instance using SSH.
- TCP Port 3389 is used to remotely connect to a Windows instance using Windows Remote Desktop.

Classic network

The default security group created by the system includes rules of:

- **Intranet:** drop inbound traffic on all ports, and accept outbound traffic on all ports.
- **Internet:** accept outbound traffic on all ports, but only accept inbound traffic on TCP Port 22, TCP Port 3389, and all ICMP ports.

VPC

A security group for VPC only has intranet rules for inbound and outbound traffic.

The default security group created by the system includes rules of:

- **Outbound:** accept outbound traffic on all ports.
- **Inbound:** Only accept inbound traffic on TCP Port 22, TCP Port 3389, and all ICMP ports.

Inbound		Outbound				
Authorization Policy	Protocol Type	Port Range	Authorization Type	Authorization Object	Description	Priority
Allow	Custom TCP	22/22	Address Field Access	0.0.0.0/0	-	110
Allow	Custom TCP	3389/3389	Address Field Access	0.0.0.0/0	-	110
Allow	All ICMP	-1/-1	Address Field Access	0.0.0.0/0	-	110

A rule of Priority 110 can be hidden by any rule that you add, the priority of which can only be set to a number from 1 to 100.

For user-defined security groups

For user-defined security groups, the default rules of the default security group are as follows:

- Allow all for outbound traffic.
- Drop all for inbound traffic, for both intranet and Internet.

Create a security group

A security group functions similarly to virtual firewalls, and is used to set network access controls for

one or more ECS instances. When creating instances, you must select a security group. You can also add security group rules to control outbound and inbound network access for all ECS instances in the security group.

To create a security group, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Security Groups**.

Select a region.

Click **Create Security Group**. In the displayed dialog box, enter the following:

Security Group Name

The length must be 2–128 characters. It can contain uppercase letters, lowercase letters, and Chinese characters. It cannot contain numbers, underscores (_), or hyphens (-).

Description

The length must be 2–256 characters. Do not start with http:// or https://.

- Network Type

There are two network types, Classic network and VPC. If you select VPC, you must select a specific VPC. If no VPCs have been created in the current region, you must create one first.

Click **OK**.

You can add security group rules to enable or disable access to the Internet, intranet, or private networks of ECS instances in the security group:

- **VPC network:** You only need to set outbound and inbound rules, and do not need different rules for private networks and Internet.
- **Classic network:** It is required to set outbound and inbound rules for Internet and intranet respectively.

Changes to the security group rules automatically be applied to ECS instances in the security group.

Prerequisites

You have created a security group. See [Create a security group](#).

You know which Internet, intranet, or private network requests to allow or disallow for your instance.

Procedure

To add a security group rule, perform these steps:

Log on to the ECS console.

In the left-side navigation pane, select **Networks & Security > Security Groups**.

Select a region.

Find the security group to add authorization rules, in the **Action** column, click **Configure Rules**.

On the **Security Group Rules** page, click **Add Security Group Rules**.

(Optional) If you do not need to enable or disable all ports for all protocols, ICMP, or GRE, you can select **Quickly Create Rules**.

In the dialog box, set the following parameters:

- **NIC:**
 - If the security group is for VPC, you do not need to select NIC.
 - If your instances can access the Internet, you set rules for both Internet and private networks.
 - If your instances cannot access the Internet, you set rules for private networks only.
 - If the security group is for Classic network, you can select **Internet** or **Intranet**.
- **Rule Direction:**
 - **Outbound:** ECS instances access other ECS instances over intranet, private networks, or through Internet resources.
 - **Inbound:** Other ECS instances in the intranet or private networks, and Internet resources access the ECS instance.
- **Authorization Policy:** Select **Allow** or **Drop**.

Note: **Drop** policy discards the data packet without returning response. If two security groups are the same except for the authorization policy, the **Drop** rule takes effect, and the **Allow** rule does not.
- **Protocol Type** and **Port Range:** The port range setting is affected by the selected protocol type. The following table shows the relationship between protocol types

and port ranges.

Protocol type	Port range	Application scenarios
All	Shown as -1/-1, indicating all ports.	Used in scenarios where both applications are fully mutually trusted.
All ICMP	Shown as -1/-1, indicating all ports.	Used to detect the instance' s network connection status by using the Ping tool.
All GRE	Shown as -1/-1, indicating all ports.	Used for VPN service.
Custom TCP	For custom port range, the valid port value is 1–65536, and the valid port range format is Start Port/End Port . Port range needs to be set using valid format even if setting only one port, for example, using 80/80 to indicate port 80.	Used to allow or deny one or several successive ports.
Custom UDP		
SSH	Shown as 22/22, the default SSH port 22.	Used for remotely connecting to Linux instances.
TELNET	Shown as 23/23.	Used for Telnet to remotely log on to instances.
HTTP	Shown as 80/80.	The instance is used as a server for a website or a web application.
HTTPS	Shown as 443/443.	The instance is used as a server for a website or a web application that supports HTTPS protocol.
MS SQL	Shown as 1433/1433.	The instance is used as a MS SQL server.
Oracle	Shown as 1521/1521.	The instance is used as an Oracle SQL server.
MySQL	Shown as 3306/3306.	The instance is used as a MySQL server.
RDP	Shown as 3389/3389, the default RDP port 3389.	The instance is a Windows instance, which requires a remote desktop

		connection.
PostgreSQL	Shown as 5432/5432.	The instance is used as a PostgreSQL server.
Redis	Shown as 6379/6379.	The instance is used as a Redis server.

Port 25 is disabled by default, and it cannot be enabled by adding security group rules.

- **Authorization Type** and **Authorization Object**: The authorization object affects setting of authorization type. The following is the relationship between them.

Authorization type	Authorization object
Address Field Access	Use IP or CIDR block format, such as <i>10.0.0.0</i> or <i>192.168.0.0/24</i> . Only IPv4 addresses are supported. 0.0.0.0/0 indicates all IP addresses.
Security Group Access	Authorize the instances in a security group under your account or other account to access the instances in this security group. • Authorize This Account: Select a security group under your account. • Authorize Other Account: Enter the target security group ID and the Account ID. You can view the account ID in Account Management > Security Settings. For VPC network instances, Security Group Access works for private IP addresses only. If you want to authorize Internet IP address access, use Address Field Access.

Notice: To ensure the security of your instance, when you are configuring an intranet inbound rule for a security group of the Classic network type, **Security Group Access** is the top priority for **Authorization Type**. If you want to select **Address Field Access**, you must enter an IP address with the CIRD prefix, *"/32"* , in the format of a.b.c.d/32.

- **Priority:** 1–100. The smaller the number is, the higher the priority is. For more information on priority, see [Instructions to the priority of security group rules](#).
7. Click **OK** to successfully add a security group rule for a specified security group.

Check whether security group rules takes effect

Assume that you have installed a web service in the instance and added a security group rule in a security group: allow all IP addresses to have inbound access to TCP port 80 of the instance.

Security group rules usually take effect immediately, though some delay may occur.

Linux instances

If it is a Linux instance in the security group, perform these steps to check whether the security group rule has been activated.

Remotely connect to the ECS instance.

Run the following command to check whether TCP 80 is being listened.

```
netstat -an | grep 80
```

If the following result is returned, web service for TCP port 80 is enabled.

```
tcp      0      0 0.0.0.0:80          0.0.0.0:*          LISTEN
```

Enter `http://IP address` in the address bar of a browser. If the rule takes effect, the access is successful.

Windows instances

If it is a Windows instance in the security group, perform these steps to check whether the security group rule has been activated.

Remotely connect to the ECS instance.

Run **cmd**, and run the following command to check whether TCP 3389 is being listened.

```
netstat -aon | findstr :80
```

If the following result is returned, TCP port 3389 is enabled.

```
TCP  0.0.0.0:80      0.0.0.0:0      LISTENING      1172
```

Enter `http://IP address` in the address bar of a browser. If the rule takes effect, the access is successful.

Instruction to the priority of security group rules

The **Priority** of a security group rule can be a number from 1 to 100. The smaller the number is, the higher the priority is.

ECS instances can belong to different security groups. Either in one or multiple security groups, if security group rules have the same protocol types, port ranges, authorization types, and authorization objects, the rule that takes effect depends on the settings listed in the following table. See the **Result** column of the table for details.

No.	Security group rules	Priority	Authorization policy	Result
i	A	Same	Allow	B takes effect, which means if two security group rules differentiate from each other only in the authorization policy, the Drop rule takes effect.
	B		Drop	
ii	C	1	Allow	C takes effect, which means the rule with the higher priority takes effect.
	D	2	Drop	

View the security group list

You can view the security groups on the ECS Management Console at any time.

To view the security groups list, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Security Groups**.

Select a region. A list of all the security groups in the specified region will be displayed.

(Optional) You can select VPC ID in the filter input box, and then search for a specific ID, to list all the security groups under this VPC.

Modify security group properties

You can modify the name and description of a security group at any time.

To modify the name and description of a security group, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Security Groups**.

Select a region. A list of all the security groups in this region will be displayed.

Modify properties of your desired security group:

- Modify the name: Hover the cursor over the group's name, and then click the pen icon that appears.
- Modify the name and description: Click **Modify**, and then enter a new name and description in the dialog box.

Click **OK**.

View the rules of a security group

You can view the rules of a security group at any time.

To view the rules of a security group, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Security Groups**.

Select a region.

Select a security group.

Click **Configure Rules**. The following security group rule tabs will be displayed for Classic Networks and VPCs:

- For Classic networks
 - Internet Inbound
 - Internet Outbound
 - Intranet Inbound
 - Intranet Outbound
- For VPCs
 - Intranet Inbound
 - Intranet Outbound (Egress)

Click a tab to view the security group rules for that type.

You can delete security group rules if you no longer need them.

To delete rules in a security group, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Security Groups**.

Select a region.

Find the security group where you want to delete rules, and in the **Action** column, click **Configure Rules**.

On the security group management page, choose the rule direction and find the rule you want to delete.

- If the security group is for Classic network, the rule directions are Internet Inbound, Internet Outbound, Intranet Inbound, and Intranet Outbound.
- If the security group is for VPC network, the rule directions are Inbound and Outbound.

In the **Action** column, click **Delete**.

On **Delete Security Group Rules** dialog box, read and confirm the notes, and then click **OK**.

Delete a security group

You can delete security groups, if you no longer require them.

Note:

- Before deleting a security group, ensure it does not contain instances and is not referenced in the rules of another security group.
- Deleting a security group will delete all its rules.

To delete a security group, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Security Groups**.

Select a region. You can then view a list of all the security groups in the region.

Select one or more security groups.

Click **Delete**.

In the displayed dialog box, click **OK**.

Alibaba Cloud supports cloning a security group across regions and network types.

Application scenarios

You may need to clone a security group in the following scenarios:

You have created a security group, named SG1, in Region A, and you want to apply the same rules of SG1 to ECS instances in Region B. Then you can clone SG1 to Region B without creating a new security group in Region B.

You have created a security group for Classic the network type, named SG2, and you need to apply rules identical to SG2 to instances of VPC the network type. Then you can change the

network type to VPC when cloning SG2 to generate a security group suitable for the VPC network.

If you want to apply new security group rules to an ECS instance that are running an online business application, we recommend that you clone the security group as a backup before modifying the rules. If the new security group rules are disadvantageous to the online business application, you can restore the rules completely or partly.

Prerequisites

If you want to change the network type of a security group from Classic to VPC, you have to create a VPC and VSwitch in the target region first.

Operating procedure

Follow the steps to clone a security group.

Log on to the ECS Management Console.

In the left-side navigation pane, select **Network & Security > Security Groups**.

Select the target region in the **Security Group List** page.

Find the target security group, and in the **Action** column, click **Clone Security Group**.

In the **Clone Security Group** dialog box, set the new security group information:

- **Destination Region:** Select a region suitable for the new security group. Not all regions are supported at present. The supported regions are displayed in the drop-down list.
- **Security Group Name:** Specify a new name for the new security group.
- **Network Type:** Select a network type suitable for the new security group. If VPC is selected, you have to choose one VPC in the drop-down list.

Click **OK**.

The new security group is displayed in the Security Group List.

Restoring security group rules refers to the process of completely or partially restoring the rules in the original security group to those of a target security group. Specifically:

Completely restoring refers to moving the rules that do not exist in the target security group from the original security group and adding the rules that only exist in the target security group to the original security group. After restoration, rules in the original security group are identical with those in the target security group.

Partially restoring refers to adding the rules that only exist in the target security group to the original security group and ignoring the rules that only exist in the original group.

Restrictions

Currently, restoring security group rules has the following restrictions:

The original security group and the target security group must be in the same region.

The original security group and the target security group must be of the same network type.

If any system-level security group rules, of which the priority is 110, exist in the target security group, they will not be created during restoration. After restoration, the rules in the original security group may be different from what is expected. If you need the system-level security group rules, you have to manually create the rules and set their priority to 100.

Application scenarios

If you want to apply new security group rules to an ECS instance that is running an online business application, you can clone the former security group as a backup, and then modify the rules inside. If the new security group rules impair the online business application, you can restore the rules fully or partially.

Prerequisites

You must own at least one security group of the same network type in the same region.

Operating procedure

Follow the steps to restore your security group rules:

Log on to the ECS Management Console.

In the left-side navigation pane, select **Network & Security > Security Groups**.

On the **Security Group List** page, select the target region.

Find the security group you want to restore rules for as the original security group, and in the **Action** column, click **Restore rules**.

In the **Restore rules** dialog box, complete the following tasks:

- i. Select the **Target security group**: Select a security group as the target security group that must have different rules from the original security group.
- ii. Select a **Restore type**:
 - i. If you want the original security group to have the same rules as the target security group, select **Completely restored**.
 - ii. If you only want to add the rules that only exist in the target security group to the original security group, select **Partially restored**.
- iii. In the **Result preview** area, preview the restoration result:
 - i. Rules highlighted in green only exist in the target security group. No matter whether you choose **Completely restored** or **Partially restored**, these rules will be added to the original security group.
 - ii. Rules highlighted in red are the rules that do not exist in the target security group. If **Completely restored** is selected, the system will remove these rules from the original security group. If **Partially restored** is selected, the rules will be retained in the original security group.

Click **OK**.

The **Restore rules** dialog box will be closed automatically after successful creation. In the **Security Group List**, find the original security group you just restored the rules for. In the **Action** column, click **Configure Rules** to enter the **Security Group Rules** page and view the updated security group rules.

Key pairs

Introduction to SSH key pairs

Alibaba Cloud offers two authentication methods for remote logon to ECS instances:

- Password logon: A standard authentication method using the administrator password. It applies to both Windows instances and Linux instances.
- SSH key pair logon: This method only applies to Linux instances. If you are running Linux, it is

recommended that you choose this authentication method to protect your ECS instance's security.

What is an SSH key pair?

An SSH key pair is a pair of keys generated through an encryption algorithm: one key is intentionally available, known as the **public key**, and the other key is kept confidential, known as the **private key**.

If you have placed the public key in a Linux instance, you can use the private key to log on to the instance through using SSH commands or related tools from local computer or another instance, without the need to enter a password.

Advantages

SSH key pairs have the following advantages:

High security:

- The security strength of a key pair is much higher than that of user passwords. A key pair can hinder brute force password-cracking attacks.
- It is impossible to deduce the private key even if the public key is maliciously acquired.

Ease-of-use: You can log on to the instance remotely through simple configuration in the console and on the local client. No password is required for the logon next time. If you need to maintain multiple ECS instances in batch, this logon method is recommended.

Alibaba Cloud SSH key pairs

To generate key pairs, you can use either of the following methods:

- Use Alibaba Cloud to generate key pairs. Alibaba Cloud uses 2048-bit RSA keys by default.
- Import the public key of a key pair that has been generated by another key pair generation tool. The key pair must be one of the following types:
 - 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

If you generate your key pair using Alibaba Cloud, you must download the private key immediately after the creation and keep it secure. If you do not have the private key, you will not be able to log on to the ECS instance that is bound to this key pair.

You can allocate a key pair to an instance when you create the Linux instance, or you can allocate it after the instance is created.

If you use an SSH key pair to log on to a Linux instance, password authentication is disabled by default to improve instance security.

Limitations for using SSH keys are as follows:

- Only Linux instances are supported. Windows instances are not supported.
- An account can have a maximum of 500 key pairs in a region.
- A Linux instance can only bind one SSH key pair. If your instance has already bound a key pair, the new key pair will replace the original key pair.
- Within the lifecycle of a Linux instance, you can re-bind the SSH key pair and instance. After re-binding, the key pair will take effect without the need to restart the instance.
- All instances of any instance type family, except those I/O optimized instances of Generation I, support SSH key pairs.

Operate an SSH key pair

- If you do not have an SSH key pair, you can create an SSH key pair.
- If you have had an SSH key pair generated by another tool, you can import an SSH key pair.
- If you no longer require an SSH key pair, you can delete an SSH key pair.
- If you want to enable or disable SSH key pair authentication for logging on to a Linux ECS instance, you can bind or unbind an SSH key pair.
- Allocate an SSH key pair when creating an ECS instance.
- Log on to an instance using an SSH key pair.

Alibaba Cloud only supports the creation of 2048-bit RSA key pairs.

- Alibaba Cloud will save the public key of the key pair.
- After the key pair is created successfully, you must download the private key in order to successfully access the instance.
- The private key follows the unencrypted PEM-encoded PKCS#8 format.

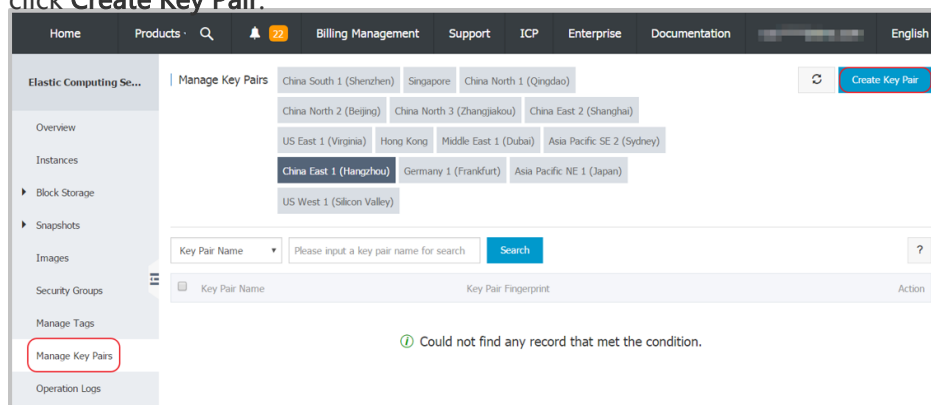
An account can have a maximum of 500 key pairs per region.

To create an SSH key pair, use the following steps.

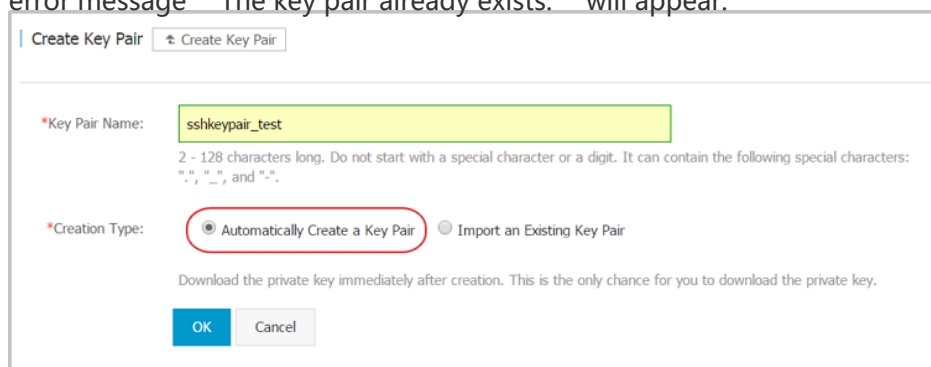
Log on to the ECS console.

In the left-side navigation pane, click **Manage Key Pairs**.

On the **Manage Key Pairs** page, select the region in which to create a key pair, and then click **Create Key Pair**.

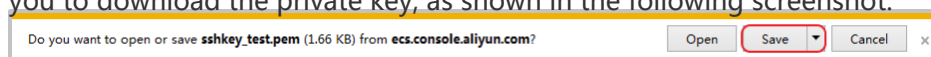


On the **Create Key Pair** page, enter a name for the key pair, and select **Automatically Create a Key Pair**. The specified key pair name must not be the same with that of an existing key pair or a key pair that was deleted when it was still bound to an instance. Otherwise, an error message “The key pair already exists.” will appear.



Click **OK** to start creating a key pair.

Download the private key. In some browsers, a pop-up window may be displayed to notify you to download the private key, as shown in the following screenshot.



Note: Alibaba Cloud will not save your private key. You must download your private key immediately after creation of the key pair, and keep it secure. If you do not have the private key, you cannot ever log on to your ECS instance that is bound to this key pair.

After creation, you can view the information, including **Key Pair Name** and **Key Pair Fingerprint**, in the key pair list.

Import an SSH key pair

If you prefer to use another key generation tool, you can use it to generate an RSA key pair and then import its public key into Alibaba Cloud. Refer to **Introduction to SSH key pairs** for the supported types of imported key pairs.

Note: To ensure your instance's security, keep the private key of the key pair secure and do not import the private key to Alibaba Cloud.

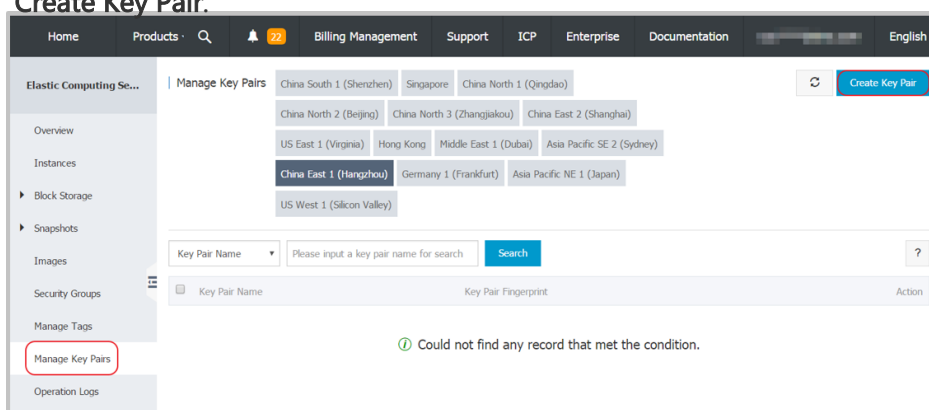
To import an SSH key pair, you must have a key pair that has been generated using another tool, and the public key to be imported into Alibaba Cloud must be Base64-encoded.

To import an SSH key pair, use the following steps.

Log on to the ECS console.

In the navigation pane, click **Manage Key Pair**.

On the **Manage Key Pair** page, select the region in which to import a key pair, and then click **Create Key Pair**.



On the **Create Key Pair** page, enter a name for the key pair, and select **Import an Existing Key Pair**. The specified key pair name must not be the same with that of an existing key pair or a key pair that was deleted when it was still bound to an instance. Otherwise, an error message “The key pair already exists.” will appear.

Create Key Pair ↶ Create Key Pair

*Key Pair Name:
 2 - 128 characters long. Do not start with a special character or a digit. It can contain the following special characters: ".", "-", and "_".

*Creation Type: ☐ Automatically Create a Key Pair ☒ Import an Existing Key Pair

*Public Key:

```
1 ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQChYaZ-jH00509dY0/uvHqo1zf8v39zYPBwxdNBL
KCMWA081yeVA/ZzYrAOCcQ6DjsReM5R4x7+sRgs8t8PFwbEPHwTKw0JFqpngZU2ipxg65rAc7zqs
sqysVSrz9ex1Io0pWp6020k7j4mrsUtpS3UAAqKPt0V6kdpBY0d+0yy4t1vRfswZJc5uaVmORqc
zQCr-iQKoIBIVH1fh1HAzFtsvTttXNASwUjOW1Ptq9i10nef0FOU95wLbf8tmxhLkdXeyD0e8bmPq
zjLlrMkoDcQEY4usqS+FWD8zs01UAo9ntGGBFQm+iLCx56Z4HEqIwH0tdc2ZF4rUV0uLUp1KD535
imported-openssh-key
```

(Base64 encoding) Import example

OK Cancel

Click **OK**.

After creation, you can view the information, including **Key Pair Name** and **Key Pair Fingerprint**, in the key pair list.

Bind or unbind an SSH key pair

You can bind a key pair to a Linux instance. One instance can only bind one key pair. You can also unbind a key pair from an instance. The unbinding will take effect in real time.

- If the instance is running, the bound key pair will take effect immediately without the need to restart the instance.
- If your instance is using password-based authentication, Alibaba Cloud will automatically disable the password authentication feature after the key pair is bound. After a key pair is unbound, you need to **reset the instance password** and restart the your instance on the console for successful login.
- If your instance is stopped, the bound key pair will take effect after the instance is restarted.
- If the instance already has a key pair, it will be replaced with the new key pair.

Bind an SSH key pair

To bind an SSH key pair to an ECS instance, use the following steps.

Log on to the ECS console.

In the left-side navigation pane, click **Manage Key Pairs**.

On the **Manage Key Pairs** page, select your desired region, find your desired key pair, and then click **Bind Instance**.

On the **Bind Instance** dialog, select your desired instances in the **Select Instance** box, and then click the > icon.

Note: In the **Select Instance** box, the instance names in gray are either non-I/O-optimized instances of Generation I or Windows instances. Those instances do not support SSH key pairs.

Click **OK**.

Unbind an SSH key pair

To unbind an SSH key pair from an ECS instance, use the following steps.

Log on to the ECS console.

In the left-side navigation pane, click **Manage Key Pairs**.

On the **Manage Key Pairs** page, select your desired region, find your desired key pair, and then click **Unbind Instance**.

On the **Unbind Instance** dialog, select your desired instances in the **Select Instance** box, and then click the > icon.

Click **OK**.

Delete an SSH key pair

If you no longer require a key pair, you can delete it. Note that a deleted key pair is not recoverable. Existing instances that have used the key pair will not be affected, and the deleted key pair name will

remain associated to the instance.

To delete one or more key pairs, use the following steps.

Log on to the ECS console.

In the left-side navigation pane, click **Manage Key Pair**.

Select your desired key pair, or key pairs, to be deleted.

Click **Delete**.

Note:

- If you delete a key pair that is still bound to an instance, its name is not available for you to create or import a key pair again. Otherwise, an error message "The key pair already exists." will appear when you are using the same name to create or import a key pair.
- If you delete a key pair that is not bound to an instance, its name is still available for you to create or import a key pair again.

Tags

As of January 2017, the ECS Management Console only supports adding tags to instances.

Each tag is composed of a key-value pair.

Up to 10 tags can be bound to a single instance.

For a single instance, the tag key for each tag must be unique. Tags with identical tag keys will be overwritten.

Tag information cannot be transmitted across regions.

A tag will be automatically deleted when it is not bound to any instance.

If you have a large number of instances, you can add tags to them to facilitate unified management and allow you to group different instances easily.

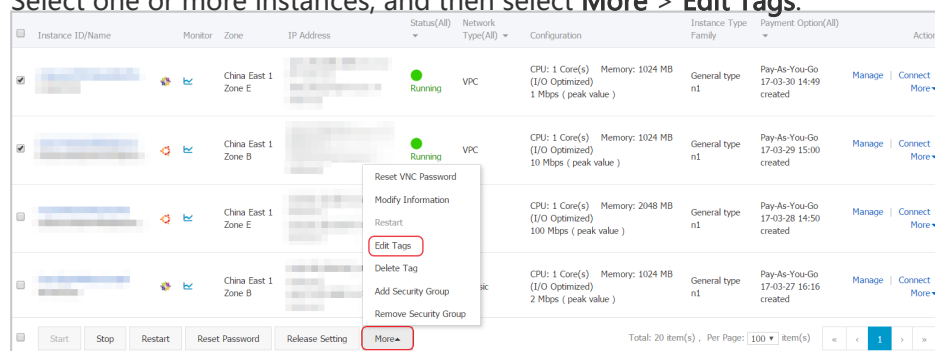
To add a tag to an instance, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Instances**.

Select a region.

Select one or more instances, and then select **More > Edit Tags**.



In the dialog box, click **Create**.

Enter your custom tag key and value, and then click **Confirm**.

Your tag will then be added to the available tag selection. Click **Confirm** to add the created tag to the instance.

Delete a tag

You can delete tags if you no longer require them.

To delete a tag, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Instances**.

Select a region.

Select the instance you want to remove tag from.

Click **More > Edit Tag**.

In the dialog box, enter the tag key you want to delete, and then click **OK**.

Click **OK** to delete the tag.

You can search for instances by tags.

To search for instances by tags, perform the following:

Log on to the ECS console.

In the left-side navigation pane, click **Instances**.

Select a region.

Click **Tag** above the instance list.

Select the tag key you want to use for the instance search.

The instances that use the specified tag are displayed.

Monitoring

You can monitor the operating statuses of instances to ensure optimal performance.

You can monitor the status of instances using the following two portals:

- Instance Details page
- CloudMonitor

Monitor the status of an instance using Instance Details page

To monitor the Status of an instance using Instance Details, perform the following:

Log on to the ECS console.

Click **Instances** in the left navigation bar.

Select your desired region.

Click the instance you want to monitor.

On the **Instance Details** page, you can view the monitoring information, including CPU usage and outbound/inbound network traffic information.

- Information about CPU monitoring:
 - For Linux instances, use the `top` command to view CPU usage details. Log on to the instance and execute the `top` command in the command line. Then, press `Shift+P` key to list programs by CPU usage to view which processes are using the most CPU resources.
 - For Windows instances, use the **Task Manager** on an instance to view the CPU usage to view which programs are using the CPU resources of the server.
- The displayed monitoring data shows the Internet traffic of the instance in Kbps (1 Mbps = 1,024 Kbps). The monitoring data shows inbound and outbound instance traffic. For 1 Mbps of bandwidth, the bandwidth is working at full capacity if the outbound network traffic reaches 1,024 Kbps.

CloudMonitor

To install a CloudMonitor, perform the following:

On the Alibaba Cloud Console, select **Products and Services > CloudMonitor**.

In the left-side navigation pane, click **Host Monitoring**, and then select the name of the instance you wish to monitor.

Click **Click to Install** to monitor the instance OS. Click **Monitoring Chart** to view basic parameters. Click **Alarm Rules** to set alarm rules.

Note: For more information about CloudMonitor, see [CloudMonitor Product Documentation](#).