

Alibaba Cloud Elastic Compute Service

Instances

Issue: 20190920

Legal disclaimer

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

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

or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.

5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).
6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

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

Style	Description	Example
<code>{}</code> or <code>{a b}</code>	It indicates that it is a required value, and only one item can be selected.	<code>switch {stand slave}</code>

Contents

Legal disclaimer.....	I
Generic conventions.....	I
1 What are ECS instances?.....	1
2 ECS instance life cycle.....	2
3 Instance families.....	6
4 Instance type families.....	51
4.1 General purpose.....	51
4.2 Compute optimized.....	56
4.3 Memory optimized.....	63
4.4 Big data instance type families.....	72
4.5 Local SSD-equipped instance type families.....	73
4.6 High clock speed instance type families.....	75
4.7 Compute optimized type family with GPU.....	75
4.7.1 Compute optimized instance families with GPU capabilities.....	75
4.7.2 Create a compute optimized instance with GPU.....	85
4.7.3 Install the GPU driver.....	101
4.7.4 Uninstall the GPU driver.....	103
4.7.5 Install a GRID driver in a GPU-equipped ECS instance (Linux).....	105
4.7.6 Install an NVIDIA GRID driver on a vgn5i instance (Windows).....	112
4.7.7 Install an NVIDIA GRID driver on a vgn5i instance (Linux).....	113
4.8 Visualization compute type family with GPU.....	116
4.8.1 GPU-equipped instance type families designed for visualization computing.....	116
4.8.2 Create a ga1 instance.....	117
4.9 Compute optimized type family with FPGA.....	118
4.9.1 Compute optimized and FPGA-equipped instance type families....	118
4.9.2 Create an f1 instance.....	118
4.9.3 Create an f3 instance.....	119
4.10 ECS bare metal instance type family	120
4.10.1 What is an ECS Bare Metal Instance?.....	120
4.10.2 Create an ECS Bare Metal instance.....	131
4.10.3 Install SGX.....	132
4.11 Super Computing Cluster instance type family.....	134
4.11.1 What is a Super Computing Cluster?.....	134
4.11.2 Create an SCC instance.....	140
4.12 Burstable performance instances.....	142
4.12.1 Overview.....	142
4.12.2 Benefits.....	149
4.12.3 CPU credit change examples.....	150

4.12.4 Billing.....	157
4.12.5 Manage performance modes of a burstable performance instance.....	163
4.12.6 Monitor a burstable performance instance.....	164
4.12.7 View bills of a burstable performance instance.....	165
4.13 Entry-level instance type families.....	166
4.14 Phased-out instance types.....	167
5 Instance purchasing options.....	176
5.1 Subscription.....	176
5.2 Pay-As-You-Go.....	182
5.3 Preemptible instances.....	187
5.3.1 Preemptible instances.....	187
5.3.2 Create a preemptible instance.....	195
5.3.3 Stop a preemptible instance.....	198
5.4 Reserved Instances.....	199
5.4.1 Reserved instance overview.....	199
5.4.2 Matching rules of Reserved Instances.....	203
5.4.3 Purchase reserved instances.....	205
5.4.4 Manage Reserved Instances.....	207
5.5 Switch billing method.....	210
5.5.1 Switch from Pay-As-You-Go to Subscription billing.....	210
5.5.2 Switch the billing method from Subscription to Pay-As-You-Go....	212
6 Create an instance.....	215
6.1 ECS instance creation overview.....	215
6.2 Create an instance by using the wizard.....	215
6.3 Create an instance by using a custom image.....	221
6.4 Create an instance of the same configuration.....	222
6.5 Create an instance by using a launch template.....	223
7 Connect to instances.....	224
7.1 Overview.....	224
7.2 Connect to Linux instances.....	226
7.2.1 Connect to a Linux instance by using the Management Terminal...	226
7.2.2 Connect to a Linux instance by using an SSH key pair.....	229
7.2.3 Connect to a Linux instance by using a password.....	237
7.2.4 Connect to an instance on a mobile device.....	240
7.3 Connect to Windows instances.....	244
7.3.1 Connect to a Windows instance by using the Management Terminal.....	244
7.3.2 Connect to a Windows instance.....	247
7.3.3 Connect to a Windows instance from a mobile device.....	251
8 Manage instances.....	253
8.1 Start or stop an instance.....	253
8.2 Restart an instance.....	256
8.3 Release an instance.....	257

8.4 Reactivate an instance.....	260
8.5 Check instance information.....	261
8.6 Reset an instance logon password.....	263
8.7 Change the instance logon password by connecting to the instance.....	267
8.8 Enable instance release protection.....	268
8.9 User-defined data and metadata.....	272
8.9.1 Metadata.....	272
8.9.2 Retrieve instance metadata.....	276
8.10 User-defined data.....	278
8.10.1 User data.....	278
8.10.2 Manage user data.....	285
8.10.3 Customize yum repository, NTP service, and DNS service.....	288
8.10.4 Create a new account with the root user privilege.....	291
8.11 Instance identity.....	293
8.12 Manage software on Linux instances.....	298
8.12.1 Add software repositories.....	298
8.12.2 Search for software packages.....	302
8.12.3 Install software packages.....	306
8.12.4 Update software.....	307
8.13 Configure time.....	310
8.13.1 Time setting: NTP servers and other public services.....	310
8.13.2 Time setting: Synchronize NTP servers for Windows instances....	311
8.13.3 Time setting: Synchronize NTP servers and change time zone for Linux instances.....	313
9 Renew Subscription instances.....	316
9.1 Renewal overview.....	316
9.2 Manual renewal.....	317
9.3 Auto-renewal.....	318
9.4 Renew for configuration downgrade.....	320
10 Change configurations.....	323
10.1 Overview of configuration changes.....	323
10.2 Instance type families that support instance type upgrades.....	327
10.3 Upgrade configurations.....	329
10.3.1 Upgrade configurations of Subscription instances.....	329
10.4 Downgrade configurations.....	332
10.4.1 预付费实例续费降配.....	332
10.4.2 Downgrade configurations of Subscription instances.....	334
10.4.3 Downgrade bandwidth configurations of Subscription instances.....	335
10.5 Change configurations of Pay-As-You-Go instances.....	336
10.5.1 Change configurations of Pay-As-You-Go instances.....	336
10.5.2 Change Internet bandwidth.....	337
11 ECS instance FAQ.....	339

1 What are ECS instances?

An ECS instance is a virtual computing environment that includes CPU, memory, operating system, bandwidth, disks, and other basic computing components.

An ECS instance is an independent virtual machine, and is the core element of ECS . Other resources, such as disks, IPs, images, and snapshots can only be used in conjunction with an ECS instance.

2 ECS instance life cycle

The life cycle of an ECS instance begins when the instance is created and ends when the instance is released.

Instance status

During its life cycle, an ECS instance may undergo several status changes, as described in the following table.

Status	Status attribute	Description	API status	Visible in the console?
Preparing	Intermediate	After an instance is created, it is in this state before it enters the Running state. If the instance is in this state for a long time, it means that an exception has occurred.	Pending	No
Starting	Intermediate	After an instance is started or restarted in the console or through an API, the instance is in this state before it enters the Running state. If the instance is in this state for a long time, it means that an exception has occurred.	Starting	Yes
Running	Stable	The instance is operating normally, and you can run your services.	Running	Yes
Expiring	Stable	A Subscription instance is in this state for 15 days before it expires. We recommend that you Renew the instance .	Running	Yes
Stopping	Intermediate	After an instance is stopped in the console or through an API, the instance is in this state before it enters the Stopped state. If the instance is in this state for a long time, it means that an exception has occurred.	Stopping	Yes
Stopped	Stable	An instance is in this state after it has been created but has not been started yet, or when it has been stopped due to normal operations. An instance in this state cannot provide external services.	Stopped	Yes

Status	Status attribute	Description	API status	Visible in the console?
Expired	Stable	A Subscription instance enters the Expired state when it expires. A Pay-As-You-Go instance enters this state if you have an overdue payment under your account. An instance in this state cannot provide external services. For information about resource status changes, see #unique_6 and #unique_7 .	Stopped	Yes
Recycling Upon Expiration	Stable	<p>Within 15 days after a VPC Subscription instance expires or is stopped due to an overdue payment, the instance stays in the Expired state for a period of time, and then enters the Recycling Upon Expiration state.</p> <ul style="list-style-type: none"> • You can renew the instance before it enters the Recycling Upon Expiration state. If the renewal is successful, all resources are retained without being affected. • After the instance enters the Recycling Upon Expiration state, its computing resources (vCPU + memory) and local disk data are no longer retained, but its cloud disks and assigned Internet IP address are retained. You can renew the instance. If the renewal is successful, the cloud disks and the assigned Internet IP address remain the same. 	Stopped	Yes

Status	Status attribute	Description	API status	Visible in the console?
Overdue and Being Recycled	Stable	<p>Within 15 days after a VPC Pay-As-You-Go instance expires or is stopped due to an overdue payment, the instance stays in the Expired state for a period of time, and then enters the Overdue and Being Recycled state.</p> <ul style="list-style-type: none"> You can add funds to your account and Restart the instance before the instance enters the Overdue and Being Recycled state. If the restart is successful, all resources are retained without being affected. After the instance enters the Overdue and Being Recycled state, its computing resources (vCPU + memory) and local disk data are no longer retained, but its cloud disks and assigned Internet IP address are retained. You can add funds to your account and restart the instance. If the instance fails to be restarted, try again later or open a ticket. When you restart the instance successfully, the cloud disks and the assigned Internet IP address remain the same. 	Stopped	Yes
Locked	Stable	An instance is in this state if you have an overdue payment under your account or your account is insecure. You can open a ticket to unlock the instance.	Stopped	Yes
Release pending	Stable	A Subscription instance is in this state after you apply for a refund before the instance expires.	Stopped	Yes

API status

You can call [DescribeInstanceStatus](#) or [DescribeInstances](#) to view the API status of an instance. The following figure shows the status conversions described in this topic.

3 Instance families

This topic describes available ECS instance families and their features, specifications, and application scenarios.

An ECS instance is the smallest unit that can provide compute capabilities and services for your business.

ECS instances are categorized into different instance families based on the business scenarios to which they can be applied. Each instance family is divided into different instance types based on their CPU and memory specifications. *ECS instance type* defines the basic properties of an ECS instance: CPU (including CPU model and clock speed) and memory. In addition to the instance type, you must also configure the block storage, image, and network type when you create an instance.



Note:

The available instance families and types varies from region to region. You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.

Enterprise scenarios have high business stability requirements. Alibaba Cloud ECS instance families are divided into enterprise-level and entry-level based on whether the instance families are suitable for enterprise scenarios. Enterprise-level instance families offer consistent performance and dedicated resources. In enterprise-level instance families, each vCPU corresponds to an Intel Xeon core hyperthread.



Note:

For the differences between enterprise-level and entry-level instance families, see [#unique_12](#).

You can upgrade or downgrade instances within or between certain instance families. For such instance families and the corresponding upgrade and downgrade rules, see [#unique_13](#).

Some instance families are no longer available for purchase. For more information, see [Phased-out instance types](#).


Alibaba Cloud ECS instances are categorized into the following instance families:

- **Enterprise-level computing instance families based on the x86 architecture:**
 - g5, general purpose instance family
 - g6, general purpose instance family
 - sn2ne, general purpose instance family with enhanced network performance
 - ic5, compute intensive instance family
 - c5, compute optimized instance family
 - c6, compute optimized instance family
 - sn1ne, compute optimized instance family with enhanced network performance
 - r5, memory optimized instance family
 - r6, memory optimized instance family
 - re4, memory optimized instance family with enhanced performance
 - re4e, memory optimized instance family with enhanced performance
 - se1ne, memory optimized instance family with enhanced network performance
 - se1, memory optimized instance family
 - d1ne, big data instance family with enhanced network performance
 - d1, big data instance family
 - i2, instance family with local SSDs
 - i2g, instance family with local SSDs
 - i1, instance family with local SSDs
 - hfc5, compute optimized instance family with high clock speed
 - hfg5, general purpose instance family with high clock speed
- **Enterprise-level heterogeneous computing instance families:**
 - vgn5i, lightweight compute optimized instance family with GPU capabilities
 - gn6i, compute optimized instance family with GPU capabilities
 - gn6v, compute optimized instance family with GPU capabilities
 - gn5, compute optimized instance family with GPU capabilities
 - gn5i, compute optimized instance family with GPU capabilities
 - gn4, compute optimized family with GPU capabilities
 - ga1, visualization and compute optimized instance family with GPU capabilities
 - f1, compute optimized instance family with FPGAs
 - f3, compute optimized instance family with FPGAs

- ECS Bare Metal Instance families and Super Computing Cluster (SCC) instance families:
 - [ebmc5s](#), compute optimized ECS Bare Metal Instance family with enhanced network performance
 - [ebmg5s](#), general purpose ECS Bare Metal Instance family with enhanced network performance
 - [ebmr5s](#), memory optimized ECS Bare Metal Instance family with enhanced network performance
 - [ebmhfg5](#), ECS Bare Metal Instance family with high clock speed
 - [ebmc4](#), compute optimized ECS Bare Metal Instance family
 - [ebmg5](#), general purpose ECS Bare Metal Instance family
 - [scch5](#), SCC instance family with high clock speed
 - [sccg5](#), general purpose SCC instance family
 - [sccgn6](#), compute optimized SCC instance family with GPU capabilities
- Entry-level computing instance families based on the x86 architecture:
 - [t5](#), burstable instance family
 - [xn4](#), previous-generation entry-level computing instance family
 - [n4](#), previous-generation entry-level computing instance family
 - [mn4](#), previous-generation entry-level computing instance family
 - [e4](#), previous-generation entry-level computing instance family

[g5](#), general purpose instance family

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:4
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Provides strong network performance based on sufficient computing capacity

- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Enterprise-level applications of various types and sizes
 - Small and medium-sized database systems, cache, and search clusters
 - Data analysis and computing
 - Computing clusters and memory-intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.g5.large	2	8.0	N/A	1.0	300	Yes	2	2	6
ecs.g5.xlarge	4	16.0	N/A	1.5	500	Yes	2	3	10
ecs.g5.2xlarge	8	32.0	N/A	2.5	800	Yes	2	4	10
ecs.g5.3xlarge	12	48.0	N/A	4.0	900	Yes	4	6	10
ecs.g5.4xlarge	16	64.0	N/A	5.0	1,000	Yes	4	8	20
ecs.g5.6xlarge	24	96.0	N/A	7.5	1,500	Yes	6	8	20
ecs.g5.8xlarge	32	128.0	N/A	10.0	2,000	Yes	8	8	20
ecs.g5.16xlarge	64	256.0	N/A	20.0	4,000	Yes	16	8	20




Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.

- For more information about these specifications, see [Instance specification metrics](#).

g6, general purpose instance family

Features

- Provides predictable stable and high performance and reduces virtualization overheads with the use of the X-Dragon architecture
- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:4
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Second Generation Intel Xeon Scalable processors, up to 3.2 GHz Turbo Boost
- Provides strong network performance based on sufficient computing capacity
- Does not support instance upgrades or downgrades
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Enterprise-level applications of various types and sizes
 - Websites and application servers
 - Game servers
 - Small and medium-sized database systems, cache, and search clusters
 - Data analysis and computing
 - Computing clusters and memory-intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burst bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI	Disk IOPS	Disk bandwidth (Gbit/s)
ecs.g6.large	2	8.0	N/A	1.0	3.0	300	Yes	2	2	6	10,000	1
ecs.g6.xlarge	4	16.0	N/A	1.5	5.0	500	Yes	4	3	10	20,000	1.5
ecs.g6.2xlarge	8	32.0	N/A	2.5	8.0	800	Yes	8	4	10	25,000	2
ecs.g6.3xlarge	12	48.0	N/A	4.0	10.0	900	Yes	8	6	10	30,000	2.5
ecs.g6.4xlarge	16	64.0	N/A	5.0	10.0	1,000	Yes	8	8	20	40,000	3
ecs.g6.6xlarge	24	96.0	N/A	7.5	10.0	1,500	Yes	12	8	20	50,000	4
ecs.g6.8xlarge	32	128.0	N/A	10.0	N/A	2,000	Yes	16	8	20	60,000	5
ecs.g6.13xlarge	52	192.0	N/A	12.5	N/A	3,000	Yes	32	7	20	100,000	8
ecs.g6.26xlarge	104	384.0	N/A	25.0	N/A	6,000	Yes	32	15	20	200,000	16



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

sn2ne, general purpose instance family with enhanced network performance

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:4
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors
- Provides strong network performance based on sufficient computing capacity
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Enterprise-level applications of various types and sizes
 - Small and medium-sized database systems, cache, and search clusters
 - Data analysis and computing
 - Computing clusters and memory-intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.sn2ne.large	2	8.0	N/A	1.0	300	Yes	2	2	6

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.sn2ne.xlarge	4	16.0	N/A	1.5	500	Yes	2	3	10
ecs.sn2ne.2xlarge	8	32.0	N/A	2.0	1,000	Yes	4	4	10
ecs.sn2ne.3xlarge	12	48.0	N/A	2.5	1,300	Yes	4	6	10
ecs.sn2ne.4xlarge	16	64.0	N/A	3.0	1,600	Yes	4	8	20
ecs.sn2ne.6xlarge	24	96.0	N/A	4.5	2,000	Yes	6	8	20
ecs.sn2ne.8xlarge	32	128.0	N/A	6.0	2,500	Yes	8	8	20
ecs.sn2ne.14xlarge	56	224.0	N/A	10.0	4,500	Yes	14	8	20


**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ic5, compute-intensive instance family

Features

- I/O optimized

- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:1
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Web front-end servers
 - Data analysis, batch compute, and video encoding
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Massively Multiplayer Online (MMO) game frontends

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ic5.large	2	2.0	N/A	1.0	300	No	2	2	6
ecs.ic5.xlarge	4	4.0	N/A	1.5	500	No	2	3	10
ecs.ic5.2xlarge	8	8.0	N/A	2.5	800	No	2	4	10
ecs.ic5.3xlarge	12	12.0	N/A	4.0	900	No	4	6	10
ecs.ic5.4xlarge	16	16.0	N/A	5.0	1,000	No	4	8	20




Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

c5, compute optimized instance family

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:2
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Web front-end servers
 - Massively Multiplayer Online (MMO) game frontends
 - Data analysis, batch compute, and video encoding
 - High-performance science and engineering applications

Instance types


Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.c5.large	2	4.0	N/A	1.0	300	Yes	2	2	6
ecs.c5.xlarge	4	8.0	N/A	1.5	500	Yes	2	3	10
ecs.c5.2xlarge	8	16.0	N/A	2.5	800	Yes	2	4	10

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.c5.3xlarge	12	24.0	N/A	4.0	900	Yes	4	6	10
ecs.c5.4xlarge	16	32.0	N/A	5.0	1,000	Yes	4	8	20
ecs.c5.6xlarge	24	48.0	N/A	7.5	1,500	Yes	6	8	20
ecs.c5.8xlarge	32	64.0	N/A	10.0	2,000	Yes	8	8	20
ecs.c5.16xlarge	64	128.0	N/A	20.0	4,000	Yes	16	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

c6, compute optimized instance family**Features**

- Provides predictable stable and high performance and reduces virtualization overheads with the use of the X-Dragon architecture
- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:2
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Second Generation Intel Xeon Scalable processors, up to 3.2 GHz Turbo Boost
- Provides strong network performance with large computing capacity

- Does not support instance upgrades or downgrades
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Web front-end servers
 - Massively Multiplayer Online (MMO) game frontends
 - Data analysis, batch compute, and video encoding
 - High-performance science and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burst bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI	Disk IOPS	Disk bandwidth (Gbit/s)
ecs.c6.large	2	4.0	N/A	1.0	3.0	300	Yes	2	2	6	10,000	1
ecs.c6.xlarge	4	8.0	N/A	1.5	5.0	500	Yes	4	3	10	20,000	1.5
ecs.c6.2xlarge	8	16.0	N/A	2.5	8.0	800	Yes	8	4	10	25,000	2
ecs.c6.3xlarge	12	24.0	N/A	4.0	10.0	900	Yes	8	6	10	30,000	2.5
ecs.c6.4xlarge	16	32.0	N/A	5.0	10.0	1,000	Yes	8	8	20	40,000	3
ecs.c6.6xlarge	24	48.0	N/A	7.5	10.0	1,500	Yes	12	8	20	50,000	4

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burst bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI	Disk IOPS	Disk bandwidth (Gbit/s)
ecs.c6.8xlarge	32	64.0	N/A	10.0	N/A	2,000	Yes	16	8	20	60,000	5
ecs.c6.13xlarge	52	96	N/A	12.5	N/A	3,000	Yes	32	7	20	100,000	8
ecs.c6.26xlarge	104	192.0	N/A	25.0	N/A	6,000	Yes	32	15	20	200,000	16

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

sn1ne, compute optimized instance family with enhanced network performance

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:2
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors
- Provides strong network performance with large computing capacity

- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Web front-end servers
 - Massively Multiplayer Online (MMO) game frontends
 - Data analysis, batch compute, and video encoding
 - High-performance science and engineering applications


Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.sn1ne.large	2	4.0	N/A	1.0	300	Yes	2	2	6
ecs.sn1ne.xlarge	4	8.0	N/A	1.5	500	Yes	2	3	10
ecs.sn1ne.2xlarge	8	16.0	N/A	2.0	1,000	Yes	4	4	10
ecs.sn1ne.3xlarge	12	24.0	N/A	2.5	1,300	Yes	4	6	10
ecs.sn1ne.4xlarge	16	32.0	N/A	3.0	1,600	Yes	4	8	20
ecs.sn1ne.6xlarge	24	48.0	N/A	4.5	2,000	Yes	6	8	20
ecs.sn1ne.8xlarge	32	64.0	N/A	6.0	2,500	Yes	8	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

r5, memory optimized instance family**Features**

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - High-performance and in-memory databases
 - Data analysis and mining, and distributed memory cache
 - Hadoop, Spark, and other memory-intensive enterprise applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.r5.large	2	16.0	N/A	1.0	300	Yes	2	2	6
ecs.r5.xlarge	4	32.0	N/A	1.5	500	Yes	2	3	10
ecs.r5.2xlarge	8	64.0	N/A	2.5	800	Yes	2	4	10

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.r5.3xlarge	12	96.0	N/A	4.0	900	Yes	4	6	10
ecs.r5.4xlarge	16	128.0	N/A	5.0	1,000	Yes	4	8	20
ecs.r5.6xlarge	24	192.0	N/A	7.5	1,500	Yes	6	8	20
ecs.r5.8xlarge	32	256.0	N/A	10.0	2,000	Yes	8	8	20
ecs.r5.16xlarge	64	512.0	N/A	20.0	4,000	Yes	16	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

r6, memory optimized instance family**Features**

- Provides predictable stable and high performance and reduces virtualization overheads with the use of the X-Dragon architecture
- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Second Generation Intel Xeon Scalable processors, up to 3.2 GHz Turbo Boost
- Provides strong network performance with large computing capacity
- Does not support instance upgrades or downgrades

- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - High-performance and in-memory databases
 - Data analysis and mining, and distributed memory cache
 - Hadoop, Spark, and other memory-intensive enterprise applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burst bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI	Disk IOPS	Disk bandwidth (Gbit/s)
ecs.r6.large	2	16.0	N/A	1.0	3.0	300	Yes	2	2	6	10,000	1
ecs.r6.xlarge	4	32.0	N/A	1.5	5.0	500	Yes	4	3	10	20,000	1.5
ecs.r6.2xlarge	8	64.0	N/A	2.5	8.0	800	Yes	8	4	10	25,000	2
ecs.r6.3xlarge	12	96.0	N/A	4.0	10.0	900	Yes	8	6	10	30,000	2.5
ecs.r6.4xlarge	16	128.0	N/A	5.0	10.0	1,000	Yes	8	8	20	40,000	3
ecs.r6.6xlarge	24	192.0	N/A	7.5	10.0	1,500	Yes	12	8	20	50,000	4
ecs.r6.8xlarge	32	256.0	N/A	10.0	N/A	2,000	Yes	16	8	20	60,000	5

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burst bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI	Disk IOPS	Disk bandwidth (Gbit/s)
ecs.r6.13xlarge	52	384	N/A	12.5	N/A	3,000	Yes	32	7	20	100,000	8
ecs.r6.26xlarge	104	768.0	N/A	25.0	N/A	6,000	Yes	32	15	20	200,000	16

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

re4, memory optimized instance family with enhanced performance

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Optimized for high-performance databases, in-memory databases, and other memory-intensive enterprise applications
- Equipped with 2.2 GHz Intel Xeon E7 8880 v4 (Broadwell) processors, up to 2.4 GHz Turbo Boost
- CPU-to-memory ratio of 1:12, up to 1920.0 GiB memory
- ecs.re4.20xlarge and ecs.re4.40xlarge have been certified by SAP HANA

- Suitable for the following scenarios:
 - High-performance databases and in-memory databases (for example, SAP HANA)
 - Memory-intensive applications
 - Big data processing engines such as Apache Spark or Presto

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.re4.20xlarge	80	960.0	N/A	15.0	2,000	Yes	16	8	20
ecs.re4.40xlarge	160	1920.0	N/A	30.0	4,500	Yes	16	8	20



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

re4e, memory optimized instance family with enhanced performance

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Optimized for high-performance databases, in-memory databases, and other memory-intensive enterprise applications

Equipped with 2.2 GHz Intel Xeon E7 8880 v4 (Broadwell) processors, up to 2.4 GHz Turbo Boost

- CPU-to-memory ratio of 1:24, up to 3840.0 GiB memory
- Suitable for the following scenarios:
 - High-performance databases and in-memory databases (for example, SAP HANA)
 - Memory-intensive applications
 - Big data processing engines such as Apache Spark or Presto

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.re4e.40xlarge	160	3840.0	N/A	30.0	4,500	Yes	16	15	20



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

se1ne, memory optimized instance family with enhanced network performance

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:8
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors
- Provides strong network performance with large computing capacity

- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - High-performance and in-memory databases
 - Data analysis and mining, and distributed memory cache
 - Hadoop, Spark, and other memory-intensive enterprise applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.se1ne.large	2	16.0	N/A	1.0	300	Yes	2	2	6
ecs.se1ne.xlarge	4	32.0	N/A	1.5	500	Yes	2	3	10
ecs.se1ne.2xlarge	8	64.0	N/A	2.0	1,000	Yes	4	4	10
ecs.se1ne.3xlarge	12	96.0	N/A	2.5	1,300	Yes	4	6	10
ecs.se1ne.4xlarge	16	128.0	N/A	3.0	1,600	Yes	4	8	20
ecs.se1ne.6xlarge	24	192.0	N/A	4.5	2,000	Yes	6	8	20
ecs.se1ne.8xlarge	32	256.0	N/A	6.0	2,500	Yes	8	8	20

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.se1ne.14xlarge	56	480.0	N/A	10.0	4,500	Yes	14	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

se1, memory optimized instance family**Features**

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:8
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - High-performance and in-memory databases
 - Data analysis and mining, and distributed memory cache
 - Hadoop, Spark, and other memory-intensive enterprise applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.se1.large	2	16.0	N/A	0.5	100	No	1	2	6
ecs.se1.xlarge	4	32.0	N/A	0.8	200	No	1	3	10
ecs.se1.2xlarge	8	64.0	N/A	1.5	400	No	1	4	10
ecs.se1.4xlarge	16	128.0	N/A	3.0	500	No	2	8	20
ecs.se1.8xlarge	32	256.0	N/A	6.0	800	No	3	8	20
ecs.se1.14xlarge	56	480.0	N/A	10.0	1,200	No	4	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

vgn5i, lightweight compute optimized instance family with GPU capabilities

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Uses NVIDIA P4 GPU computing accelerators

- Contains virtual GPUs (vGPUs), which are the result of GPU virtualization with mediated pass-through
 - Supports the 1/8, 1/4, 1/2, and 1/1 computing capacity of NVIDIA® Tesla® P4 GPUs
 - Supports 1, 2, 4, and 8 GB of GPU memory
- CPU-to-memory ratio of 1:3
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Real-time rendering for cloud gaming
 - Real-time rendering for AR and VR applications
 - AI (deep learning and machine learning) inference for the elastic deployment of Internet services
 - Educational environment of deep learning
 - Modeling experiment environment of deep learning

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs
ecs.vgn5i-m1.large	2	6	N/A	P4*1/8	1	1	300	Yes	2	2
ecs.vgn5i-m2.xlarge	4	12	N/A	P4*1/4	2	2	500	Yes	2	3
ecs.vgn5i-m4.2xlarge	8	24	N/A	P4 * 1/2	4	3	800	Yes	2	4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs
ecs.vgn5i-m8.4xlarge	16	48	N/A	P4*1	8	5	1,000	Yes	4	5

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

gn6i, compute optimized instance family with GPU capabilities

Features

- I/O optimized
- CPU-to-memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports standard SSDs, ultra disks, and enhanced SSDs that deliver millions of IOPS
- Achieves better performance with the new-generation X-Dragon compute architecture
- Uses NVIDIA T4 GPU computing accelerators
 - Powered by the new NVIDIA Turing architecture
 - Up to 320 Turing Tensor Cores
 - Up to 2,560 CUDA Cores
 - Mixed-precision Tensor Cores support 65 FP16 TFLOPS, 130 INT8 TOPS, and 260 INT4 TOPS
 - 16 GB memory capacity (320 GB/s bandwidth)
- Provides strong network performance with large computing capacity

- Suitable for the following scenarios:
 - AI (deep learning and machine learning) inference for computer vision, voice recognition, speech synthesis, natural language processing, machine translation , and reference systems
 - Real-time rendering for cloud gaming
 - Real-time rendering for AR and VR applications
 - Reload graphics computing or Graphics Workstation
 - GPU-accelerated databases
 - High-performance computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs
ecs.gn6i-c4g1.xlarge	4	15	N/A	T4*1	16	4	500	Yes	2	2
ecs.gn6i-c8g1.2xlarge	8	31	N/A	T4*1	16	5	800	Yes	2	2
ecs.gn6i-c16g1.4xlarge	16	62	N/A	T4*1	16	6	1,000	Yes	4	3
ecs.gn6i-c24g1.6xlarge	24	93	N/A	T4*1	16	7.5	1,200	Yes	6	4
ecs.gn6i-c24g1.12xlarge	48	186	N/A	T4*2	32	15	2,400	Yes	12	6

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs
ecs.gn6i-c24g1.24xlarge	96	372	N/A	T4*4	64	30	4,800	Yes	24	8

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

gn6v, compute optimized instance family with GPU capabilities

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks.
- Uses NVIDIA V100 GPU processors
- CPU-to-memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Uses NVIDIA V100 GPU computing accelerators (SXM2-based)
 - Powered by the innovative NVIDIA Volta architecture
 - 16 GB HBM2 GPU memory
 - 5,120 CUDA Cores
 - 640 Tensor Cores
 - Memory bandwidth of 900 GB/s
 - Supports up to six NVLink connections for a total bandwidth of 300 GB/s with 25 GB/s each
- Provides strong network performance with large computing capacity

- Suitable for the following scenarios:
 - Deep learning, image classification, autonomous vehicles, voice recognition, and other AI applications
 - Scientific computing applications, such as fluid dynamics, finance, molecular dynamics, and environmental analysis

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENIs	Private IP address of a single ENI
ecs.gn6v-c8g1.2xlarge	8	32.0	N/A	1 * NVIDIA V100	1 * 16	2.5	800	Yes	4	4	10
ecs.gn6v-c8g1.8xlarge	32	128.0	N/A	4 * NVIDIA V100	4 * 16	10.0	2,000	Yes	8	8	20
ecs.gn6v-c8g1.16xlarge	64	256.0	N/A	8 * NVIDIA V100	8 * 16	20.0	2,500	Yes	16	8	20





Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

gn5, compute optimized instance family with GPU capabilities

Features

- I/O optimized
- Supports standard SSDs and ultra disks

- Uses NVIDIA P100 GPU processors
- Multiple CPU-to-memory ratios
- High-performance local NVMe SSDs
-  
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Deep learning
 - Scientific computing applications, such as fluid dynamics, finance, genomics, and environmental analysis
 - High-performance computing, rendering, multimedia encoding and decoding, and other server-side GPU compute workloads

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs	Private IP address of a single ENI
ecs.gn5-c4g1.xlarge	4	30.0	440	1 * NVIDIA P100	1 * 16	3.0	300	No	1	3	10
ecs.gn5-c8g1.2xlarge	8	60.0	440	1 * NVIDIA P100	1 * 16	3.0	400	No	1	4	10
ecs.gn5-c4g1.2xlarge	8	60.0	880	2 * NVIDIA P100	2 * 16	5.0	1,000	No	2	4	10
ecs.gn5-c8g1.4xlarge	16	120.0	880	2 * NVIDIA P100	2 * 16	5.0	1,000	No	4	8	20

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENIs	Private IP address of a single ENI
ecs.gn5-c28g1.7xlarge	28	112.0	440	1 * NVIDIA P100	1 * 16	5.0	1,000	No	8	8	20
ecs.gn5-c8g1.8xlarge	32	240.0	1,760	4 * NVIDIA P100	4 * 16	10.0	2,000	No	8	8	20
ecs.gn5-c28g1.14xlarge	56	224.0	880	2 * NVIDIA P100	2 * 16	10.0	2,000	No	14	8	20
ecs.gn5-c8g1.14xlarge	54	480.0	3,520	8 * NVIDIA P100	8 * 16	25.0	4,000	No	14	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

gn5i, compute optimized instance family with GPU capabilities

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Uses NVIDIA P4 GPU processors
- CPU-to-memory ratio of 1:4

- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Deep learning inference
 - Multimedia encoding and decoding, and other server-side GPU compute workloads

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs	Private IP address of a single ENI
ecs.gn5i-c2g1.large	2	8.0	N/A	1 * NVIDIA P4	1 * 8	1.0	100	Yes	2	2	6
ecs.gn5i-c4g1.xlarge	4	16.0	N/A	1 * NVIDIA P4	1 * 8	1.5	200	Yes	2	3	10
ecs.gn5i-c8g1.2xlarge	8	32.0	N/A	1 * NVIDIA P4	1 * 8	2.0	400	Yes	4	4	10
ecs.gn5i-c16g1.4xlarge	16	64.0	N/A	1 * NVIDIA P4	1 * 8	3.0	800	Yes	4	8	20
ecs.gn5i-c16g1.8xlarge	32	128.0	N/A	2 * NVIDIA P4	2 * 8	6.0	1,200	Yes	8	8	20

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENIs	Private IP address of a single ENI
ecs.gn5i-c28g1.14xlarge	56	224.0	N/A	2 * NVIDIA P4	2 * 8	10.0	2,000	Yes	14	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

gn4, compute optimized family with GPU capabilities

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Uses NVIDIA M40 GPU processors
- Multiple CPU-to-memory ratios
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Deep learning
 - Scientific computing applications, such as fluid dynamics, finance, genomics, and environmental analysis
 - High-performance computing, rendering, multimedia encoding and decoding, and other server-side GPU compute workloads

Instance types


Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENIs	Private IP address of a single ENI
ecs.gn4-c4g1.xlarge	4	30.0	N/A	1 * NVIDIA M40	1 * 12	3.0	300	No	1	3	10
ecs.gn4-c8g1.2xlarge	8	30.0	N/A	1 * NVIDIA M40	1 * 12	3.0	400	No	1	4	10
ecs.gn4.8xlarge	32	48.0	N/A	1 * NVIDIA M40	1 * 12	6.0	800	No	3	8	20
ecs.gn4-c4g1.2xlarge	8	60.0	N/A	2 * NVIDIA M40	2 * 12	5.0	500	No	1	4	10
ecs.gn4-c8g1.4xlarge	16	60.0	N/A	2 * NVIDIA M40	2 * 12	5.0	500	No	1	8	20
ecs.gn4.14xlarge	56	96.0	N/A	2 * NVIDIA M40	2 * 12	10.0	1,200	No	4	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmc5s, compute optimized ECS Bare Metal Instance family with enhanced network performance

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:2
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96 vCPUs, up to 2.7 GHz Turbo Boost
- High network performance: 4.5 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
 - Containers (includes but is not limited to Docker, Clear Containers, and Pouch)
 - Video encoding, decoding, and rendering

Instance types


Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmc5s	96	192.0	N/A	30.0	4,500	No	8	32	10
• 24xlarge									

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmg5s, general purpose ECS Bare Metal Instance family with enhanced network performance

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:4
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96 vCPUs, up to 2.7 GHz Turbo Boost
- High network performance: 4.5 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation
- Suitable for the following scenarios:
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
 - Containers (includes but is not limited to Docker, Clear Containers, and Pouch)
 - Enterprise-level applications, such as large and medium-sized databases
 - Video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmg5s.24xlarge	96	384.0	N/A	30.0	4,500	No	8	32	10

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmr5s, memory optimized ECS Bare Metal Instance family with enhanced network performance

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:8
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96 vCPUs, up to 2.7 GHz Turbo Boost
- High network performance: 4.5 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation

- Suitable for the following scenarios:
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
 - Containers (includes but is not limited to Docker, Clear Containers, and Pouch)
 - High-performance databases and in-memory databases
 - Data analysis and mining and distributed memory cache
 - Hadoop, Spark, and other memory-intensive enterprise applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmr5s	96	768.0	N/A	30.0	4,500	No	8	32	10
• 24xlarge									



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmhfg5, ECS Bare Metal Instance family with high clock speed

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:4

- Equipped with 3.7 GHz Intel Xeon E3-1240v6 (Skylake) processors, 8 vCPUs, up to 4.1 GHz Turbo Boost
- High network performance: 2 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation
- Does not support automatic recovery
- Supports Intel SGX
- Suitable for the following scenarios:
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Gaming or finance applications requiring high performance
 - High-performance Web servers
 - Enterprise-level applications such as high-performance databases

Instance types



Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmhfg5.2xlarge	8	32.0	N/A	6.0	2,000	No	8	6	8



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmc4, compute optimized ECS Bare Metal Instance family**Features**

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:2
-   Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors, up to 2.9 GHz Turbo Boost
- High network performance: 4 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation
- Suitable for the following scenarios:
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
 - Containers (includes but is not limited to Docker, Clear Containers, and Pouch)
 - Enterprise-level applications, such as large and medium-sized databases
 - Video encoding

Instance types


Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmc4.xlarge	32	64.0	N/A	10.0	4,000	No	8	12	10

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmg5, general purpose ECS Bare Metal Instance family

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:4
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96 vCPUs, up to 2.7 GHz Turbo Boost
- High network performance: 4 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation
- Suitable for the following scenarios:
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
 - Containers (includes but is not limited to Docker, Clear Containers, and Pouch)
 - Enterprise-level applications, such as large and medium-sized databases
 - Video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmg5.24xlarge	96	384.0	N/A	10.0	4,000	No	8	32	10

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

scch5, SCC instance type family with high clock speed

Features

- I/O optimized
- Supports only standard SSDs and ultra disks
- Supports both RoCE and VPC networks, of which RoCE is dedicated to RDMA communication
- With all features of ECS Bare Metal Instances
- Equipped with 3.1 GHz Intel® Xeon® Gold 6149 (Skylake)
- CPU-to-memory ratio of 1:3
- Suitable for the following scenarios:
 - Large-scale machine learning training
 - Large-scale high performance scientific computing and simulations
 - Large-scale data analysis, batch computing, and video encoding

Instance types

Instance type	vCPU	Physical core	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.scch5.16xlarge	64	32	192.0	N/A	10.0	4,500	46	No	8	32	10

**Note:**

- ecs.scch5.16xlarge provides 64 logical processors on 32 physical cores.
- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

sccg5, general purpose SCC instance family**Features**

- I/O optimized
- Supports only standard SSDs and ultra disks
- Supports both RoCE and VPC networks, of which RoCE is dedicated to RDMA communication
- With all features of ECS Bare Metal Instances
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- CPU-to-memory ratio of 1:4
- Suitable for the following scenarios:
 - Large-scale machine learning training
 - Large-scale high performance scientific computing and simulations
 - Large-scale data analysis, batch computing, and video encoding

Instance types

Instance type	vCPU	Physical core	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.sccg5.24xlarge	96	48	384.0	N/A	10.0	4,500	46	No	8	32	10

**Note:**

- ecs.sccg5.24xlarge provides 96 logical processors on 48 physical cores.
- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

sccgn6, compute optimized SCC GPU instance family**Features**

- I/O-optimized
- CPU-to-memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- With all features of ECS Bare Metal Instances
- Storage:
 - Supports enhanced SSDs (million-level IOPS), standard SSDs, and ultra disks
 - Supports high performance Cloud Parallel File System (CPFS)
- Networking:
 - Supports VPC networks equipped with two 25 Gbps ports
 - Supports RoCE v2 networks, which is dedicated to RDMA communication

- Uses NVIDIA V100 GPU processors (with the SXM2 module):
 - Based on the new NVIDIA Volta architecture
 - 16 GB HBM2 memory capacity
 - Up to 5,120 CUDA Cores
 - Up to 640 Tensor Cores
 - Offers a 900 GB/s memory bandwidth
 - Supports up to six NVLink connections and total bandwidth of 300 GB/s (25 GB/s per connection)
- Suitable for the following scenarios:
 - Ultra-large-scale machine learning training on a distributed GPU cluster
 - Large-scale high performance scientific computing and simulations
 - Large-scale data analysis, batch computing, and video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.sccgn6.24xlarge	96	384	N/A	8 × V100	30	4,500	2 × 25	Yes	8	32	10



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

Instance specification metrics

Metric	Description
<i>Local storage</i>	Local storage, also called cache disks or local disks, refers to the disks attached to the physical servers where ECS instances are hosted. Local storage provides temporary block storage for instances. Local storage capacity is measured in GiB. Data stored in the local disks may be lost in some cases, such as when the compute resources (vCPU and memory) of an instance are released or an instance is migrated because the physical server is down. For more information, see Local disks .
<i>Bandwidth</i>	The maximum bandwidth in one direction. Inbound bandwidth and outbound bandwidth are calculated separately.
<i>Packet forwarding rate</i>	The maximum sum of inbound and outbound packet forwarding rates. For details about how to test the packet forwarding rate, see Test network performance .
<i>NIC queues</i>	The maximum number of NIC queues that the primary NIC of an instance supports. If your instance type is not of an ECS Bare Metal Instance family, the maximum number of NIC queues supported by the secondary NIC is the same as that supported by the primary NIC.
<i>ENIs</i>	Enterprise-level instance types that are equipped with two or more vCPUs support Elastic Network Interfaces (ENIs). Entry-level instance types that are equipped with four or more vCPUs support ENIs. For more information, see ENI overview .


4 Instance type families

4.1 General purpose

This topic describes the g5 and g6 general purpose instance families, and sn2ne general purpose instance family with enhanced network performance and lists their instance types.

g6, general purpose instance family

Features

- Provides predictable stable and high performance and reduces virtualization overheads with the use of the X-Dragon architecture
- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:4
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Second Generation Intel Xeon Scalable processors, up to 3.2 GHz Turbo Boost
- Provides strong network performance based on sufficient computing capacity
- Does not support instance upgrades or downgrades
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Enterprise-level applications of various types and sizes
 - Websites and application servers
 - Game servers
 - Small and medium-sized database systems, cache, and search clusters
 - Data analysis and computing
 - Computing clusters and memory-intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burst bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI	Disk IOPS	Disk bandwidth (Gbit/s)
ecs.g6.large	2	8.0	N/A	1.0	3.0	300	Yes	2	2	6	10,000	1
ecs.g6.xlarge	4	16.0	N/A	1.5	5.0	500	Yes	4	3	10	20,000	1.5
ecs.g6.2xlarge	8	32.0	N/A	2.5	8.0	800	Yes	8	4	10	25,000	2
ecs.g6.3xlarge	12	48.0	N/A	4.0	10.0	900	Yes	8	6	10	30,000	2.5
ecs.g6.4xlarge	16	64.0	N/A	5.0	10.0	1,000	Yes	8	8	20	40,000	3
ecs.g6.6xlarge	24	96.0	N/A	7.5	10.0	1,500	Yes	12	8	20	50,000	4
ecs.g6.8xlarge	32	128.0	N/A	10.0	N/A	2,000	Yes	16	8	20	60,000	5
ecs.g6.13xlarge	52	192.0	N/A	12.5	N/A	3,000	Yes	32	7	20	100,000	8
ecs.g6.26xlarge	104	384.0	N/A	25.0	N/A	6,000	Yes	32	15	20	200,000	16




Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

g5, general purpose instance family

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:4
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Provides strong network performance based on sufficient computing capacity
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Enterprise-level applications of various types and sizes
 - Small and medium-sized database systems, cache, and search clusters
 - Data analysis and computing
 - Computing clusters and memory-intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.g5.large	2	8.0	N/A	1.0	300	Yes	2	2	6
ecs.g5.xlarge	4	16.0	N/A	1.5	500	Yes	2	3	10
ecs.g5.2xlarge	8	32.0	N/A	2.5	800	Yes	2	4	10



Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.g5.3xlarge	12	48.0	N/A	4.0	900	Yes	4	6	10
ecs.g5.4xlarge	16	64.0	N/A	5.0	1,000	Yes	4	8	20
ecs.g5.6xlarge	24	96.0	N/A	7.5	1,500	Yes	6	8	20
ecs.g5.8xlarge	32	128.0	N/A	10.0	2,000	Yes	8	8	20
ecs.g5.16xlarge	64	256.0	N/A	20.0	4,000	Yes	16	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

sn2ne, general purpose instance family with enhanced network performance

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:4
- Ultra high packet forwarding rate
-   Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors
- Provides strong network performance based on sufficient computing capacity

- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Enterprise-level applications of various types and sizes
 - Small and medium-sized database systems, cache, and search clusters
 - Data analysis and computing
 - Computing clusters and memory-intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.sn2ne.large	2	8.0	N/A	1.0	300	Yes	2	2	6
ecs.sn2ne.xlarge	4	16.0	N/A	1.5	500	Yes	2	3	10
ecs.sn2ne.2xlarge	8	32.0	N/A	2.0	1,000	Yes	4	4	10
ecs.sn2ne.3xlarge	12	48.0	N/A	2.5	1,300	Yes	4	6	10
ecs.sn2ne.4xlarge	16	64.0	N/A	3.0	1,600	Yes	4	8	20
ecs.sn2ne.6xlarge	24	96.0	N/A	4.5	2,000	Yes	6	8	20
ecs.sn2ne.8xlarge	32	128.0	N/A	6.0	2,500	Yes	8	8	20

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.sn2ne.14xlarge	56	224.0	N/A	10.0	4,500	Yes	14	8	20

**Note:**


- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

4.2 Compute optimized

This topic describes the ic5 compute-intensive instance family, c5 and c6 compute optimized instance families, and sn1ne compute optimized instance family with enhanced network performance and lists their instance types.

ic5, compute-intensive instance family

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:1
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Provides strong network performance with large computing capacity

- Suitable for the following scenarios:
 - Web front-end servers
 - Data analysis, batch compute, and video encoding
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Massively Multiplayer Online (MMO) game frontends

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ic5.large	2	2.0	N/A	1.0	300	No	2	2	6
ecs.ic5.xlarge	4	4.0	N/A	1.5	500	No	2	3	10
ecs.ic5.2xlarge	8	8.0	N/A	2.5	800	No	2	4	10
ecs.ic5.3xlarge	12	12.0	N/A	4.0	900	No	4	6	10
ecs.ic5.4xlarge	16	16.0	N/A	5.0	1,000	No	4	8	20



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

c6, compute optimized instance family

Features

- Provides predictable stable and high performance and reduces virtualization overheads with the use of the X-Dragon architecture
- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:2
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Second Generation Intel Xeon Scalable processors, up to 3.2 GHz Turbo Boost
- Provides strong network performance with large computing capacity
- Does not support instance upgrades or downgrades
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Web front-end servers
 - Massively Multiplayer Online (MMO) game frontends
 - Data analysis, batch compute, and video encoding
 - High-performance science and engineering applications

Instance types


Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burst bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI	Disk IOPS	Disk bandwidth (Gbit/s)
ecs.c6.large	2	4.0	N/A	1.0	3.0	300	Yes	2	2	6	10,000	1
ecs.c6.xlarge	4	8.0	N/A	1.5	5.0	500	Yes	4	3	10	20,000	1.5

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burst bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI	Disk IOPS	Disk bandwidth (Gbit/s)
ecs.c6.2xlarge	8	16.0	N/A	2.5	8.0	800	Yes	8	4	10	25,000	2
ecs.c6.3xlarge	12	24.0	N/A	4.0	10.0	900	Yes	8	6	10	30,000	2.5
ecs.c6.4xlarge	16	32.0	N/A	5.0	10.0	1,000	Yes	8	8	20	40,000	3
ecs.c6.6xlarge	24	48.0	N/A	7.5	10.0	1,500	Yes	12	8	20	50,000	4
ecs.c6.8xlarge	32	64.0	N/A	10.0	N/A	2,000	Yes	16	8	20	60,000	5
ecs.c6.13xlarge	52	96	N/A	12.5	N/A	3,000	Yes	32	7	20	100,000	8
ecs.c6.26xlarge	104	192.0	N/A	25.0	N/A	6,000	Yes	32	15	20	200,000	16

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

c5, compute optimized instance family**Features**

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:2
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Web front-end servers
 - Massively Multiplayer Online (MMO) game frontends
 - Data analysis, batch compute, and video encoding
 - High-performance science and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.c5.large	2	4.0	N/A	1.0	300	Yes	2	2	6
ecs.c5.xlarge	4	8.0	N/A	1.5	500	Yes	2	3	10
ecs.c5.2xlarge	8	16.0	N/A	2.5	800	Yes	2	4	10
ecs.c5.3xlarge	12	24.0	N/A	4.0	900	Yes	4	6	10
ecs.c5.4xlarge	16	32.0	N/A	5.0	1,000	Yes	4	8	20
ecs.c5.6xlarge	24	48.0	N/A	7.5	1,500	Yes	6	8	20

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.c5.8xlarge	32	64.0	N/A	10.0	2,000	Yes	8	8	20
ecs.c5.16xlarge	64	128.0	N/A	20.0	4,000	Yes	16	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region page](#) to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

sn1ne, compute optimized instance family with enhanced network performance

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:2
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Web front-end servers
 - Massively Multiplayer Online (MMO) game frontends
 - Data analysis, batch compute, and video encoding
 - High-performance science and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.sn1ne.large	2	4.0	N/A	1.0	300	Yes	2	2	6
ecs.sn1ne.xlarge	4	8.0	N/A	1.5	500	Yes	2	3	10
ecs.sn1ne.2xlarge	8	16.0	N/A	2.0	1,000	Yes	4	4	10
ecs.sn1ne.3xlarge	12	24.0	N/A	2.5	1,300	Yes	4	6	10
ecs.sn1ne.4xlarge	16	32.0	N/A	3.0	1,600	Yes	4	8	20
ecs.sn1ne.6xlarge	24	48.0	N/A	4.5	2,000	Yes	6	8	20
ecs.sn1ne.8xlarge	32	64.0	N/A	6.0	2,500	Yes	8	8	20



Note:


- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

4.3 Memory optimized

This topic describes the r5 and r6 memory optimized instance families, re4 memory optimized instance family with enhanced performance, re4e memory optimized instance family with enhanced performance, se1ne memory optimized instance family with enhanced network performance, and se1 memory optimized instance family and lists their instance types.

r6, memory optimized instance family

Features

- Provides predictable stable and high performance and reduces virtualization overheads with the use of the X-Dragon architecture
- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- Ultra high packet forwarding rate
-  Equipped with 2.5 GHz Second Generation Intel Xeon Scalable processors, up to 3.2 GHz Turbo Boost
- Provides strong network performance with large computing capacity
- Does not support instance upgrades or downgrades
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - High-performance and in-memory databases
 - Data analysis and mining, and distributed memory cache
 - Hadoop, Spark, and other memory-intensive enterprise applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burst bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI	Disk IOPS	Disk bandwidth (Gbit/s)
ecs.r6.large	2	16.0	N/A	1.0	3.0	300	Yes	2	2	6	10,000	1
ecs.r6.xlarge	4	32.0	N/A	1.5	5.0	500	Yes	4	3	10	20,000	1.5
ecs.r6.2xlarge	8	64.0	N/A	2.5	8.0	800	Yes	8	4	10	25,000	2
ecs.r6.3xlarge	12	96.0	N/A	4.0	10.0	900	Yes	8	6	10	30,000	2.5
ecs.r6.4xlarge	16	128.0	N/A	5.0	10.0	1,000	Yes	8	8	20	40,000	3
ecs.r6.6xlarge	24	192.0	N/A	7.5	10.0	1,500	Yes	12	8	20	50,000	4
ecs.r6.8xlarge	32	256.0	N/A	10.0	N/A	2,000	Yes	16	8	20	60,000	5
ecs.r6.13xlarge	52	384	N/A	12.5	N/A	3,000	Yes	32	7	20	100,000	8
ecs.r6.26xlarge	104	768.0	N/A	25.0	N/A	6,000	Yes	32	15	20	200,000	16



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

r5, memory optimized instance family

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - High-performance and in-memory databases
 - Data analysis and mining, and distributed memory cache
 - Hadoop, Spark, and other memory-intensive enterprise applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.r5.large	2	16.0	N/A	1.0	300	Yes	2	2	6
ecs.r5.xlarge	4	32.0	N/A	1.5	500	Yes	2	3	10
ecs.r5.2xlarge	8	64.0	N/A	2.5	800	Yes	2	4	10
ecs.r5.3xlarge	12	96.0	N/A	4.0	900	Yes	4	6	10

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.r5.4xlarge	16	128.0	N/A	5.0	1,000	Yes	4	8	20
ecs.r5.6xlarge	24	192.0	N/A	7.5	1,500	Yes	6	8	20
ecs.r5.8xlarge	32	256.0	N/A	10.0	2,000	Yes	8	8	20
ecs.r5.16xlarge	64	512.0	N/A	20.0	4,000	Yes	16	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

re4, memory optimized instance family with enhanced performance

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Optimized for high-performance databases, in-memory databases, and other memory-intensive enterprise applications
- Equipped with 2.2 GHz Intel Xeon E7 8880 v4 (Broadwell) processors, up to 2.4 GHz Turbo Boost
- CPU-to-memory ratio of 1:12, up to 1920.0 GiB memory
- ecs.re4.20xlarge and ecs.re4.40xlarge have been certified by SAP HANA

- Suitable for the following scenarios:
 - High-performance databases and in-memory databases (for example, SAP HANA)
 - Memory-intensive applications
 - Big data processing engines such as Apache Spark or Presto

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.re4.20xlarge	80	960.0	N/A	15.0	2,000	Yes	16	8	20
ecs.re4.40xlarge	160	1920.0	N/A	30.0	4,500	Yes	16	8	20



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

re4e, memory optimized instance family with enhanced performance

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Optimized for high-performance databases, in-memory databases, and other memory-intensive enterprise applications

Equipped with 2.2 GHz Intel Xeon E7 8880 v4 (Broadwell) processors, up to 2.4 GHz Turbo Boost

- CPU-to-memory ratio of 1:24, up to 3840.0 GiB memory
- Suitable for the following scenarios:
 - High-performance databases and in-memory databases (for example, SAP HANA)
 - Memory-intensive applications
 - Big data processing engines such as Apache Spark or Presto

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.re4e.40xlarge	160	3840.0	N/A	30.0	4,500	Yes	16	15	20



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

se1ne, memory optimized instance family with enhanced network performance

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:8
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors
- Provides strong network performance with large computing capacity

- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - High-performance and in-memory databases
 - Data analysis and mining, and distributed memory cache
 - Hadoop, Spark, and other memory-intensive enterprise applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.se1ne.large	2	16.0	N/A	1.0	300	Yes	2	2	6
ecs.se1ne.xlarge	4	32.0	N/A	1.5	500	Yes	2	3	10
ecs.se1ne.2xlarge	8	64.0	N/A	2.0	1,000	Yes	4	4	10
ecs.se1ne.3xlarge	12	96.0	N/A	2.5	1,300	Yes	4	6	10
ecs.se1ne.4xlarge	16	128.0	N/A	3.0	1,600	Yes	4	8	20
ecs.se1ne.6xlarge	24	192.0	N/A	4.5	2,000	Yes	6	8	20
ecs.se1ne.8xlarge	32	256.0	N/A	6.0	2,500	Yes	8	8	20

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.se1ne.14xlarge	56	480.0	N/A	10.0	4,500	Yes	14	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

se1, memory optimized instance family**Features**

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:8
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - High-performance and in-memory databases
 - Data analysis and mining, and distributed memory cache
 - Hadoop, Spark, and other memory-intensive enterprise applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.se1.large	2	16.0	N/A	0.5	100	No	1	2	6
ecs.se1.xlarge	4	32.0	N/A	0.8	200	No	1	3	10
ecs.se1.2xlarge	8	64.0	N/A	1.5	400	No	1	4	10
ecs.se1.4xlarge	16	128.0	N/A	3.0	500	No	2	8	20
ecs.se1.8xlarge	32	256.0	N/A	6.0	800	No	3	8	20
ecs.se1.14xlarge	56	480.0	N/A	10.0	1,200	No	4	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

4.4 Big data instance type families

This topic describes the big data instance type family d1 and the big data instance type family with enhanced network performance d1ne, and lists the specific instance types within each instance type family.

Overview

The big data instance type families d1 and d1ne are designed to compute and store massive amounts of data on the cloud, allowing you to achieve big data solutions at an enterprise level. Moreover, these instance family types can be used to build a Hadoop distributed computing architecture that is supplemented by self-hosted storage at your on-premises data center. This way, you can build a Hadoop cluster at costs similar to that of building a self-hosted cluster in your on-premises data center, while at the same time also guarantee increased storage space with improved performance.

d1ne and d1 instances have the following benefits:

- Provide stable computing power by using enterprise-level architecture to ensure efficient processing of computing operations.
- Achieve higher network performance (specifically, greater maximum intranet bandwidth and packet forwarding rate per instance) to achieve optimal data transfer among instances during peak periods.
- Support 190 MiB/s of sequential read/write speeds for each disk and 5 GiB/s of storage throughput for each instance, reducing the overall HDFS file read/write time. Note that disks need to be fully initialized for optimal performance when an instance is newly created.
- Provide local storage at a price 97% lower than that of SSD cloud disks, greatly reducing the cost to build a Hadoop cluster.

d1ne and d1 instances have the following limits:

- The configurations of d1 and d1ne instances cannot be modified after they are created.
- Downtime migration is not supported currently.
- You cannot purchase a local disk separately. You can only purchase a local disk when you create a d1 or d1ne instance. The number and capacity of local disks that you can purchase is determined by the instance type you choose.

- Snapshots are not supported currently. Therefore, if you need to create a full image for a d1 or d1ne instance, we recommend that you create one by combining the system disk snapshot and data disk (only cloud disks are supported) snapshots.
- Currently you cannot create a full image based on an instance ID, so you cannot create a d1 or d1ne instance with a custom image either.
- The data disks of d1 and d1ne instances are local disks that are subject to data loss. If your application cannot implement the data reliability architecture, we recommend the following:
 - Use cloud disks for your instances, rather than local disks to store business data that needs to be stored long term.
 - Back up your data regularly and adopt a high-availability architecture in which SSD cloud disks are attached your d1 and d1ne instances.
- The following table shows how the operations on instances with local disks impact the disk data.

Operations	Status of disk data	Description
Restart the OS, restart an instance in the console , and force restart an instance	Retained	Local disk volumes and data are retained.
Shut down the OS, stop an instance in the console, and force stop an instance	Retained	Local disk volumes and data are retained.
Release an instance in the console	Erased	Local disk volumes and data are erased.

4.5 Local SSD-equipped instance type families

This topic describes the local SSD-equipped instance type families i2, i2g, and i1, and lists the specific instance types within each of the instance type families.

Characteristics and issues to note of i1

i1 is an I/O-optimized instance type family with local SSD disks, and is designed for scenarios that require high storage I/O performance. It is ideal for enterprises that provide online services, such as Internet games, e-commerce, live webcast, and Internet media, fully meeting the requirements for low latency and high I/O performance imposed by I/O-intensive applications. i1 can be used for businesses

that require high storage I/O performance and have a high-availability architecture at the application layer, such as NoSQL non-relational databases, MPP data stores, and distributed file systems.

i1 has the following characteristics:

- For large database applications, it offers tens of thousands of or even hundreds of thousands of low-latency random I/O reads and writes per second.
- For large data set applications such as big data and parallel computing, it offers the sequential read and write throughput of up to several GiBs.
- Based on local NVMe SSD disks, it provides up to hundreds of thousands of random I/O reads and writes while keeping a latency of μ s.

When you use i1 instances, note the following:

- i1 does not support changing configurations or downtime migration currently . Please check the official website notice regularly for the availability of these features.
- Snapshots are not supported for local disks currently. Therefore, if you need to create a full image for an i1 instance, we recommend that you create it by combining the system disk snapshot and data disk (cloud disks only) snapshots. Currently you cannot create a full image based on an instance ID.
- i1 instances are designed for scenarios that require high storage I/O performance. Local disks are attached to specific instance types and cannot be purchased separately. You cannot detach a local disk and then attach it to another instance.
- The data disks of i1 instances are local disks that are subject to data loss (such as when NVMe SSD disks are faulty or the host is down). If your application cannot implement the data reliability architecture, we strongly recommend that you use cloud disks for your instances.
- The following table shows how the operations on instances with local disks impact the disk data.

Operation	Status of disk data	Description
Restart the OS/restart an instance in the console/force restart an instance	Retained	Local disk volumes and data are retained.
Shut down the OS/stop an instance in the console/force stop an instance	Retained	Local disk volumes and data are retained.

Operation	Status of disk data	Description
Release an instance in the console	Erased	Local disk volumes and data are erased.

- Do not use local disks to store business data that needs to be stored for a long time. Back up your data in a timely manner and adopt a high-availability architecture. For long-term storage of business data, cloud disks are recommended.
- In VPCs, if an i1 instance had overdue payment or was stopped upon expiration before March 1st, 2018, its local disks were retained for 15 days. After March 1st, 2018, if an i1 instance has overdue payment or is stopped upon expiration, its local disks are released along with vCPUs and memory in 15 days.

4.6 High clock speed instance type families

This topic describes the compute optimized instance type family with high clock speed hfc5 and the general-purpose instance type family with high clock speed hfg5, and lists the specific instance types within each of these two instance type families.

4.7 Compute optimized type family with GPU

4.7.1 Compute optimized instance families with GPU capabilities

This topic describes the vgn5i, gn6i, gn6v, gn5, gn5i, and gn4 compute optimized instance families with GPU capabilities and lists their instance types.

vgn5i, lightweight compute optimized instance family with GPU capabilities

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Uses NVIDIA P4 GPU computing accelerators
- Contains virtual GPUs (vGPUs), which are the result of GPU virtualization with mediated pass-through
 - Supports the 1/8, 1/4, 1/2, and 1/1 computing capacity of NVIDIA® Tesla® P4 GPUs
 - Supports 1, 2, 4, and 8 GB of GPU memory
- CPU-to-memory ratio of 1:3

- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Real-time rendering for cloud gaming
 - Real-time rendering for AR and VR applications
 - AI (deep learning and machine learning) inference for the elastic deployment of Internet services
 - Educational environment of deep learning
 - Modeling experiment environment of deep learning

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs
ecs.vgn5i-m1.large	2	6	N/A	P4*1/8	1	1	300	Yes	2	2
ecs.vgn5i-m2.xlarge	4	12	N/A	P4*1/4	2	2	500	Yes	2	3
ecs.vgn5i-m4.2xlarge	8	24	N/A	P4 * 1/2	4	3	800	Yes	2	4
ecs.vgn5i-m8.4xlarge	16	48	N/A	P4*1	8	5	1,000	Yes	4	5




Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

gn6i, compute optimized instance family with GPU capabilities

Features

- I/O optimized
- CPU-to-memory ratio of 1:4
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports standard SSDs, ultra disks, and enhanced SSDs that deliver millions of IOPS
- Achieves better performance with the new-generation X-Dragon compute architecture
- Uses NVIDIA T4 GPU computing accelerators
 - Powered by the new NVIDIA Turing architecture
 - Up to 320 Turing Tensor Cores
 - Up to 2,560 CUDA Cores
 - Mixed-precision Tensor Cores support 65 FP16 TFLOPS, 130 INT8 TOPS, and 260 INT4 TOPS
 - 16 GB memory capacity (320 GB/s bandwidth)
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - AI (deep learning and machine learning) inference for computer vision, voice recognition, speech synthesis, natural language processing, machine translation , and reference systems
 - Real-time rendering for cloud gaming
 - Real-time rendering for AR and VR applications
 - Reload graphics computing or Graphics Workstation
 - GPU-accelerated databases
 - High-performance computing

Instance types



Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs
ecs.gn6i-c4g1.xlarge	4	15	N/A	T4*1	16	4	500	Yes	2	2
ecs.gn6i-c8g1.2xlarge	8	31	N/A	T4*1	16	5	800	Yes	2	2
ecs.gn6i-c16g1.4xlarge	16	62	N/A	T4*1	16	6	1,000	Yes	4	3
ecs.gn6i-c24g1.6xlarge	24	93	N/A	T4*1	16	7.5	1,200	Yes	6	4
ecs.gn6i-c24g1.12xlarge	48	186	N/A	T4*2	32	15	2,400	Yes	12	6
ecs.gn6i-c24g1.24xlarge	96	372	N/A	T4*4	64	30	4,800	Yes	24	8

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

gn6v, compute optimized instance family with GPU capabilities

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks.
- Uses NVIDIA V100 GPU processors
- CPU-to-memory ratio of 1:4
-  
 - Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Uses NVIDIA V100 GPU computing accelerators (SXM2-based)
 - Powered by the innovative NVIDIA Volta architecture
 - 16 GB HBM2 GPU memory
 - 5,120 CUDA Cores
 - 640 Tensor Cores
 - Memory bandwidth of 900 GB/s
 - Supports up to six NVLink connections for a total bandwidth of 300 GB/s with 25 GB/s each
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Deep learning, image classification, autonomous vehicles, voice recognition, and other AI applications
 - Scientific computing applications, such as fluid dynamics, finance, molecular dynamics, and environmental analysis

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENIs	Private IP address of a single ENI
ecs.gn6v-c8g1.2xlarge	8	32.0	N/A	1 * NVIDIA V100	1 * 16	2.5	800	Yes	4	4	10

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENIs	Private IP address of a single ENI
ecs.gn6v-c8g1.8xlarge	32	128.0	N/A	4 * NVIDIA V100	4 * 16	10.0	2,000	Yes	8	8	20
ecs.gn6v-c8g1.16xlarge	64	256.0	N/A	8 * NVIDIA V100	8 * 16	20.0	2,500	Yes	16	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

gn5, compute optimized instance family with GPU capabilities

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Uses NVIDIA P100 GPU processors
- Multiple CPU-to-memory ratios
- High-performance local NVMe SSDs
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Provides strong network performance with large computing capacity

· Suitable for the following scenarios:

- Deep learning
- Scientific computing applications, such as fluid dynamics, finance, genomics, and environmental analysis
- High-performance computing, rendering, multimedia encoding and decoding, and other server-side GPU compute workloads

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs	Private IP address of a single ENI
ecs.gn5-c4g1.xlarge	4	30.0	440	1 * NVIDIA P100	1 * 16	3.0	300	No	1	3	10
ecs.gn5-c8g1.2xlarge	8	60.0	440	1 * NVIDIA P100	1 * 16	3.0	400	No	1	4	10
ecs.gn5-c4g1.2xlarge	8	60.0	880	2 * NVIDIA P100	2 * 16	5.0	1,000	No	2	4	10
ecs.gn5-c8g1.4xlarge	16	120.0	880	2 * NVIDIA P100	2 * 16	5.0	1,000	No	4	8	20
ecs.gn5-c28g1.7xlarge	28	112.0	440	1 * NVIDIA P100	1 * 16	5.0	1,000	No	8	8	20

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENIs	Private IP address of a single ENI
ecs.gn5-c8g1.8xlarge	32	240.0	1,760	4 * NVIDIA P100	4 * 16	10.0	2,000	No	8	8	20
ecs.gn5-c28g1.14xlarge	56	224.0	880	2 * NVIDIA P100	2 * 16	10.0	2,000	No	14	8	20
ecs.gn5-c8g1.14xlarge	54	480.0	3,520	8 * NVIDIA P100	8 * 16	25.0	4,000	No	14	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

gn5i, compute optimized instance family with GPU capabilities

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Uses NVIDIA P4 GPU processors
- CPU-to-memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Provides strong network performance with large computing capacity

- Suitable for the following scenarios:
 - Deep learning inference
 - Multimedia encoding and decoding, and other server-side GPU compute workloads

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs	Private IP address of a single ENI
ecs.gn5i-c2g1.large	2	8.0	N/A	1 * NVIDIA P4	1 * 8	1.0	100	Yes	2	2	6
ecs.gn5i-c4g1.xlarge	4	16.0	N/A	1 * NVIDIA P4	1 * 8	1.5	200	Yes	2	3	10
ecs.gn5i-c8g1.2xlarge	8	32.0	N/A	1 * NVIDIA P4	1 * 8	2.0	400	Yes	4	4	10
ecs.gn5i-c16g1.4xlarge	16	64.0	N/A	1 * NVIDIA P4	1 * 8	3.0	800	Yes	4	8	20
ecs.gn5i-c16g1.8xlarge	32	128.0	N/A	2 * NVIDIA P4	2 * 8	6.0	1,200	Yes	8	8	20
ecs.gn5i-c28g1.14xlarge	56	224.0	N/A	2 * NVIDIA P4	2 * 8	10.0	2,000	Yes	14	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

gn4, compute optimized family with GPU capabilities

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- Uses NVIDIA M40 GPU processors
- Multiple CPU-to-memory ratios
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Provides strong network performance with large computing capacity
- Suitable for the following scenarios:
 - Deep learning
 - Scientific computing applications, such as fluid dynamics, finance, genomics, and environmental analysis
 - High-performance computing, rendering, multimedia encoding and decoding, and other server-side GPU compute workloads

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs	Private IP address of a single ENI
ecs.gn4-c4g1.xlarge	4	30.0	N/A	1 * NVIDIA M40	1 * 12	3.0	300	No	1	3	10

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPU	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENIs	Private IP address of a single ENI
ecs.gn4-c8g1.2xlarge	8	30.0	N/A	1 * NVIDIA M40	1 * 12	3.0	400	No	1	4	10
ecs.gn4.8xlarge	32	48.0	N/A	1 * NVIDIA M40	1 * 12	6.0	800	No	3	8	20
ecs.gn4-c4g1.2xlarge	8	60.0	N/A	2 * NVIDIA M40	2 * 12	5.0	500	No	1	4	10
ecs.gn4-c8g1.4xlarge	16	60.0	N/A	2 * NVIDIA M40	2 * 12	5.0	500	No	1	8	20
ecs.gn4.14xlarge	56	96.0	N/A	2 * NVIDIA M40	2 * 12	10.0	1,200	No	4	8	20

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

4.7.2 Create a compute optimized instance with GPU

This topic describes how to create a compute optimized instance with GPU (hereinafter referred to as GPU instance) and configure the GPU driver to be automatically installed at instance creation. GPU instances can be used only after the GPU driver is installed. If you do not configure automatic installation of the GPU

driver when you create a GPU instance, you can install it manually after you create the GPU instance.

Limits

If you configure automatic installation of the GPU driver, the following limits apply:

- Only public images of Linux OSs support automatic installation of the GPU driver.
- Automatic installation of the GPU driver takes 4 to 10 minutes, depending on the intranet bandwidth and the CPU count of different instance types. During automatic installation, GPU cannot be used, the instance cannot be operated, and other GPU software cannot be installed. Otherwise, the automatic installation fails and the instance becomes unavailable.
- If you change the operating system after you create a GPU instance, you must make sure that you use the same image or [Images that support automatic installation of the GPU driver](#). Otherwise, automatic installation of the GPU driver fails.



Note:

You can [connect to the target instance](#) and check the installation progress and result by using the installation log.

- If you select Auto-install GPU Driver, the installation log is stored in the `/ root / nvidia_ins tall . log` directory.
- If you configure the `nvidia_install_v2.0` installation script in User Data, the installation log is stored in the `/ root / nvidia / nvidia_ins tall . log` directory.

Procedure

The following procedure describes how to configure a GPU instance. For information on how to configure general instances, see [Create an instance](#).

1. Go to the [ECS purchase page](#).
2. Complete the following Basic Configurations.
 - Region: Select the target region. For information about regions and zones, see [Regions and zones that provide GPU instances](#). If the regions and zones listed in this table are different from those displayed on the Basic Configurations page,

the regions and zones displayed on the page are the actual regions and zones available.

- **Instance Type:** Choose Heterogeneous Computing > Compute Optimized Type with GPU, and then select an instance type as needed.
- **Image:** Some Public Image items of Linux OSs support automatic installation of the CUDA Toolkit and the GPU driver. For more information, see [Images that support automatic installation of the GPU driver](#).

If you select an image that supports automatic installation of the preceding drivers, select Auto-install GPU driver, and select a GPU driver version. If you want to use the GPU instance for a new service system, we recommend that you select the latest GPU driver version.

If you do not select Auto-install GPU driver, or if the selected image does not support automatic installation of the GPU driver, you need to configure the installation script in User Data, or `#unique_70` after you install the GPU instance. For information on how to configure the installation script, see [Script version](#).



Note:

If you call the [RunInstances](#) API action to create a GPU instance, you need to use the `UserData` parameter to upload the installation script, which must be Base64-encoded.

3. Complete the Networking configurations as follows:

- **Network Type:** Select VPC.
- **Network Billing Method:** Select a bandwidth as needed.



Note:

If you select Windows 2008 R2 or an earlier image on the Basic Configurations page, the GPU instance cannot be accessed by using the [Management Terminal](#)

after the GPU driver takes effect after installation. To resolve this issue, you must select Assign Public IP, or [bind EIP](#) after you create the instance. In this way, you can connect to the instance by using other protocols, such as RDP, PCOIP, and XenDesktop HDX 3D. However, RDP does not support such applications as DirectX or OpenGL. If you require RDP to be the protocol, you must install the VNC service and the VNC client.

4. Complete the System Configurations as follows:

- **Log on Credentials:** Select Key Pair or Password. If you select Set Later, you must associate an SSH key pair or reset your password when you log on to the instance by using the Management Terminal, and then restart the instance to make the settings take effect. If the GPU driver is not installed completely, the restart operation will result in installation failure.
- **User Data:**
 - If you select Auto-install GPU Driver in the Image area on the Basic Configurations page, the Shell script and the precautionary information regarding the installation of the CUDA Toolkit and the GPU driver are displayed in this area.

Advanced (based on instance RAM roles or cloud-init)

RAM Role: [See details](#) | [Create instance RAM role](#)

User Data: ☐ The text is Base64-encoded

This script automatically installs the GPU driver. If you do not want to run this script, go back to the previous step and uncheck "Auto-install GPU Driver". Auto-install GPU driver performs the following:

- 1.The NVIDIA GPU Driver and the CUDA Library will be automatically downloaded from the intranet repository.
- 2.The installation process may take 4 to 10 minutes to complete, depending on the instance type, instance specification and intranet bandwidth. The GPU will be unavailable during installation. To prevent interrupting the installation and causing instance failure, do not perform any instance operations while the GPU driver is installing, and do not install any other GPU related software.
- 3.The instance will reboot after installation.
- 4.Persistence Mode will be automatically enabled for the GPU, and will be added as a system startup item, to ensure increased GPU performance stability.
- 5.If you have decided to change the system disk, the new system disk image must be the same as the previous system disk image that you installed the GPU driver with, to prevent system failure.

```
#!/bin/sh
log="/root/nvidia_install.log"
driver_version="390.46"
```

Windows guest operating systems support bat and powershell formats. Before you base64-encode the content, the first line must be [bat] or [powershell]. Linux guest operating systems support shell scripts. [cloud-init](#) | [Learn more](#).

- If you do not select Auto-install GPU Driver, you can configure the installation script in the User Data area. For an installation script example, see [User Data installation script](#).

5. Complete the Grouping configurations and confirm your order on the Preview page.



Note:

- If you configure an automatic installation script, the GPU driver is automatically installed after you install the GPU instance. After the GPU driver is installed

, the GPU instance is automatically restarted. Then, the GPU driver can operate properly.

- The GPU driver operates more reliably in the Persistence mode. The installation script automatically enables the Persistence mode of the GPU driver and adds this setting to the automatic startup script. This mode is then enabled by default whenever the instance is restarted.

Regions and zones that provide GPU instances

The following table describes the regions and zones where GPU instance type families are available.

Instance type	Region and zone
gn4	<ul style="list-style-type: none"> • China (Beijing) zone A, China (Shanghai) zone B • China (Shenzhen) zone C
gn5	<ul style="list-style-type: none"> • China (Beijing) zones C and E, China (Hohhot) zone A • China (Hangzhou) zones G and F, China (Shanghai) zones D, B, and E • China (Shenzhen) zone D • China (Hong Kong) zones C and B • Singapore zones B and A, Australia zone A, Malaysia zone A, Indonesia zone A • US (Silicon Valley) zones B and A, US (Virginia) zones B and A • Germany (Frankfurt) zone A
gn5 (NGC environment)	The instance type family gn5 does not fully support the NVIDIA GPU CLOUD (NGC) environment. For more information, see #unique_74 .
gn5i	<ul style="list-style-type: none"> • China (Beijing) zones C, E, and A • China (Hangzhou) zone B, China (Shanghai) zones D and B • China (Shenzhen) zone A
gn6v	China (Shanghai) zone F

Images that support automatic installation of the GPU driver

The following table describes the images that support automatic installation of the CUDA Toolkit and the GPU driver.

Image	Version
Public image	<p>The following versions are supported:</p> <ul style="list-style-type: none"> CentOS 64-bit (All available versions are supported.) Ubuntu 16.04 64-bit SUSE Linux Enterprise Server 12 SP2 64-bit
Alibaba Cloud Marketplace	<p>Depending on your image requirements, obtain the required image by using either of the following methods:</p> <ul style="list-style-type: none"> Search for NVIDIA and select the required image. Currently, only CentOS 7.3 is supported. If you want to use the GPU instance for deep learning, you can select an image that has a pre-installed deep learning frame. To do so, search for Deep Learning and select the required image. Currently, only CentOS 7.3 is supported.

Script version

When the GPU instance is restarted for the first time, the cloud-init tool automatically runs the Shell script to install the CUDA Toolkit and the GPU driver.

- If you select Auto-install GPU Driver, the GPU instance uses `nvidia_istall_v1 . 0`. The following table describes the available CUDA Toolkit versions of GPU driver versions.

CUDA	GPU driver	Supported instance type
9.1.85	390.46	<ul style="list-style-type: none"> gn5 gn5i gn6v gn4
9.0.176	<ul style="list-style-type: none"> 390.46 384.125 384.111 	<ul style="list-style-type: none"> gn5 gn5i gn6v gn4
8.0.61	<ul style="list-style-type: none"> 390.46 384.125 384.111 	<ul style="list-style-type: none"> gn5 gn5i gn4

- If you configure the installation script in the User Data area, we recommend that you use `nvidia_install_v2.0`. For more information, see [User Data installation script](#). The `nvidia_install_v2.0` script has the following benefits:
 - It provides the latest CUDA Toolkit, GPU driver, and cuDNN library.
 - After you log on to the instance, a bar is displayed to indicate the installation progress of the GPU driver. Alternatively, you can see the installation result when installation is completed (NVIDIA INSTALL OK or NVIDIA INSTALL FAIL).

If you use the `nvidia_install_v2.0` script, you need to specify the version number of the GPU driver, CUDA Toolkit, and cuDNN library. For example:

```
driver_version="410.79"
cuda_version="9.0.176"
cudnn_version="7.4.2"
```

The following table describes the supported CUDA Toolkit, GPU driver, and cuDNN library versions.

CUDA	GPU driver	cuDNN library
10.0.130	410.79	<ul style="list-style-type: none"> - 7.4.2 - 7.3.1
9.2.148	<ul style="list-style-type: none"> - 410.79 - 396.44 	<ul style="list-style-type: none"> - 7.4.2 - 7.3.1 - 7.1.4
9.0.176	<ul style="list-style-type: none"> - 410.79 - 396.44 - 390.46 	<ul style="list-style-type: none"> - 7.4.2 - 7.3.1 - 7.1.4 - 7.0.5
8.0.61	<ul style="list-style-type: none"> - 410.79 - 396.44 - 390.46 	<ul style="list-style-type: none"> - 7.1.3 - 7.0.5

User Data installation script

If you install the GPU driver by using User Data, we recommend that you use the `nvidia_install_v2.0` script as follows:

```
#!/bin/sh
driver_version=$1
cuda_version=$2
cudnn_version=$3
```

```

NVIDIA_DIR="/ root / nvidia "
log=${ NVIDIA_DIR }"/ nvidia_ins tall . log "
PROCESS_NAME="/ var / lib / cloud / instance / scripts / part -
001 "
DRIVER_PROCESS_NAME=${ NVIDIA_DIR }"/ NVIDIA - Linux - x86_64 "
CUDA_PROCESS_NAME=${ NVIDIA_DIR }"/ cuda "
CUDA_PROCESS_NAME=${ NVIDIA_DIR }"/ cudnn "
DOWNLOAD_PROCESS_NAME=" wget "
SUCCESS_STR=" NVIDIA INSTALL OK "
DOWNLOAD_SUCCESS_STR=" Download OK "
DRIVER_FAIL_STR=" Driver INSTALL FAIL "
CUDA_FAIL_STR=" CUDA INSTALL FAIL "
CUDA_FAIL_STR=" CUDNN INSTALL FAIL "
DOWNLOAD_FAIL_STR=" Download FAIL "
install_notes=" The script automatically downloads and
installs a NVIDIA GPU driver and CUDA / CUDNN Toolkit .
1 . The installation takes 6 to 10 minutes ,
depending on the intranet bandwidth and the quantity
of vCPU cores of the instance . Please do not
operate the GPU or install any GPU - related software
until the GPU driver is installed successfully .
2 . After the GPU is installed successfully , the
instance will restarts automatically ."
check_install ()
{
    b=''
    if [ "$ 1 " = " NVIDIA " ]; then
        ProcessName=$ DRIVER_PROCESS_NAME
        t = 2
    elif [ "$ 1 " = " cuda " ]; then
        ProcessName=$ CUDA_PROCESS_NAME
        t = 2 . 5
    elif [ "$ 1 " = " cudnn " ]; then
        ProcessName=$ CUDNN_PROCESS_NAME
        t = 0 . 5
    fi
    i = 0
    while true
    do
        pid_num=$( ps - ef | grep $ ProcessName | grep - v
grep | wc - l )
        if [ $ pid_num - eq 0 ]; then
            str=$( printf "%- 100s " "#")
            b=$( echo "$ str " | sed ' s / /# / g ' )
            printf "| %- 100s | % d %% \ r \ n " "$ b " " 100 ";
            break
        fi
        i=$(( $ i + 1 ))
        str=$( printf "%-${ i } s " "#")
        b=$( echo "$ str " | sed ' s / /# / g ' )
        printf "| %- 100s | % d %% \ r " "$ b " "$ i ";
        sleep $ t
    done
    echo
    return 0
}
check_download ()
{
    name=$ 1
    i = 0
    b=''
    filesize = 0
    percent = 0
    sleep 0 . 5

```

```

while true
do
    pid_num=$( ps - ef | grep wget | grep $ name | grep
- v grep | wc - l )
    if [ $ pid_num - eq 0 ]; then
        filesize=$( du - sk / root / nvidia / ${ name } * | awk
' { print $ 1 } ' )
        str=$( printf "%- 100s " "#")
        b=$( echo "$ str " | sed ' s / # / g ' )
        printf "%- 8s | %- 100s | % d %% \ r \ n " "${ filesize
} K " "$ b " " 100 ";
        break
    fi
    line=$( tail - 2 / root / nvidia / nvidia_ins tall . log
)
    filesize=$( echo $ line | awk - F ' ' ' { print $ 1 } ' )
    percent=$( echo $ line | awk - F '%' ' { print $ 1 } '
| awk - F ' ' ' { print $ NF } ' )
    if [ "$ percent " - ge 0 ] 2 > / dev / null ; then
        str=$( printf "%- ${ percent } s " "#")
        b=$( echo "$ str " | sed ' s / # / g ' )
        printf "%- 8s | %- 100s | % d %% \ r " "${ filesize }"
"$ b " "$ percent ";
    else
        continue
    fi
    sleep 0 . 5
done
return 0
}
check_inst all_log ()
{
    if [ ! - f "$ log " ]; then
        echo " NVIDIA install log $ log not exist !
Install may fail !"
        echo
        exit 1
    fi
    if [ "$ 1 " = " NVIDIA " ]; then
        succstr=$ SUCCESS_ST R
        str2=$( cat $ log | grep " INSTALL_ER ROR ")
        echo
        if [ - n "$ succstr " ] && [ - z "$ str2 " ]; then
            echo "$ succstr !"
            echo
            return 0
        else
            echo " NVIDIA install may have some
INSTALL_ER ROR , please check log $ log !"
            return 1
        fi
    fi
    if [ "$ 1 " = " DRIVER " ]; then
        failstr=$ DRIVER_FAI L_STR
    elif [ "$ 1 " = " CUDA " ]; then
        failstr=$ CUDA_FAIL_ STR
    elif [ "$ 1 " = " CUDNN " ]; then
        failstr=$ CUDNN_FAIL _STR
    fi
    str1=$( cat $ log | grep "$ failstr ")
    if [ - n "$ str1 " ]; then
        echo
        echo " NVIDIA $ failstr ! please check install log
$ log !"

```

```

        return 1
    fi
}
check_inst all_processes ()
{
    echo "CHECKING NVIDIA INSTALL , PLEASE WAIT ....."
    echo "$install_name"
    echo
    while true
    do
        pid_num=$( ps -ef | grep $PROCESS_NAME | grep -v
grep | grep -v check | wc -l )
        if [ $pid_num -eq 0 ]; then
            check_inst all_log "NVIDIA "
            return 0
        else
            pid_num=$( ps -ef | grep $DOWNLOAD_PROCESS_NAME
E | grep driver | grep -v grep | wc -l )
            if [ $pid_num -gt 0 ]; then
                echo "Driver-${1} downloading , need 10
seconds . Remaining installation time 360 - 600 seconds
!"
                check_down load "NVIDIA "
            fi
            pid_num=$( ps -ef | grep $DOWNLOAD_PROCESS_NAME
E | grep cuda | grep -v grep | wc -l )
            if [ $pid_num -gt 0 ]; then
                echo "CUDA-${2} downloading , need 150
or more seconds . Remaining installation time 350 -
590 seconds !"
                while true
                do
                    check_down load "cuda "
                    sleep 1
                    pid_num=$( ps -ef | grep $DOWNLOAD_P
PROCESS_NAME E | grep cuda | grep -v grep | wc -l )
                    if [ $pid_num -eq 0 ]; then
                        break
                    fi
                done
            fi
            pid_num=$( ps -ef | grep $DOWNLOAD_PROCESS_NAME
E | grep cudnn | grep -v grep | wc -l )
            if [ $pid_num -gt 0 ]; then
                echo "cuDNN-${3} downloading , need about
30 seconds . Remaining installation time 200 - 430
seconds !"
                check_down load "cudnn "
            fi
            pid_num=$( ps -ef | grep $DRIVER_PRO cess_NAME |
grep -v grep | wc -l )
            if [ $pid_num -gt 0 ]; then
                echo
                echo "Driver-${1} installing , need 30 -
160 seconds . Remaining installation time 160 - 400
seconds !"
                check_inst all "NVIDIA "
                check_inst all_log "DRIVER "
            fi
            pid_num=$( ps -ef | grep $CUDA_PROCE SS_NAME |
grep -v grep | wc -l )
            if [ $pid_num -gt 0 ]; then

```



```

        echo " CUDA-${ 2 } installing , need 80 - 200
seconds . Remaining installati on time 90 - 220 seconds
!"
        check_inst all " cuda "
        check_inst all_log " CUDA "
    fi
    pid_num=$( ps - ef | grep $ CUDNN_PROC ESS_NAME |
grep - v grep | wc - l )
    if [ $ pid_num - gt 0 ]; then
        echo " cuDNN-${ 3 } installing , need 10
seconds . Installati on will be successful soon , please
wait ....."
        check_inst all " cudnn "
        check_inst all_log " CUDNN "
    fi
    fi
    sleep 1
done
}
create_nvdi a_repo_c entos ()
{
    baseurl_ce ntos=$( cat / etc / yum . repos . d / CentOS - Base
. repo | grep baseurl | head - 1 | awk - F '[' ']' '{ print $
1 "/"$ 3 }')
    if [ - z "$ baseurl_ce ntos " ]; then
        url =" http :// mirrors . cloud . aliyuncs . com "
    fi
    cudaurl =$ baseurl_ce ntos "/ opsx / ecs / linux / rpm / cuda /
${ version }/$ basearch /"
    driverurl =$ baseurl_ce ntos "/ opsx / ecs / linux / rpm /
driver /${ version }/$ basearch /"
    echo "[ ecs - cuda ]" > / etc / yum . repos . d / nvidia . repo
    echo " name = ecs cuda - \"$ basearch " >> / etc / yum . repos
. d / nvidia . repo
    echo $ cudaurl >> / etc / yum . repos . d / nvidia . repo
    echo " enabled = 1 " >> / etc / yum . repos . d / nvidia . repo
    echo " gpgcheck = 0 " >> / etc / yum . repos . d / nvidia .
repo
    echo "[ ecs - driver ]" >> / etc / yum . repos . d / nvidia .
repo
    echo " name = ecs driver - \"$ basearch " >> / etc / yum .
repos . d / nvidia . repo
    echo $ driverurl >> / etc / yum . repos . d / nvidia . repo
    echo " enabled = 1 " >> / etc / yum . repos . d / nvidia . repo
    echo " gpgcheck = 0 " >> / etc / yum . repos . d / nvidia .
repo
    yum clean all >> $ log 2 >& 1
    yum makecache >> $ log 2 >& 1
}
disable_no uveau_cent os ()
{
    echo " blacklist nouveau " > / etc / modprobe . d / blacklist
- nouveau . conf
    echo " options nouveau modeset = 0 " >> / etc / modprobe . d
/ blacklist - nouveau . conf
    echo "*** exec \" dracut -- force \" to regenerate the
kernel initramfs "
    dracut -- force
}
disable_no uveau_ubun tu ()
{
    echo " blacklist nouveau " > / etc / modprobe . d / blacklist
- nouveau . conf

```

```

    echo " options    nouveau    modeset = 0 " >> / etc / modprobe . d
/ blacklist - nouveau . conf
    echo "*** exec \" update - initramfs - u \" to    regenerate
the    kernel    initramfs \"
    update - initramfs - u
}
install_kernel_centos ()
{
    kernel_version=$(uname -r)
    kernel_dev_num=$(rpm -qa | grep kernel - devel |
grep $kernel_version | wc -l)
    if [ $kernel_dev_num -eq 0 ]; then
        echo "***** exec \" yum install -y kernel - devel -$
kernel_version \"
        yum install -y kernel - devel -$kernel_version
        if [ $? -ne 0 ]; then
            echo " INSTALL_ERROR : install kernel - devel
fail !!!"
            return 1
        fi
    fi
    return 0
}
install_kernel_suse ()
{
    kernel_version=$(uname -r | awk -F '-' '{ print $1 "-"$
2 }')
    kernel_dev_num=$(rpm -qa | grep kernel - default -
devel | wc -l)
    if [ $kernel_dev_num -eq 0 ]; then
        echo "*** exec \" zypper install -y kernel - default
- devel -$kernel_version \"
        zypper install -y kernel - default - devel =$
kernel_version
        if [ $? -ne 0 ]; then
            echo " error : install kernel - default - devel
fail !!!"
            return 1
        fi
    fi
}
install_kernel_ubuntu ()
{
    kernel_version=$(uname -r)
    linux_headers_num=$(dpkg --get-libs | grep linux - headers
| grep $kernel_version | wc -l)
    if [ $linux_headers_num -eq 0 ]; then
        echo "*** exec \" apt-get install -y --allow-
unauthenticated linux - headers -$kernel_version \"
        apt-get install -y --allow-unauthenticated
linux - headers -$kernel_version
        if [ $? -ne 0 ]; then
            echo " error : install linux - headers fail !!!"
            return 1
        fi
    fi
}
download ()
{
    download_url="${baseurl}/opsx/ecs/linux/binary/
nvidia "
    wget ${download_url}/driver/${driver_file}
    if [ $? -ne 0 ]; then

```

```

        echo "INSTALL_ER ROR : Download driver fail !!!"
return : $?"
    return 1
fi
    cudafileli st=$( curl ${download_u rl }/ cuda /${
cuda_versi on }/ | grep " cuda_ ${ cuda_versi on }" | awk - F
'>' '{ print $ 2 }' | awk - F '<' '{ print $ 1 }')
    if [ - z "$ cudafileli st " ]; then
        echo "INSTALL_ER ROR : Download CUDA fail !!! get
cuda -${ cuda_versi on } filename fail !!!"
        return 1
    fi
    mkdir / root / nvidia / cuda
    cd / root / nvidia / cuda
    echo $ cudafileli st
    for cudafile in $ cudafileli st
    do
        sleep 1
        wget ${download_u rl }/ cuda /${ cuda_versi on }/$
cudafile
        if [ $? - ne 0 ]; then
            echo "INSTALL_ER ROR : Download CUDA fail !!!"
            wget $ cudafile fail ! return : $?"
            return 1
        fi
    done
    chmod + x / root / nvidia / cuda /*
    cd / root / nvidia
    wget ${download_u rl }/ cudnn /${ cuda_big_v ersion }/${
cudnn_file }
    if [ $? - ne 0 ]; then
        echo "INSTALL_ER ROR : Download cuDNN fail !!!"
        return :$?"
        return 1
    fi
    chmod + x / root / nvidia /*
    echo "$ DOWNLOAD_S UCCCESS_STR !"
    return 0
}
install_driver ()
{
    / root / nvidia /$ driver_fil e -- silent
    if [ $? - ne 0 ]; then
        echo "INSTALL_ER ROR : driver install fail !!!"
        return 1
    fi
    echo "$ DRIVER_SUC CCESS_STR !"
    return 0
}
install_cuda ()
{
    cd / root / nvidia / cuda
    cuda_file=$( ls - S | grep cuda | grep $ cuda_versi on
| head - 1 )
    echo " cuda file : "$ cuda_file
    if [ - z "$ cuda_file " ]
    then
        echo "INSTALL_ER ROR : cuda file is null , cuda
install fail !!!"
        return 1
    fi
    / root / nvidia / cuda /$ cuda_file -- silent -- toolkit --
samples -- samplespat h =/ root
    if [ $? - ne 0 ]; then

```

```

        echo "INSTALL_ER ROR : cuda install fail !!!"
        return 1
    fi
    cuda_patch file=$(ls | grep cuda | grep $ cuda_versi
on | grep -v ${ cuda_file })
    for cuda_patch in $ cuda_patch file
    do
        echo "install cuda patch file : "$ cuda_patch
        / root / nvidia / cuda /$ cuda_patch -- silent -- installdir
=/ usr / local / cuda -- accept - eula
        if [ $? -ne 0 ]; then
            echo "INSTALL_ER ROR : cuda patch install fail
!!!"
            return 1
        fi
    done
    echo "$ CUDA_SUCCE SS_STR !"
    return 0
}
install_cu_dnn ()
{
    tar zxvf / root / nvidia /$ cudnn_file -C / usr / local
    if [ $? -ne 0 ]; then
        echo "INSTALL_ER ROR : CUDNN INSTALL FAIL !!!"
        return 1
    fi
    echo "$ CUDNN_SUCC ESS_STR !"
    return 0
}
enable_pm ()
{
    echo "#!/ bin / bash " > / etc / init . d / enable_pm . sh
    echo " nvidia - smi - pm 1 " >> / etc / init . d / enable_pm
. sh
    echo " exit 0 " >> / etc / init . d / enable_pm . sh
    chmod + x / etc / init . d / enable_pm . sh
    str=$( tail -1 $ filename | grep " exit ")
    if [ -z "$ str " ]; then
        echo "/ etc / init . d / enable_pm . sh " >> $ filename
    else
        sed -i '$ i \ / etc / init . d / enable_pm . sh ' $
filename
    fi
    chmod + x $ filename
}
issue=$( cat / etc / issue | grep Ubuntu )
if [ -n "$ issue " ]; then
    os=" ubuntu "
    profile_file="/ root /. profile "
    filename="/ etc / rc . local "
else
    issue=$( cat / etc / issue | grep SUSE )
    if [ -n "$ issue " ]; then
        os=" suse "
        filename="/ etc / init . d / after . local "
    else
        os=" centos "
        filename="/ etc / rc . d / rc . local "
    fi
    profile_file="/ root /. bash_profi le "
fi
if [ "$ 1 " = " check " ]; then

```

```

    check_inst_all_processes $driver_version $cuda_version
$ cudnn_version
    sed -i '/ part - 001 / d' $profile_file
    exit 0
else
    mkdir $NVIDIA_DIR
    echo "begin to install, driver:$driver_version,
cuda:$cuda_version, cudnn:$cudnn_version" >> $log 2
>& 1
    driver_file="NVIDIA - Linux - x86_64 -"${driver_version}
}".run"
    cuda_big_version=$(echo $cuda_version | awk -F '.' '{
print $1" "$2}')
    cudnn_file="cudnn -"${cuda_big_version}"- linux - x64 - v
"${cudnn_version}".tgz"
    echo "sh /var/lib/cloud/instance/scripts/part-001
check" | tee -a $profile_file
fi
echo "os:$os" >> $log 2 >& 1
if [ "$os" = "ubuntu" ]; then
    disable_no_ubuntu >> $log 2 >& 1
    if [ -f "/etc/apt/sources.list.d/sources-aliyun-
0.list" ]; then
        repo_file="/etc/apt/sources.list.d/sources-
aliyun-0.list"
    else
        repo_file="/etc/apt/sources.list"
    fi
    baseurl=$(cat $repo_file | grep "^deb" | head -1 |
awk -F '[/]' '{print $1 "/"$3}' | awk -F ' ' '{print $2
}')
    if [ -z "$baseurl" ]; then
        baseurl="http://mirrors.cloud.aliyuncs.com"
    fi
elif [ "$os" = "suse" ]; then
    baseurl=$(cat /etc/zypp/repos.d/SLES12-SP2-0.
repo | grep baseurl | head -1 | awk -F '[=]' '{print $2
"/"$4}')
    if [ -z "$baseurl" ]; then
        baseurl="http://mirrors.cloud.aliyuncs.com"
    fi
elif [ "$os" = "centos" ]; then
    baseurl=$(cat /etc/yum.repos.d/CentOS-Base.repo
| grep baseurl | head -1 | awk -F '[/]' '{print $1 "/"$
3}' | awk -F '=' '{print $2}')
    if [ -z "$url" ]; then
        baseurl="http://mirrors.cloud.aliyuncs.com"
    fi
    if [ ! -f "/usr/bin/lsb_release" ]; then
        pkgname=$(yum provides /usr/bin/lsb_release |
grep centos | grep x86_64 | head -1 | awk -F : '{print $
1}')
        if [ -z "$pkgname" ]; then
            echo "INSTALL ERROR: /usr/bin/lsb_release
pkg not exists!" >> $log 2 >& 1
            exit 1
        fi
        yum install -y $pkgname >> $log 2 >& 1
    fi
    if [ ! -f "/usr/bin/gcc" ]; then
        yum install -y gcc
    fi
    disable_no_ubuntu_centos >> $log 2 >& 1
    str=$(lsb_release -r | awk -F '[:.]' '{print $2}')

```

```

        version=$( echo $ str | sed ' s / // g ')
        create_nvi dia_repo_c entos
    fi
    install_kernel ${ os } >> $ log 2 >& 1
    if [ $? -ne 0 ]; then
        echo " INSTALL_ERROR : kernel - devel install fail !!!"
    >> $ log 2 >& 1
        exit 1
    fi
    cd / root / nvidia
    begin_download=$( date '+% s ')
    download >> $ log 2 >& 1
    if [ $? -ne 0 ]; then
        exit 1
    fi
    end_download=$( date '+% s ')
    time_download=$(( end_download - begin_download ))
    echo " NVIDIA download OK ! Using time $ time_download s
    !" >> $ log 2 >& 1
    begin=$( date '+% s ')
    install_driver >> $ log 2 >& 1
    if [ $? -ne 0 ]; then
        exit 1
    fi
    end=$( date '+% s ')
    time_install=$(( end - begin ))
    echo " NVIDIA install driver OK ! Using time $ time_install
    s !" >> $ log 2 >& 1
    begin=$( date '+% s ')
    install_cuda >> $ log 2 >& 1
    if [ $? -ne 0 ]; then
        exit 1
    fi
    end=$( date '+% s ')
    time_install=$(( end - begin ))
    echo " NVIDIA install cuda OK ! Using time $ time_install
    s !" >> $ log 2 >& 1
    begin=$( date '+% s ')
    install_cudnn >> $ log 2 >& 1
    if [ $? -ne 0 ]; then
        exit 1
    fi
    end=$( date '+% s ')
    time_install=$(( end - begin ))
    echo " NVIDIA install cudnn OK ! Using time $ time_install
    s !" >> $ log 2 >& 1
    enable_pm
    echo " reboot ....." >> $ log 2 >& 1
    sleep 2
    reboot

```

What to do next

- If you do not configure automatic installation of the GPU driver when you create the GPU instance, you must install it manually. For more information, see [Install the GPU driver](#).

- If the selected gn5, gn5i, gn6v, vgn5i, or gn6i type instance need to support OpenGL, you must install the GRID driver. For more information, see [Install the GRID driver on a gn5/gn5i/gn6v instance](#).
- You can uninstall the GPU driver and the CUDA Toolkit. For more information, see [Uninstall the GPU driver](#).
- You can view GPU monitoring data by using the CloudMonitor console or call the API. For more information, see [GPU monitoring](#).


4.7.3 Install the GPU driver

This topic describes how to download and install the GPU driver. If you do not configure your GPU driver to automatically install when you create the GPU instance, you must install it manually after the GPU instance is created.

Download the GPU driver

To download a GPU driver, follow these steps:

1. Go to [NVIDIA website](#).
2. Locate the drivers for your NVIDIA products, and click Search. The following table provides relevant metrics.

Item	gn4	gn5	gn5i	gn6v	gn6i	
Product type	Tesla	Tesla	Tesla	Tesla	Tesla	
Product series	M-Class	P-Series	P-Series	V-Series	T-Series	
Product	M40	Tesla P100	Tesla P4	Tesla V100	Tesla T4	
Operating system (OS)	Select a release according to the image of your instance. <div>  Note: <ul style="list-style-type: none"> • If your OS is not in the drop-down list, click Select All Operating Systems at the bottom of the list. • For Debian, select Linux 64-bit. </div>					

3. Confirm the information is correct, then click Download.

Install the GPU driver for a Linux instance

To install the GPU driver, follow these steps:

1. Download and install the kernel-devel and kernel-header packages based on your kernel version.
2. Run the `sudo rpm -qa | grep $(uname -r)` command to verify that the kernel-devel and kernel-header packages are successfully installed.

For example, if you are using CentOS 7.3, the installation is successful if the following information is displayed.

```
kernel - 3 . 10 . 0 - 514 . 26 . 2 . el7 . x86_64
kernel - headers - 3 . 10 . 0 - 514 . 26 . 2 . el7 . x86_64
kernel - tools - libs - 3 . 10 . 0 - 514 . 26 . 2 . el7 . x86_64
python - perf - 3 . 10 . 0 - 514 . 26 . 2 . el7 . x86_64
kernel - tools - 3 . 10 . 0 - 514 . 26 . 2 . el7 . x86_64
```

3. Install the GPU driver according to the description in the ADDITIONAL INFORMATION area of the GPU download page.

The following figure uses Linux 64-bit Ubuntu 14.04 for example.



Note:

- The preceding operations do not apply to a vgn5i instance. For information about how to install a GRID driver on a vgn5i Linux instance, see [#unique_79](#).
- The version of kernel-devel and kernel-header must match your kernel version. Otherwise, driver compilation errors will occur during driver rpm installation. You can run `rpm -qa | grep kernel` in your instance to check version consistency before installing the driver.

Install the GPU driver for a Windows instance

Double-click the executable file to install the driver as instructed.



Note:

- The preceding operation does not apply to a vgn5i instance. For information about how to install a GRID driver on a vgn5i Windows instance, see [#unique_80](#).
- For Windows instances, the Windows RDP protocol does not support DirectX or OpenGL after the GPU driver is successfully installed. In this case, you need to

install the VNC service and client or other protocols (for example, PCoIP, and XenDesktop HDX 3D).

4.7.4 Uninstall the GPU driver

This topic describes how to uninstall the GPU driver from a GPU instance. Depending on the installation script and the image type, the command needed to uninstall the GPU driver varies.

Background

This topic uses Driver 390.46 and CUDA 9.0.176 in the example. The GPU driver can only be uninstalled by the `root` user. If you are a common user, you must run the `sudo` command to obtain the privileges of the `root` user.

If you use the `nvidia_install_v1.0` package, the format is either `rpm` or `deb`. If you use `nvidia_install_v2.0` package, the format is `runfile`.

Uninstall the GPU driver installed by using `nvidia_install_v1.0`

- If the operating system of your GPU instance is Ubuntu, follow these steps:

1. Run the following command to uninstall the GPU driver:

```
apt - get remove -- purge nvidia - 390 nvidia - modprobe  
nvidia - settings  
apt - get remove nvidia - diag - driver - local - repo -  
ubuntu1604 - 390 . 46
```

2. Optional. Run the following command to uninstall X server.



Note:

If you install the GPU driver by using the `deb` format package, an error message is displayed when you use the `runfile` package to install the GPU

driver. You need to uninstall or stop X server and restart the GPU instance before you can use the `runfile` package to install the GPU driver.

```
apt - get remove xserver - xorg xserver - xorg - core
```

3. Run the following command to uninstall the CUDA Toolkit.

```
apt autoremove -- purge cuda - 9 . 0
```

4. Run the following command to restart the instance.

```
reboot
```

- If the operating system of your GPU instance is CentOS 7, follow these steps:

1. Run the following command to uninstall the GPU driver:

```
yum remove xorg - x11 - drv - nvidia nvidia - kmod cuda -  
drivers  
yum remove nvidia - diag - driver - local - repo - rhel7 -  
390 . 46
```

2. Run the following command to uninstall the CUDA Toolkit.

```
yum remove / usr / local / cuda - 9 . 0
```

3. Run the following command to restart the instance.

```
reboot
```

- If the operating system of your GPU instance is CentOS 6, follow these steps:

1. Run the following command to uninstall the GPU driver:

```
yum remove xorg - x11 - drv - nvidia nvidia - kmod cuda -  
drivers  
yum remove nvidia - diag - driver - local - repo - rhel6 -  
390 . 46
```

2. Run the following command to uninstall the CUDA Toolkit.

```
yum remove / usr / local / cuda - 9 . 0
```

3. Run the following command to restart the instance.

```
reboot
```

Uninstall the GPU driver installed by using `nvidia_install_v2.0`

To uninstall the GPU driver, follow these steps:

1. Run the following command to uninstall the GPU driver:

```
/usr/bin/nvidia-uninstall
```

2. Run the following command to uninstall the CUDA Toolkit and the cuDNN library.

```
/usr/local/cuda/bin/uninstall_cuda_9.0.pl  
rm -rf /usr/local/cuda-9.0
```

4.7.5 Install a GRID driver in a GPU-equipped ECS instance (Linux)

This topic describes how to install a GRID driver and build a desktop environment in a GPU-equipped ECS instance running Linux of the following instance type families: gn5, gn5i, gn6v, gn6i, and vgn5i.

Background information

If your GPU-equipped instance requires OpenGL, you must install the GRID driver on the instance. This is because the NVIDIA GRID license granted to NVIDIA GPUs (such as Tesla P100, Tesla P4 and Tesla V100) of the supported instances cannot meet the graphics requirements of OpenGL.

Prerequisites

- A GPU compute instance that can access the Internet is created. For more information, see [Create a compute optimized instance with GPU](#).



Note:

We recommend that you select a Public Image. If you select a Marketplace Image that is pre-installed with an NVIDIA driver, after the instance is created, you must disable the Nouveau driver through this process: create a file named `nouveau.conf` in the `/etc/modprobe.d` directory, and then add `blacklist nouveau` to it.

- VNC software (such as VNC Viewer, which is used in this example) is installed.

Install a GRID driver in Ubuntu 16.04 64-bit

To install the GRID driver, follow these steps:

1. [Connect to the Linux instance](#).

2. Run the following commands in sequence to upgrade the system and install the KDE.

```
apt - get      update
apt - get      upgrade
apt - get      install    kubuntu - desktop
```

3. Run `reboot` to restart the system.
4. [Connect to the Linux instance](#) again, and then run the following commands to download and decompress the NVIDIA GRID driver package.

The NVIDIA GRID driver package contains the drivers for various operating systems. For Linux OS, select NVIDIA-Linux-x86_64-390.57-grid.run.

```
wget http://nvidia-driver-410.oss-cn-shenzhen-
aliyuncs.com/NVIDIA-Linux-x86_64-410.39-grid.run
```

5. Run the following commands in sequence, then follow the prompts to install the NVIDIA GRID driver.

```
chmod 777 NVIDIA-Linux-x86_64-410.39-grid.run
./NVIDIA-Linux-x86_64-410.39-grid.run
```

6. Run `nvidia-smi` to verify the installation.

If the following message appears, the driver is installed successfully.

```
$nvidia-smi
Wed Apr 10 10:24:58 2019

+-----+
| NVIDIA-SMI 410.39                Driver Version: 410.39          |
+-----+-----+
| GPU  Name            Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+
|  0   Tesla P4             On      | 00000000:05:00.0 Off  |             Off      |
| N/A   34C    P8      10W /  75W   |  25MiB /  8191MiB |      0%    Default   |
+-----+-----+
|  1   Tesla P4             On      | 00000000:87:00.0 Off  |             Off      |
| N/A   32C    P8      11W /  75W   |  25MiB /  8191MiB |      0%    Default   |
+-----+-----+

+-----+
| Processes:                                     GPU Memory |
|  GPU       PID    Type    Process name                     Usage    |
+-----+-----+
| No running processes found                     |
+-----+
```

7. Add License Server to activate the license:

- a. Switch to the /etc/nvidia directory: `cd / etc / nvidia .`
- b. Create a file named gridd.conf: `cp gridd . conf . template gridd . conf .`
- c. Add the following lines about License Server to the gridd.conf file.

```
ServerAddress = IP address of the License server
ServerPort = Port of the License server ( defaulted to
7070 )
FeatureType = 2
EnableUI = TRUE
```

8. Run the following command to install x11vnc.

```
apt - get install x11vnc
```

9. Run `lspci | grep NVIDIA` to check GPU BusID.

In this example, the GPU BusID is `00 : 07 . 0 .`

10. Configure the X Server environment and restart the system:

- a. Run `nvidia - xconfig -- enable - all - gpus -- separate - x - screens .`
- b. Edit /etc/X11/xorg.conf: Add your GPU BusID to the Section " Device ". In this example, `BusID " PCI : 0 : 7 : 0 "` is added .

```
Section "Device"
    Identifier      "Device0"
    Driver          "nvidia"
    VendorName      "NVIDIA Corporation"
    BoardName       "Tesla P4"
    BusID           "PCI:0:7:0"
EndSection
```

- c. Run `reboot` to restart the system.

Verify installation

To verify the installation of the GRID driver, follow these steps:

1. Run the following command to install the GLX application.

```
apt - get install mesa - utils
```

2. Run `startx` to start X Server.



Note:

- If the `startx` command cannot be found, run `apt - get install xinit` to install it.
- Running `startx` may result in the `hostname : Name or service not known` error. This error has no effect on starting X Server. Run `hostname` to obtain the host name of the instance, and then modify the `/etc/hosts` file by replacing the `hostname`, which is preceded by `127 . 0 . 0 . 1`, with the actual host name of your instance.

```
root@iz[REDACTED]:~# startx
hostname: Name or service not known
xauth: (stdin):1: bad display name "iz[REDACTED]:1" in "add" command
```

3. Start a new terminal session of the SSH client and run the following command to start `x11vnc`.

```
x11vnc - display : 1
```

If the following message appears, `x11vnc` has been restarted successfully. At this stage, you can connect to the instance by using a VNC application. In this example, VNC Viewer is used.

```
The VNC desktop is:      localhost:0
PORT=5900
```

4. Log on to the ECS console, and [add security group rules](#) to the target security group to allow inbound traffic from the Internet through the TCP 5900 port.
5. On the local machine, start VNC Viewer and enter the `public IP address` of the instance : 5900 to connect to the instance and enter the KDE desktop.
6. Run `glxinfo` to view the configurations supported by the current GRID driver:
 - a. Start a new terminal session of the SSH client.
 - b. Run `export DISPLAY =: 1`.
 - c. Run `glxinfo - t` to list the configurations supported by the current GRID driver.

7. Run `glxgears` to test the GRID driver:

- a. On the KDE desktop, right-click the desktop and select Run Command.
- b. Run `glxgears` to start the testing application. If the following figure appears, the GRID driver works normally.

Install a GRID driver in CentOS 7.3 64-bit

To install the GRID driver, follow these steps:

1. [Connect to the Linux instance](#).
2. Run the following commands in sequence to upgrade the system and install the KDE.

```
yum update
yum install kernel - devel
yum groupinstall "KDE Plasma Workspaces"
```

3. Run `reboot` to restart the system.
4. [Connect to the Linux instance](#) again, and then run the following commands to download and decompress the NVIDIA GRID driver package.

The NVIDIA GRID driver package contains the drivers for various operating systems. For Linux OS, select `NVIDIA-Linux-x86_64-390.57-grid.run`.

```
wget http://nvidia-driver-410.oss-cn-shenzhen.aliyuncs.com/NVIDIA-Linux-x86_64-410.39-grid.run
```

5. Disable the nouveau driver:

- a. Run `vim /etc/modprobe.d/blacklist.conf`, and add `blacklist nouveau` to the file.
- b. Run `vim /lib/modprobe.d/dist-blacklist.conf` and add the following lines.

```
blacklist nouveau
options nouveau modeset = 0
```

- c. Run `mv /boot/initramfs-$(uname -r).img /boot/initramfs-$(uname -r)-nouveau.img`.
- d. Run `dracut /boot/initramfs-$(uname -r).img $(uname -r)`.

6. Run `reboot` to restart the system.

7. Run the following commands in sequence, then follow the prompts to install the NVIDIA GRID driver.

```
chmod 777 NVIDIA - Linux - x86_64 - 410 . 39 - grid . run
./ NVIDIA - Linux - x86_64 - 410 . 39 - grid . run
```

8. Run `nvidia - smi` to verify the installation.

If the following message appears, the driver is installed successfully.

```
$nvidia-smi
Wed Apr 10 10:24:58 2019

+-----+
| NVIDIA-SMI 410.39                Driver Version: 410.39          |
+-----+-----+
| GPU  Name            Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+
|    0   Tesla P4             On      | 00000000:05:00.0 Off |           Off        |
| N/A   34C    P8      10W /  75W |  25MiB /  8191MiB |      0%      Default  |
+-----+-----+
|    1   Tesla P4             On      | 00000000:87:00.0 Off |           Off        |
| N/A   32C    P8      11W /  75W |  25MiB /  8191MiB |      0%      Default  |
+-----+-----+

+-----+
| Processes:                                     GPU Memory               |
|  GPU       PID    Type    Process name                       Usage                     |
+-----+-----+
| No running processes found                       |
+-----+
```

9. Add License Server to activate the license:

- Run `cd / etc / nvidia` to switch to the `/etc/nvidia` directory.
- Run `cp gridd . conf . template gridd . conf` to create a file named `gridd.conf`.
- Add the following lines about License Server to the `gridd.conf` file.

```
ServerAddr  ess = IP    address    of the License server
ServerPort  = Port    of the License server ( defaulted to
7070 )
FeatureType e = 2
EnableUI    = TRUE
```

10. Run the following command to install `x11vnc`.

```
yum install x11vnc
```

11. Run `lspci | grep NVIDIA` to check GPU BusID.

In this example, the GPU BusID is `00 : 07 . 0`.

12. Configure the X Server environment:

- a. Run `nvidia - xconfig -- enable - all - gpus -- separate - x - screens .`
- b. Edit `/etc/X11/xorg.conf`: Add your GPU BusID to the `Section " Device "`. In this example, `BusID " PCI : 0 : 7 : 0 "` is added

13. Run `reboot` to restart the system.

Verify installation

To verify the installation of the GRID driver, follow these steps:

1. Run `startx` to start X Server.
2. Start a new terminal session of the SSH client and run the command to start `x11vnc`.

```
x11vnc - display : 0
```

If the following message appears, `x11vnc` has been restarted successfully. Now you can connect to the instance by using a VNC application. In this example, VNC Viewer is used.

3. Log on to the ECS console, and [add security group rules](#) to the target security group to allow inbound traffic from the Internet through TCP 5900 port.
4. On the local machine, start VNC Viewer and enter the `public IP address of the instance : 5900` to connect to the instance and enter the KDE desktop.
5. Run `glxinfo` to view the configurations supported by the current GRID driver:
 - a. Start a new terminal session of the SSH client.
 - b. Run `export DISPLAY =: 0 .`
 - c. Run `glxinfo - t` to list the configurations supported by the current GRID driver.

6. Run `glxgears` to test the GRID driver:
 - a. On the VNC Viewer, right-click the desktop and select Run Command.
 - b. Run `glxgears` to start the testing application. If the following image appears, the GRID driver works normally.

4.7.6 Install an NVIDIA GRID driver on a vgn5i instance (Windows)

This topic describes how to install an NVIDIA GRID driver for instances running Windows Server 2016(64-bit) of the vgn5i instance type family. You must install an NVIDIA GRID driver if your GPU-equipped compute optimized instance requires Open Graphics Library (OpenGL). By default, the NVIDIA GRID license granted to NVIDIA GPUs is not activated, but you can activate the license by using a trial license to build a license server.

Prerequisites

- A vgn5i instance that can access the Internet is created. For more information, see [#unique_83](#).



Note:

We recommend that you select a Public Image when you create an instance.

- The Remote Desktop Connection tool or a remote connection tool such as VNC Viewer is installed on your local machine.
- An NVIDIA GRID license is obtained. If you need to obtain a license, you can:
 - Visit the [NVIDIA official website](#) and register an account to apply for a 90-day trial license.
 - Obtain a temporary license from Alibaba Cloud technical support.
 - Buy a license from NVIDIA.

Procedure

1. [Connect to the instance](#).

2. Download the NVIDIA GRID driver package.

- If your instance is located in China (Shenzhen) Zone C or Zone D, run the following command to download [GRID 7.0 Guest](#).
- If your instance is located in other regions or zones, run the following command to download GRID 8.0 Guest based on your Windows operating system:
 - [GRID 8.0 for Windows 10 and Windows Server 2016](#)
 - [GRID 8.0 for Windows 8, Windows 7, Windows Server 2012R2, and Windows Server 2008R2](#)

3. Install the NVIDIA GRID driver.



Note:

For Windows operating systems, the Remote Desktop Protocol (RDP) may not support DirectX or OpenGL after the GPU driver is installed. In this case, you need to manually install the VNC server and client, or use other protocols that support the preceding applications, for example, PCoIP or XenDesktop HDX 3D.

4. Restart the instance.

5. Add a license server and activate the license.

- a) [Connect to the instance](#).
- b) Right-click on the desktop and choose NVIDIA Control Panel.
- c) In the left-side navigation pane, choose Licensing > Manage License.
- d) Enter the IP address and port number of the license server, and click Apply.

If Your system is licensed for GRID vGPU is displayed, the license is activated.

4.7.7 Install an NVIDIA GRID driver on a vgn5i instance (Linux)

This topic describes how to install an NVIDIA GRID driver for instances running 64-bit Ubuntu 16.04 of the vgn5i instance type family. You must install an NVIDIA GRID driver if your GPU-equipped compute optimized instance requires Open Graphics Library (OpenGL). By default, the NVIDIA GRID license granted to NVIDIA GPUs is not activated, but you can activate the license by using a trial license to build a license server.

Prerequisites

- A vgn5i instance that can access the Internet is created. For more information, see [#unique_83](#).

**Note:**

We recommend that you select a Public Image when you create the instance.

- A remote connection tool such as VNC Viewer is installed on your local machine.
- An NVIDIA GRID license is obtained. If you need to obtain a license, you can:
 - Visit the [NVIDIA official website](#) and register an account to apply for a 90-day trial license.
 - Obtain a temporary license from Alibaba Cloud technical support.
 - Buy a license from NVIDIA.

Procedure**1. Disable Nouveau.**

a) [Connect to the instance](#).

b) Check whether the `blacklist - nouveau . conf` file exists.

```
ls / etc / modprobe . d / blacklist - nouveau . conf
```

c) If the file exists, skip this step. If the file does not exist, run the `vim / etc / modprobe . d / blacklist - nouveau . conf` command to create the file and then add the following content to the file:

```
blacklist nouveau
blacklist lbm - nouveau
options nouveau modeset = 0
```

d) Remove Nouveau and generate kernel initramfs.

```
rmmod nouveau
```

```
update - initramfs - u
```

e) Restart the instance.

```
reboot
```

2. Download the NVIDIA GRID driver package.

a) [Connect to the instance](#).

b) Download the NVIDIA GRID driver package.

- If your instance is located in China (Shenzhen) Zone C or Zone D, run the following command to download GRID 7.0 Guest:

```
wget http://nvidia-driver-410.oss-cn-shenzhen.aliyuncs.com/NVIDIA-Linux-x86_64-410.39-grid.run
```

- If your instance is located in other regions or zones, run the following command to download GRID 8.0 Guest:

```
wget http://nvidia-418.oss-cn-shenzhen.aliyuncs.com/NVIDIA-Linux-x86_64-418.70-grid.run
```

3. Install the NVIDIA GRID driver.

```
chmod +x NVIDIA-Linux-x86_64-410.39-grid.run
./NVIDIA-Linux-x86_64-410.39-grid.run
```

4. Check whether the NVIDIA GRID driver is installed.

```
nvidia-smi
```

The driver is installed if GRID driver information similar to the following figure is displayed.

```
root@~:~# nvidia-smi
Mon May 6 13:44:37 2019
+-----+
| NVIDIA-SMI 410.39                Driver Version: 410.39 |
+-----+-----+
| GPU Name      Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+
|   0  GRID P4-4Q      On   | 00000000:00:07.0 Off |          N/A         |
| N/A   N/A   P8     N/A /  N/A | 272MiB / 4096MiB |      0%      Default |
+-----+-----+
+-----+
| Processes: |
| GPU       PID   Type   Process name                      | GPU Memory |
+-----+-----+
|          |          |          |          |          |          |
```

5. Add a license server.

- a) Switch to the `/ etc / nvidia` directory.

```
cd / etc / nvidia
```

- b) Create a `gridd . conf` file.

```
cp gridd . conf . template gridd . conf
```

- c) Add license server information to the created `gridd . conf` file.

```
ServerAddr ess = IP address of the license server
ServerPort = Port of the license server ( default port
: 7070 )
FeatureType = 1
```

6. Restart the instance for the license server configuration to take effect.

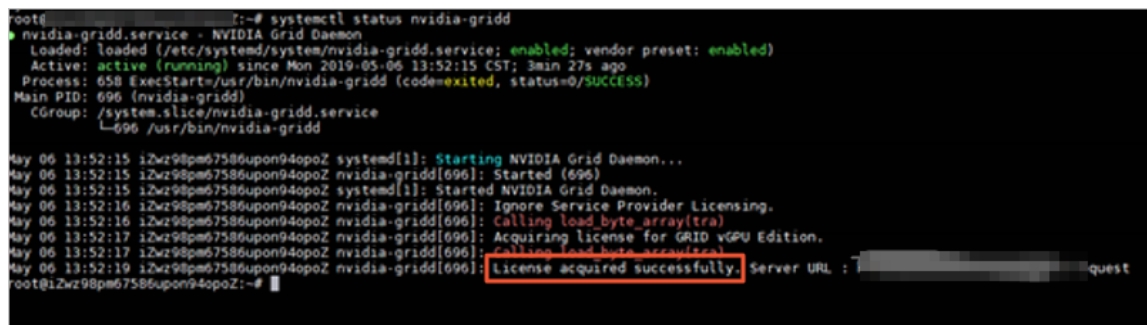
```
reboot
```

7. Check whether the license is activated.

- a) [Connect to the instance.](#)
- b) Check the license status.

```
systemctl status nvidia - gridd
```

The license is activated if License acquired successfully is displayed.



```
root@i2wz98pm67586upon940poZ:~# systemctl status nvidia-gridd
nvidia-gridd.service - NVIDIA Grid Daemon
Loaded: loaded (/etc/systemd/system/nvidia-gridd.service; enabled; vendor preset: enabled)
Active: active (running) since Mon 2019-05-06 13:52:15 CST; 3min 27s ago
Process: 658 ExecStart=/usr/bin/nvidia-gridd (code=exited, status=0/SUCCESS)
Main PID: 696 (nvidia-gridd)
CGroup: /system.slice/nvidia-gridd.service
└─696 /usr/bin/nvidia-gridd

May 06 13:52:15 i2wz98pm67586upon940poZ systemd[1]: Starting NVIDIA Grid Daemon...
May 06 13:52:15 i2wz98pm67586upon940poZ nvidia-gridd[696]: Started (696)
May 06 13:52:15 i2wz98pm67586upon940poZ systemd[1]: Started NVIDIA Grid Daemon.
May 06 13:52:16 i2wz98pm67586upon940poZ nvidia-gridd[696]: Ignore Service Provider Licensing.
May 06 13:52:16 i2wz98pm67586upon940poZ nvidia-gridd[696]: Calling load_byte_array(trs)
May 06 13:52:17 i2wz98pm67586upon940poZ nvidia-gridd[696]: Acquiring license for GRID vGPU Edition.
May 06 13:52:17 i2wz98pm67586upon940poZ nvidia-gridd[696]: Calling load_byte_array(trs)
May 06 13:52:19 i2wz98pm67586upon940poZ nvidia-gridd[696]: License acquired successfully. Server URL : [redacted] quest
root@i2wz98pm67586upon940poZ:~#
```

4.8 Visualization compute type family with GPU

4.8.1 GPU-equipped instance type families designed for visualization computing

This topic describes the instance type family `ga1`, which is GPU-equipped and designed for visualization computing, and lists the specific instance types within the instance type family.

4.8.2 Create a ga1 instance

This topic describes how to create a ga1 instance.

Image description

GPU visualization computing ga1 instances use the AMD S7150 series GPU. Alibaba Cloud and AMD work together to optimize GPU drivers, so you can select the following images with preinstalled drivers from Image Marketplace:

- Ubuntu16.04 with AMD GPU driver preinstalled
- Windows Server 2016 English version with AMD GPU driver preinstalled
- Ubuntu16.04 with AMD GPU driver and KDE preinstalled

Procedure

You can create a ga1 instance by following the instructions provided in [creating an instance](#). Use the following configurations when configuring an instance:

- Network Type: Select VPC. The current GPU-rendered ga1 instances only support VPCs.
- Instance: Select Heterogeneous Computing > Visualization Compute Optimized Type with GPU > GPU Rendering Type ga1.
- Image: Click Marketplace Image, and then click Select from image market (including operating system). Enter GPU or AMD in the search box to search images.



Note:

It is recommended that you purchase or subscribe to these images because you can select images later for another instance directly from Purchased Images or Subscribed Images.

Precautions

- The driver used for GPU visualization computing ga1 instances is optimized jointly by Alibaba Cloud and AMD. It is currently available only in images provided by Alibaba Cloud. No driver download link is provided, so you cannot install the driver on your own.

- If the driver does not work due to uninstallation or deletion of GPU driver-related components, [change a system disk](#) to restore the GPU functions.

**Warning:**

Changing a system disk may cause data loss.

- If you use an unspecified image when creating a GPU visualization computing ga1 instance, the instance driver will not work. In this case, [change a system disk](#) to use the image with an AMD GPU driver preinstalled.
- For Windows, after the GPU driver takes effect, the Connect function cannot be used, and the Management Terminal page displays a black screen or the startup page. In this case, use another protocol to enter the OS, for example, Windows embedded Remote Desktop Protocol (RDP).
- RDP does not support DirectX, OpenGL, or other related applications. You need to install the VNC service and client or use other protocols, such as XenDesktop HDX 3D and PCOIP.

4.9 Compute optimized type family with FPGA

4.9.1 Compute optimized and FPGA-equipped instance type families

This topic describes the compute optimized instance type families with FPGAs f1 and f3, and lists the specific instance types within each of these two instance type families.

4.9.2 Create an f1 instance

This topic describes how to create an f1 instance.

Prerequisites

You must use an image that is pre-installed with the Intel development environment to create an f1 instance. To obtain the image, [open a ticket](#).

Procedure

Follow the steps described in [create an ECS instance](#). The following configurations must be selected:

- Region: Select China (Hangzhou) > Zone F.

- **Instance Type:** Select Heterogeneous Computing > Compute Optimized Type with FPGA, and then select the appropriate f1 instance type.
- **Image:** Select Shared Image, and then select the shared image.

**Note:**

You must use an image that is pre-installed with the Intel development environment to create an f1 instance. This image is not available in the Alibaba Cloud Marketplace directly. To obtain the image, please find quartus17.0, vcs2017.3, dcp sdk in the `opt` directory.

- **Network Type:** Select VPC, and select a created VPC and VSwitch.

After an f1 instance is created, [connect to the instance](#) and run the following command to check whether the licence is configured.

```
echo $ LM_LICENSE _FILE # Check whether the variable is  
set .
```

Best practices

See best practices of f1 instances:

- [Use OpenCL on an f1 instance](#)
- [Use f1 RTL \(Register Transfer Level\)](#)

4.9.3 Create an f3 instance

This topic describes how to create an f3 instance.

Procedure

For more information about how to create an f3 instance, see [create an instance by using the wizard](#). However, the following configurations are recommended:

- **Billing Method:** Select Pay-As-You-Go or Subscription.

**Note:**

f3 instances are not available as preemptible instances.

- **Region:** Select China (Shanghai).
- **Instance Type:** Select Heterogeneous Computing > Compute Optimized Type with FPGA, and then select your required instance type.

- **Image:** Click Shared Image, and then select the specified image.



Note:

A Xilinx image is available for use (recommended). The image is only available as a Shared image. To obtain the image, open a ticket.

- **System Disk:** Allocate a 200 GiB Ultra Disk for the system image.
- **Network Type:** Select VPC.

Best practices

[#unique_97](#)

[#unique_98](#)

4.10 ECS bare metal instance type family

4.10.1 What is an ECS Bare Metal Instance?

This topic describes ECS Bare Metal Instances and the following instance families and their instance types: ebmc5s compute optimized instance family with enhanced network performance, ebmg5s general purpose instance family with enhanced network performance, ebmr5s memory optimized instance family with enhanced network performance, ebmhfg5 instance family with high clock speed, ebmc4 compute optimized instance family, and ebmg5 general purpose instance family.

ECS Bare Metal Instance overview

ECS Bare Metal Instances are a computing service that combines the elasticity of virtual machines and the performance and features of physical machines. ECS Bare Metal Instances are designed based on the state-of-the-art virtualization 2.0 technology developed by Alibaba Cloud. The virtualization used by ECS Bare Metal Instances is optimized to support common ECS instances and nested virtualization, maintaining the elastic performance and user experience of physical machines.

ECS Bare Metal Instances combine the strengths of both physical machines and ECS instances to deliver powerful and robust computing capabilities. With the virtualization 2.0 technology, ECS Bare Metal Instances provide your business applications with direct access to the processor and memory resources of the underlying servers without virtualization overheads. ECS Bare Metal Instances keep the hardware feature sets (such as Intel® VT-x) and resource isolation capabilities of physical machines, ideal for applications that need to run in non-virtualization environments.

By virtue of the independently developed chips and hypervisor system software and the redefined server hardware architecture, ECS Bare Metal Instances deeply integrate features from both physical and virtual machines. ECS Bare Metal Instances can seamlessly connect with other Alibaba Cloud products for storage, networking, and database tasks. ECS Bare Metal Instances are fully compatible with ECS instance images. These properties allow you to build resources to suit your business requirements.

ECS Bare Metal Instances provide the following benefits through technological innovation:

- Exclusive computing resources

ECS Bare Metal Instances provide better performance and resource isolation than standard physical machines and can ensure the exclusivity of computing resources without virtualization overhead or feature loss. ECS Bare Metal Instances support ultrahigh-frequency instances and can contain 8, 32, or 96 vCPUs. An ECS Bare Metal Instance with eight vCPUs can provide a core frequency of 3.7 to 4.1 GHz for better performance and faster response for gaming and finance businesses than peer services.

- Chip-level security

®

ECS Bare Metal Instances implement Intel® SGX, which allows ECS Bare Metal Instances to compute only encrypted data in a safe and trusted environment, to provide advanced security for cloud data. For more information, see [Install SGX](#).

- Compatible with multiple private clouds

ECS Bare Metal Instances can address the needs of high-performance computing and help you build new hybrid clouds. Thanks to the flexibility, elasticity, and other strengths inherited from the mix of physical and virtual machines, ECS Bare Metal Instances can implement re-virtualization. Offline private clouds can be seamlessly migrated to Alibaba Cloud without the performance overheads that may arise from nested virtualization, giving you a new method to move businesses onto the cloud.

- Support for heterogeneous instruction set processors

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

When you use ECS Bare Metal Instances, note the following items:

- ECS Bare Metal Instances cannot be upgraded or downgraded.
- If a hardware fault occurs to an ECS Bare Metal Instance, a failover occurs and data is stored in the disks of the ECS Bare Metal Instance.

Comparison of ECS Bare Metal Instances, physical machines, and virtual machines

The following table shows a comparison of features. In this table, Y means supported, N means not supported, and N/A means not applicable.

Feature type	Feature	ECS Bare Metal Instance	Physical machine	Virtual machine
Automated O&M	Delivery in minutes	Y	N	Y
Computing	Zero performance loss	Y	Y	N
	Zero feature loss	Y	Y	N
	Zero resource preemption	Y	Y	N
Storage	Compatible with ECS disks	Y	N	Y
	Startup from disks (system disks)	Y	N	Y
	Quick reset of system disks	Y	N	Y
	Compatible with ECS images	Y	N	Y

Feature type	Feature	ECS Bare Metal Instance	Physical machine	Virtual machine
	Cold migration between physical and virtual machines	Y	N	Y
	Free of OS installation	Y	N	Y
	Free of local RAID, and stronger protection of data in disks	Y	N	Y
Network	Compatible with ECS VPC networks	Y	N	Y
	Compatible with ECS classic networks	Y	N	Y
	Free of communication bottlenecks between physical and virtual machine clusters in VPC	Y	N	Y
Management	Compatible with existing ECS management systems	Y	N	Y
	Consistent user experiences on VNC and other features with that of virtual machines	Y	N	Y

Feature type	Feature	ECS Bare Metal Instance	Physical machine	Virtual machine
	Out-of-band (OOB) network security	Y	N	N/A

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmc5s, compute optimized ECS Bare Metal Instance family with enhanced network performance

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:2
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors, 96 vCPUs, up to 2.7 GHz Turbo Boost
- High network performance: 4.5 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation
- Suitable for the following scenarios:
 - Scenarios where large volumes of packets are received and transmitted, such as bullet screen and re-transmission of telecommunication information
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
 - Containers (includes but is not limited to Docker, Clear Containers, and Pouch)
 - Video encoding, decoding, and rendering

Instance types


Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmc5s.24xlarge	96	192.0	N/A	30.0	4,500	No	8	32	10

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmc5s, general purpose ECS Bare Metal Instance family with enhanced network performance

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:4
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96 vCPUs, up to 2.7 GHz Turbo Boost
- High network performance: 4.5 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation

- Suitable for the following scenarios:
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
 - Containers (includes but is not limited to Docker, Clear Containers, and Pouch)
 - Enterprise-level applications, such as large and medium-sized databases
 - Video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmg5s.24xlarge	96	384.0	N/A	30.0	4,500	No	8	32	10



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmr5s, memory optimized ECS Bare Metal Instance family with enhanced network performance

Features

- I/O optimized
- Supports enhanced SSDs, standard SSDs, and ultra disks
- CPU-to-memory ratio of 1:8
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96 vCPUs, up to 2.7 GHz Turbo Boost

- High network performance: 4.5 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation
- Suitable for the following scenarios:
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
 - Containers (includes but is not limited to Docker, Clear Containers, and Pouch)
 - High-performance databases and in-memory databases
 - Data analysis and mining and distributed memory cache
 - Hadoop, Spark, and other memory-intensive enterprise applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmr5s · 24xlarge	96	768.0	N/A	30.0	4,500	No	8	32	10





Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmhfg5, ECS Bare Metal Instance family with high clock speed

Features

- I/O optimized

- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:4
-  Equipped with 3.7 GHz Intel Xeon E3-1240v6 (Skylake) processors, 8 vCPUs, up to 4.1 GHz Turbo Boost
- High network performance: 2 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation
- Does not support automatic recovery
-  Supports Intel SGX
- Suitable for the following scenarios:
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Gaming or finance applications requiring high performance
 - High-performance Web servers
 - Enterprise-level applications such as high-performance databases

Instance types



Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmhfg5.xlarge	8	32.0	N/A	6.0	2,000	No	8	6	8



Note:

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmc4, compute optimized ECS Bare Metal Instance family**Features**

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:2
-   Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors, up to 2.9 GHz Turbo Boost
- High network performance: 4 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation
- Suitable for the following scenarios:
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
 - Containers (includes but is not limited to Docker, Clear Containers, and Pouch)
 - Enterprise-level applications, such as large and medium-sized databases
 - Video encoding

Instance types



Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmc4.xlarge	32	64.0	N/A	10.0	4,000	No	8	12	10

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

ebmg5, general purpose ECS Bare Metal Instance family

Features

- I/O optimized
- Supports standard SSDs and ultra disks
- CPU-to-memory ratio of 1:4
-  
Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96 vCPUs, up to 2.7 GHz Turbo Boost
- High network performance: 4 million pps packet forwarding rate
- Supports VPC networks only
- Provides dedicated hardware resources and physical isolation
- Suitable for the following scenarios:
 - Workloads that require direct access to physical resources, or scenarios that require a license to be bound to the hardware
 - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
 - Containers (includes but is not limited to Docker, Clear Containers, and Pouch)
 - Enterprise-level applications, such as large and medium-sized databases
 - Video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forward rate (Kpps)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.ebmg5.24xlarge	96	384.0	N/A	10.0	4,000	No	8	32	10

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

Billing methods

ECS Bare Metal Instances support pay-as-you-go and subscription billing methods. For more information, see [#unique_102](#).

4.10.2 Create an ECS Bare Metal instance

This topic describes how to create an ECS Bare Metal instance.

Follow the steps in [creating an instance by using the wizard](#) to create an ECS Bare Metal instance. However, the following configurations are recommended:

- **Region:** Currently, ECS Bare Metal instances are available in the following regions and zones: China (Shanghai), Zone D, China (Beijing), Zone C, China (Hangzhou), Zone G, and China (Shenzhen), Zone D.
- **Instance Type:** In ebmhfg5, ebmc4, and ebmg5 type families are available. For more information about instance types, see [instance type families](#).

- Image: The following public images are supported.

Operating system	Image
Linux	<ul style="list-style-type: none">- CentOS 7.2/7.3/7.4/6.9/6.8 64-bit- Ubuntu 14.04/16.04 64-bit- Debian 8.9/9.2 64-bit- openSUSE 42.3 64-bit- SUSE Linux Enterprise Server 12 SP2 64-bit
Windows	<ul style="list-style-type: none">- 2016 Data Center Edition 64-bit Chinese Edition- 2016 Data Center Edition 64-bit English Edition- 2012 R2 Data Center Edition 64-bit Chinese Edition- 2012 R2 Data Center Edition 64-bit English Edition

- Storage: ECS Bare Metal instances support up to 16 data disks. You can add a data disk during or after instance creation, and then [mount the data disk](#).
- Network Type: Only VPC is supported.

4.10.3 Install SGX

This topic describes what Intel Software Guard Extension (Intel SGX) is and how to install it.

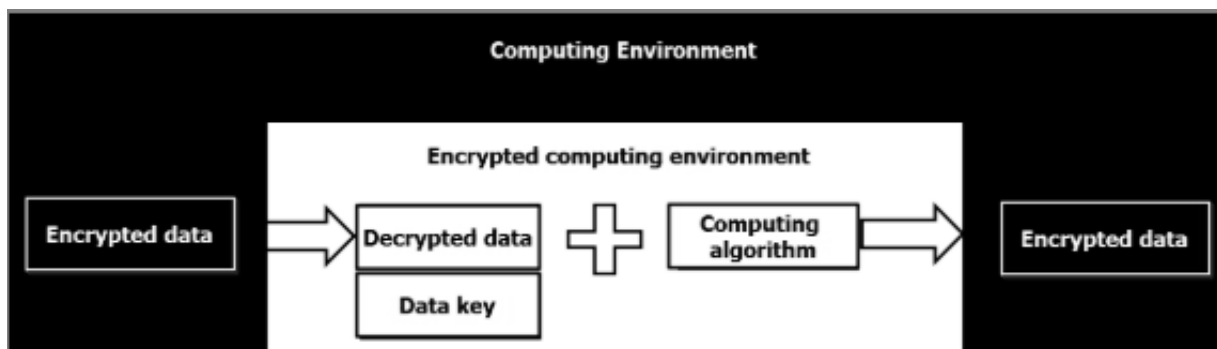
What is SGX?

SGX is an Intel architecture extension designed to increase the security of application code and data. You can partition your application into processor-hardened enclaves or protected areas of execution in memory that increase security even on compromised platforms.

SGX sets aside one or more ranges of physical memory as the Enclave Page Cache (EPC) and encrypts the data stored in the EPC using the Memory Encryption Engine (MEE). The data stored in the EPC is only decrypted inside the CPU. SGX offers CPU-based security controls. Data remains protected even when the OS, VMM, or BIOS are compromised.

Application

You can encrypt sensitive data, pass the encrypted data to the enclave in the cloud, and provide the corresponding key to the enclave through remote attestation. Then, you can compute over the fully encrypted data protected by the CPU, and the result is returned to you in an encrypted version. In this way, you can make use of the powerful cloud computing infrastructures with reduced risk of data disclosure.



Enclave Definition Language (EDL)

EDL is the fundamental part of SGX. It defines all enclave interface functions. During the build process, the Edger8r tool generates trusted and untrusted proxy/bridge functions and performs security checks.

Enclave interface functions can be divided into Enclave Calls (ECALLs) and Outside Calls (OCALLs).

- **ECALL:** A call from the application into an interface function within the enclave, which is defined as a trusted environment.
- **OCALL:** A call made from within the enclave to the application, which is defined as an untrusted environment.

```
// demo .edl
enclave {
    // Add your definition of "secret_t" here
    trusted {
        public void get_secret ([ out ] secret_t *
secret );
    };
    untrusted {
        // This OCALL is for illustrati on purposes only
        .
        // It should not be used in a real enclave ,
        // unless it is during the developmen t phase
        // for debugging purposes .
        void dump_secre t ([ in ] const secret_t * secret );
    };
}
```

```
};
```

Install SGX using the installer file

You can install SGX using the installer file, which includes the SGX driver, SGX Platform Software (PSW), and SDK. You must also install corresponding Linux kernel header files. The procedure is as follows:



Note:

The default directory for the Makefile in the example is `/opt/intel/`.

1. [Download](#) the SGX installer file.
2. Follow the steps in the [Installation guide](#).

Install SGX using the source code

You can install SGX using the source code, which includes the SGX driver, SGX Platform Software (PSW), and SDK. You must also install corresponding Linux kernel header files. The procedure is as follows:



Note:

The default directory for the Makefile in the example is `/opt/intel/`.

1. Download the source code from [Github](#).
2. Compile the source code according to the README.md file.

4.11 Super Computing Cluster instance type family

4.11.1 What is a Super Computing Cluster?

This topic describes Super Computing Clusters (SCCs) and the following SCC instance families and their instance types: scch5 instance family with high clock speed, sccg5 general purpose instance family, and sccgn6 compute optimized GPU instance family.

Overview

SCCs are based on ECS Bare Metal Instances. With the high-speed interconnects of Remote Direct Memory Access (RDMA) technology, SCCs greatly improve network performance and the acceleration ratio of large-scale clusters. Therefore, SCCs have all the benefits of ECS Bare Metal Instances and offer high-quality network performance featuring high bandwidth and low latency.

SCCs are released by Alibaba Cloud to meet the demands of applications such as high performance computing, artificial intelligence, machine learning, scientific and engineering computing, data analysis, and audio and video processing. In the clusters, nodes are connected over RDMA networks featuring high bandwidth and low latency, which guarantee the parallel efficiency demanded by applications that require high performance computing. The RDMA over Convergent Ethernet (RoCE) rivals an InfiniBand network in terms of connection speed and supports more extensive Ethernet-based applications.

The combination of SCCs and other Alibaba Cloud computing products such as ECS and GPU instances provides [Elastic HPC](#) with the ultimate high performance parallel computing resources, making supercomputing on the cloud a reality.

Comparison of SCCs, physical machines, and virtual machines

The following table shows a comparison of features. In this table, Y means supported, N means not supported, and N/A means not applicable.

Feature type	Feature	SCC	Physical machine	Virtual machine
Automated O&M	Delivery in minutes	Y	N	Y
Computing	Zero performance loss	Y	Y	N
	Zero feature loss	Y	Y	N
	Zero resource preemption	Y	Y	N
Storage	Compatible with ECS disks	Y	N	Y
	Startup from disks (system disks)	Y	N	Y
	Quick reset of system disks	Y	N	Y
	Compatible with ECS images	Y	N	Y

Feature type	Feature	SCC	Physical machine	Virtual machine
	Cold migration between physical and virtual machines	Y	N	Y
	Free of OS installation	Y	N	Y
	Free of local RAID, and stronger protection of data in disks	Y	N	Y
Network	Compatible with ECS VPC networks	Y	N	Y
	Compatible with ECS classic networks	Y	N	Y
	Free of communication bottlenecks between physical and virtual machine clusters in VPC	Y	N	Y
Management	Compatible with existing ECS management systems	Y	N	Y
	Consistent user experiences on VNC and other features with that of virtual machines	Y	N	Y

Feature type	Feature	SCC	Physical machine	Virtual machine
	Out-of-band (OOB) network security	Y	N	N/A

scch5, SCC instance type family with high clock speed

Features

- I/O optimized
- Supports only standard SSDs and ultra disks
- Supports both RoCE and VPC networks, of which RoCE is dedicated to RDMA communication
- With all features of ECS Bare Metal Instances
- Equipped with 3.1 GHz Intel Xeon Gold 6149 (Skylake)
- CPU-to-memory ratio of 1:3
- Suitable for the following scenarios:
 - Large-scale machine learning training
 - Large-scale high performance scientific computing and simulations
 - Large-scale data analysis, batch computing, and video encoding

Instance types

Instance type	vCPU	Physical core	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.scch5.16xlarge	64	32	192.0	N/A	10.0	4,500	46	No	8	32	10



Note:

- ecs.scch5.16xlarge provides 64 logical processors on 32 physical cores.

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

sccg5, general purpose SCC instance family

Features

- I/O optimized
- Supports only standard SSDs and ultra disks
- Supports both RoCE and VPC networks, of which RoCE is dedicated to RDMA communication
- With all features of ECS Bare Metal Instances
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- CPU-to-memory ratio of 1:4
- Suitable for the following scenarios:
 - Large-scale machine learning training
 - Large-scale high performance scientific computing and simulations
 - Large-scale data analysis, batch computing, and video encoding

Instance types

Instance type	vCPU	Physical core	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.sccg5.24xlarge	96	48	384.0	N/A	10.0	4,500	46	No	8	32	10




Note:

- ecs.sccg5.24xlarge provides 96 logical processors on 48 physical cores.

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

sccgn6, compute optimized SCC GPU instance family

Features

- I/O-optimized
- CPU-to-memory ratio of 1:4
-  Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- With all features of ECS Bare Metal Instances
- Storage:
 - Supports enhanced SSDs (million-level IOPS), standard SSDs, and ultra disks
 - Supports high performance Cloud Parallel File System (CPFS)
- Networking:
 - Supports VPC networks equipped with two 25 Gbps ports
 - Supports RoCE v2 networks, which is dedicated to RDMA communication
- Uses NVIDIA V100 GPU processors (with the SXM2 module):
 - Based on the new NVIDIA Volta architecture
 - 16 GB HBM2 memory capacity
 - Up to 5,120 CUDA Cores
 - Up to 640 Tensor Cores
 - Offers a 900 GB/s memory bandwidth
 - Supports up to six NVLink connections and total bandwidth of 300 GB/s (25 GB/s per connection)
- Suitable for the following scenarios:
 - Ultra-large-scale machine learning training on a distributed GPU cluster
 - Large-scale high performance scientific computing and simulations
 - Large-scale data analysis, batch computing, and video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queue	ENI	Private IP address of a single ENI
ecs.sccgn6.24xlarge	96	384	N/A	8 × V100	30	4,500	2 × 25	Yes	8	32	10

**Note:**

- You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.
- For more information about these specifications, see [Instance specification metrics](#).

Billing methods

SCC supports monthly and yearly subscription billing methods. For more information, see [#unique_102](#).

4.11.2 Create an SCC instance

This topic describes how to create a Super Computing Cluster (SCC) instance.

Super Computing Cluster (SCC) is based on the ECS Bare Metal instance type. By utilizing the high-speed interconnectivity of RDMA (Remote Direct Memory Access) technology, SCC greatly improves network performance and increases the acceleration ratio of large-scale clusters. SCC offers all the advantages of ECS Bare Metal instances, and provides high-quality network performance featuring high bandwidth and low latency. For more information, see [ECS Bare Metal instance and Super Computing Clusters](#).

Procedure

You can create an SCC instance by following the instructions described in [Create an ECS Instance](#).

However, the following configurations must be considered:

- **Region:** Select the region and zone according to the table [Regions and zones where SCC instances are available](#). Note that the purchase page displays the latest region and zone information, which may differ from information provided in this topic.
- **Network Type:** Only VPC is supported.
- **Instance Type:** Currently, the instance type families scch5, sccg5, and sccgn6 are available.
- **Image:** Select Public Image. Currently, only a custom Linux CentOS 7.5 image for SCC is supported.

**Note:**

The custom image supports the RDMA RoCE driver and OFED stack. You can use the RDMA functions through IB verbs programming or implement RDMA communication through MPI.

- **Storage:** SCC supports up to 16 data disks. You can [add a data disk](#) during or after instance creation, and then [mount the data disk](#).

Regions and zones where SCC instances are available

The following table shows the regions and zones where SCC instances are available.

Instance type family	Region and zone
scch5	China (Shanghai) Zones D and B
sccg5	China (Shanghai) Zones D and B
sccgn6	<ul style="list-style-type: none">• China (Shanghai)• China (Beijing)• China (Zhangjiakou)

What to do next

If you not only need to use the RDMA feature, but also need to use the HPC scheduler and the cluster scaling service, you can create an SCC instance by creating an SCC cluster through the [E-HPC console](#).



Note:

For more information about the availability of SCC instances, see [Regions and zones where SCC instances are available](#). Note that only the Subscription billing method is supported.

4.12 Burstable performance instances

4.12.1 Overview

This topic describes burstable performance instances, including their features, application scenarios, and concepts such as baseline performance, CPU credits, and performance modes. Burstable performance instances are able to burst above baseline CPU performance by using CPU credits.

Application scenarios

When you use other types of Alibaba Cloud ECS instances, the number and performance metrics of the corresponding vCPUs are fixed, which means that you may be billed for unused computing resources. To resolve this issue, Alibaba Cloud offers burstable performance instances to help you build high-performance servers, at a low cost, for a wide range of scenarios that experience traffic spikes in specific time periods. Examples include stress test applications, light-load applications, microservices, and web application servers.



Note:

For the types of burstable performance instances provided by Alibaba Cloud, see [t5 instance type family](#). If the instance you initially select does not meet performance requirements, you can [change the instance type](#).

Features

A burstable performance instance continuously accrues CPU credits. In this way, when higher CPU performance is required, the accrued CPU credits can be used without affecting the resource usage of applications deployed on the instance. For

more information, see [Baseline performance](#), [Earn CPU credits](#), and [CPU credit consumption](#).

CPU credits allow you to allocate computing resources according to the overall service landscape. The computing capability during off-peak hours can be reserved for peak hours, to make full use of CPU resources. If your network experiences traffic spikes, you can choose to enable the unlimited mode to ensure high performance. In the unlimited mode, if CPU credits are used in advance or overdrawn, additional fees may be charged. In the standard mode, if the instance has no remaining CPU credits, its CPU usage will not burst above the baseline performance. For more information, see [Standard mode](#) and [Unlimited mode](#).

t5 instance type family

Baseline performance

Baseline performance enables a burstable performance instance to maintain steady CPU performance according to the selected instance type. That is, when the CPU usage equals the baseline performance, the instance earns the same CPU credits as it consumes. For more information, see the *Avg baseline CPU performance* column in [t5 instance type family](#).

Earn CPU credits

CPU credits can be applied to adjust the computing capability of a burstable performance instance. The more credits that a burstable performance instance accrues, the more time it can burst beyond its baseline performance when higher performance is needed.

To ensure that you have sufficient CPU credits for running an instance, each vCPU is allocated 30 CPU credits after you create a burstable performance instance (called *launch CPU credit*). For example, an `ecs.t5-lc1m2.large` instance that is configured with two vCPUs receives 60 launch CPU credits after being created.

Each burstable performance instance continuously earns a fixed number of CPU credits per hour, depending on the instance type. The number of CPU credits that an instance earns per hour per vCPU corresponds to its baseline performance. For more information, see the *CPU credit/hour* column in [t5 instance type family](#).



Note:

For example, an `ecs.t5-c1m1.large` instance with a baseline performance of 15% means that the CPU credits the instance can earn per hour per vCPU allow the instance to run at 15% usage for one hour or run at 100% usage for nine minutes ($60 \times 15\%$). Therefore, based on the baseline performance, each vCPU can earn nine CPU credits per hour. In this case, an `ecs.t5-c1m1.large` instance that is configured with two vCPUs earns 18 CPU credits per hour.

If a burstable performance instance earns more CPU credits than it consumes, the unused CPU credits are accrued in the *CPU credit balance*. The CPU credit balance can be retained for up to 24 hours to achieve a dynamic balance, which has an upper limit. For more information, see the *Max. CPU credit balance* column in [t5 instance type family](#). For example, an `ecs.t5-c1m1.large` instance can earn 18 CPU credits per hour. Its maximum CPU credit balance is 432 (18×24 hours).

Stopping an instance affects CPU credits in the following ways:

- If an instance is stopped but the No fees for stopped instances (VPC-Connected) feature is not enabled, the current CPU credit balance is unaffected and CPU credits continue to accrue.
- If an instance is stopped and the No fees for stopped instances (VPC-Connected) feature is enabled, the accrued CPU credits are lost and the instance stops accruing CPU credits until you restart the instance.
- If a Subscription instance expires, the current CPU credit balance is unaffected but the instance stops accruing CPU credits until you reactivate the instance.
- If the payment of a Pay-As-You-Go instance is overdue, the current CPU credit balance is unaffected but the instance stops accruing CPU credits until you clear the overdue payment.

CPU credit consumption

The number of CPU credits that a burstable performance instance consumes per hour corresponds to the number of vCPUs, CPU usage, and running time of the instance.

One CPU credit is consumed in the following scenarios:

- One vCPU running at 100% usage for one minute
- One vCPU running at 50% usage for two minutes
- Two vCPUs running at 25% usage for two minutes

A burstable performance instance starts to consume CPU credits after being started. Launch CPU credits are consumed first and cannot be accrued in the process of being

used up. The instance then uses CPU credits that it has accrued in its CPU credit balance.

- If the CPU usage is lower than the baseline performance, the unused CPU credits are accrued in the CPU credit balance.
- If the CPU usage equals the baseline performance, the CPU credit balance remains unchanged.
- If the CPU usage is higher than the baseline performance, the instance consumes the accrued CPU credits.

Standard mode

In the standard mode, the performance of a burstable performance instance is governed by the CPU credits it has accrued. If the instance consumes all the credits it has accrued, it cannot burst above the baseline performance. If the CPU credits are insufficient, performance is gradually lowered to the baseline performance level within 15 minutes, so that the instance does not experience reduced performance when its CPU credit balance is used up.



Note:

For the relationship between instance performance and CPU credits in the standard mode, see [#unique_115](#).

A burstable performance instance running in the standard mode is suitable for scenarios where you do not usually, but occasionally require high CPU performance, such as lightweight web servers, development and testing environments, and low or mid-performance databases.

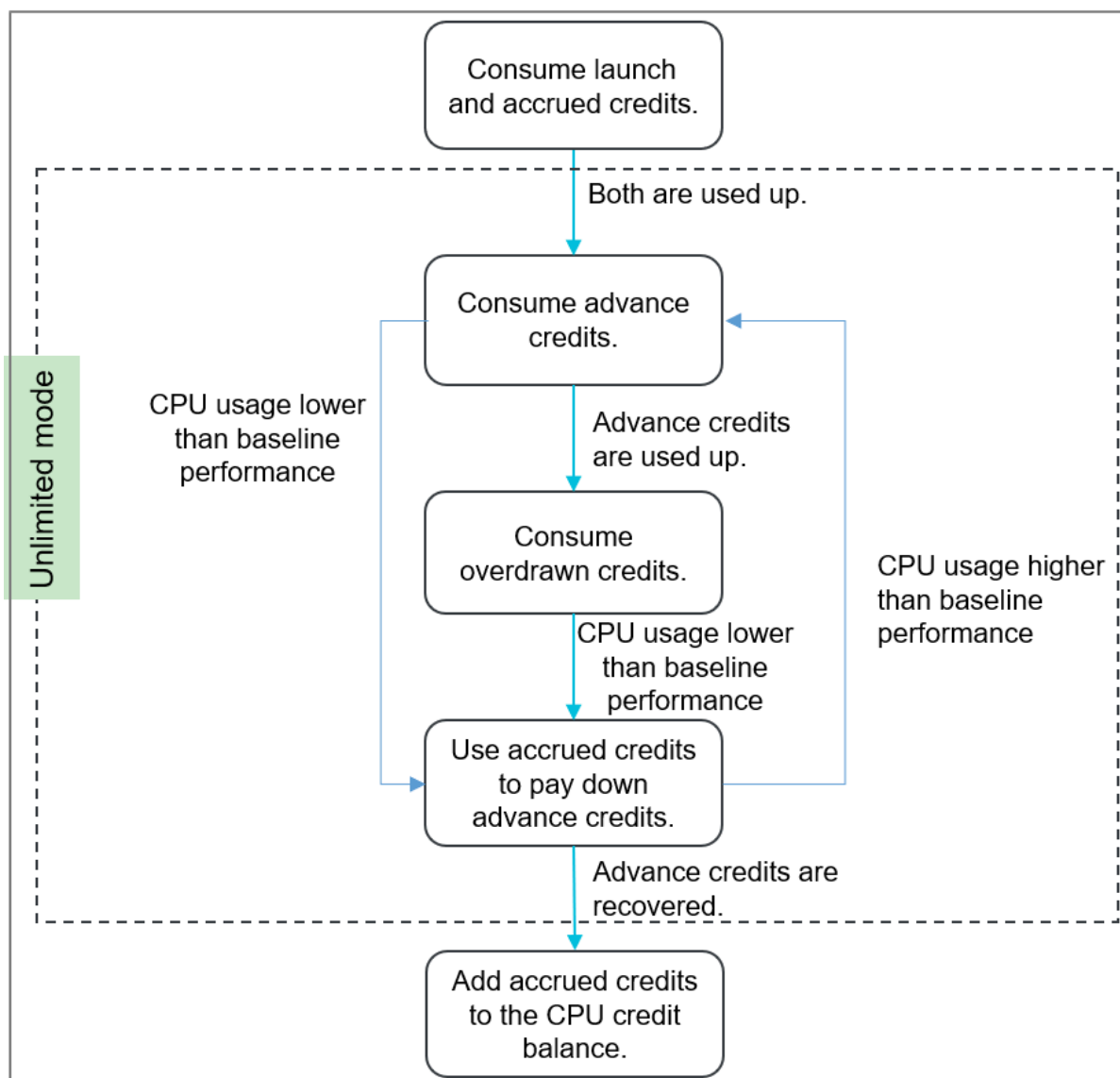
Unlimited mode

In the unlimited mode, a burstable performance instance can use advance CPU credits or overdrawn CPU credits to maintain high CPU performance whenever required, without being limited to the baseline CPU performance.

- **Advance CPU credit:** CPU credits that are used in advance but should be obtained within the next 24 hours. Advance CPU credits may incur fees. When an unlimited instance runs out of its CPU credit balance, advance CPU credits are used to deliver high CPU performance. When the CPU usage is lower than the baseline performance, the earned CPU credits are used to pay down (offset) the advance credits. For more information, see [Impact of performance modes on billing](#).

- **Overdrawn CPU credit:** Additional CPU credits that are used to sustain high CPU usage after advance CPU credits are used up. Overdrawn CPU credits incur additional fees. For more information, see [Impact of performance modes on billing](#)

The following figure illustrates how the unlimited mode works.



Note:

For the relationship between instance performance and CPU credits in the unlimited mode, see [#unique_115](#).

The unlimited mode is suitable for the following scenarios:

- Heavy-load occasions such as launch of new product functions or large-scale promotion on e-commerce platforms. You can enable the unlimited mode in peak hours and disable it in off-peak hours to reduce costs.
- Some websites experience traffic peaks in specific time periods, but the average CPU usage over a 24-hour period is lower than the baseline performance. You can keep the unlimited mode enabled to ensure use experience in peak hours. If CPU credits accrued in off-peak hours can pay down advance CPU credits, the overall website access experience can be guaranteed without incurring additional fees.

**Note:**

By default, the standard mode is used. For information about how to switch to the unlimited mode, see [#unique_116](#).

Create a burstable performance instance

For information about how to create a burstable performance instance, see [#unique_117](#). Note the following setting items when you create a burstable performance instance:

- Instance type
 1. Select x86-Architecture and Entry-Level (Shared).
 2. Select an appropriate [type of burstable performance instance](#).
 3. Select Enable Unlimited Mode for T5 Instances. You can also [enable the unlimited mode](#) after instance creation.

- Image type: The smallest t5 instance `ecs.t5-lc2m1.nano` has a memory of 0.5 GiB. It supports only Linux variants or the Windows Server Version 1809 operating system. For more information, see [#unique_118](#).
- Network type: Only VPC is supported.

Change the instance type

You can replace your current burstable performance instance with other types of burstable performance instances or enterprise-level instances at any time. For more information, see [#unique_119](#).

The change operations vary according to the billing method:

- For information about how to change the configuration of a Subscription instance, see [#unique_120](#).
- For information about how to change the configuration of a Pay-As-You-Go instance, see [#unique_121](#).

Basic concepts related to CPU credits

Basic concept	Description	Reference
Launch CPU credit	Each burstable performance instance is allocated 30 launch CPU credits for each vCPU after you create a burstable performance instance.	For more information, see Earn CPU credits .
CPU credit balance	If a burstable performance instance earns more CPU credits than it consumes, the unused CPU credits are accrued in the CPU credit balance. If a burstable performance instance needs to burst above the baseline performance, it consumes the accrued credits.	For more information, see Earn CPU credits .
Max. CPU credit balance	The maximum number of credits that an instance can accrue within 24 hours. The CPU credit balance is retained for up to 24 hours to achieve a dynamic balance.	For more information, see Earn CPU credits .
Advance CPU credit	CPU credits that are used in advance but should be obtained within the next 24 hours. Advance CPU credits can be used only in the unlimited mode.	For more information, see Unlimited mode .

Basic concept	Description	Reference
Overdrawn CPU credit	Additional credits required to maintain high performance after the advance CPU credits are used up. Overdrawn CPU credits incur fees and can be used only in the unlimited mode.	For more information, see Unlimited mode .

4.12.2 Benefits

Burstable performance instances allow you to use CPU credits to maximize the performance of your instances while remaining cost-effective.

Cost-effective

Burstable performance instances can cost nearly 50% less than other shared instances of the same configuration, depending on the instance type you selected.



Note:

The actual performance of a burstable performance instance corresponds to its accumulation and usage of CPU credits. For more information, see [Burstable performance instances](#).

Multiple specifications

Alibaba Cloud provides a variety of instance type specifications for burstable performance instances that you can select to meet a wide range of scenarios. The smallest burstable performance instance specification starts with only one vCPU core and 0.5 GiB, allowing you to flexibly combine instances. For more information, see [t5 instance type family](#).

Predictable CPU performance

Burstable performance instances are designed to provide a baseline level of CPU performance (*baseline performance*). You can predict the maximum possible performance required by your instance based on its baseline performance and accrued CPU credits, and select compute capabilities for a variety of scenarios as needed.

High network performance

Burstable performance instances use the most advanced network-based technologies to reduce network latency to a third of the level compared to the last-generation instances.

4.12.3 CPU credit change examples

This topic describes how the CPU credits change when the instance is running in different performance modes. After a burst performance instance is created, its CPU credits change according to the relationship between its CPU usage and baseline performance.

Background information

The examples given in this topic are only for your reference. Please understand that these may vary from your actual situation. We recommend that before you continue to read the rest of this topic, make sure that you are familiar with related concepts, including t5 instance type family, CPU credits, baseline performance, [performance modes](#), and [instance configuration change](#).

We also recommend that you are familiar with the following information:

- Each vCPU is allocated 30 CPU credits every time you create a burstable performance instance, which is called launch CPU credits.
- The number of CPU credits that a burstable performance instance uses on an hourly basis depends on the number of vCPUs, CPU usage, and running time of the instance. One CPU credit is equal to one vCPU running at full capacity for one minute. Additionally, other combinations of number of vCPUs, CPU usage, and running time can also equate to one CPU credit.
- An instance that runs at baseline performance earns the same CPU credits as it consumes. For more information, see [Earn CPU credits](#) and [CPU credit consumption](#).

Standard mode

In the standard mode, the performance of a burstable performance instance is governed by the CPU credits it has accrued. If the instance has consumed all the accrued credits, it cannot burst above the baseline performance.

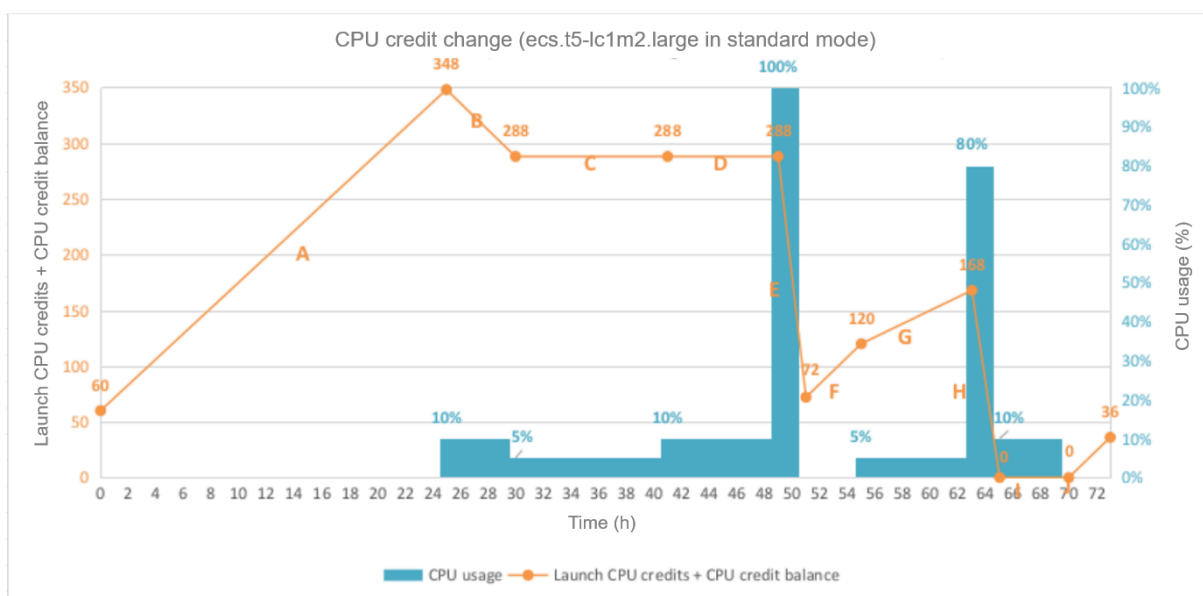
This section uses the `ecs.t5-lc1m2.large` instance for example.

**Note:**

The ecs.t5-lc1m2.large instance has the following features:

- It is configured with 2 vCPUs and 4 GiB memory.
- It receives 60 launch CPU credits upon creation (30 CPU credits for each vCPU).
- It has a baseline performance of 10%.
- It earns 12 CPU credits per hour and accrues a maximum of 288 CPU credits over a 24-hour period. For more information, see [t5 instance type family](#).
- It consumes 12 CPU credits per hour when running at baseline performance.

The following figure shows how the CPU credits of the ecs.t5-lc1m2.large instance change in the standard mode.



As shown in the preceding figure, CPU credit changes over time with the following distinct phrases:

- 0–24 hours

Phase A: The instance earns 60 launch CPU credits after being started. The initial CPU usage is 0% and the CPU credits continue to accrue until they reach the maximum value at hour 24.

At the end of this phase, the available CPU credits reach 348.

Available CPU credits (348) = Launch CPU credits (60) + Maximum CPU credit balance (288).

- 25–48 hours

1. Phase B: The CPU usage is 10%, which is equal to the baseline performance.

Launch CPU credits are preferentially used and will not be recovered after being used up. The instance consumes 12 CPU credits per hour when it starts to run.

At the end of this phase, the available CPU credits reach 288.

Available CPU credits (288) = CPU credits at the end of phase A (348) – Launch CPU credits (60).

2. Phase C: The CPU usage is 5%. Although the CPU usage is lower than the baseline performance, the CPU credit balance has reached the upper limit and remains unchanged.

At the end of this phase, the CPU credit balance reaches the maximum value 288

.

3. Phase D: The CPU usage is 10%, which is equal to the baseline performance. The instance earns the same CPU credits as it consumes and the CPU credit balance remains unchanged.

At the end of this phase, the CPU credit balance reaches the maximum value 288

.

- 48–72 hours

1. Phase E: The instance runs at full capacity for 2 hours and consumes 120 CPU credits per hour. The baseline performance cannot meet requirements, and the instance starts to consume CPU credits accrued in the CPU credit balance.

At the end of this phase, the CPU credit balance is 72.

CPU credit balance (72) = Maximum CPU credit balance (288) – 2 × CPU credits consumed per hour (120) + 2 × CPU credits earned per hour (12)

2. Phase F: The CPU usage is 0%. The instance is idle for 4 hours and earns 12 CPU credits per hour. All of the earned CPU credits are accrued in the CPU credit balance.

At the end of this phase, the CPU credit balance is 120.

CPU credit balance (120) = CPU credit balance at the end of phase E (72) + 4 × CPU credits earned per hour (12)

3. Phase G: The instance runs at 5% CPU usage for 8 hours and consumes 6 CPU credits per hour. The remaining CPU credits are accrued in the CPU credit balance.

At the end of this phase, the CPU credit balance is 168.

CPU credit balance (168) = CPU credit balance at the end of phase F (120) – 8 × CPU credits consumed per hour (6) + 8 × CPU credits earned per hour (12)

4. Phase H: The CPU usage is 80% and the baseline performance cannot meet the requirements. The instance runs for 2 hours and consumes 96 CPU credits per hour. The CPU credit balance is used up and the instance cannot burst above the baseline performance.



Note:

If CPU credits are insufficient, performance is gradually lowered to the baseline performance level within 15 minutes, so that the instance does not experience sharp performance drop-off when its CPU credit balance is used up.

At the end of this phase, the CPU credit balance is 0.

CPU credit balance (0) = CPU credit balance at the end of phase G (168) – 2 × CPU credits consumed per hour (96) + 2 × CPU credits earned per hour (12)

5. Phase I: The CPU usage is 10%, which is equal to the baseline performance. The instance earns the same CPU credits as it consumes, and the CPU credit balance remains unchanged.

At the end of this phase, the CPU credit balance is 0.

CPU credit balance (0) = CPU credit balance at the end of phase H (0) – 5 × CPU credits consumed per hour (12) + 5 × CPU credits earned per hour (12)

6. Phase J: The CPU usage is 0%. The instance is idle for 3 hours and earns 12 CPU credits per hour. All of the earned CPU credits are accrued in the CPU credit balance.

At the end of this phase, the CPU credit balance is 36.

CPU credit balance (36) = CPU credit balance at the end of phase I (0) + 3 × CPU credits earned per hour (12)

Unlimited mode

In the unlimited mode, a burstable performance instance can use advance CPU credits or overdrawn CPU credits to maintain high CPU performance whenever required, without being limited to the baseline CPU performance.

This section uses the `ecs.t5-c1m1.xlarge` instance for example.

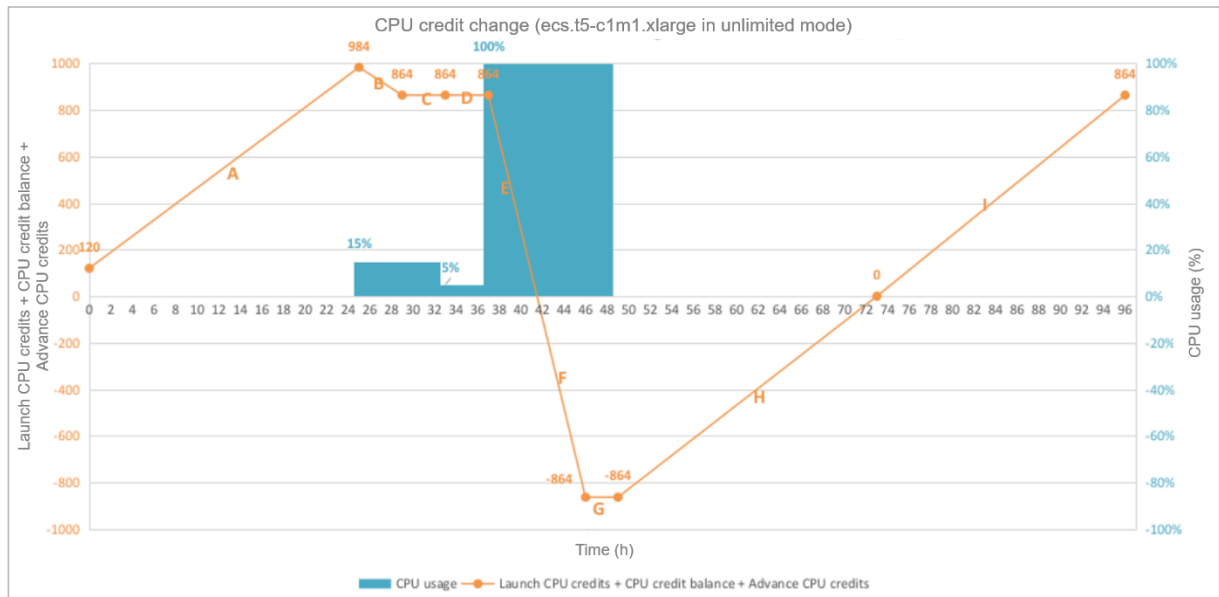


Note:

The `ecs.t5-c1m1.xlarge` instance has the following features:

- It is configured with 4 vCPUs and 4 GiB memory.
- It receives 120 launch CPU credits upon creation (30 CPU credits for each vCPU).
- It has a baseline performance of 15%.
- It earns 36 CPU credits per hour and accrues a maximum of 864 CPU credits over a 24-hour period. For more information, see [t5 instance type family](#).
- It consumes 36 CPU credits per hour when running at baseline performance.

The following figure shows how the CPU credits of the `ecs.t5-c1m1.xlarge` instance change in the unlimited mode.



As shown in the preceding figure, CPU credit changes over time with the following distinct phrases:

- 0–24 hours

Phase A: The instance earns 120 launch CPU credits after being started. The initial CPU usage is 0% and the CPU credits continue to accrue until they reach the maximum value at hour 24.

At the end of this phase, the available CPU credits reach 984.

Available CPU credits (984) = Launch CPU credits (120) + Maximum CPU credit balance (864)

- 25–48 hours

1. Phase B: The CPU usage is 15%, which is equal to the baseline performance.

Launch CPU credits are preferentially used and will not be recovered after being used up. The instance consumes 36 CPU credits per hour when it starts to run.

At the end of this phase, the available CPU credits reach 864.

Available CPU credits (864) = CPU credits at the end of phase A (984) – Launch CPU credits (120)

2. Phase C: The CPU usage is 15%, which is equal to the baseline performance. The instance earns the same CPU credits as it consumes and the CPU credit balance remains unchanged.

At the end of this phase, the CPU credit balance reaches the maximum value 864

.

3. Phase D: The CPU usage is 5%. Although the CPU usage is lower than the baseline performance, the CPU credit balance has reached the upper limit and remains unchanged.

At the end of this phase, the CPU credit balance reaches the maximum value 864

.

4. Phase E: The instance runs at full capacity and consumes 240 CPU credits per hour. The baseline performance cannot meet requirements, and the instance starts to consume CPU credits accrued in the CPU credit balance.

At the end of this phase, the CPU credit balance is used up.

5. Phase F: The instance runs at full capacity and consumes 240 CPU credits per hour. The baseline performance cannot meet requirements, and the instance starts to consume advance CPU credits. For more information, see [Unlimited mode](#).

At the end of this phase, all of the 864 advance CPU credits are used up.

6. Phase G: The instance runs at full capacity and consumes 240 CPU credits per hour. The advance CPU credits are used up and the instance starts to consume overdrawn CPU credits, which incur additional fees. For more information, see [Unlimited mode](#).

At the end of this phase, the available CPU credits remain unchanged and the instance overdraws 864 CPU credits.

- 48–72 hours

Phase H: The CPU usage is 0%. The earned CPU credits are first used to pay down the advance CPU credits until the advance CPU credits are cleared at hour 72.

At the end of this phase, the advance CPU credits are recovered and the CPU credit balance remains at 0.

- 73–96 hours

Phase I: The CPU usage is 0%. The instance is idle for 24 hours and earns 36 CPU credits per hour. All of the earned CPU credits are accrued in the CPU credit balance until the CPU credit balance reaches the maximum value at hour 96.

At the end of this phase, the CPU credit balance reaches the maximum value 864.

4.12.4 Billing

This topic describes the billing methods of a burstable performance instance.

The performance of a burstable performance instance is governed by its CPU credits. Specifically, the CPU credits allow you to control the cost of an instance. If your instance consumes additional CPU credits, additional fees may be incurred. Therefore, the price of a burstable performance instance comprises of the instance purchase fee and additional fees.

Instance purchase fee

Burstable performance instances support the Pay-As-You-Go and Subscription billing methods. For information about how the billing methods compare with each other, see [#unique_102](#).

For the price of a burstable performance instance, see [Pricing](#).

A burstable performance instance can be a preemptible instance. For more information, see [#unique_126](#).

After creating a Pay-As-You-Go instance, you can purchase a Reserved Instance (RI) and use it to generate a billing discount. An RI is a discount coupon with specific attributes. For more information, see [#unique_127](#). The following limitations apply if you use RIs for a burstable performance instance:

- You can only purchase zonal RIs.
- You cannot merge, split, or change the scope of RIs.
- RIs do not match preemptible instances.

Impact of performance modes on billing

The following table describes how the performance mode affects the billing of a burstable performance instance.

Performance mode	Instance purchase fee	Additional fee
Standard mode	The fee is determined by the billing method, not the performance mode. For more information, see Instance purchase fee .	None

Performance mode	Instance purchase fee	Additional fee
Unlimited mode	The fee is determined by the billing method, not the performance mode. For more information, see Instance purchase fee .	<p>If your instance consumes additional CPU credits, you may need to pay additional fees:</p> <ul style="list-style-type: none"> · Overdrawn CPU credits are billed by hour. · If you use advance CPU credits and switch to the standard mode before the advance CPU credits are cleared, a one-off fee is charged for the advance CPU credits and the current CPU credit balance remains unchanged. · If you use advance CPU credits and stop or release the instance before the advance credits are cleared, a one-off fee is charged for the advance CPU credits. Stopping an instance affects the CPU credits. For more information, see Earn CPU credits.

The following table describes how fees are collected in unlimited mode.

Region	Windows instance (USD/CPU credit)	Linux instance (USD/CPU credit)
Mainland China	0.0008	0.0008
Outside Mainland China	0.0016	0.0008

Price comparison between burstable performance instances and enterprise-level instances

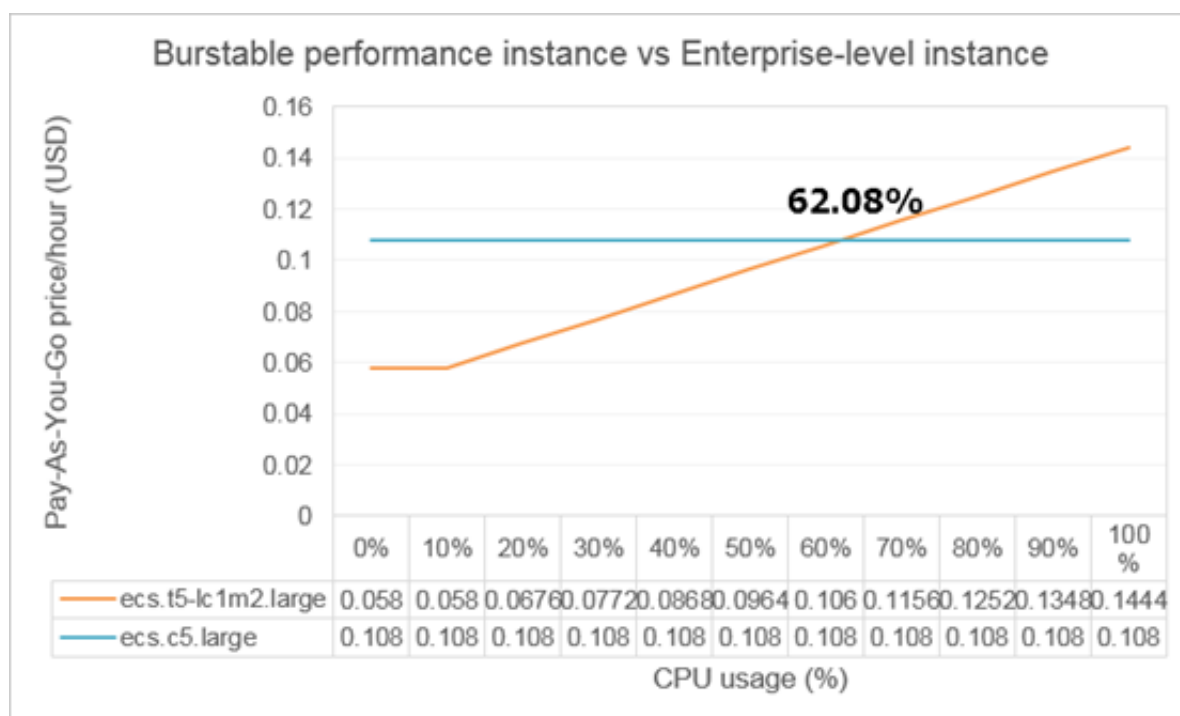
Overdrawn CPU credits only take effect in unlimited mode. If a burstable performance instance uses excess overdrawn CPU credits, the total price of the instance may equal or exceed enterprise-level instances of equivalent configurations. For more information, see [Unlimited mode](#).



Note:

The following description uses instance prices on April 30th, 2019 for example. For the latest prices, see [Pricing](#).

The following figure compares the price of a burstable performance instance `ecs.t5-lc1m2.large` with an enterprise-level instance `ecs.c5.large`. Both the instances have two vCPUs and 4 GiB memory. The `ecs.t5-lc1m2.large` instance uses overdrawn CPU credits.



In the preceding figure, the hourly price is calculated based on the purchase fee of `ecs.t5-lc1m2.large` (10% baseline performance) and the fee of overdrawn CPU credits (Linux instance in China Beijing). The formula is the same for different instance types but the CPU usage threshold may vary.

The price of the two instances varies according to the following scenarios:

- If the average CPU usage is lower than 62.08%, `ecs.t5-lc1m2.large` costs less.
- If the average CPU usage equals 62.08%, the two instances cost the same.

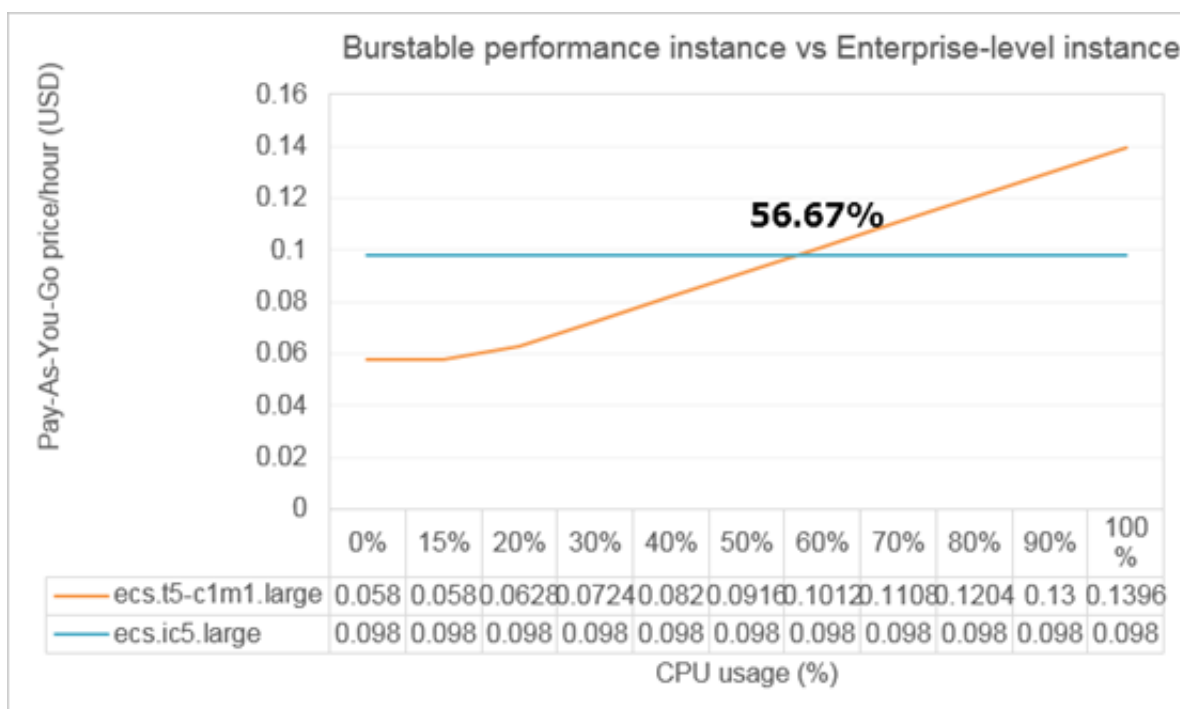
- If the average CPU usage is higher than 62.08%, the ecs.c5.large instance costs less.

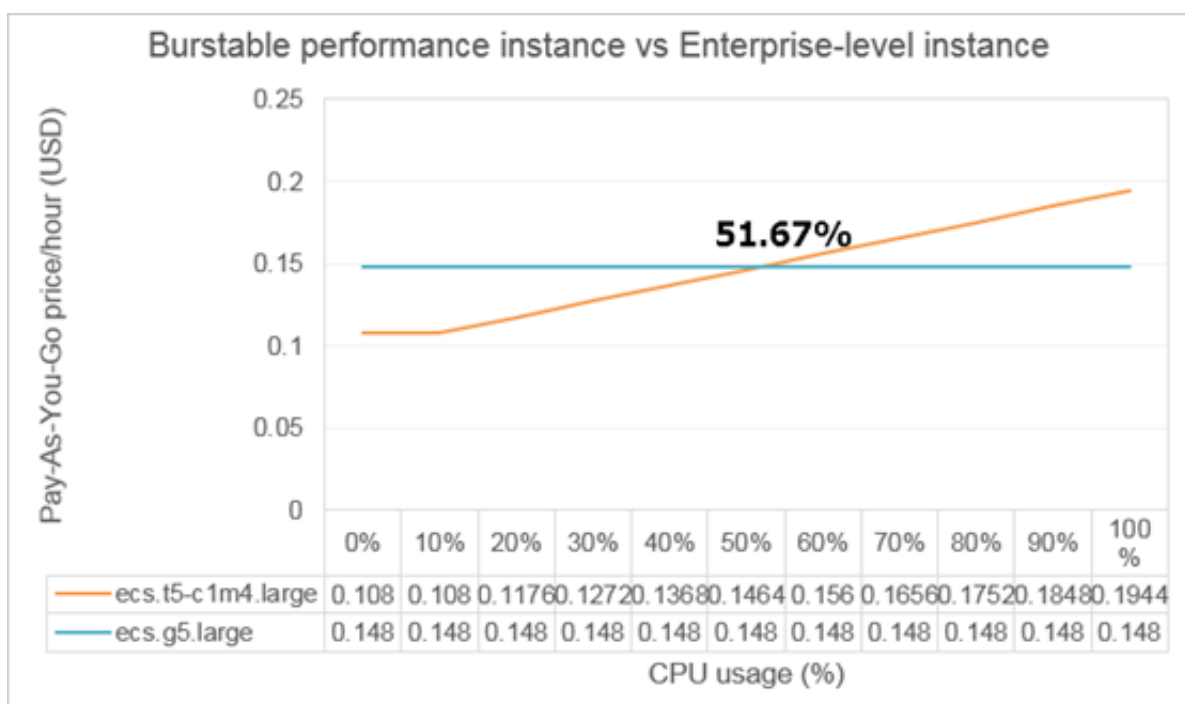
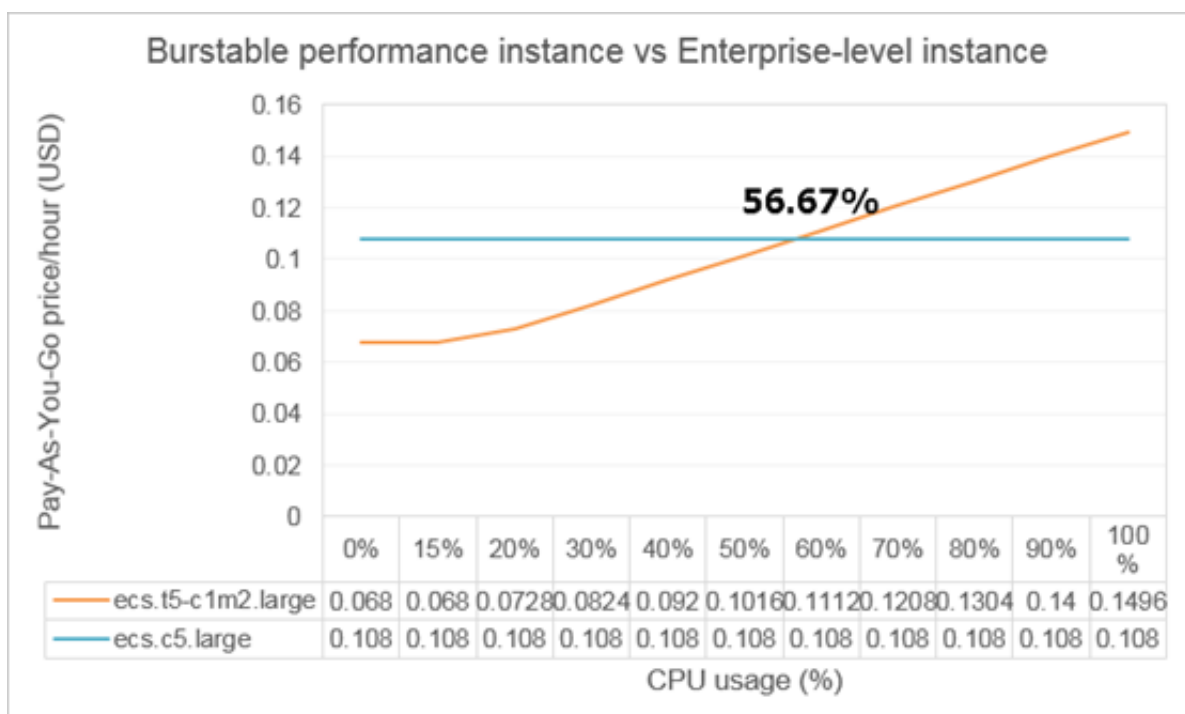


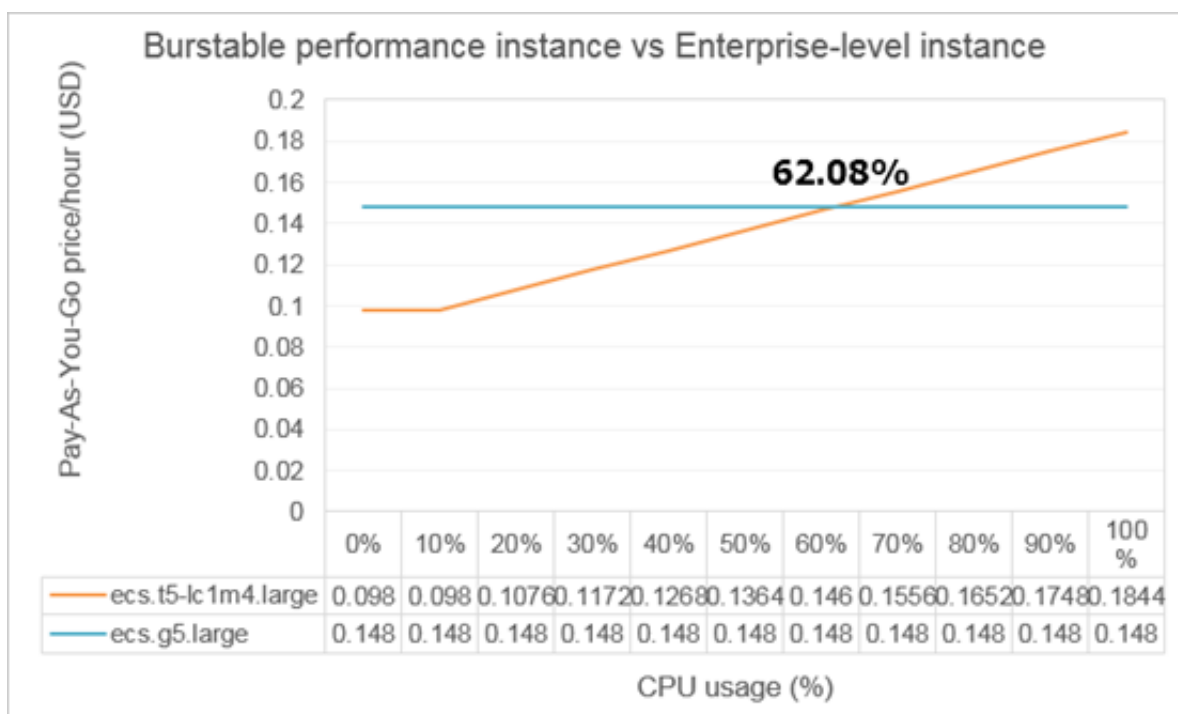
Note:

If you can accurately estimate the performance requirements of your instances, you may choose an appropriate instance type based on the CPU usage threshold. This helps you achieve high performance at a low cost.

The following figures show how the price change trend of several best-selling burstable performance instances. For the price change trends of other instance types, [open a ticket](#).







4.12.5 Manage performance modes of a burstable performance instance

This topic describes how to query the performance mode of a burstable performance instance and switch between performance modes. A burstable performance instance can run in either the standard or unlimited mode. You can switch between performance modes by enabling or disabling the selected mode at any time.

Background information

In standard mode, if an instance has no remaining CPU credits, its CPU usage will not burst above the baseline performance. However, in unlimited mode, an instance can use advance or overdrawn CPU credits to maintain high CPU usage whenever required, but this may incur additional fees. For more information, see [Unlimited mode](#).

- By default, the standard mode is enabled at instance creation.
- If a burstable performance instance is in the Stopped status and the No fees for stopped instances (VPC-Connected) feature is enabled, the instance enters the standard mode by default after it is started.
- If a burstable performance instance is in the Stopped status and the No fees for stopped instances (VPC-Connected) feature is not enabled, the performance mode used before the instance is stopped continues to take effect after the instance is started.

- If your account has an overdue payment, the unlimited mode will be automatically disabled and will not be enabled until you clear the overdue payment.

Query the performance mode of a burstable performance instance

1. Find the target instance and click the instance ID.
2. In the Basic Information area, view Instance Type.
 - Standard indicates that the instance is in standard mode.
 - Unlimited indicates that the instance is in unlimited mode.

Enable the unlimited mode

Before enabling the unlimited mode, ensure that:

- The instance is in the Running status.
 - The standard mode is enabled.
1. Find the target instance and click the instance ID.
 2. In the Basic Information area, choose More > Enable Unlimited Mode.
 3. In the displayed Enable Unlimited Mode dialog box, click OK.

Disable the unlimited mode

Before disabling the unlimited mode, ensure that:

- The instance is in the Running status.
 - The unlimited mode is enabled.
1. Find the target instance and click the instance ID.
 2. In the Basic Information area, choose More > Disable Unlimited Mode.
 3. In the displayed Disable Unlimited Mode dialog box, click OK.

4.12.6 Monitor a burstable performance instance

This topic describes how to monitor the CPU usage and CPU credits of a burstable performance instance. When an instance is running, the CPU usage changes dynamically. As a result, the CPU credits vary according to both the CPU usage and the baseline performance of the instance.

Procedure

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Instances.
3. Find the target instance and click the instance ID.

4. On the Instance Details page, view the monitoring information of the instance in the Monitoring Information area.

The following table describes trend charts in the Monitoring Information area.

Trend chart	Description
CPU	Shows the trend of CPU usage.
Burstable Instance-CPU Credit Consumption	Shows the trend of CPU credit consumption. The trend corresponds to the CPU usage. For more information, see CPU credit consumption .
Burstable Instance-CPU Credit Balance	Shows the trend of CPU credit balance, that is, CPU credits that can be used to burst above the baseline performance. For more information, see Earn CPU credits .
Burstable Instance-Overdrawn CPU Credits	Shows the trend of overdrawn CPU credits. Overdrawn CPU credits can be used only when you enable the unlimited mode. For more information, see Unlimited mode .
Burstable Instance-Advance CPU Credits	Shows the trend of advance CPU credits. Advance CPU credits can be used only when you enable the unlimited mode. For more information, see Unlimited mode .

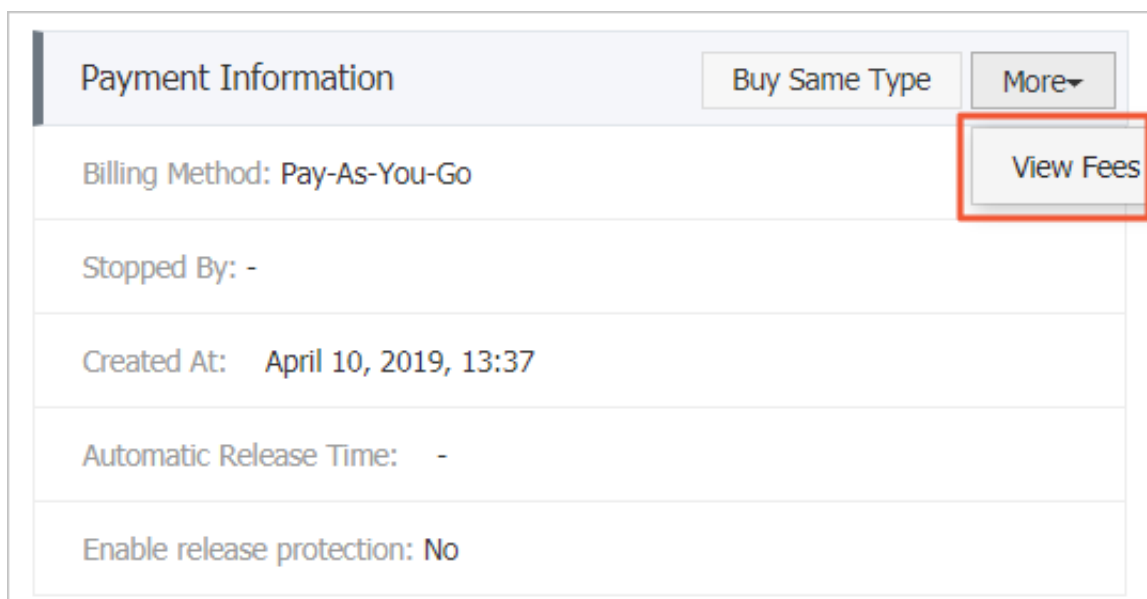
4.12.7 View bills of a burstable performance instance

This topic describes how to view the bills of a burstable performance instance that uses overdrawn CPU credits.

Procedure

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Instances.
3. Find the target instance and click the instance ID.

4. In the Payment Information area, choose More > View Fees.



Payment Information

Buy Same Type More ▾

Billing Method: Pay-As-You-Go

Stopped By: -

Created At: April 10, 2019, 13:37

Automatic Release Time: -

Enable release protection: No

View Fees

5. On the Bills page, click Detail in the Actions column of the target instance.

6. In the Summary of billing items area, view SurplusCredit.

Summary of billing items		
Cloud disk : ¥0	Cloud server configuration : ¥62.087301	Outgoing traffic : ¥0
Local SSD disk : ¥0	SurplusCredit : ¥0	Os Disk : ¥7.493129
Local disk : ¥0		

You can also view SurplusCredit in the Cost details area to check the cost details of a single purchase.

Cost details			
Purchase time	Billing item	Usage volume	Payable amount
2019-04-26 14:00:00 To 2019-04-26 15:00:00	Cloud disk	0.000GB - hour	¥0
2019-04-26 14:00:00 To 2019-04-26 15:00:00	Local SSD disk	0.000GB - hour	¥0
2019-04-26 14:00:00 To 2019-04-26 15:00:00	Cloud server configuration (ecs.t5-c1m2.large)	1Hour	¥0.161111
2019-04-26 14:00:00 To 2019-04-26 15:00:00	Local disk	0.000GB - hour	¥0
2019-04-26 14:00:00 To 2019-04-26 15:00:00	Os Disk	40.000GB - hour	¥0.019444
2019-04-26 14:00:00 To 2019-04-26 15:00:00	Outgoing traffic	0.000GB	¥0
2019-04-26 14:00:00 To 2019-04-26 15:00:00	SurplusCredit	0	¥0

4.13 Entry-level instance type families

This topic describes the entry-level instance type families xn4, n4, mn4, and e4, and lists the specific instance types within each of these instance type families.

4.14 Phased-out instance types

This topic describes all phased-out instance types. These instance types are no longer available for purchase at Chinese sites. However, the instance types sn2, sn1, n1, n2, and e3 are still available for purchase at international sites.

Indicator description

- The packet forwarding rate refers to the maximum number of inbound and outbound packets that can be processed by an ECS instance. For more information, see [Test network performance](#).
- NIC queues refer to the maximum number of NIC queues that an instance type supports. If your instance runs CentOS 7.3, the maximum number of NIC queues is used by default.
- Enterprise-level instances with two or more vCPU cores and entry-level instances with four or more vCPU cores support elastic network interfaces (ENIs). For more information, see [#unique_57](#).

Instance configuration upgrade

Some phased-out instance type families support upgrading to a newer instance type. For more information, see [#unique_119](#).

sn2, general-purpose type family

Features

- vCPU : Memory = 1:4
- 2.5 GHz Intel Xeon, E5-2682 v4 (Broadwell), or E5-2680 v3 (Haswell) processors
- The network performance of an instance matching the computing type (the more advanced the computing type, the more powerful the network performance)
- Ideal for:
 - Enterprise-class applications of various types and sizes
 - Medium and small database systems, cache, and search clusters
 - Data analysis and computing
 - Computing clusters, and data processing depending on memory

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (1000 pps)	NIC queues	ENIs (including 1 primary ENI)
ecs.sn2.medium	2	8.0	N/A	0.5	100	1	2
ecs.sn2.large	4	16.0	N/A	0.8	200	1	3
ecs.sn2.xlarge	8	32.0	N/A	1.5	400	1	4
ecs.sn2.3xlarge	16	64.0	N/A	3.0	500	2	8
ecs.sn2.7xlarge	32	128.0	N/A	6.0	800	3	8
ecs.sn2.13xlarge	56	224.0	N/A	10.0	1,200	4	8

sn1, compute optimized type family

Features

- vCPU : Memory = 1:2
- 2.5 GHz Intel Xeon, E5-2682 v4 (Broadwell), or E5-2680 v3 (Haswell) processors
- The network performance of an instance matching the computing type (the more advanced the computing type, the more powerful the network performance)
- Ideal for:
 - Web front-end servers
 - Front ends of Massively Multiplayer Online (MMO) games
 - Data analysis, batch compute, and video coding
 - High performance science and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (1000 pps)	NIC queues	ENIs (including 1 primary ENI)
ecs.sn1.medium	2	4.0	N/A	0.5	100	1	2
ecs.sn1.large	4	8.0	N/A	0.8	200	1	3
ecs.sn1.xlarge	8	16.0	N/A	1.5	400	1	4
ecs.sn1.3xlarge	16	32.0	N/A	3.0	500	2	8
ecs.sn1.7xlarge	32	64.0	N/A	6.0	800	3	8

c4/ce4/cm4, compute optimized type families with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- Stable performance
- 3.2 GHz Intel Xeon E5-2667 v4 (Broadwell) processors
- Higher computing specifications matching higher network performance
- Ideal for:
 - High performance Web front-end servers
 - High performance science and engineering applications
 - Massively Multiplayer Online (MMO) games and video coding

c4

Instance type	vCPU	Memory (GiB)	Local disks (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (1000 pps)	NIC queues	ENIs
ecs.c4.xlarge	4	8.0	N/A	1.5	200	1	3
ecs.c4.2xlarge	8	16.0	N/A	3.0	400	1	4
ecs.c4.3xlarge	12	24.0	N/A	4.5	600	2	6
ecs.c4.4xlarge	16	32.0	N/A	6.0	800	2	8

ce4

Instance type	vCPU	Memory (GiB)	Local disks (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (1000 pps)	NIC queues	ENIs
ecs.ce4.xlarge	4	32.0	N/A	1.5	200	1	3
ecs.ce4.2xlarge	8	64.0	N/A	3.0	400	1	3

cm4

Instance type	vCPU	Memory (GiB)	Local disks (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (1000 pps)	NIC queues	ENIs
ecs.cm4.xlarge	4	16.0	N/A	1.5	200	1	3
ecs.cm4.2xlarge	8	32.0	N/A	3.0	400	1	4

Instance type	vCPU	Memory (GiB)	Local disks (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (1000 pps)	NIC queues	ENIs
ecs.cm4.3xlarge	12	48.0	N/A	4.5	600	2	6
ecs.cm4.4xlarge	16	64.0	N/A	6.0	800	2	8
ecs.cm4.6xlarge	24	96.0	N/A	10.0	1,200	4	8

n1/n2/e3, Entry-level instance type families

Features

- 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors
- Higher computing specifications matching higher network performance
- I/O-optimized
- Supporting the following disk types:
 - SSD Cloud Disks
 - Ultra Cloud Disks

Type family	Features	vCPU : Memory	Idea for
n1	General entry-level instances	1:2	<ul style="list-style-type: none"> • Small and medium-sized web servers • Batch processing • Distributed analysis • Advertisement services

Type family	Features	vCPU : Memory	Idea for
n2	Balanced entry-level instances	1:4	<ul style="list-style-type: none"> • Medium-sized Web servers • Batch processing • Distributed analysis • Advertisement services • Hadoop clusters
e3	Memory entry-level instances	1:8	<ul style="list-style-type: none"> • Cache, Redis • Search • Memory databases • Databases with high I/O. For example, Oracle and MongoDB • Hadoop clusters • Computing scenarios that involve massive data processing

n1

Instance type	vCPU	Memory (GiB)	Local disks (GiB)	ENIs (including 1 primary ENI)
ecs.n1.tiny	1	1.0	N/A	1
ecs.n1.small	1	2.0	N/A	1
ecs.n1.medium	2	4.0	N/A	1
ecs.n1.large	4	8.0	N/A	2
ecs.n1.xlarge	8	16.0	N/A	2
ecs.n1.3xlarge	16	32.0	N/A	2
ecs.n1.7xlarge	32	64.0	N/A	2

n2

Instance type	vCPU	Memory (GiB)	Local disks (GiB)	ENIs (including 1 primary ENI)
ecs.n2.small	1	4.0	N/A	1
ecs.n2.medium	2	8.0	N/A	1
ecs.n2.large	4	16.0	N/A	2
ecs.n2.xlarge	8	32.0	N/A	2
ecs.n2.3xlarge	16	64.0	N/A	2
ecs.n2.7xlarge	32	128.0	N/A	2

e3

Instance type	vCPU	Memory (GiB)	Local disks (GiB)	ENIs (including 1 primary ENI)
ecs.e3.small	1	8.0	N/A	1
ecs.e3.medium	2	16.0	N/A	1
ecs.e3.large	4	32.0	N/A	2
ecs.e3.xlarge	8	64.0	N/A	2
ecs.e3.3xlarge	16	128.0	N/A	2

Instance types of Generation I

Instance types of Generation I include t1, s1, s2, s3, m1, m2, c1, and c2. All these instance types are legacy shared instance types. They are still categorized by the number of cores (1, 2, 4, 8, and 16 cores) and are not sensitive to type families.

**Note:**

For information about instance type families of Generation II and Generation III, see [Instance type families](#).

Features

- 1.9 GHz Intel Xeon E5-2420 processors or higher
- The latest DDR3 memory
- [I/O-optimized](#) and [non I/O-optimized](#) at your choice

I/O-optimized instance types

I/O-optimized instances support two types of disks:

- SSD cloud disk
- Ultra cloud disk

Specification types	Type code	vCPU	Memory (GiB)
Standard	ecs.s2.large	2	4
	ecs.s2.xlarge	2	8
	ecs.s2.2xlarge	2	16
	ecs.s3.medium	4	4
	ecs.s3.large	4	8
High Memory	ecs.m1.medium	4	16
	ecs.m2.medium	4	32
	ecs.m1.xlarge	8	32
High CPU	ecs.c1.small	8	8
	ecs.c1.large	8	16
	ecs.c2.medium	16	16
	ecs.c2.large	16	32
	ecs.c2.xlarge	16	64

Non I/O-optimized instance types

Non I/O-optimized instances only support basic cloud disks.

Specification types	Instance type	vCPU	Memory (GiB)
Tiny	ecs.t1.small	1	1
Standard	ecs.s1.small	1	2
	ecs.s1.medium	1	4
	ecs.s1.large	1	8
	ecs.s2.small	2	2
	ecs.s2.large	2	4
	ecs.s2.xlarge	2	8
	ecs.s2.2xlarge	2	16
	ecs.s3.medium	4	4
	ecs.s3.large	4	8
High Memory	ecs.m1.medium	4	16

Specification types	Instance type	vCPU	Memory (GiB)
	ecs.m2.medium	4	32
	ecs.m1.xlarge	8	32
High CPU	ecs.c1.small	8	8
	ecs.c1.large	8	16
	ecs.c2.medium	16	16
	ecs.c2.large	16	32
	ecs.c2.xlarge	16	64

5 Instance purchasing options

5.1 Subscription

Subscription is a billing method that allows you to use resources only after payment for them is received.

Applicable resources

Currently, Subscription-based billing is applicable to the following ECS resources:

- ECS instances, including CPU configuration and memory capacity
- Images
- System disks and data disks

If you want to create a Subscription ECS instance, the Total price shown in the lower left corner of the page includes the costs of the preceding resources.

You can use the ECS [TCO Calculator](#) to quickly analyze your cloud migration costs.

Payment methods

You can pay for Subscription resources by using either of the following methods:

- Your credit card or PayPal account associated to your Alibaba Cloud account. For more information, see [Add a payment method](#) in *Account Management*.
- Coupons that can be used to pay for Subscription resources.



Note:

You can log on to the ECS console, and then choose Expenses center > Coupon Management to view the Applicable Scenarios to determine whether your coupons apply to Subscription resources.

Billing cycle

Subscription resources are billed on a monthly basis. The billing cycle is based on UTC+8:00, starting at the time when the resources are activated, and ending at 00:00:00 on the second day after the month or the year (depending on the billing cycle that is selected).

For example, if you activated a monthly Subscription ECS instance at 13:23:56 on March 12, 2017, the first billing cycle ended at 00:00:00 on April 13, 2017.

The price unit varies depending on different resources. The following table lists the price units for available resources.

Resource	Price unit
ECS instance	USD/month
Image	USD/month
System disk	USD/(GiB x month)
Data disk	USD/(GiB x month)

After a billing cycle expires, you can renew your ECS instance to continue using the resources. For more information, see [Manual renewal](#) or [Auto-renewal](#).

Resource status changes after expiration

After a Subscription instance expires, the status of its corresponding resources changes depending on whether you have enabled [auto-renewal](#).

- If you have not enabled automatic renewal:

The Subscription instance will become unavailable at any time within 24 hours from 00:00:00 of the expiration day to 00:00:00 of the next day. If you do not renew

the instance within 15 days after it expires, the status of relevant resources will change as follows.

Time period	ECS instance and image	System disk	Data disk	Internet IP address	Snapshot
On the expiration day	The instance will become unavailable* and the image will be disabled.	The system disk will become unavailable, but the data will be retained.	The data disk will become unavailable, but the data will be retained.	For a Classic network instance, the assigned Internet IP address will be retained. For a VPC instance, the assigned Internet IP address and the attached EIP address will be retained.	Snapshots will not be affected.

Time period	ECS instance and image	System disk	Data disk	Internet IP address	Snapshot
More than 15 days after expiration	The ECS instance and the image will be automatically released.	The system disk will be released along with the instance . The data cannot be restored.	By default , the cloud disk will be released along with the instance . If you have set the cloud disk not to be released along with the instance , the disk will stop operating. The local disk will be released along with the instance . The shared block storage will be automatically detached .	For a Classic network instance, the assigned Internet IP address will be released. For a VPC instance, the assigned Internet IP address will be released and the attached EIP address will be detached.	Automatic snapshots will be automatically deleted , whereas snapshots created manually will not be affected.

* After the instance is unavailable, you cannot connect to the instance remotely, and websites deployed on the instance cannot be accessed. Service errors will occur.



Note:

Auto-renewal cannot be enabled for expired instances.

- If you have enabled automatic renewal:

If you do not renew the instance within the specified time period, the status of resources corresponding to the instance will change as follows.

Time period	ECS instance and image	System disk	Data disk	Internet IP address	Snapshot
Within 15 days after expiration	The ECS instance and the image will operate normally*.	The system disk will operate normally.	The data disk will operate normally.	The Internet IP address will not be affected.	Snapshots will not be affected.
More than 15 days after expiration	The instance will become unavailable** at any time within 24 hours from 00:00:00 of the 15th day after expiration to 00:00:00 of the 16th day. After that, the image will be disabled.	The system disk will become unavailable, but the data will be retained.	The data disk will become unavailable, but the data will be retained.	For a Classic network instance, the assigned Internet IP address will be retained. For a VPC instance, the assigned Internet IP address and the attached EIP address will be retained.	Snapshots will not be affected.

Time period	ECS instance and image	System disk	Data disk	Internet IP address	Snapshot
More than 30 days after expiration	The ECS instance and the image will be automatically released.	The system disk will be released along with the instance. The data cannot be restored.	By default, the cloud disk will be released along with the instance. If you have set the cloud disk not to be released along with the instance, they will stop operating. The local disk will be released along with the instance. The shared block storage will be automatically detached.	For a Classic network instance, the assigned Internet IP address will be released. For a VPC instance, the assigned Internet IP address will be released and the attached EIP address will be automatically detached.	Automatic snapshots will be automatically deleted, whereas snapshots created manually will not be affected.

* When the instance operates normally, it means that you can start or stop the instance and connect to the instance by using the management terminal of the ECS console.

** After the instance becomes unavailable, you cannot connect to the instance remotely, and websites deployed on the instance cannot be accessed. Service errors will occur.

5.2 Pay-As-You-Go

With the Pay-As-You-Go billing method, you are charged based on the amount of resources you actually use. Pay-As-You-Go allows you to activate and release resources at any time to meet your requirements. You can purchase resources on demand, and scale up as your business grows. Costs can be reduced by 30% to 80% compared with the investment on a traditional host, with which many resources may be wasted at times.

Applicable resources

Currently, the Pay-As-You-Go billing method is applicable to the following ECS resources:

- ECS instances, including CPU configuration and memory capacity
- Images
- System disks and data disks

If you create an ECS instance that uses the Pay-As-You-Go billing method, the Instance Cost displayed at the bottom of the instance creation page is the total fee for the preceding three types of resources.

You can make the following changes after activating Pay-As-You-Go resources:

- Resource configuration change

You can change the instance types, including CPU configuration and memory capacity, after you create an instance. For more information, see [#unique_121](#).

- Billing method change

You can change the billing method of instances, system disks, and data disks from Pay-As-You-Go to Subscription. For more information, see [Switch from Pay-As-You-Go to Subscription billing](#).

Payment methods

You can use a credit card or a PayPal account linked to your Alibaba Cloud account to pay for Pay-As-You-Go resources. For more information, see [Add a payment method](#).



Note:

If you are using PayPal as the payment method after activating a Pay-As-You-Go resource and place an order, Alibaba Cloud preauthorizes on your PayPal account.

Billing cycle

A Pay-As-You-Go resource is billed by second after it is created, and billing stops after it is released. For a VPC instance, you can enable the [#unique_142](#) feature to ensure that the VPC instance is not billed when it is in the Stopped status. This feature is available only for instances.

The minimum charge for the life cycle of an ECS instance (from creation to release) is USD 0.01.

Item	Instance + Image	System disk	Data disk
Billing cycle	One second	One second	One second
Price unit	USD/hour	USD/(GiB * hour)	USD/(GiB * hour)

Settlement period

Pay-As-You-Go resources are billed by second, but settled by hour. Note the following:

- Payments for Pay-As-You-Go resources are settled together with other products under your account that are billed after you use them.
- Generally, if the cumulative monthly consumption amount of your account is less than 1,000 USD, fees are deducted on the first day of the following month.

- If you have a quota agreement with Alibaba Cloud, fees are deducted when the cumulative consumption amount of your account exceeds the quota.

Cumulative consumption amount	Due date (T)	Fee deduction day	Description
Cumulative monthly consumption amount less than 1,000 USD.	The first day of the following month.	T, T+7, and T+14	<ul style="list-style-type: none"> - In the event of deduction failure on the due date (T), the system attempts to deduct fees again on the day T+7 and the day T+14. - If fee deduction fails three times, the instance goes out of service on the day T+15. In this case, the instance stops running but data is retained. Billing stops when the instance is out of service. - When your instance is out of service, you must open a ticket to clear the overdue payment. After the overdue payment is cleared, you must reactivate the instance before the day T+30. Otherwise, the instance is released automatically. - If the overdue payment fails to be cleared before the day T+30, the instance is released and data cannot be recovered.
Agreed quota	On the day when the quota is exceeded		

Instructions for settlement

- Settlement period
 - For ECS instances: The active duration is the time during which the ECS instance runs properly, starting from when the ECS instance is activated to when it is released or expires. If the instance is out of service during the active duration due to an overdue payment, billing stops until the payment is cleared.
 - For system disks and data disks: The active duration is the time during which the disks run properly, starting from when the disks are activated to when they are released. Fees are charged by hour.
- Release rules
 - If payment for an ECS instance is overdue, usage of Pay-As-You-Go cloud disks is restricted, and the cloud disks cannot process I/O read and write requests properly, affecting the normal running of the ECS instance. The impact includes

but is not limited to reduced performance of application read/write, serious time-out of some operations, and power-off or restart failure for some operating system versions.

- ECS instances configured with the automatic release time are automatically released at a specified time.
- Notification of release: In the event of service expiration or overdue payment, the system notifies you by email.

Resource status when an instance is out of service

If you fail to pay for Pay-As-You-Go resource fees three times in one settlement period, the instance is out of service on the day T+15. When your instance is out of service, you cannot use the resources normally until you clear the overdue payment. Once the payment is cleared, you must reactivate the instance within the specified period. The following table lists the status of the related resources once the instance is out of service.

Period	ECS instance and image	System disk and data disk	Internet IP address	Snapshot
Within 15 days after the instance goes out of service (T+15 to T+30)	Both stop working.	When the instance is out of service*, the capability of related cloud disks and local disks is limited. But the data on them is retained.	<ul style="list-style-type: none">• For instances of the classic network type: The assigned Internet IP address is retained.• For VPC instances: If an Internet IP address is assigned, it is retained. If an elastic IP (EIP) address is bound to the instance, it is retained.	Retained.

Period	ECS instance and image	System disk and data disk	Internet IP address	Snapshot
15 days after the instance goes out of service (T+30)	Released automatically. You are notified in advance by emails if your resources will be released.	<ul style="list-style-type: none"> • All cloud disks, including system disks and data disks are released automatically regardless of whether they are attached to the instance and whether they are created separately or at the same time as the instance. Data on the cloud disks cannot be recovered. • Local disks are released automatically, and data on them cannot be recovered. • If Shared Block Storage is attached to the instance, it is detached automatically, and data on it is retained. 	<ul style="list-style-type: none"> • For instances of the classic network type: The assigned Internet IP address is released. • For VPC instances: If an Internet IP address is assigned, it is released. If an EIP address is bound to the instance, it is unbound automatically. 	Automatic snapshots are deleted automatically. Snapshots that are manually created are retained.
				Issue: 20190920

* When a Pay-As-You-Go instance is out of service, you cannot remotely connect to the instance or access websites hosted on the instance.

FAQ

If a Pay-As-You-Go ECS instance is out of service or has stopped running, are fees still incurred?

An instance stops working and is rendered out-of-service when a payment is overdue. When a Pay-As-You-Go instance is out of service, it is in the Expired status, and no fees are incurred.

A stopped instance is stopped through the ECS console or by calling `StopInstance`. Billing of a stopped instance varies according to the network type of the instance:

- VPC: You can enable the [no fees for stopped instances](#) feature. If this feature is enabled, the instance is not billed when it is in the Stopped status. This feature is available only for instances.
- Classic: An instance continues to be billed even after it is in the Stopped status.

5.3 Preemptible instances

5.3.1 Preemptible instances

Preemptible instances are a type of on-demand instances.

When you create a preemptible instance, you can set a maximum price per hour to bid for a specified instance type. If your bid is higher than or equal to the current market price, your instance is created and billed according to the current market price. You can hold a preemptible instance without interruption for at least one hour. After one hour, your bid is compared with the market price. When the market price exceeds your bid or the resource stock is insufficient, the instance is automatically released.



Note:

After an instance is released, its data cannot be recovered. We recommend that you [create a snapshot](#) for an instance to back up its data before releasing it.

Scenarios

Preemptible instances are ideal for stateless applications, such as scalable Web services, figure rendering, big data analysis, and massively parallel computing.

Furthermore, applications that require a higher level of distribution, scalability, and fault tolerance capabilities, benefit from preemptible instances in terms of costs and throughput.

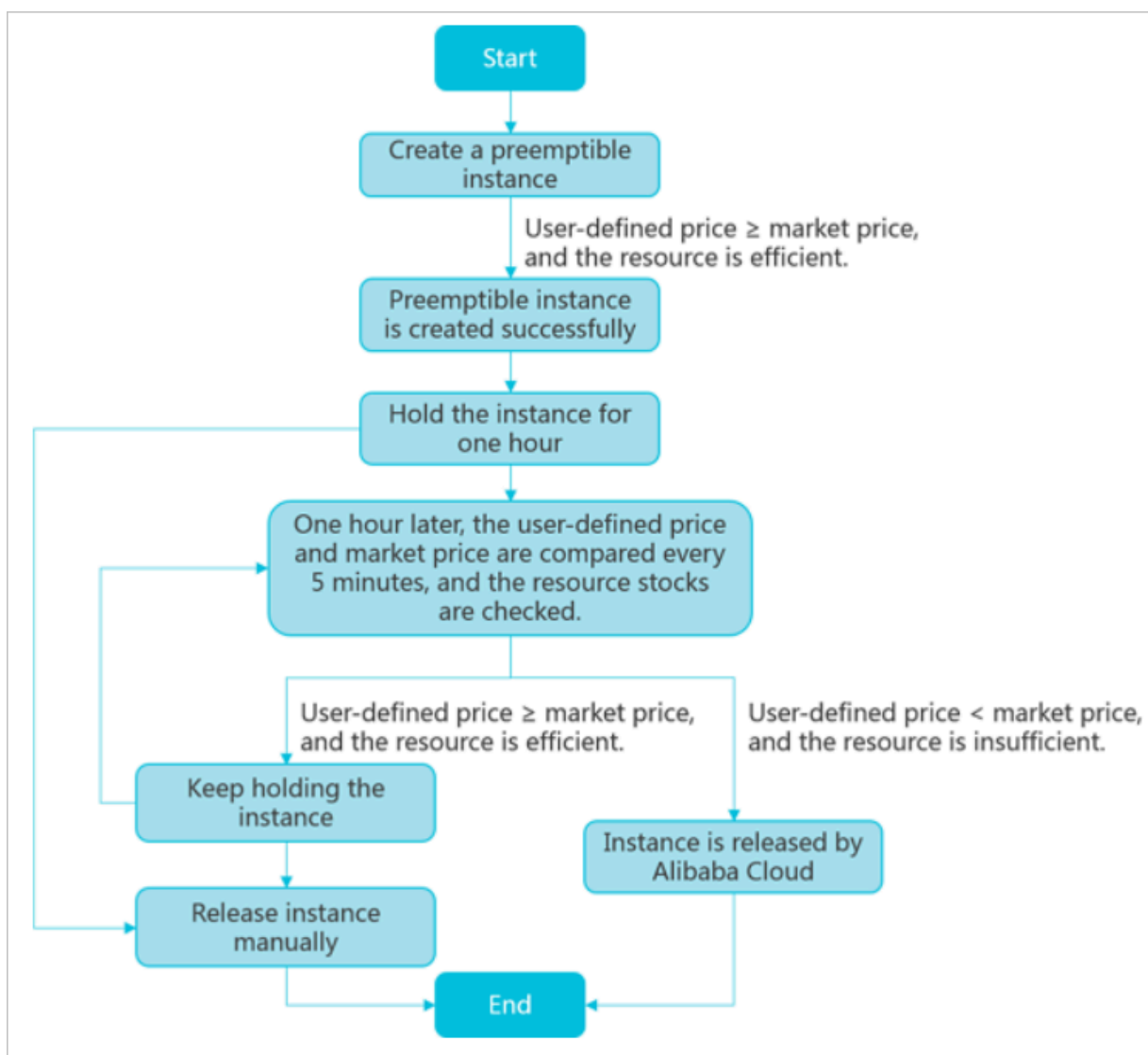
You can use preemptible instances for the following scenarios:

- Real-time analysis
- Big data
- Geological surveys
- Image coding and media coding
- Scientific computing
- Scalable Web sites and Web crawlers
- Image and media coding
- Testing

Preemptible instances are not suitable for stateful applications, such as databases, because it is difficult to store application states if the instance is released because of a failed bid or other reasons.

Lifecycle

The following figure shows the lifecycle of a preemptible instance.



Bidding modes

You can bid for a preemptible instance only once. Moreover, you can bid in either of the following bidding modes:

- SpotAsPriceGo

This mode means you always set the real-time market price as the bidding price. Moreover, there is an upper limit for your bidding price, namely the price of the corresponding Pay-As-You-Go instance type.

If you use the [ECS purchase page](#) to create an instance, select Use Automatic Bid. If you use the [#unique_71](#) interface to create an instance, set the `SpotStrategy` parameter to `SpotAsPriceGo`.

- **SpotWithPriceLimit**

This mode means you must set the highest price you are willing to pay for a specified instance type.

If you use the [ECS purchase page](#) to create an instance, select Set Custom Maximum Price (Per Instance/Hour). If you use the [#unique_71](#) interface to create an instance, set the `SpotStrategy` parameter to `SpotWithPriceLimit`.

Currently, the maximum bid of a preemptible instance is the price of a Pay-As-You-Go instance of the same configuration. When creating a preemptible instance, you can set a price according to the displayed price range, business features, and the estimated future price fluctuation. When the market price is lower than or equal to your bid, and the resource stock is sufficient, the instance can be created. If your bid takes into account the estimated future price fluctuation, you can hold the instance even after the one-hour [guaranteed duration](#). Otherwise, your instance may be automatically released at any time after that duration.

Guaranteed duration

When a preemptible instance is created, it has a guaranteed duration of one hour, namely, the first hour after it is created. During this period, the instance will not be released because of stock shortage, and you can run services on the instance as normal. Beyond the guaranteed duration, the market price and stock is checked every five minutes. If the market price at any given point of time is higher than your bid or the instance type stock is insufficient, your preemptible instance will be automatically released.

Price and billing

Preemptible instance price and billing considerations:

- **Price**

The preemptible instance price applies to the instance type only, including vCPUs and memory, but does not include system disks, data disks, or network bandwidth.

Instead, system disks and data disks are billed according to the [Pay-As-You-Go](#) billing method. Network bandwidth is billed according to the bandwidth billing rules of Pay-As-You-Go instances. For more information, see [#unique_148](#).

- **Billing method**

Preemptible instances are billed by second. When a preemptible instance is created successfully, the market price is an hourly rate and you only need to divide it by 3,600 to get the price per second.

The cost incurred from creating a preemptible instance to releasing it is accurate to two decimal places. An accrued cost of less than USD 0.01 is not charged.

- **Billing duration**

Instances are billed according to the actual period of use. The actual period of use is the duration from instance creation to instance release. After an instance is released, it is no longer billed. If you stop an instance by using [#unique_149](#) or in the [ECS console](#), the instance continues to be billed.

- **Market price**

During creation of a preemptible instance, it runs when your bid is higher than the current market price and the resource stock is sufficient.

In the first hour of its running, the instance is billed according to the initial market price. After that, it is billed according to the real-time market price.

The market price of a preemptible instance fluctuates according to the changes in the demand and supply of a given instance type. Therefore, we recommend that you pay attention to and evaluate the market price fluctuations to ensure you can take advantage of lower computing costs and increased throughput when purchasing preemptible instances.

Quota

Whether you can purchase a preemptible instance depends on your ECS instance resource usage.

For more information about the preemptible instance quota, see [#unique_151](#).

Create a preemptible instance

You can purchase a preemptible instance by using the [#unique_71](#) interface.

After a preemptible instance is created, it can be used in the same way as a Pay-As-You-Go instance. You can also use it with other cloud products, such as cloud disks or EIP addresses.

Stop a preemptible instance

You can stop a preemptible instance in the [ECS console](#) or by using the [#unique_149](#) interface. VPC preemptible instances support the [No fees for stopped VPC instances](#) feature.

The network type and the bidding mode of a preemptible instance determine whether it can start after it is stopped, as shown in the following table.

Network type + bidding mode	Stop instance	Restart instance
VPC + SpotWithPriceLimit	Select two options: Stop and Keep Stopped Instances and Continue Billing	During the guaranteed duration, the instance can be restarted successfully. After the guaranteed duration: <ul style="list-style-type: none"> · If your bid is not lower than the market price and the resource stock is sufficient, the instance can be restarted successfully. · If your bid is lower than the market price or the resource stock is insufficient, the instance cannot be restarted.
Classic network + SpotWithPriceLimit	N/A	
VPC + SpotAsPriceGo	Select two options: Stop and Keep Stopped Instances and Continue Billing	During the guaranteed duration, the instance can be restarted successfully. After the guaranteed duration: <ul style="list-style-type: none"> · If the resource stock is sufficient, the instance can be restarted successfully. · If the resource stock is insufficient, the instance cannot be restarted.
Classic network + SpotAsPriceGo	N/A	

Network type + bidding mode	Stop instance	Restart instance
VPC + SpotWithPriceLimit	Select one option only: Stop	<p>During the guaranteed duration, the instance can be restarted successfully only if the resource stock is sufficient. After the guaranteed duration:</p> <ul style="list-style-type: none"> · If your bid is not lower than the market price and the resource stock is sufficient, the instance can be restarted successfully. · If your bid is lower than the market price or the resource stock is insufficient, the instance cannot be restarted successfully.
VPC + SpotAsPriceGo	Select one option only: Stop	<p>During the guaranteed duration, the instance can be restarted successfully only if the resource stock is sufficient. After the guaranteed duration:</p> <ul style="list-style-type: none"> · If the resource stock is sufficient, the instance can be restarted successfully. · If the resource stock is insufficient, the instance cannot be restarted.

Release a preemptible instance

When the guaranteed period ends, the instance is automatically released according to market price changes or if there is insufficient stock. Additionally, you can independently [release the instance](#).

When a preemptible instance is released because of market price changes or insufficient stock, the instance enters the Pending Release status. Then, the instance is released in about five minutes. You can use [instance metadata](#) or the `OperationLocks` information returned by calling the `#unique_10` interface to check if an instance is in the Pending Release status.

**Note:**

Although you can check if a preemptible instance is in the Pending Release status by using the API and save a small amount of data while the instance is in this status, we recommend that you design your applications in such a way that your services can continue even if the preemptible instance is immediately released. You can release a preemptible instance manually so as to test whether your applications run normally if an instance is automatically released.

Generally, preemptible instances are released in the order of bidding price, from low to high. If multiple preemptible instances have the same bidding price, they are randomly released.

Best practices

When using a preemptible instance, consider the following:

- Set a proper bidding price. That is, we recommend that you evaluate and submit a competitive price and take into account the estimated market price fluctuations. By doing so, your request can be accepted and your created preemptible instances will not be released easily due to price fluctuations. Additionally, we recommend that you take into account your pricing expectations based on your own business assessment.
- The image must have all the software configurations that your applications need, making sure that you can run your business immediately after the instance is created. Additionally, you can use `#unique_154` to run commands upon startup.
- Store your business data on storage products that are independent from preemptible instances, such as cloud disks that are not set to release together with instances, OSS, or RDS.
- Split your tasks by using grids, Hadoop, queue-based architecture, or check points, thus making it easy to store computing results frequently.

- Use release notifications to monitor the status of a preemptible instance. You can use [instance metadata](#) to check the instance status every minute. The metadata of a preemptible instance is updated five minutes before it is released automatically.
- Test your applications to make sure that they can handle the accidental release of instances. You can test your applications as follows: run the applications on a Pay-As-You-Go instance, release the instance, and then check how the applications handle the release.

For more information, see [FAQ about preemptible instances](#).

For more information about using APIs to create preemptible instances, see [Use APIs to manage preemptible instances](#).

5.3.2 Create a preemptible instance

This topic describes how to create a preemptible instance and view its bills in the ECS console.

Prerequisites

The following requirements must be met before you create a preemptible instance:

- An appropriate bid has been made for the preemptible instance you want to purchase. For more information about an appropriate bid, see [#unique_126](#).
- The image used for creating a preemptible instance must contain the configuration of all the required software. You can also use user data to run commands at instance startup. For more information, see [User data](#).
- Applications that can withstand accidental instance release are set up.



Note:

You can run your applications on a Pay-As-You-Go instance and release the instance to verify whether your applications can properly handle automatic instance release.

Precautions

We recommend that you perform the following actions before you create a preemptible instance:

- To prevent any data loss caused by instance release, save important data in storage media such as separately created cloud disks, OSS, or RDS.

- Break down your jobs into small tasks by using grids, Hadoop, queue-based architecture, or checkpoints to save calculation results on demand.
- You can monitor the status of a preemptible instance by checking the instance release notifications issued by Alibaba Cloud ECS.

**Note:**

Alibaba Cloud ECS updates the instance metadata five minutes before releasing a preemptible instance. You can obtain the status of a preemptible instance every minute by checking the [instance metadata](#).

Create a preemptible instance

1. On the Instances page, click Create Instance.
2. Set Billing Method to Preemptible Instance.
3. In the Maximum Price for Instance Type Per Instance area, type your bid in the text box.

**Note:**

- You can create a preemptible instance only if your bid is higher than the market price and the resource stock is sufficient.
- You can bid for a preemptible instance only once.
- The following two bidding modes are supported:
 - Use Automatic Bid: The real-time market price is used as the bidding price.
 - Set Custom Maximum Price (Per Instance/Hour): The highest price you are willing to pay for a specified instance type.

In the displayed price range, the highest price is the price for the Pay-As-You-Go instance of the same configuration. Your bid must be based on the displayed price range, your service needs, and the estimated future price fluctuation. If your bid takes into account the estimated future price fluctuation, you can hold the instance even after the one-hour guaranteed duration. Otherwise, your instance will be automatically released at any time after that duration.

4. Select or enter the quantity of instances you want to purchase.
5. Complete other settings.

For the description of other parameters, see [#unique_69](#).

6. After the order is confirmed, click Create Instance.

After a preemptible instance is created, you can view its information in the instance list. A preemptible instance is marked as a Pay-As-You-Go-Preemptible Instance. On the Instance Details page, you can view the bidding policy configured during instance creation in the Payment Information area.

You can also create a preemptible instance by calling the [RunInstances](#) API action through Alibaba Cloud CLI, OpenAPI Explorer, or Alibaba Cloud SDK.

**Note:**

If you select the Use Automatic Bid bidding mode, set the `SpotStrategy` parameter in this API to `SpotAsPriceGo`. If you select the Set Custom Maximum Price (Per Instance/Hour) bidding mode, set this parameter to `SpotWithPriceLimit`.

View bills of a preemptible instance

Unlike Pay-As-You-Go instances, the price of a preemptible instance is the concluded price.

To view the bills of a preemptible instance on the Instance Details page, following these steps:

1. On the Instances page, click the ID of the target preemptible instance or click **Manage** in the Actions column.
2. On the Instance Details page, choose **More > View Fees** in the Payment Information area.
3. On the Bills page, click **Detail** in the Action column.

To view the bills of a preemptible instance on the Billing Management page, following these steps:

1. Choose **Billing Management > Billing Management**.
2. On the Billing Management page, choose **Spending Summary > Instance Spending Details**.
3. On the Instance Spending Details page, find the target preemptible instance and click **Detail** in the Action column.

**Note:**

You can filter instances by billing cycle, product name, and status.

5.3.3 Stop a preemptible instance

This topic describes how to stop a preemptible instance and whether it can start successfully after being stopped in different conditions.

Context

Only preemptible instances in a VPC support the [#unique_142](#) feature. The network type, bidding mode, and stop mode of an instance combine to determine whether a preemptible instance can restart successfully. The following table provides the details.

Network type	Bidding mode	Stop mode	Can the stopped instance be restarted?
Classic network	SpotWithPriceLimit	Keep Stopped Instances and Continue Billing	During the guaranteed duration, the instance can be restarted successfully. However, after the guaranteed duration, the instance can only be restarted successfully if your bid is not lower than the market price and if the number of resources is sufficient.
	SpotAsPriceGo	Keep Stopped Instances and Continue Billing	During the guaranteed duration, the instance can be restarted successfully. However, after the guaranteed duration, the instance cannot be restarted if the number of resources is insufficient.
VPC	SpotWithPriceLimit	Keep Stopped Instances and Continue Billing	During the guaranteed duration, the instance can be restarted successfully. However, after the guaranteed duration, the instance can only be restarted successfully if your bid is not lower than the market price and if the number of resources is sufficient.
		Stop/Force Stop (no fees will be incurred if either of the options is selected)	During the guaranteed duration, the instance can be restarted successfully as long as the number of resources is sufficient. However, after the guaranteed duration, the instance can only be restarted successfully if your bid is not lower than the market price and if the number of resources is sufficient.

Network type	Bidding mode	Stop mode	Can the stopped instance be restarted?
	SpotAsPriceGo	Keep Stopped Instances and Continue Billing	During the guaranteed duration, the instance can be restarted successfully. However, after the guaranteed duration, the instance cannot be restarted if the number of resources is insufficient.
		Stop/Force Stop (no fees will be incurred if either of the options is selected)	During the guaranteed duration, the instance can be restarted successfully as long as the number of resources is sufficient. However, after the guaranteed duration, the instance cannot be restarted if the number of resources is insufficient.

Procedure

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Instances.
3. On the Instances page, find the preemptible instance to be stopped. In the Actions column, choose More > Instance Status > Stop.
4. In the Stop Instance dialog box, click OK.

More information

[#unique_159](#)

5.4 Reserved Instances

5.4.1 Reserved instance overview

A reserved instance is a discount coupon that can be applied automatically to one or more pay-as-you-go instances that belong to your account, excluding preemptible instances. A reserved instance also can be used to reserve instance resources. A combination of reserved instances and pay-as-you-go instances is more flexible and cost effective than subscription instances.

Release notes

The reserved instance feature has been formally released. You can purchase reserved instances on the [ECS Reserved Instances](#) page. For more information, see [#unique_162](#).

Comparison between reserved instances, pay-as-you-go instances and subscription instances

The following table lists differences between reserved instances, pay-as-you-go instances and subscription instances.

Item	Reserved instance	Pay-as-you-go instance	Subscription instance
Form	A discount coupon.	An instance using the #unique_7 billing method, equivalent to a virtual machine.	An instance using the #unique_6 billing method, equivalent to a virtual machine.
Purpose	Reserved instances cannot be used independently. They must match pay-as-you-go instances to generate a discount .	Pay-as-you-go instances can be managed independently. They can be used as simple Web servers, or used in combination with other Alibaba Cloud services to deliver powerful solutions.	Subscription instances can be managed independently. They can be used as simple Web servers, or used in combination with other Alibaba Cloud services to deliver powerful solutions.

Payment options, terms, and instance count

When purchasing a reserved instance, you can specify the payment option, term, and count based on your budget.

- Three payment options are available:

All Upfront, Partial Upfront, and No Upfront. For more information, see [Reserved instance billing](#).



Note:

Whether you can use the No Upfront payment option depends on your ECS instance resource usage.

- Reserved instances come in the following terms:

1 year and 3 years.



Note:

After a reserved instance expires, the corresponding pay-as-you-go instances still run normally, but they will be billed without a discount.

- **Instance count:**

The number of pay-as-you-go instances that a reserved instance can match at the same time.

Reserved instance attributes

A reserved instance has specific attributes that allow it to automatically match corresponding pay-as-you-go instances. You can also split a reserved instance, merge multiple reserved instances, or change the scope of a reserved instance to better match pay-as-you-go instances. These attributes include:

- **Operating system:** Reserved instances can match pay-as-you-go instances running the Linux and Windows operating systems.



Note:

Windows reserved instances can be used to pay for image bills of pay-as-you-go instances.

- **Instance type:** the type of a reserved instance, which includes the instance family and instance size. A reserved instance can match pay-as-you-go instances of the same type.
- **Scope:** the matching scope of a reserved instance. In terms of matching scope, reserved instances are classified into regional reserved instances and zonal reserved instances.



Note:

We recommend that you purchase both zonal and regional reserved instances. Choose zonal reserved instances when you are very sure about in which zones you want to use them. Otherwise, choose regional reserved instances to meet uncertain requirements.

- **Computing power:** the maximum computing resources that a reserved instance can match. The computing power is determined jointly by the instance type and instance count.

Limits

Reserved instances have the following limits:

- Maximum number of reserved instances
 - Maximum number of regional reserved instances: Each account can have up to 20 regional reserved instances in all regions.
 - Maximum number of zonal reserved instances: Each account can have up to 20 zonal reserved instances in each zone.

For example, you can purchase 10 regional reserved instances in China (Hangzhou) and 10 regional reserved instances China (Qingdao), but the total number of regional reserved instances cannot exceed 20. You can purchase 20 zonal reserved instances in Zone B of China (Hangzhou) and 20 zonal reserved instances Zone H of China (Hangzhou). If you need more reserved instances, you can [submit a ticket](#).

- Instance categories

Reserved instances can only match pay-as-you-go instances, excluding [preemptible instances](#).

- Instance families

Reserved instances support the following instance families: sn1ne, sn2ne, se1ne, ic5, c5, g5, r5, c6, g6, r6, i2, i2g, hfc5, hfg5, and t5. [Burstable instances \(t5\)](#) can only match zonal reserved instances. They cannot match regional reserved instances. Reserved instances that match burstable instances cannot be merged, split, or have their matching scope changed.

Billing

For more information, see [Reserved instance billing](#).

Related topics

For more information about matching rules of reserved instances, see [Matching rules of reserved instances](#).

For more information about how to purchase reserved instances, see [Purchase reserved instances](#).

For more information about how to manage reserved instances, see [Manage reserved instances](#).

For more information about how to call an API operation to purchase reserved instances, see [PurchaseReservedInstancesOffering](#).

For more information about how to call an API operation to query reserved instances, see [DescribeReservedInstances](#).

For more information about how to call an API operation to manage reserved instances, see [ModifyReservedInstances](#).

For more information about reserved instance FAQ, see [#unique_12](#).

5.4.2 Matching rules of Reserved Instances

Reserved Instances (RIs) provide a billing benefit only when they match Pay-As-You-Go instances. This topic describes the matching rules of RIs and provides some examples.

Matching rules of RIs

The matching status between an RI and a Pay-As-You-Go instance cannot be manually managed. After you purchase an RI, the RI automatically matches one or more Pay-As-You-Go instances that have certain attributes within its term. The matching elements include operating system, instance type, and scope.

If you do not have any Pay-As-You-Go instances under your account, the RI will be idle while continuing to incur fees. After you purchase one or more applicable Pay-As-You-Go instances, the RI will automatically match with the instances immediately. Successful matching leads to an immediate billing discount to your Pay-As-You-Go instances. For more information, see [Reserved Instance billing](#).

An RI takes effect and is billed on the hour upon successful purchase. It expires at 00:00:00 the day after the term end date. For example, you purchased an RI on February 26, 2019 13:45:00 PM, with a term of one year. The RI took effect on 2019-02-26 13:00:00, and its billing also started from that time. It will expire on 2020-02-27 00:00:00. If you had matchable instances when you purchased the RI, the billing discount first applied to the bill generated from 13:00 to 14:00 on February 26, 2019 till the expiration of the RI.

We recommend that you purchase RIs in advance according to your business needs. You can also manage your RIs flexibly to maximize your billing discount.

The following table describes the features of regional RIs and zonal RIs.

Feature	Regional RI	Zonal RI	Example
Instance size flexibility	Supported . A regional RI can match different sizes of Pay-As-You-Go instances that are of the same instance type family in the same region.	Not supported . A zonal RI must match one or more Pay-As-You-Go instances of a specified size.	<p>You have the following running Pay-As-You-Go instances:</p> <p>Two ecs.c5.xlarge Linux instances in China (Qingdao). The instance names are C5PAYG-1 and C5PAYG-2 respectively.</p> <p>You purchase the following RI:</p> <p>One regional ecs.c5.2xlarge RI in China (Qingdao). The RI name is C5RI.</p> <p>After the purchase, C5RI matches C5PAYG1 and C5PAYG2 simultaneously to provide a billing discount.</p>
Zone flexibility	Supported . A regional RI can match all Pay-As-You-Go instances in the same region.	Not supported . A zonal RI must match one or more Pay-As-You-Go instances in a specified zone.	<p>You have the following running Pay-As-You-Go instance:</p> <p>One ecs.c5.xlarge Linux instance in zone B of China (Qingdao). The instance name is C5PAYG-b.</p> <p>You purchase the following RI:</p> <p>One regional ecs.c5.xlarge RI in China (Qingdao). The RI name is C5RI.</p> <p>After the purchase, C5RI matches C5PAYG-b to provide a billing discount.</p> <p>You release C5PAYG-b, and then start another Linux instance named C5PAYG-c, which is of the same instance type as C5PAYG-b, in zone C. C5RI then matches C5PAYG-c to provide the same billing discount.</p>

Feature	Regional RI	Zonal RI	Example
Resource Reservation	Not supported . If there is a shortage of available instances, you may need to wait for instances to become available.	Supported. A specified number of Pay-As-You-Go instances are reserved so that these instances can be created successfully.	You purchase the following RIs: Five zonal ecs.c5.xlarge RIs in zone B of China (Qingdao). The term is 1 year and the instance count is 2. Ten ecs.c5.xlarge instances will be reserved in zone B of China (Qingdao) for one year.

5.4.3 Purchase reserved instances

This topic describes how to purchase reserved instances in the ECS console.

Prerequisites

- Before purchasing reserved instances, make sure that the pay-as-you-go instances you want to match meet requirements to apply reserved instances. For more information, see [Reserved instance overview](#).
- You cannot manually manage how reserved instances and pay-as-you-go instances are matched. Ensure that you understand the matching rules for reserved instances. For more information, see [#unique_174/unique_174_Connect_42_section_pqc_yyq_dgb](#).

Context

You can purchase reserved instances on the [ECS Reserved Instances](#) page or in the [ECS console](#). This topic demonstrates how to purchase reserved instances in the ECS console.

Procedure

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Reserved Instances.
3. In the top navigation bar, select a region.
4. Click Purchase Reserved Instance.

5. Configure region-related parameters.

- a) Set Region.
- b) Set Resource Reservation to Reserved or Not Reserved.



Note:

You can only set Resource Reservation to Reserved for zonal reserved instances. Regional reserved instances can be applied to pay-as-you-go instances in different zones within the same region.

- c) Set Zone.

6. Configure instance-related parameters.

- a) Set Instance Type.



Note:

You must select an instance size when you purchase a regional reserved instance. However, the regional reserved instance can match any pay-as-you-go instances of the specified instance family within the specified region regardless of size.

- b) Set Operating System Platform.

Linux and Windows operating systems are available.



Note:

A reserved instance only matches pay-as-you-go instances that use the selected operating system type. You cannot change the operating system type of a reserved instance after you purchase it.

To apply a reserved instance to pay-as-you-go instances created from Bring Your Own License (BYOL) images, you must [submit a ticket](#).

- c) Set Payment Option.

All Upfront, Partial Upfront, and No Upfront are available. For more information, see [Reserved instance billing](#).

7. Configure purchase parameters.

a) Optional: Set Name.

b) Set Term.

1 Year and 3 Years are available.

c) Set Instance Count.

A reserved instance can match the specified number of pay-as-you-go instances of the specified instance type. For example, if the instance type is `ecs.g5.large` and Instance Count is set to 3, the reserved instance can match three pay-as-you-go instances of the `ecs.g5.large` instance type.

8. Select *ECS Terms of Service* . Click Purchase.

9. In the dialog box that appears, confirm the parameters and click Create Order.

10. Confirm the payment information and click Pay.

What's next

After you purchase a reserved instance, discounts will be applied when the reserved instance matches one or more pay-as-you-go instances. You can also manage reserved instances to cope with configuration changes of pay-as-you-go instances. For more information, see [#unique_167](#).

5.4.4 Manage Reserved Instances

This topic describes how to split, merge, and modify the scope of your Reserved Instances (RIs). Such actions allow you to benefit from billing discounts of more Pay-As-You-Go instance types.



Note:

To make this topic easier to understand, RIs to be split, merged, or modified are hereinafter referred to as *original RIs*, while split, merged, or modified RIs are hereinafter referred to as *target RIs*.

Prerequisites

Before you split, merge, or modify RIs, make sure that the following conditions are met:

- You have successfully purchased original RIs and they are within a valid term.
- There is no ongoing splitting, merging, or modification request.

- The RI to be modified only requires its size to be adjusted. The instance type family of an RI cannot be modified.

After you submit a splitting, merging, or modification request:

- The original RI changes to changing status, which will be automatically refreshed after the request is processed.
- A request in progress cannot be changed or canceled. If you want to roll back your changes, you must submit another request.

After an RI is split, merged, or modified:

- The target RI becomes valid on the hour. If it matches one or more new Pay-As-You-Go instances, the billing discount is applied within the same hour.
- The original RI becomes invalid on the hour, and its price is updated to USD 0.
- If the target RI is a zonal RI, the type of resource reservation is also updated automatically.

For example, you successfully split an ecs.g5.2xlarge zonal RI (RI1) into two ecs.g5.xlarge zonal RIs (RI2 and RI3) at 2019-02-26 13:45:00. In this case, RI1 becomes invalid at 2019-02-26 13:00:00, while RI2 and RI3 take effect also at 2019-02-26 13:00:00. Starting 2019-02-26 13:00:00, the reserved instance type eligible for billing discount is also changed from ecs.g5.2xlarge to ecs.g5.xlarge. If RI2 and RI3 match instances immediately after they take effect, the hourly bill discount for ecs.g5.xlarge instances is also applied starting 2019-02-26 13:00:00.

If the original RI fails to be split, merged, or modified, it will remain valid.

Split an RI

You can split an RI into multiple RIs of less computing power. The smaller RIs can then match applicable Pay-As-You-Go instances to better distribute your service traffic.

1. In the left-side navigation pane, choose **Instances & Images > Reserved Instances**.
2. On the **Reserved Instances** page, click **Split** in the **Actions** column of the original RI.
3. On the **Split Reserved Instance** page, set the name, instance type, and instance quantity of the target RIs.



Note:

The total computing power of the target RIs must be equal to that of the original RI.

4. Click OK.

Merge RIs

If traffic to your instances increases, you can merge multiple RIs into one RI that has greater computing power to match larger Pay-As-You-Go instances.



Notice:

Before merging RIs, you must verify that the following conditions are met:

- The expiration date of the original RIs must be the same.
- The original RIs have been purchased using the same currency.
- If the original RIs are regional RIs, they must be in the same region. If the original RIs are zonal RIs, they must be in the same zone.

1. In the left-side navigation pane, choose **Instances & Images** > **Reserved Instances**.
2. On the **Reserved Instances** page, click **Merge** in the **Actions** column of the original RI.
3. On the **Merge Reserved Instances** page, select the original RIs, and then set the name, instance type, and instance quantity of the target RI.



Note:

The computing power of the target RI must be equal to that of all selected original RIs, and the target RI must be of an existing instance type. For example, two `ecs.g5.2xlarge` RIs can be merged into one `ecs.g5.4xlarge` RI, but one `ecs.g5.xlarge` RI and two `ecs.g5.2xlarge` RIs cannot be merged into one `ecs.g5.5xlarge` RI.

4. Click OK.

Modify the scope of an RI

If your service requirements change, you can modify the scope of your RIs. Specifically, you can:

- Modify a regional RI to a zonal RI.
- Modify a zonal RI to a regional RI.
- Modify the zone of an RI in the same region.

You cannot modify the scope of an RI across regions. For example, if you have a zonal RI in zone B of China (Hangzhou), you can modify it as a zonal RI in another zone of China (Hangzhou), or as a regional RI in China (Hangzhou). However, you cannot modify it as a regional or zonal RI in another region.

1. In the left-side navigation pane, choose **Instances & Images > Reserved Instances**.
2. On the **Reserved Instances** page, click **Modify** in the **Actions** column of the original RI.
3. On the **Modify Reserved Instance** Page, modify the parameters as needed.
4. Click **OK**.

5.5 Switch billing method

5.5.1 Switch from Pay-As-You-Go to Subscription billing

You can switch the billing method of your instance from Pay-As-You-Go to Subscription in the ECS console. Switching the billing method incurs a fee.

Limits

You can switch up to 20 Pay-As-You-Go instances to Subscription instances at one time.

Prerequisites

The ECS instance you want to switch the billing method for must meet the following requirements:

- The instance type is not a **Generation I** type.
- The instance belongs to your account.
- The instance is in a running or stopped status.

If an order to switch the billing method has been placed successfully when the ECS instance is in a running or stopped status, but the instance status changes so it no longer meets the preceding requirement when payment is attempted for the order, the order fails and the billing method is unchanged. You can go to the billing center and pay for the order when the instance is in a running or stopped status again.

- [No timed release is set for the instance.](#)

If the release time has been set for an instance, you need disable the timed release configuration and then switch the billing method.

- There is no unpaid switch order for the instance.

If an unpaid switch order exists, you must cancel the unpaid order and then place another order to switch the billing method.

Procedure

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, click Instances.
3. Select a region.
4. Select one or more Pay-As-You-Go instances, and under the instance list, click Switch to Subscription.
5. On the Switch to Subscription page, click Batch Change.
6. In the dialog box, set the Subscription Plan, including:
 - **Duration:** You can set the length of service time for the Subscription instance as 1 month or 1 year. Instances executed in batch must have the same length of service time.
 - **Data Disk (optional):** If Pay-As-You-Go data disks are mounted or attached to the selected instances, you can set whether to also switch their billing method to Subscription.
7. Click OK to place an order and go to the payment page.

After you make the payment, the operation is complete.

FAQ

What can I do if placing an order fails?

You may be prompted with any of the following error messages:

- The current instance status is not supported to switch.
- Switch is not allowed because the release time has been set for the instance.
- Switch is not allowed because instance information has changed.
- A switch order for the instance has not been paid.

If you are prompted with any of the preceding messages, adjust the instance accordingly.

How long does it take to switch the billing method after I pay for the order?

Currently, one to four seconds are required to switch the billing method of 1 to 20 instances. After the switch, the billing method is changed to Subscription in the console.

What can I do if the switch fails?

Please [Open a ticket](#).

Does the billing method for bandwidth change after the switch?

No. Only the billing method of an instance and data disk can be switched.

If I upgrade the configurations of my ECS instance that has an unpaid switch order, is the order still valid?

The order is invalid. A new order is placed when you switch the billing method of an instance from Pay-As-You-Go to Subscription. This new order must be paid. If the instance is upgraded when the order remains unpaid, the order payment cannot then be made because the instance components change and the order does not meet the requirements to switch the billing method. If you still want to change the billing method of the instance, you must cancel the unpaid order and place a new switch order.

5.5.2 Switch the billing method from Subscription to Pay-As-You-Go

This topic describes how to switch the billing method from Subscription to Pay-As-You-Go. After you create a Subscription instance, you can convert its billing method to Pay-As-You-Go if you want to pay only for the actual usage of your resources.

Before converting the billing method, note the following:

- The billing method conversion applies to the system disk and the data disks of the instance, regardless of whether the data disks are created along with the instance.
- The billing method of Internet bandwidth is unchanged.

After the conversion, make sure your account has sufficient funds to pay the bill. Otherwise, overdue payment will affect the running of your services. If you no longer require the instance, you can set it to auto release or release it manually. For more information, see [Release an instance](#).

Conversion notes

- Whether you can convert the billing method is determined by your ECS instance resource usage.
- Each account is limited by a maximum monthly refund amount. For more information, see the conversion page. You cannot claim additional refunds once the monthly limit is reached. The remaining amount will be cleared automatically on the 1st day of the next month.

To calculate the refund amount resulting from a billing method conversion, the calculation uses the number of vCPUs and the remaining hours in the current billing cycle. For example, 1 refund unit = $1 \text{ vCPU} \times 1 \text{ hour}$.

In this example, assume you have purchased a Subscription instance with four vCPUs for six months. Four months later, you convert the billing method to Pay-As-You-Go. In this case, the refund amount for this instance is $4 \text{ (vCPUs)} \times 60 \text{ (remaining days)} \times 24 \text{ (hours / day)} = 5760 \text{ (refund amount)}$.

- If the instance involves renewal or upgrade orders that have not taken effect, a full refund will be made. For orders that have already taken effect, only a partial refund will be made.
- After the billing method is converted, the duration of usage given for reasons such as ICP filing, failure, or IDC migration is automatically invalidated.

Prerequisites

- The relevant ECS instance and cloud disks must use the Subscription billing method.
- The relevant ECS instance is in the Running or Stopped state.

Procedure

1. Select the relevant instance. In the Actions column, choose More > Configuration Change > Switch to Pay-As-You-Go.
2. On the conversion page, read the notes and confirm you agree to the *ECS Service Terms*.
3. Click Confirm Conversion.

After the billing method is converted successfully, you can see that the Billing Method of the instance is changed to Pay-As-You-Go in the ECS console.

Alternatively, click the instance ID to enter the Instance Details page. In the left-side navigation pane, click Disks. You can see that the Billing Method of the system disk and data disks (if any) of the instance have changed to Pay-As-You-Go.

FAQ

What can I do if the conversion fails?

Conversion failure may be due to:

- The current instance status does not support conversion.
- Conversion is not allowed because the instance has expired.
- Conversion is not allowed because instance information has changed.

If any of the preceding error messages appears, adjust the instance accordingly. If the problem persists, [open a ticket](#).

6 Create an instance

6.1 ECS instance creation overview

This topic describes several methods by which to create an ECS instance, from basic creation operations to advanced customization operations.

Generally, we recommend that you create an ECS instance by following the instructions prompted by the wizard, which allows you to choose configurations flexibly. For more information, see [Create an instance through the wizard](#).



Note:

If you require custom configurations (such as a specific operating system or application), you can create a custom image and then select that image during the creation of an instance to improve configuration efficiency. For more information, see [Create an instance by using a custom image](#).

If you need a new instance to have the same configuration as the current instance, you can create an instance of the same configuration directly. For more information, see [Create an instance of the same configuration](#).

You can also create a launch template in advance, and then use it to create a new instance in one click as needed. For more information, see [Use a launch template](#) and [Launch templates](#).

6.2 Create an instance by using the wizard

This topic describes how to create an instance by using the ECS console wizard.

Prerequisites

- You have completed the [preparation work](#).
- To specify an SSH key pair when creating a Linux instance, you must [create an SSH key pair](#) in the target region.
- To set the user-defined data, you must prepare [User Data](#).
- To authorize an instance to assume a role, you must [create an instance RAM role and grant it permissions](#).

Procedure

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Instances.
3. On the Instances page, click Create Instance.
4. Complete Basic Configurations as follows:
 - a) Select a Billing Method: Subscription, Pay-As-You-Go, or [Preemptible Instance](#).



Note:

For information about how to create a preemptible instance, see [Create a preemptible instance](#).

- b) Select a region and zone.

By default, a zone is assigned randomly. You can select a zone that better meets your needs. For more information about regions and zones, see [Regions and zones](#)



Note:

- After an instance is created, you cannot change its region and zone.
- For the instance types that you can purchase in each region, see [Instance types available for each region](#).

- c) Select an instance type and specify the quantity of instances.

The availability of an instance type family is determined by the selected region. For the scenarios of each instance type, see [Instance type families](#).



Note:

- The quota of Pay-As-You-Go or preemptible instances for your account is shown on the page.
- To use Elastic Network Interfaces (ENIs), select an enterprise-level instance type with at least two vCPUs or an entry-level instance type with at least four vCPUs. For more information about the maximum number of ENIs that can be attached to one instance, see [Instance type families](#).

- To use an SSD Cloud Disk, select an I/O-optimized instance.

d) Select an image. You can select a public image, custom image, shared image, or Marketplace image.



Note:

- To use an SSH key pair, select a Linux image.
- To set user-defined data, select an image as instructed in [#unique_154](#).
- Public images only include the initial system environment, and more images are available in the image Marketplace.

e) Select storage devices:

- System Disk: mandatory. A system disk is required for installing the operating system. Specify the cloud disk category and size for the system disk:
 - Cloud disk category: The available categories are determined by the selected region.
 - Size: The default size is 40 GiB. If the selected image file is greater than 40 GiB, the size is defaulted to the image file size. The available size range varies with the selected image, as described in the following table.

Image	Available size range (GiB)
Linux (excluding CoreOS) and FreeBSD	[max{20, ImageSize}, 500]
CoreOS	[max{30, ImageSize}, 500]
Windows	[max{40, ImageSize}, 500]

- Data Disk: optional. If you create a cloud disk as a data disk at instance creation, you must select the disk type, size, and quantity, and set whether to [encrypt](#) it. Alternatively, you can attach a [Pay-As-You-Go cloud disk](#) to the instance after instance creation. You can also create an empty data disk or create a data disk from a snapshot. Up to 16 data disks can be added.



Note:

The data disks added here have the following features:

- The billing method is the same as that of the instance.

- A Subscription data disk must be released at the same time as its corresponding instance, while a Pay-As-You-Go data disk can be released separately or at the same time as the corresponding instance.

- If you have selected an instance family that has local disks (such as i1, d1, or d1ne), the local disk information is displayed. You cannot specify the quantity or category of local disks, which are determined by the selected instance type. For information about the local disks corresponding to various instance types with local disk, see [Instance type families](#).

5. Click Next: Networking to finish the network and security group configuration:

a) Select a network:

- VPC: You must select a VPC and a VSwitch. If you do not have a VPC and a VSwitch, you can use the default ones.
- Classic network: If you purchased the ECS instance for the first time after June 16, 2016, 12:00 (UTC + 8), you can no longer select a classic network.

b) Configure the Network Billing Method:

- To assign a public IP address to the instance, select Assign Public IP Address. Then, select Pay-By-Traffic as the network billing method and specify the bandwidth. For public IP addresses assigned in this way, you cannot detach them from the instance. For more information about Internet Bandwidth billing, see [#unique_148](#).
- If your instances do not need to access the Internet or your VPC instances [use an Elastic IP \(EIP\) address to access the Internet](#), you do not need to assign a public IP address. You can detach an EIP address from an instance.

c) Select a security group. If you have not created a security group, you can use the default security group. For the rules of the default security group, see [#unique_193](#).

d) Add an ENI. If your selected instance type supports ENIs, you can add one and specify a VSwitch for it.



Note:

By default, the ENI is released along with the instance. You can detach it from the instance in the [ECS console](#) or by using the [DetachNetworkInterface](#) API.

6. (Optional) Click Next: System Configurations to finish the following configuration:

- Select and set logon credentials. You can choose [Set Later](#) or set it now. Select a credential based on the image:
 - Linux: You can select a password or SSH key pair as a logon credential.
 - Windows: You can only select a password as a logon credential.
- Specify the instance name, which is displayed in the ECS console, and the host name, which is displayed inside the guest operating system.
- Set the advanced options:
 - Instance RAM role: Assign a RAM role to the instance.
 - User Data: Customize the startup behaviors of an instance or pass data into an instance.

7. (Optional) Click Next: Grouping to manage instances by group. You can add tags to instances to simplify future management.

8. Confirm the order:

- In the Configurations Selected area, confirm all the configurations. You can also click the edit icon to re-edit the configuration.
 - (Optional) Click Save as launch template to save your configuration as a launch template for future use. For more information, see [Instance launch template](#).
 - (Optional) Click View Open API to acquire the API best practices about how to create instances. On the left side, API Workflow explains the related APIs and request parameter values for the current operation. On the right side, the programming language-specific samples are given for you to use. Currently,

Java and Python samples are provided. For more information, see *ECS API Reference* [Overview](#).

- (Optional) If the billing method is Pay-As-You-Go, you can set the Auto Release Schedule.
- (Optional) If the billing method is Subscription, you can set the duration and select whether to enable Auto renewal.
- Confirm the configuration costs. The billing methods for an instance and its Internet bandwidth determine the displayed cost information, as described in the following table.

Instance billing method	Estimated fee
Pay-As-You-Go or preemptible instance	Internet traffic fee + configuration fee. The configuration fees include: the instance type (vCPU and memory), the system disk, data disks (if any), and local disks (if any).
Subscription	Internet traffic fee + configuration fee. Configuration fees include: the instance type (vCPU and memory), the system disk, data disks (if any), and local disks (if any).

- Read and confirm you agree to the ECS Service Level Agreement.

9. Click Create Instance.

Result

After the instance is activated, click Console to view the instance details in the console. In the Instances list of the relevant region, you can view the information of the new instance, including the instance name, the Internet IP address, and the private IP address.

What's next

- You can create an FTP site on the instance for transferring files. For more information, see [Build an FTP site on an ECS instance](#).

- To secure your instance after creation, we recommend that you perform security compliance inspection and configuration:
 - Linux instances: See [Harden operating system security for Linux](#) in *Security Advisories* .
 - Windows instances: See [Harden operating system security for Windows](#) in *Security Advisories* .
- If a data disk is created along with the instance, you must partition and format the disk before use. For more information, see [#unique_199](#) or [#unique_200](#).

6.3 Create an instance by using a custom image

This topic describes how to create an instance by using a custom image. If you want to create an ECS instance that has the same operating system, software applications, and data as an existing instance, you can create a custom image and use it to create the new ECS instance. This method improves the deployment efficiency.

Context

- If the image and the instance are in the same region, create a custom image by using one of the following methods:
 - [Import an image](#)
 - [#unique_203](#)
 - [#unique_204](#)
- If the custom image and the instance are in different regions, copy the custom image to the target region. For more information, see [#unique_205](#).
- If the image to be used is owned by another account, it must be shared with you first. For more information, see [#unique_206](#).

Procedure

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Instances.
3. In the upper-right corner of the Instances page, click Create Instance.

4. Follow the steps when you [create an instance by using the wizard](#). When creating an ECS instance, note the following:
 - **Region:** Select the region where the image is located.
 - **Image:** Select Custom Image or Shared Image, and then select an image from the drop-down list.

**Note:**

If the selected custom image contains more than one data disk snapshot, an equal number of cloud disks are automatically created to function as data disks. By default, the size of each data disk is equal to that of the source snapshot. You can only increase the size of a data disk.

5. Confirm the order.

6.4 Create an instance of the same configuration

This topic describes how to duplicate ECS instances of the same configurations by using the Buy Same Type feature.

Procedure

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Instances.
3. In the top navigation bar, select a region.
4. Find your ECS instance and, in the Actions column, choose More > Buy Same Type.
5. On the Buy the Same Configuration page, confirm the selected configurations in the Overview section. If you want to modify any configurations, select View More to change the billing method, security group, network billing method, bandwidth, logon credential, or instance name.
6. To purchase a Subscription ECS instance, you can change the Purchase time.
7. Set the Amount.
8. Read and confirm you agree to the ECS Service Terms and Product Terms of Service.
9. Confirm the order.

6.5 Create an instance by using a launch template

This topic describes how to use an existing launch template to quickly create an ECS instance.

Prerequisites

You have [created a template](#) or [created a template version](#).

Procedure

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Deployment & Elasticity > Launch Templates.
3. Find the template or version that you want to use, and then click Create Instance in the Actions column.
4. On the ECS purchase page, select the required template and version.



Note:

You can click the edit icon next to the target launch template to modify its configurations.

5. If you want to create an instance using the Subscription billing method, select a subscription duration, read and confirm you agree with the Terms of Service, and then click Create Order. After you complete the payment, you can view the newly created instance in the ECS console.

If you want to create an instance using the Pay-As-You-Go billing method, read and confirm you agree with the Terms of Service, and then click Create Instance. After the instance is created successfully, you can view its details in the ECS console.

7 Connect to instances

7.1 Overview

Based on the network type and operating system of your ECS instance, and the operating system of your local machine, use one of the following methods to connect to an ECS instance.

Connect to a Linux instance

The following table details different methods by which to remotely connect to a Linux instance.

Is Internet access required ?	Operating system of the local machine	Connection method
Yes/No	Windows or Unix-like OS	Connect to an instance by using the Management Terminal.
Yes	Windows	Use a remote connection tool to create remote connection: <ul style="list-style-type: none">• Use an SSH key pair as the credential. For details, see connect to a Linux instance by using an SSH key pair.• Use a password as the credential. For details, see connect to a Linux instance by using a password.

Is Internet access required ?	Operating system of the local machine	Connection method
Yes	Linux, Mac OS, or other Unix-like OS	Use commands to create remote connection: <ul style="list-style-type: none"> • Use an SSH key pair as the credential. For details, see connect to a Linux instance by using an SSH key pair. • Use a password as the credential. For details, see connect to a Linux instance by using a password.
Yes	iOS or Android	User apps, such as SSH Control Lite or JuiceSSH, to create remote connection. For details, see connect to an instance on a mobile device .

Connect to a Windows instance

The following table details different methods by which to remotely connect to a Windows instance.

Is Internet access required ?	Operating system of the local machine	Connection method
Yes/No	Windows or Unix-like OS	Connect to an instance by using the Management Terminal .
Yes	Windows	Use mstsc to create remote connection. For details, see connect to a Windows instance .
Yes	Linux	Use a remote connection tool, such as rdesktop, to create remote connection. For details, see connect to a Windows instance .

Is Internet access required ?	Operating system of the local machine	Connection method
Yes	Mac OS	Use Microsoft Remote Desktop Connection for Mac to create remote connection. For details, see connect to a Windows instance .
Yes	iOS or Android	Use Microsoft Remote Desktop to create a remote connection. For details, see connect to an instance on a mobile device .

7.2 Connect to Linux instances

7.2.1 Connect to a Linux instance by using the Management Terminal

This topic describes how to connect to a Linux ECS instance by using the Management Terminal (also known as VNC) in the ECS console and how to complete some related operations.

Scenarios

You can access your Linux instance by using the Management terminal in the ECS console when other remote access software (such as PuTTY, Xshell, or SecureCRT) does not work properly

The Management Terminal can be used to:

- Check the status of your instance if it boots slowly (for example, self-check upon startup).
- Reconfigure your instance (for example, disable the firewall) if a remote connection fails due to incorrect settings.
- Terminate abnormal processes that consume excessive CPU or bandwidth.



Note:

The Management Terminal can be used to connect to an instance even if no public IP address is assigned to your instance.

Prerequisites

- An ECS instance is created.
- The logon password for the ECS instance is set. If not, you can [reset the password](#).

Procedure

The following figure illustrates how to use the Management Terminal to connect to an ECS instance.

To connect to the ECS instance by using the Management Terminal, follow these steps :

1. In the instance list, find your instance and then, in the Actions column, click **Connect**.
2. In the Management Terminal page, follow the instructions to connect to the Management Terminal:
 - If you log on as an Alibaba Cloud account to connect to the Management Terminal for the first time, follow these steps:
 - a. In the VNC Connection Password dialog box, copy the password and click **Close**.



Note:

- The VNC password appears only once. You must save the password immediately and store it securely for future use. If you need to change the VNC password, see [change the VNC connection password](#).

- If you log on as a RAM user to connect to the Management Terminal for the first time, you will not see this dialog box.
- b. In the Enter VNC Password dialog box, paste the VNC connection password that you have copied, and click OK.
- If you log on as a RAM user to connect to the Management Terminal for the first time, or if you have forgotten your VNC connection password, follow these steps:
 - [Change the VNC connection password](#).
 - In the upper-left corner of the Management Terminal page, select Send Remote Command > Connect to Management Terminal.
 - In the Enter VNC Password dialog box, enter the new password and click OK.
 - If this is not your first connection to the Management Terminal, enter the VNC connection password in the Enter VNC Password dialog box and click OK.
3. To log on to the ECS instance, follow these steps according to the operating system:
- For a Linux instance: Enter the user name (`root`) and the logon password.

**Note:**

- The logon password input is invisible.
- If you want to perform additional operations within the instance, in the upper-left corner of the Management Terminal page, choose Send Remote Command > CTRL + ALT + Fx, of which Fx can be any key from F1 to F10, to switch the interfaces for different operations.
- If a black screen prompts, the Linux instance may be in sleep mode. To exit sleep mode, click your mouse or press any key.

Change the VNC connection password

If you want a simple password or forget about your password, follow these steps to change the password:

**Note:**

To connect to a non-I/O-optimized instance, you must restart your instance in the ECS console to activate the new VNC password. The restart operation stops your instance and interrupts your services. Therefore, proceed with caution.

1. Open the Management Terminal page.
2. Close the VNC Password or Enter VNC Password dialog box that displays.

3. In the upper right corner of the Management Terminal page, click Modify VNC Password.
4. In the Modify VNC Password dialog box that displays, enter a new password, and click OK to close the dialog box.
5. Activate the new password:
 - For an I/O-optimized instance, the new password takes effect immediately.
 - For a non-I/O-optimized instance, [restart the instance](#) in the ECS console.

Input commands

To connect to a Linux instance, you can use the Input Commands function to enter long texts, such as a complex command or a URL.

1. Open the Management Terminal page.
2. In the upper right corner of the Management Terminal page, click Input Commands.
3. In the Input Commands dialog box that displays, enter a command, and click OK to copy the command to the command line interface of your Linux instance.

References

- If you are using a PC, following instructions in
 - [#unique_213](#).
 - [#unique_214](#).
- If you are using a mobile device, following instructions in [#unique_215](#).

7.2.2 Connect to a Linux instance by using an SSH key pair

This topic describes how to use an SSH key pair to log on to a Linux instance in the Windows and Linux operating systems.

Prerequisites

Before you use an SSH key pair to connect to your Linux instance, the following conditions must be met:

- An SSH key pair is created in the ECS console and the `.pem` private key file is downloaded. For more information, see [Create an SSH key pair](#).

- An SSH key pair is linked with a Linux instance. You can [allocate a key pair when creating an ECS instance](#) or [attach an SSH key pair](#) to an existing instance.

**Note:**

If an ECS instance is in the Running state, you must restart it in the ECS console or by using the API to activate the key pair after attaching it to the instance.

- Security group rules are added to the security groups to which the instance belongs. For more information, see [#unique_224](#). The following table describes the security group rules.

Network type	NIC type	Rule direction	Authorization policy	Protocol type	Port range	Authorization type	Authorization object	Priority
VPC	Not required	Inbound	Allow	SSH (22)	22/22	IP address segment-based access	0.0.0.0/0	1
Classic network	Internet							

Local Windows OS

A private key file in `.pem` format is automatically generated after you create an SSH key pair in the ECS console. This section describes how to use PuTTYgen to convert the `.pem` private key file and how to use PuTTY to log on to a Linux instance through the SSH remote access tool on a Windows OS.

1. Download and install PuTTYgen and PuTTY.

The download links are as follows:

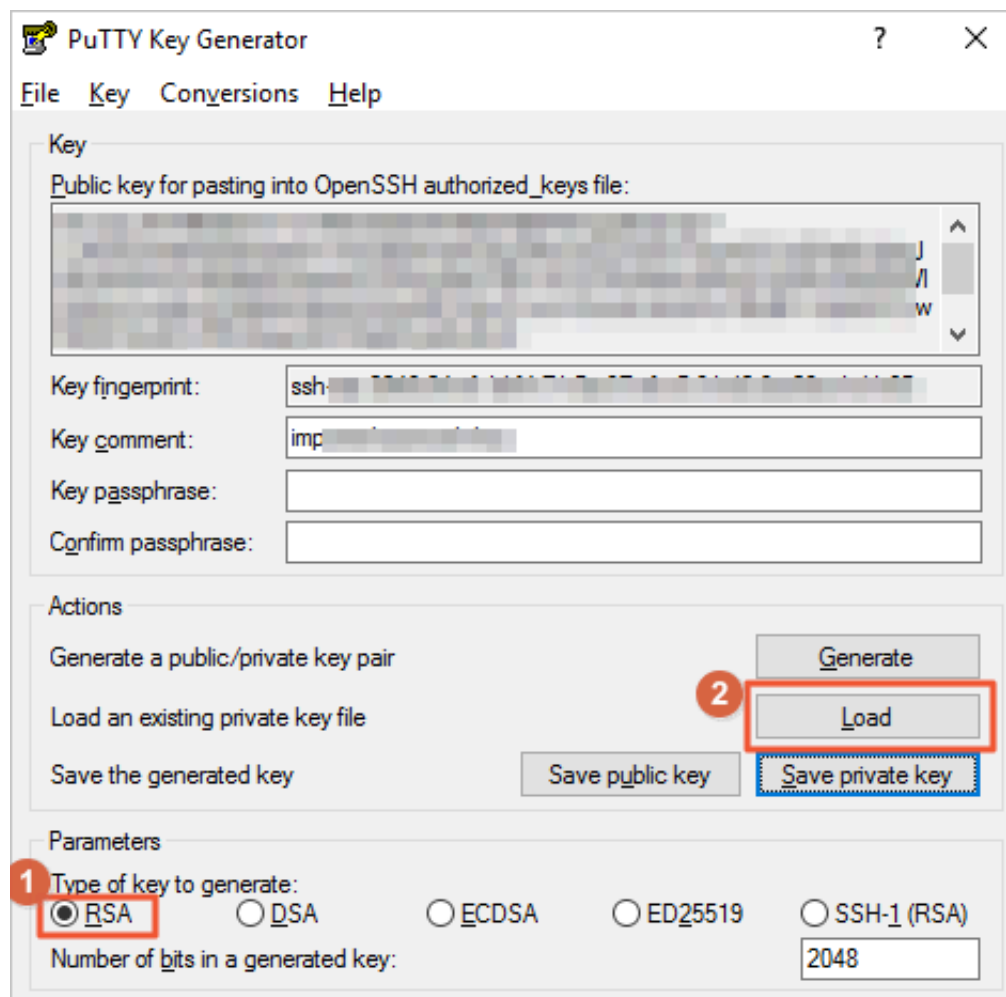
- [PuTTYgen](#)
- [PuTTY](#)

2. Convert the . pem private key file to a . ppk key file.

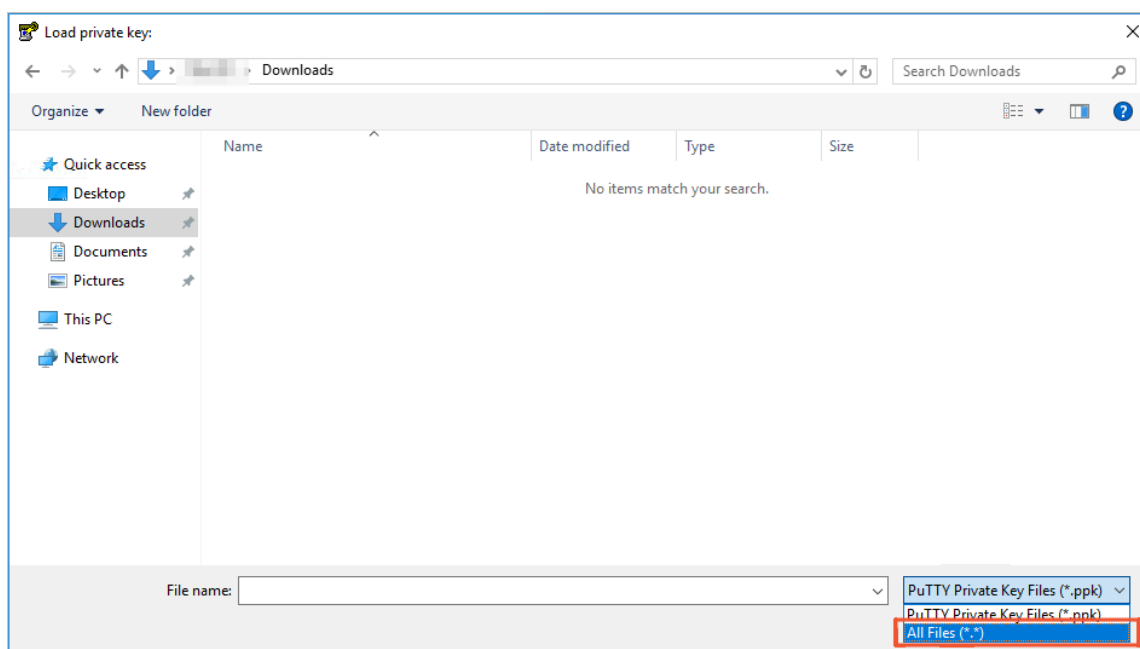
a. Start PuTTYgen.

PuTTYgen 0.71 is used as an example.

b. In the Parameters area, set Type of key to generate to RSA and then click Load.

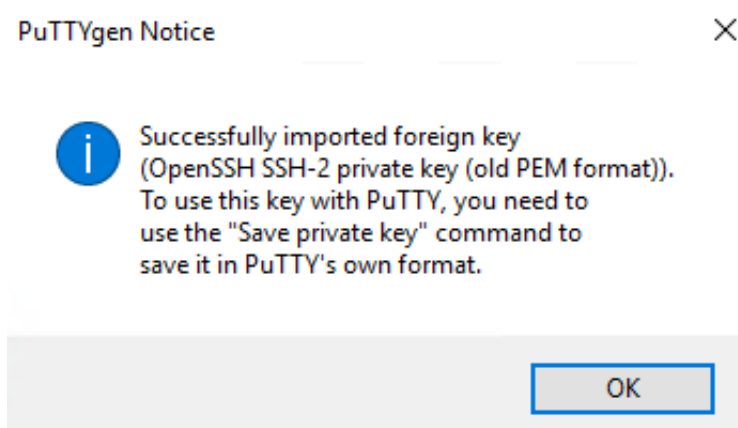


c. Select All Files (*.*) from the drop-down list.

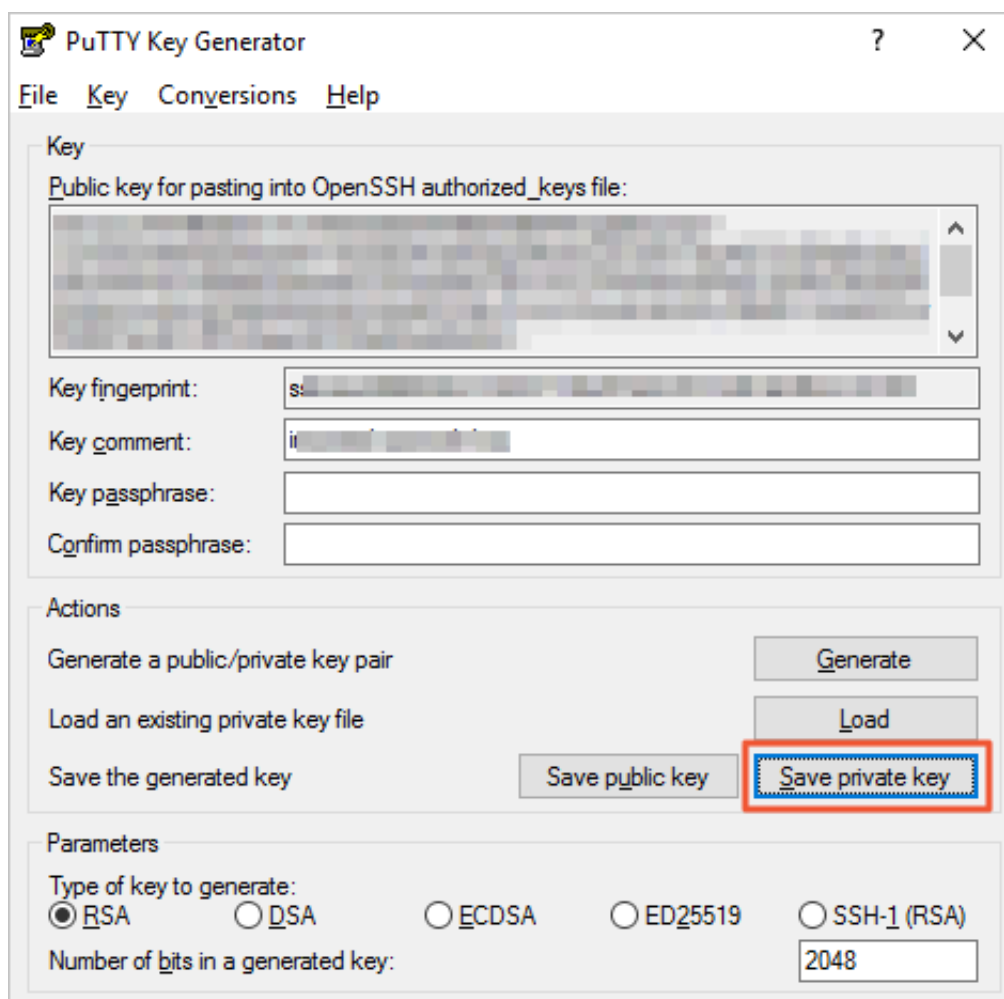


d. Select the .pem private key file to be converted.

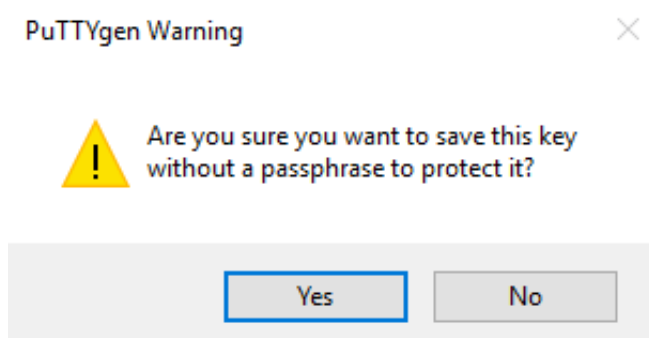
e. In the PuTTYgen Notice dialog box, click OK.



f. Click Save private key.



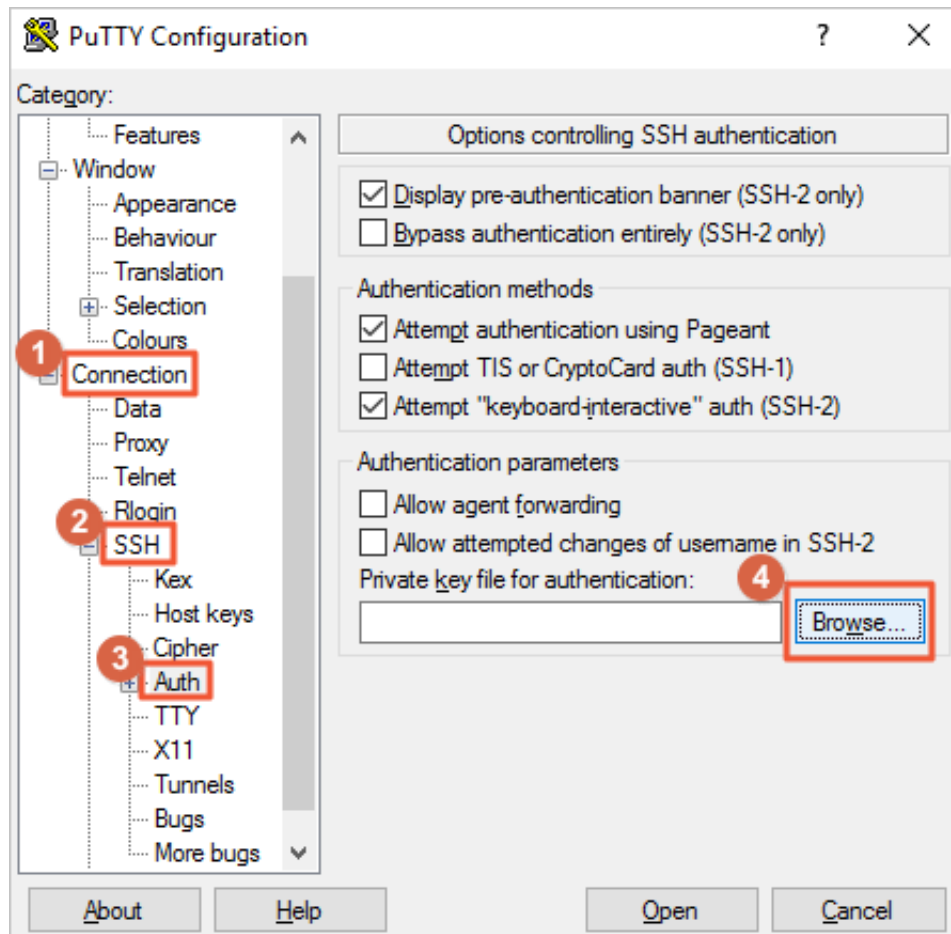
g. In the PuTTYgen Warning dialog box, click Yes.



h. Specify the name of the `.ppk` file and click Save.

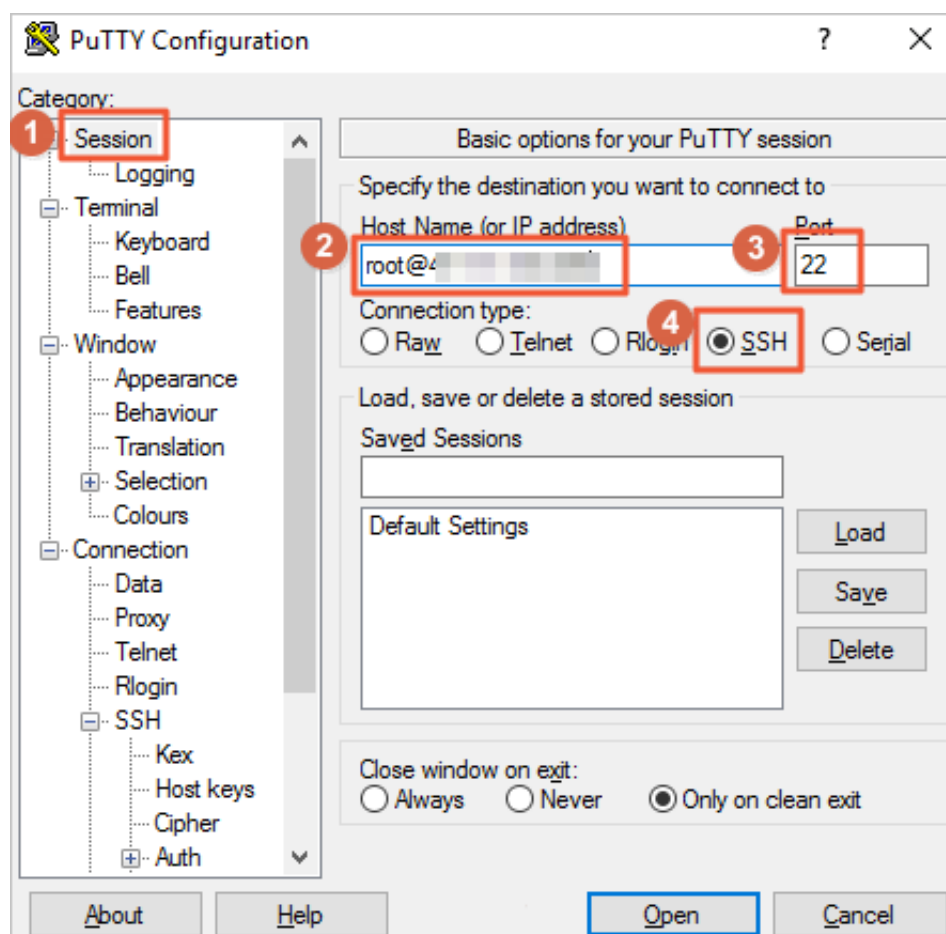
3. Start PuTTY.

4. Configure the private file used for authorization.
 - a. In the left-side navigation pane, choose Connection > SSH > Auth.
 - b. In the right pane, click Browse...
 - c. Select the . *ppk* private key file.



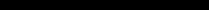
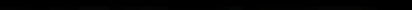
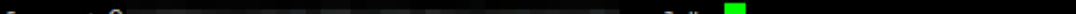

5. Configure information required for connecting to the Linux instance.

- a. In the left-side navigation pane, click Session.
- b. In the right pane, enter your account and the Internet IP address of the instance to be connected in the Host Name (or IP address) text box. The format is root @ IP address .
- c. In the Port text box, enter the port number 22 .
- d. Set Connection type to SSH.



6. Click Open to start connecting to your Linux instance.

When the following message is displayed, you have successfully logged on to the instance by using the SSH key pair.

```
Using username "root".  
Authenticating with public key "  
Last login:    
  
[root@ ~]#
```

Local Linux OS or other OSs supporting SSH commands

This section describes how to use an SSH key pair to log on to a Linux instance on a Linux OS or an OS supporting SSH commands, for example, Windows MobaXterm.

- Configure required information and connect to the Linux instance by using commands.
 1. Find the directory for saving the . pem private key file on your PC, for example, / root / mysshkey . pem .
 2. Run the following command to modify the attribute of the private key file:

```
chmod 400 [ Directory for saving the .pem private
key file on your local PC ]
```

For example,

```
chmod 400 / root / mysshkey . pem
```

- 3. Run the following command to connect to the instance:**

```
ssh -i [ Directory for saving the .pem private
key file on your local PC ] root @[ Internet IP
address ]
```

For example,

```
ssh -i /root/mysshkey.pem root@10.10.xx.xxx
```

- Configure required information by using the `config` file and connect to the instance by running commands.

1. Go to the `ssh` directory in the root directory and do the following to modify the `config` file:

```
Host     ecs      // Set the name of your ECS instance
HostName 192 . *. *. * // Enter the Internet IP
address  of your ECS instance .
```

```
Port 22 / Enter the port number , which is 22
by default .
User Root // Enter your logon account .
IdentityFile ~/.ssh / ecs . pem // Enter the directory
for saving the . pem private key file on your
PC .
```

2. Save the *config* file

3. Restart SSH.

4. Run `ssh [ECS name]` to connect to your ECS instance, for example, `ssh ecs`.

References

For information about how to connect to a Linux instance, see [#unique_214](#) and [#unique_72](#).

7.2.3 Connect to a Linux instance by using a password

You can connect to a Linux instance by using different authentication methods:

- If you are using an SSH key pair, see [connect to a Linux instance by using an SSH key pair](#).
- If you are using a password, you can [connect to an instance by using the Management Terminal](#) or by using software applications or command lines.

Prerequisites

- The instance must be in the Running status. If not, [start it](#).
- You have set a logon password for the instance. If the password is lost, you can [reset the password](#).
- The instance can access the Internet:
 - In a VPC, a public IP address is assigned to the instance or [an EIP address is bound to the instance](#).
 - In the classic network, a public IP address is assigned to the instance by using either of the following methods:
 - For a Subscription or a Pay-As-You-Go instance, you can select Assign public IP when creating the instance.
 - For a Subscription instance without a public IP address, you can assign one by [upgrading the bandwidth](#).

- The following security group rules must be added to the security group that the instance joins. For more information, see [add security group rules](#).

Network type	NIC	Rule direction	Authorization policy	Protocol type	Port range	Authorization type	Authorization object	Priority
VPC	N/A	Inbound	Allow	SSH (22)	22/22	Address Field Access	0.0.0.0/0	1
Classic	Internet							

Procedure

Based on the operating system of your local machine, use one of the following methods to connect to a Linux instance by using the SSH protocol:

- [Windows OS](#)
- [Linux or Mac OS X](#)
- [Android or iOS](#)

Windows OS

If your local machine is running Windows OS, you can use a remote connection tool, such as PuTTY, to connect to a Linux instance. In this article, we use PuTTY as an example to describe how to connect to a Linux instance by using the password authentication method. Before you start, download [PuTTY](#).

Follow these steps to connect to a Linux instance:

1. Start putty.exe.
2. In the left-side navigation pane, click **Session** , and configure the following parameters:
 - **Host Name** : Type the public IP address or EIP address of the instance.
 - **Port** : Type 22 .
 - **Connection Type** : Select SSH .
 - (Optional) **Saved Session** : If you do not want to repeat the configurations during the next logon, add a name for the session, and click Save.
3. Click Open to connect, and in the PuTTY Security Alert dialog box, click Yes.



Note:

For the first connection to an ECS instance, you have the PuTTY Security Alert as follows, which means PuTTY cannot guarantee the instance is the one that you think it is, so it can only provide the public key fingerprint of the instance for you to decide to trust the instance or not. If you select Yes, the public key will be added to the PuTTY's cache and you will not be alerted again during your next connection. If you select Yes but are alerted again, a [man-in-the-middle attack \(MITM\)](#) may occur. For more information, see [PuTTY User Manual](#).

4. Enter the user name and password for the Linux instance, and then press Enter.



Note:

The password is not displayed on screen.

If you are successfully connected to the instance, the following message is displayed.

```
Welcome to Alibaba Cloud Elastic Compute Service !
```

Linux or Mac OS X

If your local machine is running Linux OS or Mac OS X, follow these steps:

1. Run the command `ssh root @[Public IP address or EIP address of the instance]`.
2. Type the password and then press Enter.

If you are successfully connected to the instance, the following message is displayed.

```
Welcome to Alibaba Cloud Elastic Compute Service !
```

Android or iOS

If your local machine is running Android OS or iOS, see [connect to an instance on a mobile device](#).

Reference

You can run a script to install a graphical desktop on an instance running CentOS. For more information, see [automatic installation tool for Linux instance](#).

7.2.4 Connect to an instance on a mobile device

This article describes how to connect to an ECS instance on a mobile device.

Depending on the operating system of your instance, select the required method as follows.

- [Connect to a Linux instance](#): This example uses SSH Control Lite to describe how to connect to a Linux instance on an iOS device, and JuiceSSH to describe how to connect to a Linux instance on an Android device.
- [Connect to Windows instances](#): This example uses Microsoft Remote Desktop to describe how to connect to a Windows instance on an iOS or Android device.

Connect to a Linux instance

Prerequisites

- The instance is Running .
- The instance has a public IP address and is accessible from the Internet.
- You have set the logon password for the instance. If the password is lost, you can [reset the instance password](#).
- The security group of the instance has the [the following security group rules](#):

Network type	NIC	Rule direction	Authorization policy	Protocol type	Port range	Authorization type	Authorization object	Priority
VPC	No configuration required	Inbound	Allow	SSH(22)	22/22	Address Field Access	0.0.0.0/0	1
Classic	Internet							

- You have downloaded and installed the appropriate app:
 - For an iOS device, install SSH Control Lite.
 - For an Android device, install JuiceSSH.

Procedure

For iOS devices, see [use SSH Control Lite to connect to a Linux instance](#). In this example, a user name and password are used for authentication.

For Android devices, see [use JuiceSSH to connect to a Linux instance](#). In this example, a user name and password are used for authentication.

Use SSH Control Lite to connect to a Linux instance

1. Start SSH Control Lite, and tap Hosts.
2. Tap the + icon in the upper left corner of the Hosts page.
3. In the action sheet, tap Connection.
4. On the Connection page, set the connection information and tap . The following connection information is required:
 - Name : Specify the Host name. DocTest is used in this example. .
 - Protocol : Use the default value SSH.
 - Host : Type the public IP address of the Linux instance to connect to.
 - Port : Type the port number for SSH protocol. 22 is used in this example.
 - Username : Type root for the user name.
 - Password : Type the logon password of the instance.
5. In the tool bar, tap Remote Controls.
6. On the Remote Controls page, tap the + icon in the upper left corner to create a remote connection session. New remote is used in this example.

The following figure shows Steps 1 through 6.

7. On the New remote page, tap Host1.
8. In the action sheet, tap Bind.
9. Select the new Linux instance. In this example, select DocTest .
10. On the New remote page, tap Done to switch it to the Edit mode, and then tap DocTest.
11. In the action sheet, tap Connect.

The following figure shows Steps 7 through 11.


12. In the action sheet, select Yes, Once or Yes, Permanently. Once the connection is successful, the indicator in front of DocTest turns green.
13. On the New remote page, tap DocTest.
14. In the action sheet, tap Console to open Linux instance console.

The following figure shows Steps 12 through 14:

Use JuiceSSH to connect to a Linux instance

1. Start JuiceSSH, and tap Connections.

2. Under the Connections tab, tap the + icon.

3. On the New Connection page, add the connection information and then tap the 

icon. The following connection information is required:

- Nickname : Specify the name of the connection session. DocTest is used in this example.
- Type : Use the default value SSH.
- Address : Type the public IP address of the Linux instance to connect to.
- To set Identity , follow these steps:

a. Tap Identity, and tap New in the drop-down list.

b. On the New Identity page, add the connection information and then tap the



icon. The following connection information is required:

- Nickname : Optional. DocTest is used in this example.
 - Username : Type root for the user name.
 - Password : Tap SET(OPTIONAL), and type the logon password of the instance.
-
- Port : Type the port number for SSH protocol. In this example, 22 is used.

4. Confirm the message, and tap ACCEPT.

5. (Optional) For a first-time connection, the app will show a prompt of helpful tips.

Tap OK - I'VE GOT IT!.

If you are successfully connected to the Linux instance, the following screen is displayed.

Connect to Windows instances

In this section, Microsoft Remote Desktop is used as an example to describe how to use an app to connect to a Windows instance on a mobile device.

Prerequisites

- The instance is Running.
- The instance has a public IP address and is accessible from the Internet.
- You have set the logon password for the instance. If the password is lost, you must [reset the instance password](#).
- The security group of the instance has [the following security group rules](#):

Network type	NIC	Rule direction	Authorization policy	Protocol type	Port range	Authorization type	Authorization object	Priority
VPC	No configuration required	Inbound	Allow	RDP(3389)	3389/3389	Address field access	0.0.0.0/0	1
Classic	Internet							

- You have downloaded and installed Microsoft Remote Desktop.
 - For iOS devices, download the app from iTunes.
 - For Android devices, download the app from Google Play.

Procedure

To connect to a Windows instance by using Microsoft Remote Desktop, follow these steps:

1. Start RD Client. In the navigation bar, tap the + icon.
2. On the Add New page, select Desktop.

3. On the Edit Desktop page, type the connection information and tap Save. The following connection information is required:

- **PC Name** : Type the public IP address of the Windows instance to connect to.
- **User Account** : Type the account name administrator and the logon password of the Windows instance.

4. On the Remote Desktop page, tap the icon of a Windows instance.

5. On the confirmation page, confirm the message and tap Accept.

If you are successfully connected to the Windows instance, the following screen is displayed.

7.3 Connect to Windows instances

7.3.1 Connect to a Windows instance by using the Management Terminal

This topic describes how to connect to a Windows instance by using the Management Terminal (also known as VNC) in the ECS console and how to complete some related operations.

Scenarios

You can access your Windows instance by using the Management terminal in the ECS console when other remote access software (such as PuTTY, Xshell, or SecureCRT) does not work properly.

The Management Terminal can be used to:

- Check the status of your instance if it boots slowly (for example, self-check upon startup).
- Reconfigure your instance (for example, disable the firewall) if a remote connection fails due to incorrect settings.
- Terminate abnormal processes that consume excessive CPU or bandwidth.

Prerequisites

- An ECS instance is created.
- The logon password for the ECS instance is set. If not, you can [reset the password](#).

Procedure

The following figure illustrates how to use the Management Terminal to connect to an ECS instance.

1. Select the target region.
2. In the instance list, find the target instance. Then, click **Connect** in the **Actions** column.
3. Connect to the Management Terminal:
 - If you connect to the Management Terminal with the master account for the first time, follow these steps:
 - a. In the VNC Connection Password dialog box, copy the password.



Note:

- The VNC password appears only once. You must save the password immediately and store it securely for future use.

- If you connect to the Management Terminal as a RAM user for the first time, you will not see this dialog box.

- b. Click Close.
 - c. In the Enter VNC Password dialog box, paste the VNC connection password that you have copied, and click OK to connect to the Management Terminal.
- If you connect to the Management Terminal as a RAM user for the first time, or if you have forgotten your VNC connection password, follow these steps:
 - a. [Change the VNC connection password](#).
 - b. In the upper left corner of the Management Terminal page, select Send Remote Command > Connect to Management Terminal.
 - c. In the Enter VNC Password dialog box, enter the new password.
 - d. Click OK to connect to the Management Terminal.
 - If you connect to the Management Terminal again as a RAM user or by using the master account, enter the password in the Enter VNC Password dialog box, and click OK to connect to the Management Terminal.
4. Enter the user name and password to log on to the ECS instance.

**Note:**

In the upper left corner of the Management Terminal page, choose Send Remote Command > CTRL+ALT+DELETE to enter the logon interface of your Windows instance.

Change the VNC connection password

If you want a simple password or forget about your password, follow these steps to change the password:

**Note:**

To connect to a non-I/O-optimized instance, you must restart your instance in the ECS console to activate the new VNC password. The restart operation stops your instance and interrupts your services. Therefore, proceed with caution.

1. Open the Management Terminal page.
2. Close the VNC Password or Enter VNC Password dialog box that displays.
3. In the upper right corner of the Management Terminal page, click Modify VNC Password.

4. In the Modify VNC Password dialog box that displays, enter a new password, and click OK to close the dialog box.
5. Activate the new password:
 - For an I/O-optimized instance, the new password takes effect immediately.
 - For a non-I/O-optimized instance, [restart the instance](#) in the ECS console.

Input commands

To connect to a Windows instance, you can use the Input Commands function to enter long texts, such as a complex command or a URL.

1. Open the Management Terminal page.
2. In the upper right corner of the Management Terminal page, click Input Commands.
3. In the Input Commands dialog box that displays, enter a command, and click OK to copy the command to the command line interface of your Windows instance.

References

- If you are using a PC, following instructions in [Connect to a Windows instance](#).
- If you are using a mobile device, following instructions in [Connect to a Windows instance on a mobile device](#).

7.3.2 Connect to a Windows instance

If your Windows instance can access the Internet, you can use remote connection tools to connect to it. Otherwise, you can use the [Management Terminal](#).

Prerequisites

- The instance is in the Running status. If not, [start it](#).
- You have set a logon password for the instance. If the password is lost, you can [reset the password](#).

- The instance can access the Internet:
 - In a VPC, a public IP address is assigned to the instance or [an EIP address is bound to the instance](#).
 - In the classic network, a public IP address is assigned to the instance by using either of the following methods:
 - For a Subscription or a Pay-As-You-Go instance, you can select Assign public IP when creating the instance.
 - For a Subscription instance without a public IP address, you can assign one by [upgrading bandwidth](#).
- The following security group rules must be added to the security group that the instance joins. For more information, see [add security group rules](#).

Network Type	NIC	Rule Direction	Authorization Policy	Protocol Type	Port Range	Authorization Type	Authorization Object	Priority
VPC	N/A	Inbound	Allow	RDP(3389)	3389/3389	Address Field Access	0.0.0.0/0	1
Classic	Internet							

Procedure

Based on the operating system of your local machine, use one of the following methods to connect to a Windows instance:

- [Windows OS](#)
- [Linux](#)
- [Mac OS](#)
- [Android or iOS](#)

Windows OS

If the local machine is running Windows OS, you can use the mstsc to create a remote connection to a Windows instance.

1. Use any one of the following methods to start mstsc:
 - Select Start > icon > Remote Desktop Connection.
 - Click the Start icon and search for mstsc.
 - Press the Windows key + R to open the Run window, type `mstsc` , and then press Enter.
2. In the Remote Desktop Connection dialog box, follow these steps:
 - a. Click the Show Options drop-down box.
 - b. Type the public IP address or EIP address of the instance.
 - c. Type the user name. The default user name is `Administrator`

**Note:**

If you want to log on to the instance next time using the same credentials, select **Allow me to save credentials**.

- d. Optional. If you want to copy text or files from the local machine to the instance, click the Local Resources tab to see options for sharing local computer resources.
 - If you want to copy text only, select Clipboard.
 - If you also want to copy files, select More and select the drive letters from which you want to copy files to your instance and click OK.
- e. Optional. Click the Display tab to resize the remote desktop window. Full Screen is recommended.
- f. Click Connect.

Linux

If the local machine is running Linux OS, you can use a remote connection tool to create a remote connection to a Windows instance. This article takes `rdesktop` as an example to describe how to connect a Windows instance from a local machine running Linux.

1. Download and start rdesktop.
2. Run the command to connect to a Windows instance. Replace the parameter values with your own configurations.

```
rdesktop -u administra tor -p password -f -g 1024 *
720 192 . 168 . 1 . 1 -r clipboard : PRIMARYCLI PBOARD -r
disk : sunray =/ home / yz16184
```

The following table describes the parameters involved.

Parameters	Description
-u	The user name. The default user name for a Windows instance is Administrator .
-p	The password used to log on to the windows instance.
-f	Full screen by default. Use Ctrl+Alt+Enter to switched the mode.
-g	Resolution. Asterisks (*) are used for separation. If omitted, full-screen display is used by default.
192.168.1.1	The IP address of the server that requires remote connection. Replace it with the public IP or EIP address of your Windows instance.
-d	Domain name. For example, if the domain name is INC, then the parameter is - d inc .
-r	Multimedia reorientation. For example: <ul style="list-style-type: none"> • Turn on the sound:-r sound . • Use a local sound card:-r sound: - r sound : local . • Open the U Disk: - r disk : usb =/ mnt / usbdevice .
-r clipboard:PRIMARYCLIPBOARD	Realizes direct word copying and pasting between Linux and Windows instances of local devices. Supports Chinese words copying and pasteing.

Parameters	Description
<code>-r disk:sunray=/home/yz16184</code>	Specifies that a directory on the Linux system of a local device maps to a hard disk on a Windows instance. If this is configured, Samba and FTP are not recommended for file transfers.

For more information about parameters of the `rdesktop` command, see [rdesktop documentation](#).

Mac OS

To connect to a Windows instance from a local machine running Mac OS, see [get started with Remote Desktop on Mac](#).

Android or iOS

If your local machine is running Android OS or iOS, see [connect to an instance on a mobile device](#).

7.3.3 Connect to a Windows instance from a mobile device

This topic describes how to connect to a Windows instance from a mobile device (iOS or Android) by using Microsoft Remote Desktop.

Prerequisites

- The instance is in the Running state.
- The instance has a public IP address and is accessible from the Internet.
- The logon password for the instance is set. Moreover, if the password was lost, you can [reset the instance password](#).
- The security group of the instance has [the following security group rules](#):

Network type	NIC	Rule direction	Authorization policy	Protocol type	Port range	Authorization type	Authorization object	Priority
VPC	Configuration not required	Inbound	Allow	SSH (22)	22/22	CIDR block	0.0.0.0/0	1
Classic	Internet							

Procedure

Check that you have installed Microsoft Remote Desktop (RD).

1. Start the RD Client. In the upper right corner, click +.
2. On the Add New page, select Desktop.
3. On the Edit Desktop page, enter the connection information and click Save. The following connection information is required:
 - PC Name: Enter the public IP address of the Windows instance to be connected.
 - User Account: Enter the account name administrator and the logon password of the Windows instance.
4. On the Remote Desktop page, click the icon of the target Windows instance.
5. On the confirmation page, confirm the message and click Accept.

If you have successfully connected to the Windows instance, the following screen is displayed.

8 Manage instances

8.1 Start or stop an instance

This topic describes how to start or stop an ECS instance in the ECS console. It also describes operations related to the No fees for stopped instances (VPC-Connected) feature.

Prerequisites

- The instance you want to start must be in the Stopped state.
- The instance you want to stop must be in the Running state.

Start an instance

To start an instance, follow these steps:

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Instances.
3. In the top navigation bar, select a region.
4. Find the instance to be started and, in the Actions column, choose More > Instance Status > Start.



Note:

If you want to start multiple Stopped instances, select the required instances and then, under the instance list, click Start.

5. Read and confirm you agree to the note displayed in the dialog box by clicking OK.

After the instance is started, it enters the Running state.

You can also start an instance by calling the [StartInstance](#) API action through the Alibaba Cloud CLI, OpenAPI Explorer, or Alibaba Cloud SDK.

Stop a Subscription instance



Note:

Stopping an instance disrupts services. Exercise caution when performing this action.

To stop an instance, follow these steps:

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Instances.
3. In the top navigation bar, select a region.
4. Find the instance to be stopped and, in the Actions column, choose More > Instance Status > Stop.

**Note:**

If you want to stop multiple Running instances, select the required instances and then, under the instance list, click Stop.

5. Read and confirm you agree to the note displayed in the dialog box by clicking OK.
6. In the Stop Instance dialog box, set Stop Mode and click OK.

After the instance is stopped, it enters the Stopped state.

Stop a Pay-As-You-Go instance

**Note:**

Stopping an instance disrupts services. Exercise caution when performing this action. The procedure for stopping a preemptible instance is the same as that for a Pay-As-You-Go instance. For more information, see [#unique_249](#).

Stopping a Pay-As-You-Go instance may affect instance billing. The impact is determined by the network type and the No Fees for Stopped Instances (VPC-Connected) feature:

- **Classic network:** A Pay-As-You-Go instance in the classic network does not support the No Fees for Stopped Instances (VPC-Connected) feature and the instance still incurs fees after it is stopped. The billing stops only after you [release](#) the instance.
- **VPC:** A Pay-As-You-Go instance in the VPC network supports the No Fees for Stopped Instances (VPC-Connected) feature:
 - If this feature is not enabled, billing continues after the instance is stopped.
 - If this feature is enabled, you can decide whether to keep the instance after you stop the instance by specifying Keep Stopped Instances and Continue Billing. If you do not want to keep the instance, resources such as vCPUs, memory, and Internet IP addresses stop incurring fees. However, you are still billed for other resources. For more information, see [#unique_142](#).

To stop a Pay-As-You-Go instance, follow these steps:

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Instances.
3. In the top navigation bar, select a region.
4. Find the Pay-As-You-Go instance to be stopped and, in the Actions column, choose More > Instance Status > Stop.

**Note:**

If you want to stop multiple Running instances, select the required instances and then, under the instance list, click Stop.

5. Complete required actions according to the network type of the instance and whether No Fees for Stopped Instances (VPC-Connected) is enabled:
 - If the network type is classic or No Fees for Stopped Instances (VPC-Connected) is not enabled:
 - a. In the Stop Instance dialog box, set Stop Mode.
 - b. Click OK.
 - If No Fees for Stopped Instances (VPC-Connected) is enabled:
 - a. Read the note displayed in the Notes dialog box.
 - b. Confirm the note by clicking OK.
 - c. In the Stop Instance dialog box, set Stop Mode and Stopped By.

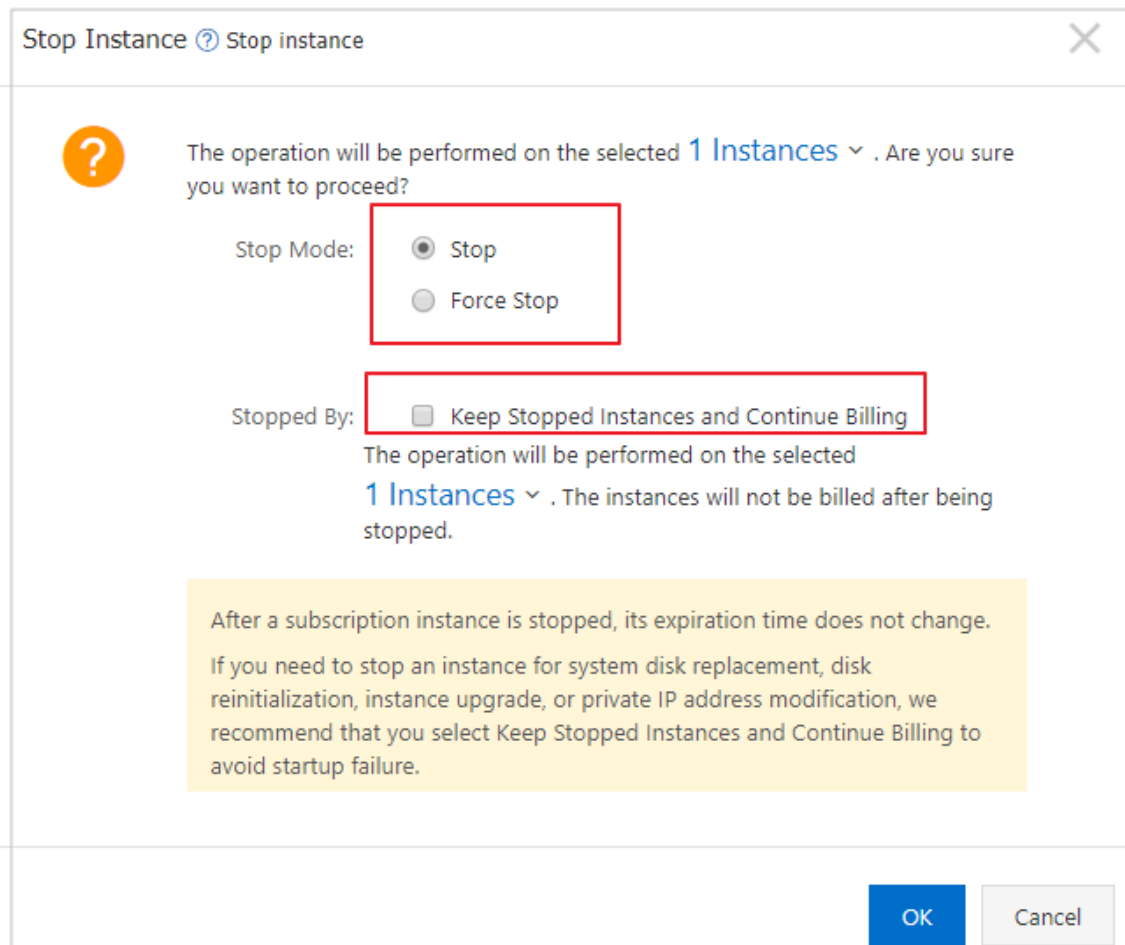
**Note:**

If Keep Stopped Instances and Continue Billing is selected, the instance continues to incur fees after being stopped. If Keep Stopped Instances and Continue Billing is not selected, the instance will not be billed after being stopped.

- d. Click OK.

**Note:**

To disable the No Fees for Stopped Instances (VPC-Connected) feature, see [#unique_142](#).



Stop Instance ? Stop instance

? The operation will be performed on the selected 1 Instances . Are you sure you want to proceed?

Stop Mode: ☒ Stop ☐ Force Stop

Stopped By: ☐ Keep Stopped Instances and Continue Billing

The operation will be performed on the selected 1 Instances . The instances will not be billed after being stopped.

After a subscription instance is stopped, its expiration time does not change. If you need to stop an instance for system disk replacement, disk reinitialization, instance upgrade, or private IP address modification, we recommend that you select Keep Stopped Instances and Continue Billing to avoid startup failure.

OK Cancel

The instance is in the Stopped state after you stop it.

You can also stop a Pay-As-You-Go instance by calling the [#unique_159](#) API action through the Alibaba Cloud CLI, OpenAPI Explorer, or Alibaba Cloud SDK.

8.2 Restart an instance

This topic describes how to restart an instance by using the ECS console.

Limits

- Only instances in the Running state can be restarted.
- Restarting an instance will stop the instance. As a result, services provided by the instance are disrupted.

Procedure

1. Find the target instance, and then choose **More > Instance Status > Restart** in the **Actions** column.

To restart multiple instances, select all required instances and then click **Restart** at the bottom of the instance list.

2. In the displayed **Restart Instance** dialog box, select a **Restart Mode**, and then click **OK**.

You can also call the [RebootInstance](#) API action to complete this task.

8.3 Release an instance

This topic describes how to release an ECS instance.

Precautions

You can release a Pay-As-You-Go instance (including a preemptible instance) when you no longer need it to avoid excess charges.



Note:

After an instance is released, its data cannot be recovered. We recommend that you [create a snapshot](#) to back up data before releasing an instance.

For a Pay-As-You-Go instance, if the [#unique_142](#) feature is not enabled, charges continue to incur until the instance is released.

For a Subscription instance, the instance is automatically released after the billing cycle expires.

You can [enable instance release protection](#) for a Pay-As-You-Go instance to prevent irreversible data loss resulting from accidental or incorrect operations during a manual release.

Release an instance immediately

To release an instance immediately, follow these steps:

1. Release the instance:

- If you want to release only one instance, find the instance that you want to release and then, in the Actions column, choose Manage > Release.
- If you want to release multiple instances, find the Pay-As-You-Go instances according to the Billing Method, select multiple instances to release, and click Release at the bottom of the list.

2. In the dialog box, select Release Now.**3. Click Next, and then click OK.**

You can also use the Alibaba Cloud CLI, OpenAPI Explorer, or SDK to call [DeleteInstance](#) to release the instance.

Enable automatic release

To enable automatic release, follow these steps:

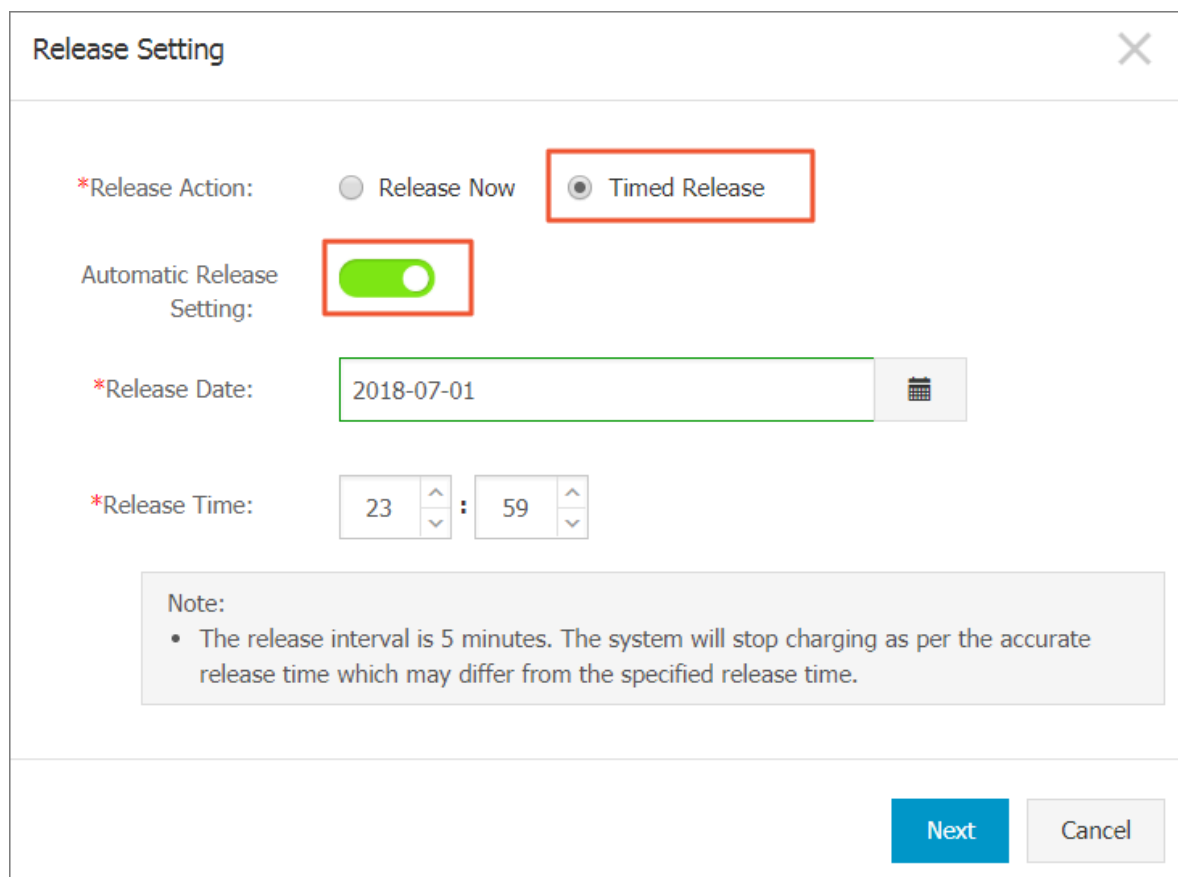
1. Release the instance:

- If you want to release only one instance, find the instance that you want to release, and in the Actions column, choose Manage > Release.
- If you want to release multiple instances, find the Pay-As-You-Go instances according to the Billing Method, select multiple instances to release, and click Release at the bottom of the list.

2. In the dialog box, select Scheduled Release.

3. Turn on the Automatic Release switch, and specify the release date and time.

The earliest setting can only be set to automatically release an instance after 30 minutes.



The image shows a 'Release Setting' dialog box with a close button (X) in the top right corner. It contains the following fields and controls:

- *Release Action:** Two radio buttons. 'Release Now' is unselected, and 'Timed Release' is selected and highlighted with a red box.
- Automatic Release Setting:** A green toggle switch is turned on and highlighted with a red box.
- *Release Date:** A text input field containing '2018-07-01' and a calendar icon to its right. The field is highlighted with a green box.
- *Release Time:** Two spinners showing '23' and '59' separated by a colon. Each spinner has up and down arrows.
- Note:** A gray box containing the text: 'Note: The release interval is 5 minutes. The system will stop charging as per the accurate release time which may differ from the specified release time.'
- Buttons:** 'Next' (blue) and 'Cancel' (gray) buttons at the bottom right.

4. Click Next, and then click OK.

You can also use the Alibaba Cloud CLI, OpenAPI Explorer, or SDK to call [ModifyInstanceAutoReleaseTime](#) to set the automatic release time for a Pay-As-You-Go instance.

Disable automatic release

If you want to cancel the automatic release schedule of a Pay-As-You-Go instance, you can disable the feature.

To disable the automatic release feature, follow these steps:

1. Release the instance:

- If you want to disable automatic release for an instance, find the instance and then, in the Actions column, choose Manage > Release.
- If you want to disable automatic release for multiple instances, find the Pay-As-You-Go instances according to the Billing Method, select the target instances, and click Release at the bottom of the list.

2. In the dialog box, select Scheduled Release.

3. Turn off the Automatic Release switch.

4. Click Next, and then click OK.

You can also use the Alibaba Cloud CLI, OpenAPI Explorer, or SDK to call [ModifyInstanceAutoReleaseTime](#) to disable automatic release for a Pay-As-You-Go instance.

8.4 Reactivate an instance

This topic describes how to reactivate an instance. After paying the overdue bill of a Pay-As-You-Go instance, you must reactivate the instance. Otherwise, the instance will be released.

Precautions

For a Pay-As-You-Go instance, if the due date of an overdue payment (T) is not settled within 15 days after the due date (T+15), the instance is stopped due to overdue payment and its status changes to Expired. You must submit a ticket to settle the payment and reactivate your instance within 30 days after the due date (T+30). Otherwise, the instance is released and the data cannot be recovered.



Note:

If you fail to reactivate the ECS instance within 30 days after the due date (T+30), the instance is automatically released 30 days after the due date and the data cannot be recovered.

Prerequisites

The Pay-As-You-Go instance is in the Expired or Expired and Being Recycled state.

You have settled the payment by opening a ticket.

Procedure

To reactivate an instance in the ECS console, follow these steps:

1. Select the instance to be reactivated, and then choose **More > Reactivate** at the bottom of the instance list.
2. Choose whether to reactivate the instance immediately or later at a specified time.

If you choose to reactivate immediately, the selected instance returns to the Running state after about 10 minutes.

Alternatively, you can complete this task by calling the [ReactivateInstances](#) API action through the Alibaba Cloud CLI, OpenAPI Explorer, or SDK.

8.5 Check instance information

This topic describes how to check instances under your account and details of a single instance.

Through the console, you can:

- [View all ECS instances under your account on the Overview page](#)
- [View the information of ECS instances on the Instances page](#)
- [View details of an ECS instance on Instance Details page](#)

View all ECS instances under your account on the Overview page

You can view information of all the ECS instances created by your account on the Overview page, including:

- Total number of ECS instances, and numbers of instances in each state.
- Number of resources in different regions and numbers of ECS instances in each state.

The homepage of the ECS console is the Overview page by default.

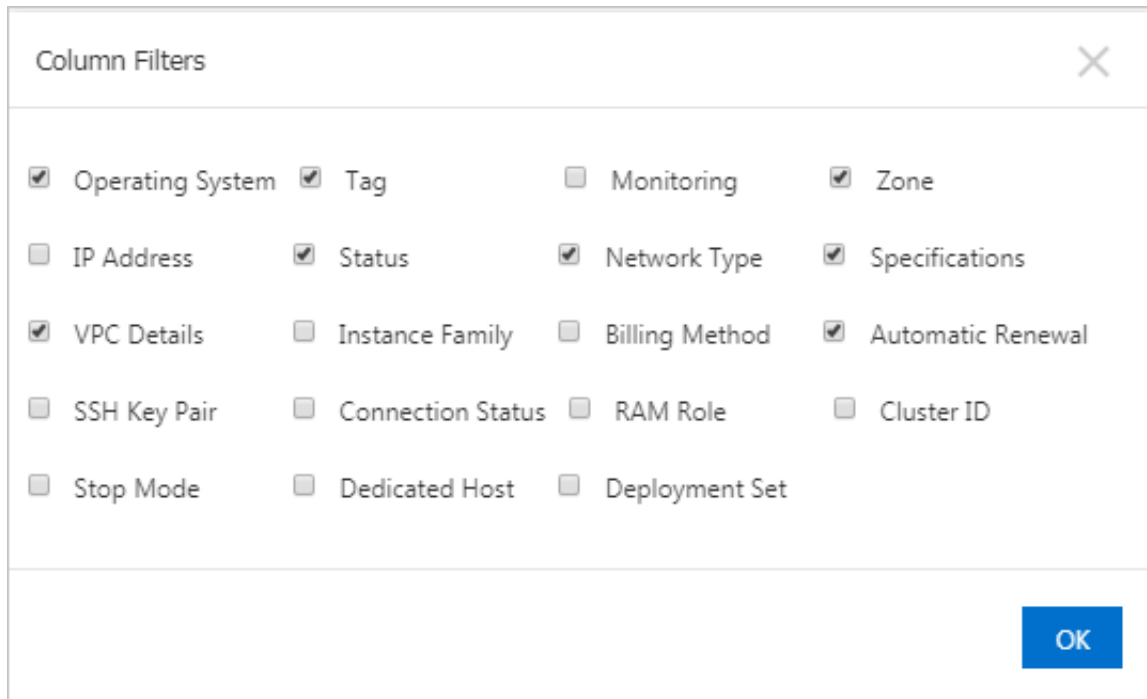
View the information of ECS instances on the Instances page

To navigate to the Instances page, follow these steps:

Here, you can see information of all the existing ECS instances in the selected region, including ECS instance ID/name, zone, IP addresses, status, network type, billing method, and actions. You can show or hide the displayed information of an instance by using the Column Filter feature.

1. In the upper-right corner of the Instance List, click the  icon.

2. Select the instance information to be displayed and then click OK.



The image shows a 'Column Filters' dialog box with a close button (X) in the top right corner. It contains a grid of checkboxes for various instance attributes. The attributes are arranged in five rows and four columns. The first four rows have four items each, and the fifth row has three items. The 'OK' button is a blue rectangle in the bottom right corner.

Column Filters			
<input checked="" type="checkbox"/> Operating System	<input checked="" type="checkbox"/> Tag	<input type="checkbox"/> Monitoring	<input checked="" type="checkbox"/> Zone
<input type="checkbox"/> IP Address	<input checked="" type="checkbox"/> Status	<input checked="" type="checkbox"/> Network Type	<input checked="" type="checkbox"/> Specifications
<input checked="" type="checkbox"/> VPC Details	<input type="checkbox"/> Instance Family	<input type="checkbox"/> Billing Method	<input checked="" type="checkbox"/> Automatic Renewal
<input type="checkbox"/> SSH Key Pair	<input type="checkbox"/> Connection Status	<input type="checkbox"/> RAM Role	<input type="checkbox"/> Cluster ID
<input type="checkbox"/> Stop Mode	<input type="checkbox"/> Dedicated Host	<input type="checkbox"/> Deployment Set	

OK

View details of an ECS instance on Instance Details page

The Instance Details displays detailed information of a selected ECS instance.

To navigate to the Instance Details page, follow these steps:

1. Find the ECS instance you want to view the details of, and then click its instance ID.

On the Instance Details page, you can view the following information:

- **Basic Information**, including the ECS instance ID, instance name, region, zone, instance type, instance type family, image ID, key pair name (applies to Linux instances only), instance RAM role, and tags.
- **Configuration Information**, including CPU, memory, I/O optimization, operating system, IP addresses, billing method for bandwidth, current bandwidth, and VPC network information (applies to VPC instances only).
- **Payment Information**, including billing method, the mode to stop an instance, creation time, and automatic release schedule (applies to Pay-As-You-Go instances only).
- **Monitoring Information**, including CPU and network usage.

You can also switch from the Instance Details page to the Disks, Instance Snapshots, or Security Groups page to view resources related to this instance.

8.6 Reset an instance logon password

This topic describes how to reset the logon password of one or more instances.

Limits

- After you reset the logon password of a running instance, you must restart the instance so that the new password can take effect. Restarting the instance may impact your services. We recommend that you reset the logon password and restart your instance during off-peak business hours to avoid service disruption.
- If your instance uses a password for authentication, the password authentication automatically becomes invalid after you [attach a key pair](#) to your instance. However, if you reset the password of your instance after you attach a key pair to the instance, you can log on to the instance by using the password or the key pair.

Prerequisites

The instance must be in a stable status, such as Stopped and Running. For more information, see [#unique_262](#).

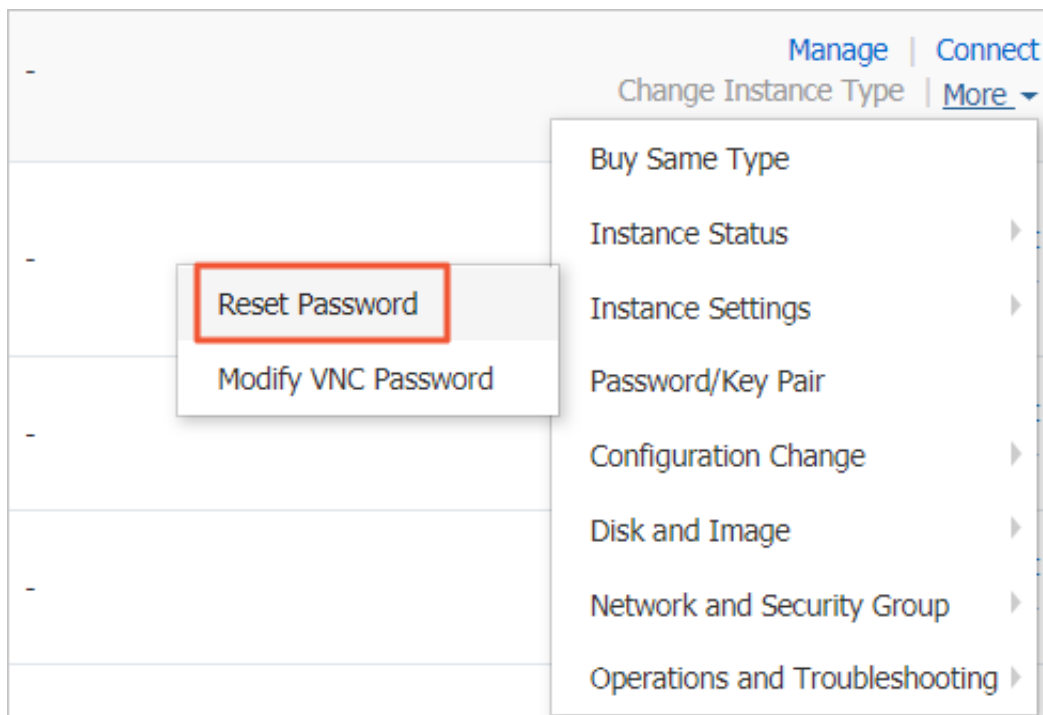
Procedure

To reset the password of one or multiple ECS instances, follow these steps:

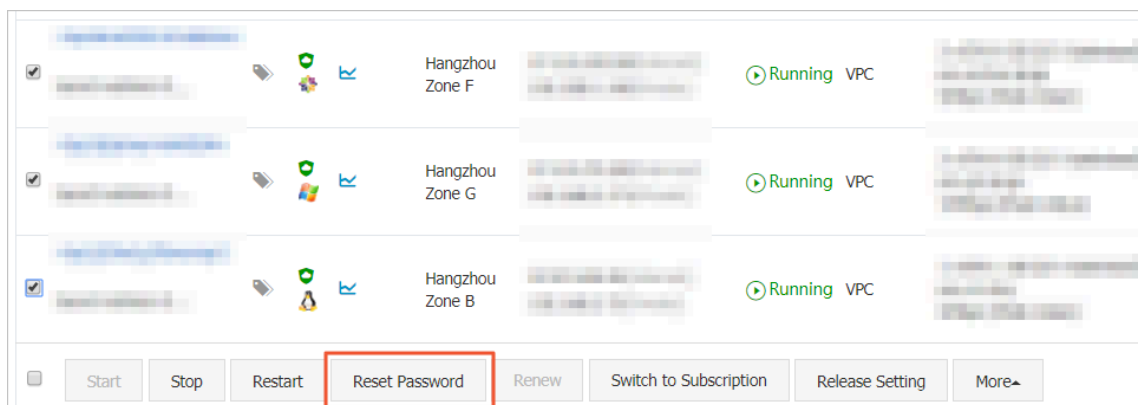
1. Log on to the [ECS console](#).
2. In the left-side navigation pane, choose Instances & Images > Instances.
3. In the top navigation bar, select a region.

4. Select one or more target instances as follows:

- To reset the password of one instance, find the target instance, and then choose **More > Password/Key Pair > Reset Password** in the Actions column.



- To reset the password of multiple instances, select the target instances, and then click **Reset Password** under the instance list.



5. Specify a new valid password, and then click OK.

Reset Password

Note: You must restart the instance in the console for the new password to take effect.

*Logon Password:

The password must be 8 to 30 characters in length.
The password must contain characters from at least three of the following categories: uppercase letters, lowercase letters, digits, and special characters (~ ()!@#\$%^&*-=|{}[];:'<>.,/?/). The password for a Windows ECS instance cannot start with a forward slash (/).

*Confirm Password:

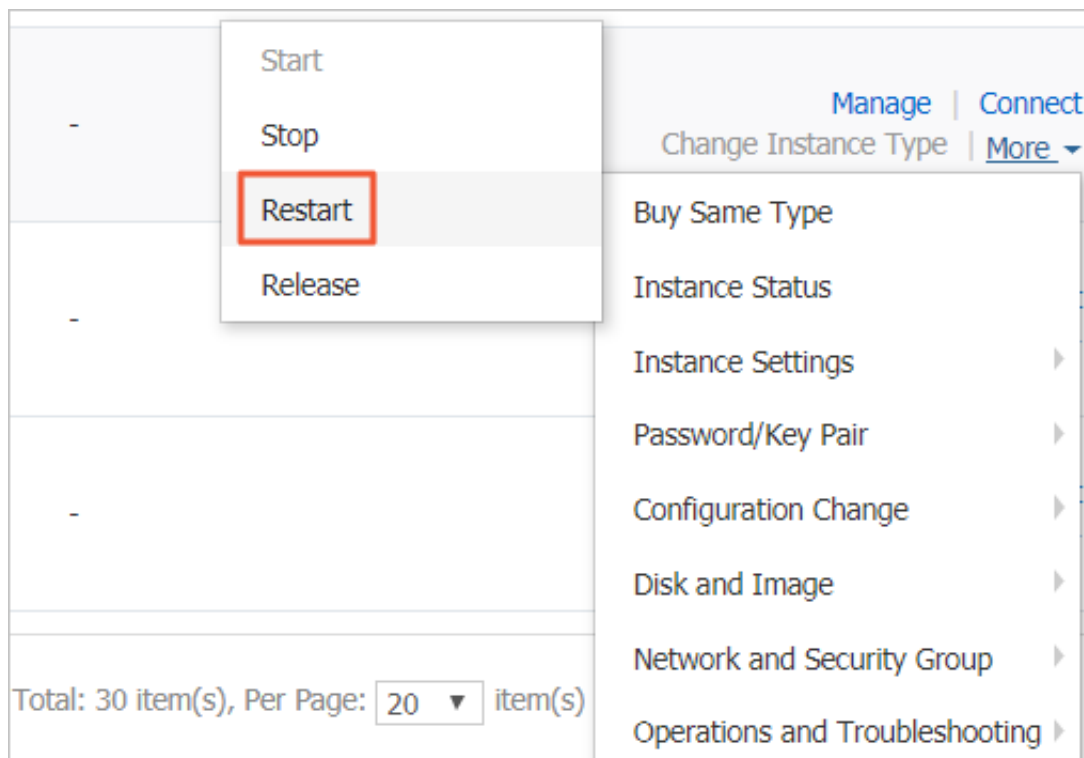
The operation will be performed on the selected **1 Instances** . Are you sure you want to proceed?

OK

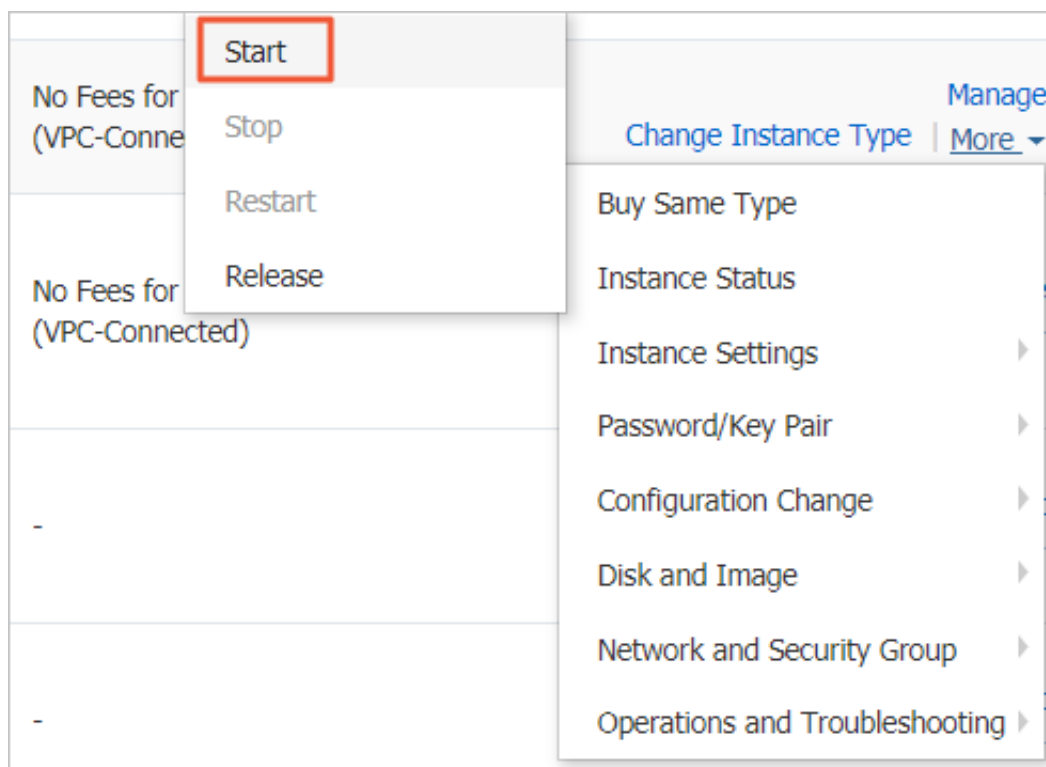
Cancel

6. Perform the following operations as needed:

- For a Running instance, [restart the instance](#) to make the new password take effect.



- For a Stopped instance, [start the instance](#) to make the new password take effect.



You can also use Alibaba Cloud CLI, OpenAPI Explorer, or Alibaba Cloud SDK to call the [ModifyInstanceAttribute](#) API action to reset the logon password.

What to do next

You can [change the password of an ECS instance](#) and then [restart the instance](#) in the ECS console to make the new password take effect.

8.7 Change the instance logon password by connecting to the instance

This topic describes how to change the logon password of an ECS instance by connecting to the instance through a remote connection. In this topic, a Linux instance running CentOS 6.8 and a Windows instance running Windows Server 2008 are used as examples.

Change the logon password of a Linux instance

1. Log on to the target instance by using a remote connection. For information about the different methods you can use to remotely connect to a Linux instance, see [#unique_68](#).
2. Run the `passwd` command (for example, `passwd root`).
3. Enter a new password.
4. Enter the new password again for the password to take effect.
5. Restart the instance in the ECS console or by calling the related API for the new password to take effect.

Change the logon password of a Windows instance

1. Log on to the target instance by using a remote connection. For information about the different methods you can use to remotely connect to a Windows instance, see [#unique_68](#).
2. Choose Start > Run, enter `compmgmt . msc` , and then press Enter.
3. In the Computer Management tool window, choose System Tools > Local Users and Groups > Users.
4. Right-click the username for which the password is to be changed (for example, Administrator).
5. Click Set Password.
6. In the Set Password for Administrator dialog box, click Proceed.

7. In the displayed dialog box, enter a new password in the New password and Confirm password text boxes, and then click OK .
8. Restart the instance in the ECS console or by calling the related API for the new password to take effect.

8.8 Enable instance release protection

If your ECS instances provide critical services, you can enable release protection for these ECS instances to prevent irreversible data loss resulting from accidental or incorrect operations during a manual release. This topic describes how to enable and disable instance release protection and view the instance release protection status.

Limits

- Instance release protection applies only to Pay-As-You-Go instances.
- An automatic release of an instance due to normal circumstances take precedence. These circumstances include, but are not limited to:
 - A payment under your account is overdue for more than 15 days.
 - The automatic release time that you set for the instance is reached.
 - The instance does not comply with the applicable security compliance policies.
 - The instance was automatically created by Auto Scaling and consequently is removed from the scaling group when the scaling group scales in.

Enable instance release protection when you create an instance



Note:

This procedure describes the instance release protection settings. For information about other settings, see [Create an instance by using the wizard](#).

To enable instance release protection when you create an instance, follow these steps:

1. On the Instances page, click Create Instance.
2. On the Basic Configurations page, set the Billing Method to Pay-As-You-Go, set other parameters as needed, and then click Next: Networking.
3. On the Networking page, set the parameters as needed, and then click Next: System Configurations.

4. On the System Configurations page, select Prevent users from releasing the instance inadvertently by using the console or API, set the parameters as needed, and then click Next: Grouping.

Basic Configurations (Required) Networking (Required) **3 System Configurations** 4 Grouping 5 Preview (Required)

Instance Name: The name can be 2 to 128 characters in length and can contain letters, Chinese characters, numbers, hyphens (-), underscores (_), and periods (.). It must start with a letter or Chinese character.

Description: The description can contain 2 to 256 characters. It cannot start with http:// or https://.

Host:

For Linux-based systems and other systems: the name can be 2 to 64 characters in length. It can contain several segments delimited by periods (.). Each segment can contain uppercase letters, lowercase letters, numbers, or hyphens. Each segment cannot contain continuous periods or hyphens. The name cannot start or end with a period or hyphen. The new hostname will take effect after the instance restarts.

Sequential Suffix: ☐ Add Sequential Suffix to Instance Name and Host Name
Sequential suffixes can be from 001 to 999. For example: LocalHost001, LocalHost002 or MyInstance001, MyInstance002.

Instance Deletion Protection: ☒ Prevent users from releasing the instance inadvertently by using the console or API

5. On the Grouping page, set the parameters as needed, and then click Next: Preview and confirm your settings.

When you create an instance through [RunInstances](#) or [CreateInstance](#), you can use the `DeletionProtection` parameter to enable or disable instance release protection.

Enable or disable instance release protection by modifying instance information

To enable or disable instance release protection by modifying the information of an instance, follow these steps:

1. On the Instances page, choose More > Instance Settings > Modify Instance Attributes in the Actions column of the instance to be modified.
2. In the Modify Instance Attributes dialog box, select Enable Instance Release Protection.



Note:

To disable the instance release protection, clear Enable Instance Release Protection.

Modify Information

* Instance Name:

It can be 2 to 128 characters in length.

Host Name:

Specify the host name of the operating system for the instance.


For Windows, the host name can be 2 to 15 characters in length and can contain letters, numbers, and hyphens (-). The host name cannot start or end with a hyphen. Do not use hyphens consecutively. The host name cannot contain only numbers.

For other operating systems, such as Linux, the host name can be 2 to 30 characters in length and can contain letters, numbers, and hyphens (-). You can separate characters in the host name with periods (.). The host name cannot start or end with a period or a hyphen. Do not use periods or hyphens consecutively.

The hostname will take effect after the instance has restarted.

Instance Description:

It can be 2 to 256 characters in length.

☒ Enable instance release protection 

Operation will be executed on the selected 1 Instances . Are you sure you want to proceed?

OK

Cancel

3. Click OK.

When you modify instance information through [ModifyInstanceAttribute](#), you can use the `DeletionProtection` parameter to enable or disable instance release protection.

View the instance release protection status

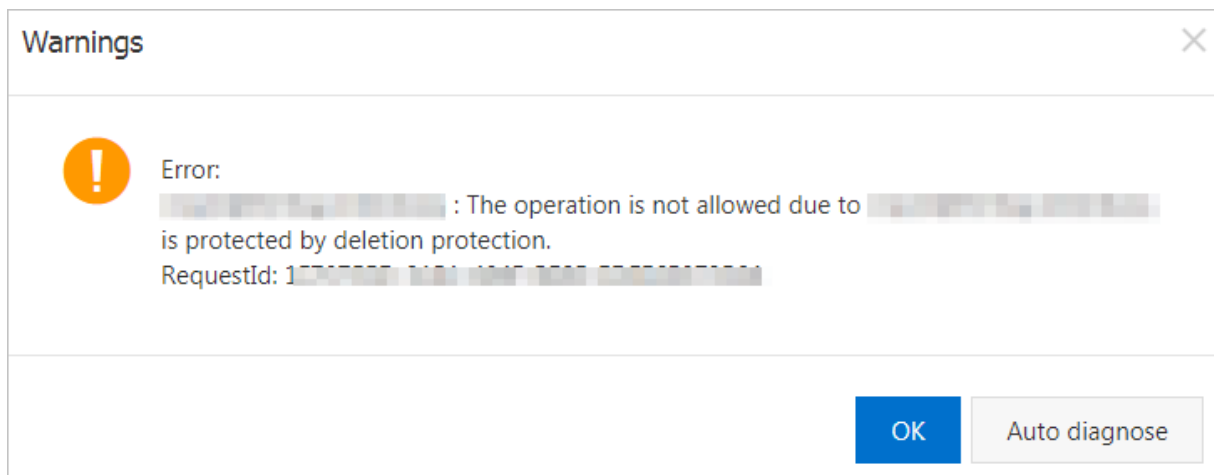
To view the release protection status of an instance, follow these steps:

1. On the Instances page, click the target instance in the Instance ID/Name column, or click Manage in the Actions column of the instance.
2. On the Instance Details page, view the status of Enable release protection in the Payment Information area.

Instance Details	Instance Type: I/O Optimized
Disks	Operating System: CentOS 7.6 64-bit
Instance Snapshots	Elastic Network Interfaces: eni-43e4a9c1
Elastic Network Inte...	Internet IP Address: 192.168.1.100
Security Groups	EIP: -
	Private IP Address: 172.31.0.100
	Secondary Private IP Addresses: Manage Secondary Private IP Address
	Bandwidth Billing Method: PayByTraffic
	Current Bandwidth: 5Mbps (Peak Value)
	VPC: vpc-43e4a9c1
	VSwitch: vsw-43e4a9c1
	Payment Information Buy Same Type More
	Billing Method: Pay-As-You-Go
	Stop Mode: -
	Created At: 15 July 2019, 17:16
	Automatic Release Time: -
	Enable release protection: Yes

Example of protection effect

After you enable release protection for an instance, a warning message will be displayed if you try to manually release the instance. If you want to manually release the instance, you must [disable instance release protection](#).



After you enable release protection for an instance, the error code `InvalidOperation.DeletionProtection` is returned if you try to use [DeleteInstance](#) to delete the instance.

8.9 User-defined data and metadata

8.9.1 Metadata

The metadata of an instance is the basic information of the ECS instance, including the Instance ID, IP Address, OS, and other related information. You can use an instance's metadata to better manage and configure an instance.



Note:

If you manually change some instance information, this change will not be reflected in the instance metadata.

Limits

The metadata is only applicable for VPC-Connected instances.

Get the metadata

Linux instance

1. [#unique_214](#).

2. Run `curl http://100.100.100.200/latest/meta-data/` to access the root directory of the metadata.

3. Add the specific metadata name to the preceding command to access the specified metadata. For example:

- Run `curl http://100.100.100.200/latest/meta-data/instance-id` to get the ID of an ECS instance.
- Run `curl http://100.100.100.200/latest/meta-data/image-id` to get the image ID of an ECS instance.

Windows instance

1. [#unique_216](#).

2. Use PowerShell to run `Invoke-RestMethod http://100.100.100.200/latest/meta-data/` to get the metadata.


3. Add the specific metadata name to the preceding command to access the specified metadata. For example:

- Run `Invoke-RestMethod http://100.100.100.200/latest/meta-data/instance-id` to get the ID of an ECS instance.
- Run `Invoke-RestMethod http://100.100.100.200/latest/meta-data/image-id` to get the image ID of an ECS instance.

List of instance metadata

Metadata name	Description	Version
dns-conf/nameservers	DNS configurations for an instance.	2016-01-01
eipv4	EIP address	2016-01-01
hostname	The OS name of an instance.	2016-01-01
image-id	ID of the image that is selected at the time of instance creation.	2016-01-01
image/market-place/product-code	Product code of the image in the Alibaba Cloud Marketplace.	2016-01-01
image/market-place/charge-type	Billing method of the image in the Alibaba Cloud Marketplace.	2016-01-01
instance-id	The instance ID	2016-01-01

Metadata name	Description	Version
dns-conf/nameservers	DNS configurations for an instance.	2016-01-01
mac	The MAC address of the instance. If multiple network interface cards exist in an instance, this metadata indicates the MAC address of eth0.	2016-01-01
network-type	Network type (only applicable for VPC network).	2016-01-01
ntp-conf/ntp-servers	The address of a NTP server.	2016-01-01
owner-account-id	The aliuid of the instance owner.	2016-01-01
private-ipv4	Private IP address.	2016-01-01
public-ipv4	Public network IP address.	2016-01-01
public-keys	A list of all public keys of the current instance.	2016-01-01
region-id	The region where the instance is located.	2016-01-01
zone-id	Zone ID of the zone where the ECS instance is located.	2016-01-01
serial-number	The serial number of an instance.	2016-01-01
source-address	The source of YUM/APT (only applicable for a Linux instance).	2016-01-01
kms-server	Activate the server (only applicable for a Windows instance).	2016-01-01
wsus-server/wu-server	Update the server (only applicable for a Windows instance).	2016-01-01
wsus-server/wu-status-server	The server that monitors the update status of an instance (only applicable for a Windows instance).	2016-01-01
vpc-id	ID of the VPC that an instance is in.	2016-01-01
vpc-cidr-block	The CIDR block of the VPC that an instance is in.	2016-01-01
vswitch-cidr-block	The CIDR block of the VSwitch that an instance is in.	2016-01-01
vswitch-id	ID of the VSwitch that an instance is in.	2016-01-01

Metadata name	Description	Version
dns-conf/nameservers	DNS configurations for an instance.	2016-01-01
ram/security-credentials/[role-name]	<p>The temporary STS credential is generated according to the policy of a RAM role. Only available when you specify a RAM role to an ECS instance. When you use this metadata to get the STS credential, [role - name] must be replaced with the actual RAM role name you create or you have created.</p> <div>  Note: A new STS credential is available 30 minutes prior to the expiration of the old one. </div>	2016-01-01
instance/spot/termination-time	The spot instance release time is based on the instance operating system time zone. It is specified in UTC format. For example, YYYY-MM-DDThh:mm:ssZ For example, 2018-04-07T17:03:00Z.	2016-01-01
network/interfaces/mac	The MAC address list of Network Interface Controllers (NICs).	2016-01-01
network/interfaces/mac/[mac]/network-interface-id	The unique ID of the NIC, [mac] must be replaced with the actual MAC address.	2016-01-01
instance/virtualization-solution	The virtualization solution, which is ECS Virt 1.0 / 2.0	2016-01-01
instance/virtualization-solution-version	The internal build version.	2016-01-01
instance/last-host-landing-time	The latest update time of the physical server to which your instance is hosted on.	2016-01-01
instance-identity/document	Instance identity Instance identity document.	2016-01-01
instance-identity/pkcs7	Instance identity signature.	2016-01-01

Other data exposing to instance metadata

- **Maintenance:** For on-going [system events](#), run the following command to get the latest data:

```
curl http://100.100.100.200/latest/maintenance/active-system-events
```

- **Signature:** For the [instance identity](#) document, run the following command to get the latest data:

```
curl http://100.100.100.200/latest/dynamic/instance-identity/document
```

- **Boot behavior configuration:** For an [instance's user data](#), run the following command to get the latest data:

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

8.9.2 Retrieve instance metadata

This topic describes how to retrieve the metadata of a Linux or Windows instance.

Limits

Only the metadata of a VPC instance can be obtained.



Note:

Manually changing instance information will not change the relevant metadata.

Retrieve the metadata of a Linux instance

1. Connect to your Linux instance. For more information, see [Overview of connecting to an ECS instance](#).

2. Run the following command to access the root directory of the metadata.

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

3. Add a specific metadata name to the preceding command to access the specified metadata. For more information, see [Instance metadata](#) and [Dynamic instance metadata](#). For example:

- Run the following command to get the ID of an ECS instance.

```
curl http://100.100.100.200/latest/meta-data/instance-id
```

- Run the following command to get the image ID of an ECS instance.

```
curl http://100.100.100.200/latest/meta-data/image-id
```

- Run the following command to get the active system events.

```
curl http://100.100.100.200/latest/maintenance/active-system-events
```

- Run the following command to get the instance identity document.

```
curl http://100.100.100.200/latest/dynamic/instance-identity/document
```

- Run the following command to get the instance user data.

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

Retrieve the metadata of a Windows instance

1. Connect to your Windows instance. For more information, see [Overview of connecting to an ECS instance](#).

2. Run the following command by using PowerShell to get instance metadata.

```
Invoke - RestMethod http :// 100 . 100 . 100 . 200 / latest /  
meta - data /
```

3. Add a specific metadata name to the preceding command to access the specified metadata. For more information, see [Instance metadata](#) and [Dynamic instance metadata](#). For example:

- Run the following command to get the ID of an ECS instance.

```
Invoke - RestMethod http :// 100 . 100 . 100 . 200 / latest /  
meta - data / instance - id
```

- Run the following command to get the image ID of an ECS instance.

```
Invoke - RestMethod http :// 100 . 100 . 100 . 200 / latest /  
meta - data / image - id
```

- Run the following command to get the active system events.

```
Invoke - RestMethod http :// 100 . 100 . 100 . 200 / latest /  
maintenanc e / active - system - events
```

- Run the following command to get the instance identity document.

```
Invoke - RestMethod http :// 100 . 100 . 100 . 200 / latest /  
dynamic / instance - identity / document
```

- Run the following command to get the instance user data.

```
Invoke - RestMethod http :// 100 . 100 . 100 . 200 / latest /  
user - data
```

8.10 User-defined data

8.10.1 User data

You can use user data of an ECS instance to customize its startup behavior and to pass data into the instance. You can specify user data when creating an instance ([#unique_71](#)) and customize startup behavior such as automatically update software packages, enable services, print logs, install dependencies, initialize web services, and more. User data of an ECS instance is implemented primarily through different types of scripts. User data can also be used as common data to be referenced in the instances.

Instructions for use

To configure instance user data, note that:

- Only VPC-Connected instances are supported.
- For [phased-out instance types](#), they must be I/O optimized. Other [instance type families](#) are not limited for I/O optimized.
- Instance user data requires Base64 encoding before being passed in, and the user data before encoding cannot exceed 16 KB.
- The instance must use an official image or a user image that is created from an official image. The operating system must be one of the following:

Windows instances	Linux instances
Windows Server 2008 R2 and later version	<ul style="list-style-type: none">- CentOS- Ubuntu- SUSE Linux Enterprise- OpenSUSE- Debian- Aliyun Linux

Module frequency

After the instance enters the Running state, use your Alibaba Cloud primary account to run the user data of the instance, followed by the initialization or `/ etc / init` information.

After you modify the instance user data, depending on the type of scripts and modules that are used, the modified user data is or is not run. For example:

- If you configure user data by using a shell script, such as a [user-data script](#), the modified user data is not run.
- If the user data configures modules such as Byobu, Set Hostname, and Set Passwords, the modified user data is not run.
- If the user data configures modules such as `bootcmd`, `update_etc_hosts`, and `yum_add_repo`, the modified user data is run.

For more information, see [modules](#).

Set user data

For this example, assume that you write user data development in a Windows environment, and you use [Upstart Job](#) to configure the user data.

1. Use an editor to create a text file, such as Notepad++.

2. Edit the script related to the user data in the text file.

**Note:**

The first line must meet the format requirements of the instance user data script, such as `#!/bin/sh`, `# cloud-config`, `# upstart - job`, `[bat]` and `[powershell]`. For more information, see [Linux instance user data](#) and [Windows instance user data](#).

3. Debug the script file to confirm that the content is valid.
4. (Optional) If you make a [Gzip compression content](#), compress the script file in .gz format.
5. (Optional) If you are creating an [Include file](#) or a [Gzip compression script](#), upload script file to available storage services, obtain the link, and set the valid period of the link.

We recommend that you use Alibaba Cloud OSS to create links. For more information, see [upload an object](#) or [set lifecycle](#).

6. Log on to the [ECS console](#).
7. Follow the instructions in [creating an instance](#) to create a Linux instance.

**Note:**

The instance must be `VPC - Connected`, and you must select a [image](#) that meets the requirement. For [phased-out instance types](#), I/O optimized instances are required. Other [instance type families](#) are not limited in terms of I/O optimized.

After creating the instance, select Advanced (based on instance RAM roles or cloud-init) use text form and enter your user data. If your user data has been encrypted by Base64 encoding, click The text is Base64-encoded.

8. Wait for the instance to be created.
9. [Connect](#) to your instance.

10. View the results of the user data. If a failure occurs, check the relevant log files.

The following is an output example of user data on a CentOS instance by using the upstart job script:

In the preceding figure, the startup job file `part - 001 . conf` is generated in the `/ etc / init` folder.

Related API: [#unique_71](#)

View user data

You can view user data of an instance from the server `100 . 100 . 100 . 200`. To do so, follow these steps:

1. Connect to the target instance.

2. In the instance, depending on your OS, run one of the following:

- For Linux, run `curl http :// 100 . 100 . 100 . 200 / latest / user - data` to view the user data.
- For Windows, run `Invoke - RestMethod http :// 100 . 100 . 100 . 200 / latest / user - data /` to view the user data.

Related API: [#unique_277](#)

Modify user data

You must stop the instance before modifying its current user data. If you need to restart a Pay-As-You-Go VPC-Connected instance immediately after you modify the user data, we recommend that you disable the No fees for stopped instances option. To modify user data of an instance, follow these steps:

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, click Instances.
3. Select the target region.
4. Select the target instance and then, in the Actions column, click Sets User Data.
5. Enter the user data and then click OK.



Note:

After you modify the user data, depending on the script type and the module type, the modified user data is or is not run.

Related API: [#unique_263](#)

Linux instance user data

Linux instance user data can be configured by several types of script, such as [User-data Script](#), [Cloud Config](#), [Include Files](#), [Gzip compression scripts](#), and [Upstart Job](#). The scripts follow the format of open source cloud-init, and reference the [#unique_157](#) for data sources. The configuration of Linux instances are automated at boot. For more information, see [formats](#).

User-data script

User-data can be a shell script. It runs once at the instance first boot. The first line is fixed as `#!/`, for example `#!/ bin / sh`. The content of user-data script before Base64 encoding cannot exceed 16 KB. The following is a User-Data script example:

```
#!/ bin / sh
echo " Hello World . The time is now $( date - R )" |
tee / root / output10 . txt
service httpd start
chkconfig httpd on
```

After the instance has been created, connect to the instance and run `cat [file]` to view the results of the user-data script.

```
[ root @ XXXXX2z ~]# cat output . txt
Hello World . The time is now Mon , 24 Jul 2017 13
: 03 : 19 + 0800 !
```

Cloud-Config

You can use Cloud-Config to configure services such as updating yum sources, importing SSH keys, installing dependency packages, and more. The first line of Cloud-Config is fixed as `# cloud - config`, and the header cannot have spaces. The file must be valid yaml syntax. Depending on the service you configured, the instance user data runs differently.

Cloud Instance user data requires Base64 encoding before being passed in, and the pre-encoding cloud config data cannot exceed 16 KB. The following is a Cloud-Config script example:

```
# cloud - config
apt :
primary :
- arches : [ default ]
uri : http:// us . archive . ubuntu . com / ubuntu /
bootcmd :
- echo 192 . 168 . 1 . 130 us . archive . ubuntu . com >> / etc
/ hosts
```

After the instance has been created, connect to the instance to view the results.

Include files

The contents of an Include File consist of a script link, with one link on one line. When the instance starts, cloud-init reads the contents of the script link in the Include File. If there is an error reading script content in a row, the instance stops performing user data. The first line of Include File is fixed as `# include` and the header cannot have spaces. The update frequency of the instance user data follows the script type configured in the include file.

Instance user data requires Base64 encoding before being passed in. The file before Base64 encoding cannot exceed 16 KB. The following is an Include File example:

```
# include
http:// ecs - image - test . oss - cn - hangzhou . aliyuncs . com /
UserData / myscript . sh
```

After the instance has been created, connect to the instance to view the results.

Gzip compressed content

The content of a [User-Data Script](#), [Cloud-Config](#), and [Include File](#) cannot exceed 16 KB. If your script content is larger than 16 KB, you can use Gzip to compress the content, then upload the compressed script to an available storage service (we recommend OSS), obtain the link, and use the Include File format to render the link. The first line of a Gzip compressed script is fixed as `# include` and the header cannot have spaces. The update frequency of the instance user data follows the script type configured in the Gzip file. The following is a Gzip compressed file example:

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

Upstart Job

Upstart service is required for an init system if you use Upstart Job to configure user data. For example, CentOS 6, Ubuntu 10/12/14, and Debian 6/7 use upstart as the init system. Upstart Job script places your instance user data into a file in `/etc/init` directory. The first line of Upstart Job script is fixed as `# upstart - job` and the header cannot have spaces. We perform the instance user data for every instance boot. The following is a Upstart Job script example:

```
# upstart - job
description "upstart test"
start on runlevel [ 2345 ]
stop on runlevel [! 2345 ]
exec echo "Hello World. The time is now $(date -R)" | tee /root/output.txt
```

Windows instance user data

Windows instance user data is supported by Alibaba Cloud ECS, and offers Windows-based instances the ability to run initialization scripts. Instance user data requires Base64 encoding before being passed in, and the pre-encoding user data cannot exceed 16 KB. Only SBC case characters are allowed. You can write Bat script or PowerShell script to configure the instance user data.

Bat scripts

The first line is fixed as `[bat]` and the header cannot have spaces. For example:

```
[ bat ]
echo "bat test" > c:\1.txt
```

After the instance has been created, connect to the instance to view the results. In the following example, a `1.txt` text file is shown under the `C:\` drive.

The first line of PowerShell scripts

is fixed as `[powershell]` and the header cannot have spaces. For example:

```
[ powershell ]
write - output "Powershell Test" | Out - File C:\2.txt
```

Reference

For more information about Linux instance user data, see cloud-init [formats](#).

For more information about the update frequency of Linux instance user data, see cloud-init [modules](#).

8.10.2 Manage user data

This topic describes how to configure generated user data for an instance in the ECS console, and how to view and modify existing user data.

Limits

The user data function has the following limits:

- User data can be configured for instances in a VPC only.
- For [phased-out instance types](#), only I/O-optimized instance types support user data. For other [instance type families](#), all instance types support user data.
- User data must be Base64-encoded before it is configured for instances. Additionally, the original user data cannot exceed 16 KB.
- An instance must use a public image or a custom image that is created from a public image. Furthermore, the following table describes the operating systems that are supported.

Platform	Operating system
Windows	Windows Server 2008 R2 or later
Linux	<ul style="list-style-type: none">- CentOS- Ubuntu- SUSE Linux Enterprise- OpenSUSE- Debian- Aliyun Linux

Configure user data for an instance



Note:

The following procedure takes a Linux instance running CentOS as an example.

1. Create a Linux instance by following the instructions in [Create an instance by using the wizard](#).



Note:

During instance creation, enter user data in the user data box below Advanced Options (Instance RAM Roles and User Data cloud-init). If your user data is Base64-encoded, select the Enter Based64 Encoded Information check box.

Advanced (based on instance RAM roles or cloud-init)

RAM Role: [See details](#) [Create instance RAM role](#)

User Data: ☒ The text is Base64-encoded

Windows guest operating systems support bat and powershell formats. Before you base64-encode the content, the first line must be [bat] or [powershell]. Linux guest operating systems support shell scripts. [cloud-init](#) | [Learn more](#).

2. After the instance starts, connect to it by following the instructions in [Overview of connecting to an ECS instance](#).
3. Check the startup results against the configured user data. If exceptions occur, check the log files. The following output is from a CentOS instance which is configured with user data through an Upstart Job script:

```
[root@ ~]# cd /etc/init
init/ init.d/ inittab
[root@ ~]# cd /etc/init/
[root@ init]# ls
part-001.conf
[root@ init]# cat part-001.conf
#upstart-job
#!/bin/sh
echo "Hello World. The time is now $(date -R)!" | tee /root/output.txt[root@ init]#
```

As shown in the preceding output, a startup task file `part - 001 . conf` is created in the folder `/ etc / init` of the instance.

You can also set user data by calling [RunInstances](#) through Alibaba Cloud CLI, OpenAPI Explorer, or Alibaba Cloud SDK. The relevant parameter is `UserData`.

View user data

You can view user data through the dedicated server 100.100.100.200. To do so, follow these steps:

1. Connect to the instance. For more information, see [Overview of connecting to an ECS instance](#).
2. Run the following commands in the instance:

- For Linux, run `curl http://100.100.100.200/latest/user-data`.
- For Windows, run `Invoke-WebRequest http://100.100.100.200/latest/user-data/`.

You can also view user data by calling [DescribeUserData](#) through Alibaba Cloud CLI, OpenAPI Explorer, or Alibaba Cloud SDK.

Modify user data

You must stop the target instance before you modify its user data. After the modification, you need to check if the modified user data needs to be run according to the script type and module type. For more information, see [User data](#).

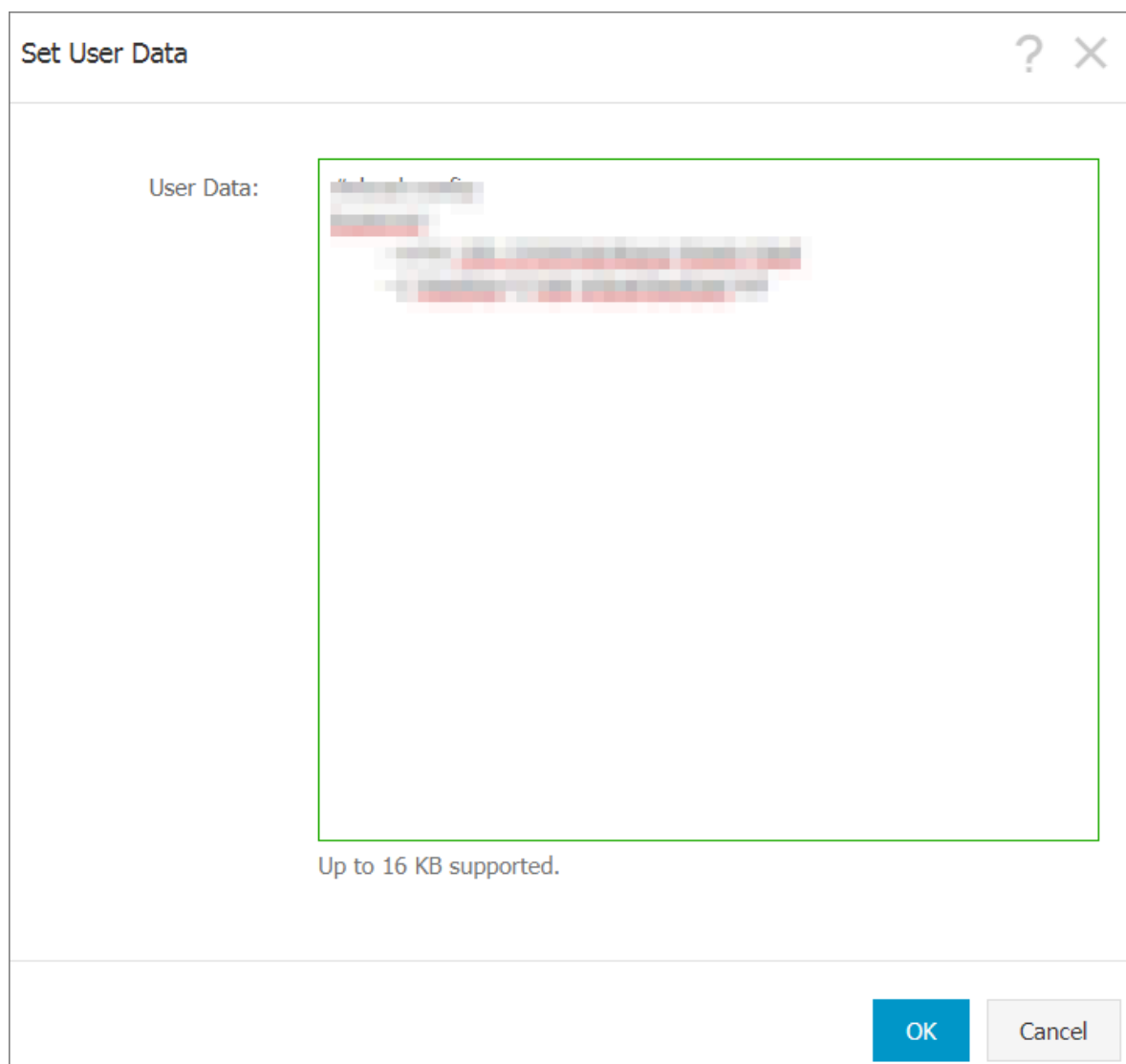


Note:

For Pay-As-You-Go VPC instances, if you need to restart an instance immediately after modifying its user data, we recommend that you turn off the option of No fees for stopped instances.

1. Find the target instance, and in the Actions column, choose Instance Settings > Set User Data.

2. In the displayed dialog box, enter the relevant information and click OK.

A screenshot of a 'Set User Data' dialog box. The title bar says 'Set User Data' with a question mark and a close button. The main area has a label 'User Data:' followed by a large text input field. Inside the input field, there is some blurred text that appears to be a script. Below the input field, it says 'Up to 16 KB supported.' At the bottom right, there are two buttons: 'OK' (blue) and 'Cancel' (gray).

You can also modify user data by calling [ModifyInstanceAttribute](#) through Alibaba Cloud CLI, OpenAPI Explorer, or Alibaba Cloud SDK.

8.10.3 Customize yum repository, NTP service, and DNS service

This topic describes how to use a user-defined script to configure your own yum repository, NTP service, and DNS service when creating a Linux instance. User-defined scripts also enable you to configure NTP service and DNS service for a Windows instance.

Scenarios

When a Linux instance is started, Alibaba Cloud automatically configures a pre-defined yum repository, NTP service, and DNS service for the instance. However, if

you want to have your own yum repository, NTP service, and DNS service, use user-defined scripts to implement them.

- If you are using a custom yum repository, Alibaba Cloud does not provide support for it.
- If you are using a custom NTP service, Alibaba Cloud does not provide time service.

Procedure

To customize your yum repository, NTP service, and DNS service for a Linux instance when creating it, follow these steps:

1. Create an instance [by using the wizard](#). Be aware of the following configuration items:

- On the Basic Configurations page, note the following customized configurations:
 - Instance Type: Select an I/O-optimized instance.
 - Image: Select a supported image type, for example, CentOS 7.2 on the Public Image tab.

2. On the Networking page, set Network Type to VPC

3. On the System Configurations page, enter the following script in the User Data text box in the Advanced area:

```
#!/bin/sh
# Modify DNS
echo "nameserver 8.8.8.8" | tee /etc/resolv.conf
# Modify yum repo and update
rm -rf /etc/yum.repos.d/*
touch myrepo.repo
echo "[base]" | tee /etc/yum.repos.d/myrepo.repo
echo "name = myrepo" | tee -a /etc/yum.repos.d/myrepo.repo
echo "baseurl = http://mirror.centos.org/centos" | tee -a /etc/yum.repos.d/myrepo.repo
echo "gpgcheck = 0" | tee -a /etc/yum.repos.d/myrepo.repo
echo "enabled = 1" | tee -a /etc/yum.repos.d/myrepo.repo
yum update -y
# Modify NTP Server
echo "server ntp1.aliyun.com" | tee /etc/ntp.conf
systemctl restart ntpd.service
```



Note:

- The first line must be `#!/`, with no leading space `. / bin / sh`, with no leading space.
- Do not add unnecessary spaces or carriage return characters in the full text.
- You can customize URLs of your own DNS server, NTP Server, and yum repository based on the instance situations.
- The preceding content applies to CentOS 7.2. If you are using other images, modify the scripts as needed.
- You can also define the yum repository in the scripts of the `cloud config` type, but it is not recommended because it is not flexible enough to get adapted to Alibaba Cloud that may pre-configure some yum repository. Scripts of `script` type is recommended for changing the yum repository.

After the instance is created, you can connect to the instance to view the implementation details, as shown in the following figure.

```
[root@ yum.repos.d]# cat /etc/resolv.conf
nameserver 8.8.8.8
[root@ yum.repos.d]# ping www.baidu.com
PING www.a.shifen.com (103.235.46.39) 56(84) bytes of data.
64 bytes from 103.235.46.39: icmp_seq=1 ttl=48 time=73.3 ms
^C64 bytes from 103.235.46.39: icmp_seq=2 ttl=48 time=74.8 ms

--- www.a.shifen.com ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1001ms
rtt min/avg/max/mdev = 73.393/74.113/74.833/0.720 ms
[root@ yum.repos.d]# cat /etc/ntp.conf
server ntp1.aliyun.com
[root@ yum.repos.d]# systemctl status ntpd.service
● ntpd.service - Network Time Service
   Loaded: loaded (/usr/lib/systemd/system/ntpd.service; enabled; vendor preset: disabled)
   Active: active (running) since Mon 2017-03-13 11:08:11 CST; 1min 58s ago
   Process: 6235 ExecStart=/usr/sbin/ntpd -u ntp:ntp $OPTIONS (code=exited, status=0/SUCCESS)
   Main PID: 6237 (ntpd)
   CGroup: /system.slice/ntpd.service
           └─6237 /usr/sbin/ntpd -u ntp:ntp -g

Mar 13 11:08:11 ntpd[6237]: 0.0.0.0 c01d 0d kern kernel time sync enabled
Mar 13 11:08:11 ntpd[6237]: ntp_io: estimated max descriptors: 1024, initial socket boundary: 16
Mar 13 11:08:11 ntpd[6237]: Listen and drop on 0 v4wildcard 0.0.0.0 UDP 123
Mar 13 11:08:11 ntpd[6237]: Listen and drop on 1 v6wildcard :: UDP 123
Mar 13 11:08:11 ntpd[6237]: Listen normally on 2 lo 127.0.0.1 UDP 123
Mar 13 11:08:11 ntpd[6237]: Listen normally on 3 eth0 172.18.48.114 UDP 123
Mar 13 11:08:11 ntpd[6237]: Listening on routing socket on fd #20 for interface updates
Mar 13 11:08:11 ntpd[6237]: 0.0.0.0 c016 06 restart
Mar 13 11:08:11 ntpd[6237]: 0.0.0.0 c012 02 freq_set kernel 0.000 PPM
Mar 13 11:08:11 ntpd[6237]: 0.0.0.0 c011 01 freq_not_set
[root@ yum.repos.d]# cat /etc/yum.repos.d/myrepo.repo
[base]
name=myrepo
baseurl=http://mirror.centos.org/centos
gpgcheck=0
enabled=1
[root@ yum.repos.d]#
```

The preceding figure shows that you have successfully customized the DNS service, the NTP service, and the yum repository.

8.10.4 Create a new account with the root user privilege

This topic describes how to use a user-defined script to create a new account with the root user privilege when creating a Linux instance. User-defined scripts can also be used to create a new account with the administrator privilege for a Windows instance.

Scenarios

Use user-defined scripts of instances if you want to achieve the following results when creating a Linux ECS instance:

- Disable the default root account that comes with a Linux ECS instance. You can use the script to customize how to disable the root user and how many root user privileges are disabled.
- Create a new account with the root user privilege and customize the account name.
- Use only SSH key pairs, but not user passwords, for remote logon to manage the instance by using the new account with the root user privilege.
- If this new account is required to perform operations that can only be done by a user with root user privilege, the `sudo` command can be used without a password for privilege escalation.

Procedure

To create a new account with the root user privilege, follow these steps:

1. Create an instance by referring to [Create an instance by using the wizard](#). Be aware of the following configuration items:
 - On the Basic Configurations page, note the following customized configurations:
 - Instance Type: Select an I/O-optimized instance.
 - Image: Select a supported image type, for example, CentOS 7.2 on the Public Image tab.
2. On the Networking page, set Network Type to VPC
3. On the System Configurations page, enter the following script in the User Data text box in the Advanced area:

```
#!/bin/sh
useradd test
echo "test ALL=(ALL) NOPASSWD : ALL" | tee -a /etc/sudoers
mkdir /home/test/.ssh
touch /home/test/.ssh/authorized_keys
echo "ssh-rsa AAAAB3NzaC1yc2EAAAABJQAAAQEAhGqhEh
/rGbIMCGItF VtYpsXPQrC aunGJKZVIW tINrGZwusL c290qDZ93K
Ceb8o6X1Ib y1Wm + psZY8THE +/ BsXq0M0Hzf kQZD2vXuhR b4xi1z98JH"
```

```
skX + 0jnbjqYGY + Brgai9BvKD XTTSyJtCYU nEKxvcK + d1ZwxbNuk2
QZ0ryHESDb SaczlnFgFQ EDxhCrvko + zWLjTVnomV UDhdMP2g6f
Z0tgFVwkJF V0bE7oob3N OVcrx2Tyhf cAjA4M2 / Ry7U2MFADD C +
EVkpoVDm0S OT / hYJgaVM1xM DLSeE7kzX7 yZbJLR1XAW V1xzZkNclY
5w1kPnW8qM YuSwphXzt4 gsF0w == rsa - key - 20170217 " | tee -
a / home / test /. ssh / authorized _keys
```

**Note:**

- The first line must be `#!/. / bin / sh`, with no leading space.
- Do not enter unnecessary spaces or carriage return characters in the text.
- The last line is your public key. You can define it.
- You can add other configuration in the script as you need.
- The example script only applies to CentOS 7.2. If you are using other images, customize the script according to the operating system types.

After the instance is started, you can use the new test user to connect to the instance by using an SSH private key. You can also escalate the permission level by using the `sudo` command and run operations that require the root user privilege, as shown in the following figure.

```
test@iZwz9bm4vhpq7275w13w7eZ:/
Using username "test".
Authenticating with public key "rsa-key-20170217"

Welcome to Alibaba Cloud Elastic Compute Service !

[test@iZwz9bm4vhpq7275w13w7eZ ~]$
[test@iZwz9bm4vhpq7275w13w7eZ ~]$
[test@iZwz9bm4vhpq7275w13w7eZ ~]$ sudo cd /root
[test@iZwz9bm4vhpq7275w13w7eZ ~]$
[test@iZwz9bm4vhpq7275w13w7eZ ~]$
[test@iZwz9bm4vhpq7275w13w7eZ ~]$
[test@iZwz9bm4vhpq7275w13w7eZ ~]$ sudo ll
sudo: ll: command not found
[test@iZwz9bm4vhpq7275w13w7eZ ~]$ sudo ls
[test@iZwz9bm4vhpq7275w13w7eZ ~]$ cd /
[test@iZwz9bm4vhpq7275w13w7eZ /]$ cd root/
-bash: cd: root/: Permission denied
[test@iZwz9bm4vhpq7275w13w7eZ /]$ sudo cd root/
[test@iZwz9bm4vhpq7275w13w7eZ /]$
```

8.11 Instance identity

Instance identity is a part of `#unique_157` that describes and validates an instance. Instance identity enables you quickly locate a target instance, and provides authentication for such actions as software updates, access control, or application activation. The signature of instance identity is encrypted by the `PKCS#7` standard.

Use cases

You can use instance identity (`instance - identity`) in scenarios such as authentication, granting access, or instance identification, as follows.

- A typical software activation (with one serial number for one device) does not work in the environment of cloud computing where the sales model of the [Alibaba Cloud Marketplace](#) is flexible. In this case, you can use instance identity to complete the software activation. For more information, see [Sample 1. No audience in the signature](#).
- When you write sensitive data in the instance, you can use instance identity to verify that the server is your instance.
- Scenarios whereby you want to confirm the source of the target server.

Feature details

Instance identity consists of a dynamically generated instance identity document (`document`) and instance identity signature (`signature`).

- Instance identity document: Describes the attributes of an instance. The following table lists instance identity document items.

Properties	Description	Can it be changed?
account-id	ID of the Alibaba Cloud account to which the instance belongs	No
create-time	Instance creation time	No
instance-id	Instance ID.	No
mac	MAC address of the instance primary network interface	No

Properties	Description	Can it be changed?
region-id	ID of the region to which the instance belongs	No
serial-number	Serial number of the instance	No
zone-id	ID of the zone to which the instance belongs	No
instance-type	Instance types	Yes. It changes after you change the instance type.
image-id	Image ID of the instance	Yes. It changes after you replace the system disk of the instance.
private-ip	Private IP of the instance	Yes. It changes after you change the private IP of a VPC-Connected instance.

- Instance identity signature: Verifies the instance identity in the cryptographic method of the PKCS#7 standard.
 - To enhance the security of the signature, you can protect it by specifying the `audience` parameter in it. However, even if you specify `audience`, another user may get information about the identity document and the identity signature. Therefore, we recommend the value of the `audience` parameter is a random string, timestamp, regularly changed data, or some output generated by a specific algorithm.
 - If you specify the `audience` parameter, you must modify the instance identity document and signature simultaneously. For example, if you have specified the `audience` parameter while obtaining the signature, before you verify the signature by using the OpenSSL commands, you must add the value of the `audience` parameter at the end of the dynamically obtained instance identity document in the format of "`audience` ":" *Value of the audience* ", and separate the parameters with a comma (,).

Usage

The instance identity is verified by using the OpenSSL commands. Make sure that you have the OpenSSL configured in your instance. Visit <https://www.openssl.org/source> to download and update OpenSSL service.

Take CentOS 7.4 as an example to use the instance identity.

1. Connect to your Linux instance.

2. Run `curl http://100.100.100.200/latest/dynamic/instance-identity/document` to query the file of instance identity document.

3. Run `curl http://100.100.100.200/latest/dynamic/instance-identity/pkcs7` or `curl http://100.100.100.200/latest/dynamic/instance-identity/pkcs7?audience=XXXX` to get the instance identity signature.

```
[root@ ~]# curl http://100.100.100.200/latest/dynamic/instance-identity/document
{"region-id":"cn-hangzhou","instance-id":"i-bp1b39...", "serial-number":
"52aelaf6-64aa-407f-88fd-...", "private-ip4":"172.17.0.174", "mac":
"00:16:3f:01:00:00", "image-id":"centos_7_04_64_20G_alibase_201701015.vhd", "zone-id":
"cn-hangzhou-g", "owner-account-id":"...", "instance-type":"ecs.g5.large"}[root@ ~]# curl http://100.100.100.200/latest/dynamic/instance-identity/pkcs7?audience=test
MIIDMwYJKoZIhvcNAQcCoIIDJCcAACAQExCzAJBgUrDgMCGGUAMIIBBQYJKoZIhvcNAQcBoIIBXGSC
BSsOAwaIaBQAwDQYJKoZIhvcNAQEBBQAEggEAUOghxG3i3hKgjPgqPX6NInYNPOZJusp//fy15Pr+gZoq
LgvxonLzOXxBG1yy1aEzjb2b2zUFZGfjuNDbk1kcVSgq3kKEbpBCXF2GNm9FaS54r0szTcw9HkNPskla
CqR9Z6LvBB/sPMTz8i3dY8pu/zhiZOulHdpvKCYYP8Q89sN+QVlsS2eQDXnqNSFBi/QS/hZOz83gIuac
H6+rWXojBf3Hs7bft4YdmNBhaTpxq8R3wl6rWvTq3W58ULraHgMzq/Kn9e1SCnSAiqETLj7i60As7h/h
d/DebOVof0QANiXAIvdhLci4KK1rKJ0HOW4BzOy44s/jc1f1ASYIsAQrug==[root@ ~]#
```

4. Verify the instance identity by using OpenSSL.

```
openssl smime -verify -in $signature -inform PEM -
content $DOCUMENT -certfile AliyunPubk ey -noverify > /
dev / null
```



Note:

- Specify the variable `$signature` with the responded [instance identity signature](#).
- Specify the variable `$DOCUMENT` with the responded [instance identity document](#).

(Optional) In [step 3](#), if you have specified the `audience` parameter, add the value of the audience parameter at the end of the dynamically obtained instance identity document in the format of "`audience`": "`Value of the audience`", and separate the parameters with a comma (,).

- Specify the variable `AliyunPubk ey` with the [Alibaba Cloud public certificate](#).

The public certificate of Alibaba Cloud in all regions is as follows.

```
----- BEGIN      CERTIFICAT E -----
MIIDdzCCA1 + gAwIBAgIEZ mbRhZANBgk qhkiG9w0BA QsFADBsMRA
wDgYDVQQGE wdV
bmtub3duMR AwDgYDVQQI EwdVbmtub3 duMRAwDgYD VQQHEwdVbm
tub3duMRAw DgYD
VQQKEwdVbm tub3duMRAw DgYDVQQLEw dVbmtub3du MRAwDgYDVQ
QEwdVbmtu b3du
MB4XDTE4MD IyMzAxMjkz OFoXDTM4MD IxODAxMjkz OFowbDEQMA
4GA1UEBhMH VW5r
bm93bjEQMA 4GA1UECBMH VW5rbm93bj EQMA4GA1UE BxMHVW5rbm
93bjEQMA4G A1UE
ChMHVW5rbm 93bjEQMA4G A1UECXMHVW 5rbm93bjEQ MA4GA1UEAx
MHVW5rbm93 bjCC
ASIwDQYJKo ZIhvcNAQEB BQADggEPAD CCAQoCggEB AIJwy5sbZD
iNyX4mvdP3 2pqM
YMK4k7 + 5lRnVR2Fky / 5uwyGSPbdd NXaXzwEm + u4wIsJiaAN 30ZgJpYIoC
Gik + 9lG
5gVAIr0 +/ 3rZ61IbeVE + vDenDd8g / m / YIdYBfC2Ib zgS9EVGAf /
gJdtD0DXrD fQj
Fk2rQsvpft VOUs3Vpl90 + jeCQLoRbZY m0c5v7jP / L2lK0Mjhiy
wPF2kpDeis MtnD
/ ArkSPIlg1q VYm3F19v3p a6ZioM2hnw Xg5DibYlgV vsIBGhvYqd
Q1KosNVcVG GQa
HCUuVGdS7v HJYp3byH0v QYYygzxUJT 2TqvK7pD57 eYMN5drc7e
19oyRQvbPQ 3kkC
AwEAAAMhMB 8wHQYDVRO0 BBYEFAwwrn HlRgFvPGo + UD5zS1xAkC
91MA0GCSqG SIb3
DQEBCwUAA4 IBAQBBLhDR gezd / 00ppuYEVNB 9 + XiJ9dNmCuH
UhjNTnjkQ Wvk / YDA
v + T2V3t9yl8L 8o61tRIVKQ ++ lDhjlVmur / mbBN25 / UNRpJllfpU
H6o0aqvQAZ e4a
nRgyTnBwVB ZkdJ0d1siv L9NZ4pKelJ F3Ylw6rp0Y MqV + cwkt /
vRtzRJ31ZE eBhs7
vKh7F6BiGC HL5ZAwEUYe 803akQwjgr MUcfuiFs4 / sAeDMnmngN6
Uq8DFEBXDp AxVN
sV / 6Hockdfinx 85RV2AUwJG fClcVcu4hM h0vKROpcH2 7xu9bBIeMu
Y0vvzP2Vy0 m
DoJeqU7qZj yCaUBkPims z / 1eRod6d4P5 qxTj
----- END      CERTIFICAT E -----
```

Sample 1. No audience in the signature

Assume that you have published an image in the image market. The following example shows you how to grant access to the instances of your customers.

1. Connect to the target instance.
2. Verify whether the image used by the instance is from the [Alibaba Cloud Marketplace](#), or from another source, by calling the metadata items of `product - code` and `charge - type`. For more information, see [#unique_157](#).

```
curl http://100.100.100.200/latest/meta-data/image/market-place/product-code
```

```
curl http://100.100.100.200/latest/meta-data/image/market-place/charge-type
```

3. Create a temporary file `cert.cer` in the working directory and save the [public certificate](#) to the file.
4. Determine the identity of the instance by running the following script.

```
#!/usr/bin/bash
function verify_sig nature_wit hout_audie nce(){
curl 100.100.100.200/latest/dynamic/instance-identity/document > document
echo "----- BEGIN CERTIFICATE -----" > signature
curl 100.100.100.200/latest/dynamic/instance-identity/pkcs7 >> signature
echo "" >> signature
echo "----- END CERTIFICATE -----" >> signature
openssl smime -verify -in signature -inform PEM -content document -certfile cert.cer -noverify > /dev/null
}
verify_sig nature_wit hout_audie nce
```

5. Once the response result shows `Verification successful`, remove the restriction and run the image in the instance.

Sample 2. Audience in the signature

Assume that you published an image in the image market. The following example shows you how to grant access to the instances of your customers by specifying an `audience` parameter during the process of validation. To make sure that the instance identity is not maliciously acquired and used, you can implement the access control at the application server by combining your audience parameter. We recommend the value of the `audience` parameter is a random string, timestamp, regularly changed data, or some output generated by a specific algorithm.

1. Connect to the target instance.
2. Verify whether the image used by the instance is from the [Alibaba Cloud Marketplace](#), or another source, by calling the metadata items of `product-code` and `charge-type`.

```
curl http://100.100.100.200/latest/meta-data/image/market-place/product-code
```

```
curl http://100.100.100.200/latest/meta-data/image/market-place/charge-type
```

3. Create a temporary file `cert.cer` in the working directory and save the **public certificate** to the file.
4. Determine the identity of the instance by running the following script.

```
#!/usr/bin/bash
function verify_sig nature_wit h_specifie d_audience(){
  audience='your audience'# Here is your audience parameter.
  document=$(curl 100.100.100.200/latest/dynamic/instance-identity/document)
  audience_json=', "audience":'${audience}'
  echo -n "${document%?} ${audience_json} > document
  echo "-----BEGIN CERTIFICATE-----" > signature
  curl 100.100.100.200/latest/dynamic/instance-identity/pkcs7?audience=${audience} >> signature
  echo "" >> signature
  echo "-----END CERTIFICATE-----" >> signature
  openssl smime -verify -in signature -inform PEM -content document -certfile cert.cer -noverify > /dev/null
}
verify_sig nature_wit h_specifie d_audience
```

5. Once the response result shows `Verification successful`, remove the restriction and run the image in the instance.

8.12 Manage software on Linux instances

8.12.1 Add software repositories

Software packages for Linux are often stored in software repositories. After you add a repository, you can use the package management tool of the Linux system to search, install, and update the software in the repository. This topic uses Alibaba Cloud repositories as an example to describe how to add repositories on different Linux distributions.



Note:

Both Alibaba Cloud users and non-users have free access to software repositories of Alibaba Cloud. You can visit the [Alibaba Open Source Mirror Site](#) to obtain the software repository of the Linux distribution.

Add software repositories on CentOS

To add a software repository on the CentOS 7 operating system, follow these steps. The operations may vary depending on the version of your operating system.

1. Connect to a Linux instance. For more information, see [#unique_68/unique_68_Connect_42_section_fjm_rgx_wdb](#).
2. Run the following command to back up the original software repository.

```
sudo mv /etc/yum.repos.d/CentOS-Base.repo /etc/yum.repos.d/CentOS-Base.repo.backup
```

3. Run any of the following commands to add a software repository on CentOS 7.

```
sudo wget -O /etc/yum.repos.d/CentOS-Base.repo http://mirrors.aliyun.com/repo/Centos-7.repo
```

```
sudo curl -o /etc/yum.repos.d/CentOS-Base.repo http://mirrors.aliyun.com/repo/Centos-7.repo
```



Note:

- To view the steps of adding Alibaba Cloud repositories on other versions of the CentOS operating system, visit the [Alibaba Open Source Mirror Site](#), find `centos`, and click Help.
- To add a software repository that is not from Alibaba Cloud, replace the URL of `http://mirrors.aliyun.com/repo/Centos-7.repo` in the commands above with the URL of the software repository that you want to add.

4. Run the following command to generate a local cache for faster searching and installation of software.

```
sudo yum clean all && sudo yum makecache
```

5. Run the `sudo yum repolist` command to check whether the software repository is added. You will see the following content after you add a software repository of Alibaba Cloud on the CentOS 7 operating system.

```
[root@i-251234567890 ~]# yum repolist
Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
* base: mirrors.cloud.aliyuncs.com
* extras: mirrors.cloud.aliyuncs.com
* updates: mirrors.cloud.aliyuncs.com
repo id      repo name    status
base/7/x86_64 CentOS-7 - Base - mirrors.aliyun.com 10,019
extras/7/x86_64 CentOS-7 - Extras - mirrors.aliyun.com 409
updates/7/x86_64 CentOS-7 - Updates - mirrors.aliyun.com 1,945
repolist: 12,373
```

Add software repositories on Ubuntu

- To add a software repository on the Ubuntu 18.04 operating system, follow these steps . The operations may vary depending on the version of your operating system.

1. Connect to a Linux instance. For more information, see [#unique_68/unique_68_Connect_42_section_fjm_rgx_wdb](#).
2. Run the following command to back up the original software repository.

```
sudo cp / etc / apt / sources . list / etc / apt / sources . list . bakup
```

3. Run the `sudo vim / etc / apt / sources . list` command to open the file. Add the following content to the file. For more information, see [#unique_286](#).

```
deb http://mirrors.cloud.aliyuncs.com/ubuntu/bionic main restricted universe multiverse
deb-src http://mirrors.cloud.aliyuncs.com/ubuntu/bionic main restricted universe multiverse
deb http://mirrors.cloud.aliyuncs.com/ubuntu/bionic-security main restricted universe multiverse
deb-src http://mirrors.cloud.aliyuncs.com/ubuntu/bionic-security main restricted universe multiverse
deb http://mirrors.cloud.aliyuncs.com/ubuntu/bionic-updates main restricted universe multiverse
deb-src http://mirrors.cloud.aliyuncs.com/ubuntu/bionic-updates main restricted universe multiverse
deb http://mirrors.cloud.aliyuncs.com/ubuntu/bionic-proposed main restricted universe multiverse
deb-src http://mirrors.cloud.aliyuncs.com/ubuntu/bionic-proposed main restricted universe multiverse
deb http://mirrors.cloud.aliyuncs.com/ubuntu/bionic-backports main restricted universe multiverse
deb-src http://mirrors.cloud.aliyuncs.com/ubuntu/bionic-backports main restricted universe multiverse
```



Note:

- To view the steps of adding software repositories of Alibaba Cloud on other versions of Ubuntu, visit the [Alibaba Open Source Mirror Site](#), find `ubuntu`, and click Help. If you want to add a software repository of Alibaba Cloud on an Elastic Computing Service (ECS) instance, replace `http://mirrors.aliyun.com/ubuntu/` in the Help page with `http://mirrors.cloud.aliyuncs.com/ubuntu`. This helps save data transfer costs.
- To add a software repository that is not from Alibaba Cloud, replace the added content above with the content of the software repository that you want to add.

4. Run the `sudo apt - get update` command to update the information about software packages.

Add software repositories on Debian

To add a software repository on the Debian 8.9 operating system, follow these steps. The operations may vary depending on the version of your operating system.

1. Connect to a Linux instance. For more information, see [#unique_68/unique_68_Connect_42_section_fjm_rgx_wdb](#).
2. Run the following command to back up the original software repository.

```
sudo cp /etc/apt/sources.list /etc/apt/sources.list.backup
```

3. Run the `sudo vim /etc/apt/sources.list` command to open the file. Add the following content to the file. For more information about how to add the content, see [#unique_286](#).

```
deb http://mirrors.cloud.aliyuncs.com/debian/jessie
main non-free contrib
deb http://mirrors.cloud.aliyuncs.com/debian/jessie
-proposed-updates main non-free contrib
deb-src http://mirrors.cloud.aliyuncs.com/debian/
jessie main non-free contrib
deb-src http://mirrors.cloud.aliyuncs.com/debian/
jessie-proposed-updates main non-free contrib
```

**Note:**

- To view the steps of adding software repositories of Alibaba Cloud on other versions of Debian, visit the [Alibaba Open Source Mirror Site](#), find `debian`, and click Help. If you want to add a software repository of Alibaba Cloud on an ECS instance, replace `http://mirrors.aliyun.com/debian/` in the Help page with `http://mirrors.cloud.aliyuncs.com/ubuntu`. This helps save data transfer costs.
- To add a software repository that is not from Alibaba Cloud, replace the added content above with the content of the software repository that you want to add.

4. Run the `sudo apt-get update` command to update the information about software packages.

Add software repositories on Fedora

1. Run the following command to back up the original software repository.

```
# sudo mv /etc/yum.repos.d/fedora.repo /etc/yum.repos.d/fedora.repo.backup
```

```
# sudo mv /etc/yum.repos.d/fedora-updates.repo /  
etc/yum.repos.d/fedora-updates.repo.backup
```

2. Run any of the following commands to add software repositories of Alibaba Cloud on Fedora.

```
sudo wget -O /etc/yum.repos.d/fedora.repo http  
://mirrors.aliyun.com/repo/fedora.repo
```

```
sudo curl -o /etc/yum.repos.d/fedora.repo http  
://mirrors.aliyun.com/repo/fedora.repo
```

**Note:**

To add a software repository that is not from Alibaba Cloud on Fedora, replace the URL of `http://mirrors.aliyun.com/repo/fedora.repo` in the commands with the URL of the software repository that you want add.

3. Run any of the following commands to add software repositories of Alibaba Cloud on fedora-updates.

```
sudo wget -O /etc/yum.repos.d/fedora-updates.repo  
http://mirrors.aliyun.com/repo/fedora-updates.repo
```

```
sudo curl -o /etc/yum.repos.d/fedora-updates.repo  
http://mirrors.aliyun.com/repo/fedora-updates.repo
```

**Note:**

To add a software repository that is not from Alibaba Cloud on fedora-updates, replace `http://mirrors.aliyun.com/repo/fedora-updates.repo` in the commands with the URL of the software repository that you want.

4. Run the following command to generate a local cache.

```
sudo yum clean all && sudo yum makecache
```

Next steps

You can install software packages after you add a software repository. For more information, see [#unique_287](#).

8.12.2 Search for software packages

If you want to install a software application but do not know the name of the software package, you can use the tools in the Linux system to search for the package. This

topic describes the commands that you can use to search for software packages and how to use these commands on different Linux distributions.

Search for software packages on CentOS

- Command

```
yum search [ keywords of software packages ]
```

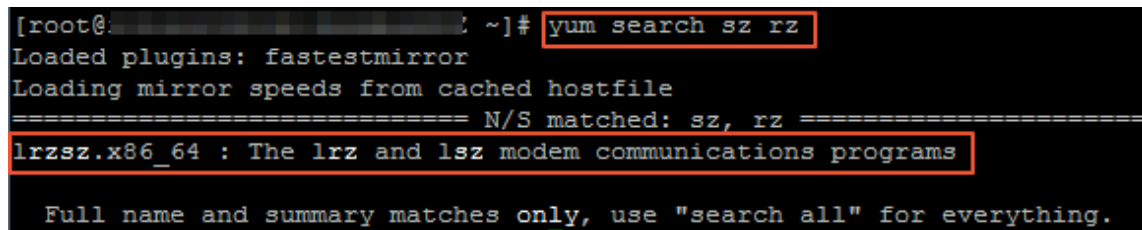
- Examples

- Example 1: Search for two software applications with the keywords "rz" and "sz".

Run the following command to search for the software packages to which the two applications belong.

```
yum search sz rz
```

The command result indicates that these two software applications belong to the `lrzsz.x86_64` package, as shown in the following figure. After you install the package, you can use the `lrz` and `lsz` software applications.



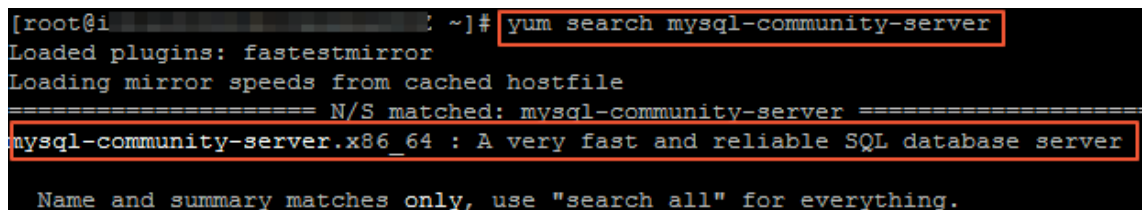
```
[root@i-xxxxx: ~]# yum search sz rz
Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
===== N/S matched: sz, rz =====
lrzsz.x86_64 : The lrz and lsz modem communications programs
Full name and summary matches only, use "search all" for everything.
```

- Example 2: Search for a software application by using the keyword "mysql-community-server".

Run the following command to search for the software package to which the application belongs.

```
yum search mysql - community - server
```

The command result indicates the package to which the software application belongs, as shown in the following figure.



```
[root@i-xxxxx: ~]# yum search mysql-community-server
Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
===== N/S matched: mysql-community-server =====
mysql-community-server.x86_64 : A very fast and reliable SQL database server
Name and summary matches only, use "search all" for everything.
```



Note:

If you cannot find the `mysql-community-server` package in the CentOS operating system, it may be because you have not added the repository of the software package. For more information about how to add the required software repository, and install and use MySQL Server, see [#unique_289](#).

Search for software packages on Debian or Ubuntu

- Command

```
apt - cache search [ keywords of software packages ]
```

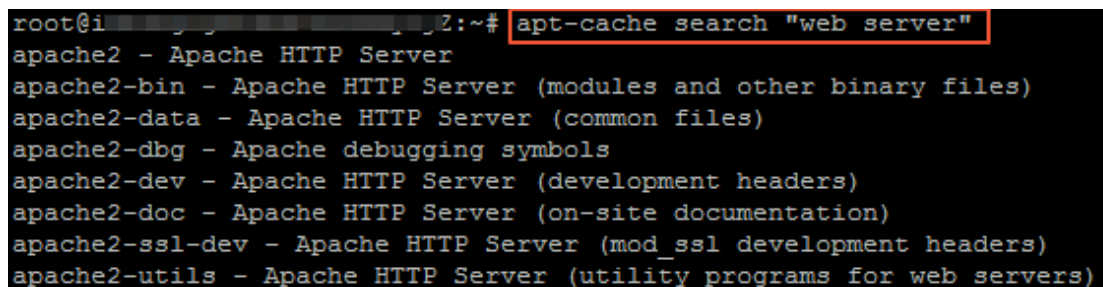
- Examples

- Example 1: Search for software packages relating to Web servers.

Run the following command to search for the software packages.

```
apt - cache search " web server "
```

The command result indicates that *apache2* is the software package of the Web server, as shown in the following figure.



```
root@i-2:~# apt-cache search "web server"
apache2 - Apache HTTP Server
apache2-bin - Apache HTTP Server (modules and other binary files)
apache2-data - Apache HTTP Server (common files)
apache2-dbg - Apache debugging symbols
apache2-dev - Apache HTTP Server (development headers)
apache2-doc - Apache HTTP Server (on-site documentation)
apache2-ssl-dev - Apache HTTP Server (mod_ssl development headers)
apache2-utils - Apache HTTP Server (utility programs for web servers)
```

- Example 2: Search for Java development environment by using the keyword "openjdk".

Run the following command to search for "openjdk".

```
apt - cache search openjdk
```

The command result shows the software packages relating to openjdk, as shown in the following figure.

```
root@i72eaddg721e0ar3dijw03:~# apt-cache search openjdk
java-package - Utility for creating Java Debian packages
jvm-7-avian-jre - lightweight virtual machine using the OpenJDK class library
freemind - Java Program for creating and viewing Mindmaps
icedtea-7-plugin - web browser plugin based on OpenJDK and IcedTea to execute Java applets
default-jdk - Standard Java or Java compatible Development Kit
default-jdk-doc - Standard Java or Java compatible Development Kit (documentation)
default-jre - Standard Java or Java compatible Runtime
default-jre-headless - Standard Java or Java compatible Runtime (headless)
jtest - Regression Test Harness for the OpenJDK platform
libreoffice - office productivity suite (metapackage)
icedtea-7-jre-jamvm - Alternative JVM for OpenJDK, using JamVM
openjdk-7-dbg - Java runtime based on OpenJDK (debugging symbols)
openjdk-7-demo - Java runtime based on OpenJDK (demos and examples)
openjdk-7-doc - OpenJDK Development Kit (JDK) documentation
openjdk-7-jdk - OpenJDK Development Kit (JDK)
openjdk-7-jre - OpenJDK Java runtime, using Hotspot JIT
openjdk-7-jre-headless - OpenJDK Java runtime, using Hotspot JIT (headless)
openjdk-7-jre-lib - OpenJDK Java runtime (architecture independent libraries)
openjdk-7-jre-zero - Alternative JVM for OpenJDK, using Zero/Shark
openjdk-7-source - OpenJDK Development Kit (JDK) source files
openjdk-7-jre-dcevm - Alternative VM for OpenJDK 7 with enhanced class redefinition
uwsgi-app-integration-plugins - plugins for integration of uWSGI and application
uwsgi-plugin-jvm-openjdk-7 - Java plugin for uWSGI (OpenJDK 7)
uwsgi-plugin-jwsgi-openjdk-7 - JWSGI plugin for uWSGI (OpenJDK 7)
```

Next steps

After you find the software packages, you may need to install them. For more information about how to install software packages, see [#unique_287](#).

8.12.3 Install software packages

After you add software repositories, you can use the package management tool of the Linux system to install the packages you need. This topic uses Apache HTTP Server in Alibaba Cloud repositories as an example to describe how to install software packages on different Linux distributions.

Prerequisites

You have added the repositories of software packages before you install them. You must add Alibaba Cloud repositories before following the steps in the examples below. For more information about how to add repositories, see [#unique_291](#).

Install software packages on CentOS

1. Connect to a Linux instance. For more information, see [#unique_68/unique_68_Connect_42_section_fjm_rgx_wdb](#).

2. Run the following command to install software packages.

```
yum install < package > # Replace < package > with the
software package you want to install .
```

For example, to install Apache HTTP Server, run the following command.

```
yum install httpd
```

Install software packages on Debian or Ubuntu

1. Connect to a Linux instance. For more information, see [#unique_68/unique_68_Connect_42_section_fjm_rgx_wdb](#).
2. Run the following command to install software packages.

```
# apt - get update
# apt - get install < package > # Replace < package > with
the software package you want to install .
```

For example, to install Apache HTTP Server, run the following command.

```
# apt - get update
# apt - get install apache2
```

Install software packages on openSUSE

1. Connect to a Linux instance. For more information, see [#unique_68/unique_68_Connect_42_section_fjm_rgx_wdb](#).
2. Run the following command to install software packages.

```
zypper install < package > # Replace < package > with the
software package you want to install .
```

For example, to install Apache HTTP Server, run the following command.

```
zypper install apache2
```

8.12.4 Update software

The software packages in Linux distributions are constantly updated to add new functions, bug fixes, and security updates. You can update software to the latest

version based on business requirements. This topic describes how to update a specific software application or all applications on different Linux distributions.

Prerequisites

You have added the repository of software packages before you update the applications. For more information about adding software repositories, see [#unique_291](#).

Update software on CentOS

1. Connect to a Linux instance. For more information, see [#unique_68/unique_68_Connect_42_section_fjm_rgx_wdb](#).
2. Update software applications.

- To update a software application, run the following command.

```
yum update < package > # Replace < package > with the software application you want to update .
```

For example, to update Apache HTTP Server, run the following command:

```
yum update httpd
```

- To update all software applications in the system, run the following command.



Notice:

If you run this command, the update process may update the kernel of the operating system and cause some issues. For example, the system may not start up or certain software applications may be incompatible with the updated kernel. Therefore, before you run this command, we recommend that you configure the system to skip kernel changes while applying updates.

```
yum update
```

After updating a software application, you can check its version. If it is the latest version, the software application has been updated.

Update software applications on Ubuntu or Debian

1. Connect to a Linux instance. For more information, see [#unique_68/unique_68_Connect_42_section_fjm_rgx_wdb](#).

2. Run the following command to obtain a list of software packages that need to be updated.

```
apt - get update
```

3. Update software applications.

- To update a software application, run the following command.

```
apt - get install < package > # Replace < package > with  
the software application you want to update .
```

For example, to update Python, run the following command:

```
apt - get install python
```

- To update all software applications in the system, run the following command.



Notice:

If you run this command, the update process may update the kernel of the operating system and cause some issues. For example, the system may not start up or certain software applications may be incompatible with the updated kernel. Therefore, before you run this command, we recommend that you configure the system to skip kernel changes while applying updates.

```
apt - get upgrade
```

After updating a software application, you can check its version. If it is the latest version, the software application has been updated.

Update software on openSUSE

1. Connect to a Linux instance. For more information, see [#unique_68/unique_68_Connect_42_section_fjm_rgx_wdb](#).

2. Run the following command to obtain a list of software packages that need to be updated.

```
zypper list - updates
```

3. Update software applications.

- To update a software application, run the following command.

```
zypper update < package > # Replace < package > with the  
software application you want to update .
```

For example, to update Python, run the following command.

```
zypper update python
```

- To update all software applications in the system, run the following command.



Notice:

If you run this command, the update process may update the kernel of the operating system and cause some issues. For example, the system may not start up or certain software applications may be incompatible with the updated kernel. Therefore, before you run this command, we recommend that you configure the system to skip kernel changes while applying updates.

```
zypper update
```

After updating a software application, you can check its version. If it is the latest version, the software application has been updated.

8.13 Configure time

8.13.1 Time setting: NTP servers and other public services

Alibaba Cloud ECS provides standard intranet NTP servers, which you can access from your instances. We also provide external NTP services for instances that need the Internet access.

Intranet and Internet NTP servers

To counterbalance the leap seconds in our world, ECS provides free of charge, highly accurate, and reliable NTP service for both classic network- and VPC-Connected instances. Among the NTP servers, the `ntp . cloud . aliyuncs . com` achieves

nearly zero difference of atomic reference by synchronizing with satellite services. See the following table for the NTP servers provided by Alibaba Cloud ECS.

Classic network intranet	VPC intranet	Internet
-	ntp.cloud.aliyuncs.com	ntp1.aliyun.com
ntp1.cloud.aliyuncs.com	ntp7.cloud.aliyuncs.com	ntp2.aliyun.com
ntp2.cloud.aliyuncs.com	ntp8.cloud.aliyuncs.com	ntp3.aliyun.com
ntp3.cloud.aliyuncs.com	ntp9.cloud.aliyuncs.com	ntp4.aliyun.com
ntp4.cloud.aliyuncs.com	ntp10.cloud.aliyuncs.com	ntp5.aliyun.com
ntp5.cloud.aliyuncs.com	ntp11.cloud.aliyuncs.com	ntp6.aliyun.com
ntp6.cloud.aliyuncs.com	ntp12.cloud.aliyuncs.com	ntp7.aliyun.com

Other public services of Alibaba Cloud ECS

See the following list for some public services provided by Alibaba Cloud ECS.

Public service	Description
Public DNS: 223.5.5.5 / 223.6.6.6	Domain name: <code>http://www.alidns.com</code>
Open source images: <code>http://mirrors.aliyun.com</code>	Update frequency: The image files are updated at everyday 02:00–04:00 (UTC+8:00), including a lot of Linux distributions and open source applications.

8.13.2 Time setting: Synchronize NTP servers for Windows instances

Network Time Protocol (NTP) is a networking protocol for clock synchronization between computer systems over networks. For highly time-sensitive applications (such as those in the communication industry), clock variation between different computers may lead to serious data inconsistencies. You can use the NTP service to synchronize clocks of all servers within the network. The current default time zone for Alibaba Cloud ECS instances across all regions is CST (China Standard Time).

This article describes how use the NTP service to synchronize the clock of a Windows ECS instance running Windows Server 2008 R2 Enterprise Edition x64.

Windows Time service is enabled by default on Windows Server. You must enable the NTP service in the instance to make sure that the NTP service can normally

synchronize time after successful NTP service configuration. To check and enable the NTP service, follow these steps:

1. [Connect to a Windows instance](#). Select Start > All Programs > Accessories > Run to open the Run dialog box, and run `services . msc` .
2. In the Services window, double click the Windows Time service.
3. In the Windows Time Properties (Local Computer) dialog box, follow these steps:
 - a. Set Startup type to Automatic.
 - b. Check if the Service status is Started. If not, click Start.

After completing the settings, click Apply, and then click OK.

Modify the default NTP server address

time.windows.com is used as the default NTP server in Windows Server, but synchronization errors may frequently occur due to network issues. When using a Windows instance, you can replace the default NTP server with the intranet NTP server provided by Alibaba Cloud. For more information, see [Internet and intranet NTP servers](#). To modify the default NTP server address, follow these steps:

1. [Connect to a Windows instance](#).
2. In the notification area of the task bar, click Date and Time, and then click Change date and time settings.
3. In the Date and Time dialog box, click the Internet Time tab, and then click Change settings.
4. In the Internet Time Settings dialog box, select Synchronize with an Internet time server, type an Alibaba Cloud intranet NTP server address (for detailed list, see [Internet and intranet NTP servers](#)), and then click Update now.

You are prompted if the synchronization is successful.

Modify NTP synchronization interval

The default NTP synchronization interval is 5 minutes. To modify the NTP synchronization interval, follow these steps:

1. [Connect to a Windows instance](#).
2. Select Start > All Programs > Accessories > Run to open the Run dialog box, and run `regedit` .

3. On the left-side navigation pane of the Registry Editor, find `HKEY_LOCAL_MACHINE / SYSTEM / CurrentControlSet / services / W32Time / TimeProviders / NtpClient`, and then double click `SpecialPollInterval`.
4. In the Edit DWORD (32-bit) Value dialog box, select Decimal as the Base, and then type the Value data as needed. The number you typed is the synchronization interval you need. Unit: seconds.

8.13.3 Time setting: Synchronize NTP servers and change time zone for Linux instances

The current default time zone for Alibaba Cloud ECS instances across all regions is CST (China Standard Time). In addition, the NTP (Network Time Protocol) service guarantees that your instances are synchronized with the standard time. Follow these steps in this topic to change the time zone for your ECS instances and configure your NTP service.

Context

Synchronizing time and the time zone is crucial for Elastic Compute Service (ECS) instances, for example, an inaccurate time may have a significant impact on business when updating your database. To avoid both business disruptions running on your instances and networking request errors, you must configure one or more instances in the same time zone, such as `Asia / Shanghai` or `America / Los Angeles`. Take CentOS 6.5 as an example to demonstrate how to change the time zone by modifying configuration file.



Note:

After you change the time zone for an instance, always run `hwclock -w` to update the real-time clock (RTC) of the instance.

Procedure

1. [Connect](#) to the Linux instance.



Note:

Only a root user can open and edit time zone configuration files, so we use the `sudo` command here.

2. Run `sudo rm /etc/localtime` to delete the local time in the instance.

3. Run `sudo vi / etc / sysconfig / clock` to edit the configuration file `/etc/sysconfig/clock`.
4. Enter `i` to add the time zone and city. For example, add `Zone = Asia / Shanghai`. Press Esc to exit the edit and enter `: wq` to save and exit.

Optional. Run `ls / usr / share / zoneinfo` to query the list of available time zones. For example, `Shanghai` is one of them.
5. Run `sudo ln -sf / usr / share / zoneinfo / XXXX / XXXXXXXX / etc / localtime` to update the time zone change, for example, run `sudo ln -sf /usr/share/zoneinfo/Asia/Shanghai /etc/localtime`.
6. Run `hwclock -w` to update the RTC.
7. Run `sudo reboot` to restart the instance.
8. Run `date -R` to check whether the new time zone is effective or not. If not, repeat the preceding steps.

What's next

The Linux instance offers the `ntpdate` and the `ntpd` two approaches of synchronizing the NTP service. The `ntpdate` can be used to force an immediate update and the `ntpd` offers a systematic approach. The `ntpdate` service can be used for new instances, whereas `ntpd` is recommended for instances that run your business. Both standard and custom NTP service configurations are provided in this section. For more information about the NTP service, see [internal and public NTP server](#).

Prerequisites

The communication port of the NTP service is UDP 123. Before configuring the service, make sure that you enabled the UDP port 123. You can use `netstat -nupl` in the Linux instance to make sure whether the UDP port 123 is enabled or not. For more information, see [add a security group rule](#).

Set up standard NTP service

1. [Connect](#) to the Linux instance.
2. Run `sudo service ntpd start` to enable the NTP service.
3. Run `chkconfig ntpd on` to enable the NTP service.
4. Run `ntostat` to check whether the NTP service is enabled or not.

5. Optional. Run `ntpq -p` to view a list of NTP service peers. Run `sudo chkconfig --list ntpd` to view the NTP service running level.

Set up custom NTP service

1. [Connect](#) to the Linux instance.
2. Run `sudo vi /etc/ntp.conf` to edit the NTP service configuration files.
3. After finding the information about `ntp server XXXX iburst`, enter `i` and start editing the file. NTP servers that are not currently needed can be hidden by adding a pound (#) at the beginning of the lines.
4. Add a new line of NTP server information in the format of `server XXXX iburst`, and the XXXX is the custom NTP endpoint. For more information, see [Internet and intranet NTP servers](#). After editing, press Esc and enter `:wq` to save and exit.
5. Run `sudo service ntpd start` to enable the customized NTP service.
6. Run `chkconfig ntpd on` to enable the NTP service.
7. Run `ntpstat` to check whether the NTP service is enabled or not.

9 Renew Subscription instances

9.1 Renewal overview

If you want to continue using a Subscription instance after it expires, you must renew the instance within the prescribed period. Otherwise, both the instance and disk are automatically released. The stored data is erased permanently and cannot be recovered. For more information about what happens when a Subscription instance expires, see [Subscription](#).



Note:

- You cannot cancel a paid renewal order.
- Pay-As-You-Go instances do not involve renewal. Make sure you have sufficient balance on your linked credit card or PayPal account to cover the costs.

Subscriptions to Alibaba Cloud ECS can be renewed by the following means:

- [Manual renewal](#)
 - The instance configurations cannot be changed at the time of renewal.
 - Renewal cycle: One month or one year.
- [Auto-renewal](#). You can enable the auto-renewal feature for an instance at or after instance creation. The instance is automatically renewed before it expires.
 - The instance configuration is unchangeable at the time of renewal.
 - Renewal cycle: Renewed based on the current billing cycle of the instance.
- [Renewal for Configuration Downgrade](#). You can downgrade the configuration of an instance and its associated resources when renewing the instance. The downgraded configuration takes effect at the start of the next billing cycle, which can reduce the costs.
 - The instance configuration is changeable at the time of renewal.
 - Renewal cycle: One month or one year.

9.2 Manual renewal

Introduction

Manual renewal only applies to instances that use the Subscription billing method.

You can manually renew your Subscription instances when they are in an `Expired` status or are shut down. You can manually renew your instance for a month or a year. Therefore, if you want to modify the service duration of your Subscription instances, you can choose manual renewal.

- Your instance will still work normally when the instance is in an `Expired` status. If manual renewal is successfully completed within 15 days of the instance expiring, the start of the next billing cycle will be the day the instance expired.

For example, if your instance expired at 00:00:00 on April 25, 2016, but you successfully renewed it for one month on May 9, 2016, the billing cycle for this renewal is from April 25, 2016 to 00:00:00 on May 25, 2016.

- If the instance fails to be renewed within 15 days of expiration, the instance will be shut down.
- After the instance is shut down:
 - Your instance will stop providing services, but your data will be retained for a further 15 days.
 - If the instance is renewed within 15 days of the instance being shut down (within 30 days of the instance expiring), your instance will enter the new billing cycle from the day of renewal, and your data will be retained for the new cycle.

For example, if your instance was shut down at 00:00:00 on May 10, 2016, but you successfully renewed it for one month at 08:09:35 on May 23, 2016, the billing cycle for this renewal is from 08:09:35 on May 23, 2016 to 00:00:00 on June 24, 2016.

- If the instance is not renewed within 15 days of the instance being shut down (within 30 days of the instance expiring), your instance will be automatically released on the 15th day. Your data will be deleted and cannot be restored.

Procedure

To manually renew your instance, follow these steps:

1. Log on to the [ECS console](#).

2. In the left-side navigation pane, click Instances.
3. On the Instance List page, select the expected region and locate the ECS instance by the instance name, instance ID, or status (Expired).
4. In the Actions column, click Renew.
5. On the renewal page,
 - a. Confirm the instance configuration.
 - b. Select the expected renewal length, which can be 1 Month or 1 Year, and click Place Order.
6. On the Pay page, confirm the order information and click Pay to complete the renewal operation.

Bulk renewal

Perform the steps below to renew multiple instances:

1. Log on to the [ECS console](#).
2. In the left-side navigation pane, click Instances.
3. Select a region.
4. Select multiple Subscription or Weekly Subscription instances.
5. At the bottom of the page, click Renew.
6. On the Bulk Instance Renewal page, click Batch Change. Select a renewal period and then click OK.
7. Click Create Order to finish the operation as instructed.

9.3 Auto-renewal

Auto-renewal service only applies to instances that use the Subscription billing method.

Introduction

If you have activated the auto-renewal service, Alibaba Cloud charges the subscription fee to your linked credit card or PayPal account when the instance expires.

The auto-renewal service can be activated after the ECS instance is purchased and before it expires. It cannot be activated after a Subscription instance expires. Auto-renewal has the following features:

- The monthly subscription service automatically renews the instance on a monthly basis when a monthly subscription instance expires.

- The annual subscription service automatically renews the instance on a yearly basis when a yearly subscription instance expires.

**Note:**

The auto-renewal service does not support switching between monthly subscription and annual subscription. If you want to change the service duration of an instance, you can choose the [Manual renewal](#) service.

After you activate the auto-renewal service:

- You are notified of the imminent expiration of your Subscription instances seven days, three days, and one day before the expiration day (T).
- Alibaba Cloud charges the subscription fee to your linked credit card or PayPal account on the expiration day (T). If payment fails, Alibaba Cloud will try to take payment again on Day 7 (T+6) and Day 15 (T+14). If all the three payment attempts fail, the instance is shut down.

- If the payment for the subscription is successful, your instance is no longer in an Expired status and the next billing cycle starts from the expiration day.

For example, if your monthly subscription instance expired at 00:00:00 on April 25, 2016, but it was successfully renewed automatically on May 9, 2016, the billing cycle for this renewal is from 00:00:01 on April 25, 2016 to 00:00:00 on May 25, 2016.

- If all the three payment attempts fail, the instance shuts down 15 days after its expiration day. If the instance shuts down, it stops providing services and you cannot log on or remotely connect to the instance. At this point, you can only choose [Manual renewal](#). If the instance is not renewed within the 15 days after the expiration day, the instance is released and the data stored is lost.
 - If you manually renew the instance before auto-renewal is attempted, your instance is renewed and no auto-renewal is attempted for the current billing cycle. The instance will then be renewed when the current billing cycle ends.
 - Alibaba Cloud sends a notification email to your linked email address for each failed auto-renewal attempt. Therefore, we recommend that you check your inbox frequently so you can keep up to date with the status of your instance and take necessary actions to avoid further business impact.
- Alibaba Cloud takes payment for the auto-renewal of instances between 08:00:00 (UTC+8) and 18:00:00 (UTC+8).

Activate auto-renewal

To activate the auto-renewal service, follow these steps:

1. Log on to the [ECS console](#).
2. At the top of the page, choose Billing Management > Renew.
3. In the left-side navigation pane, click Elastic Compute Service.
4. On the Renew page, select the Manually Renew tab.
5. Find an instance, and in the Actions column, click Enable Auto-Renew.
6. On the Enable Auto-Renew dialog box, click Enable Auto-Renew.

You can then find the instance by selecting the Auto-Renew tab.

Deactivate auto-renewal

To deactivate the auto-renewal service for an instance, follow these steps:

1. Log on to the [ECS console](#).
2. At the top of the page, choose Billing Management > Renew.
3. In the left-side navigation pane, click Elastic Compute Service.
4. On the Renew page, select the Auto-Renew tab.
5. Find the instance, and in the Actions column, click Modify Auto-Renew.
6. On the Modify Auto-Renew dialog box, select Disable Auto-Renew and click OK.

You can then find the instance by selecting the Manually Renew tab.

9.4 Renew for configuration downgrade

This topic describes how to downgrade the configuration of a Subscription instance when you renew the instance.

After a Subscription instance expires, if renewal has not been completed in the required time, the instance is released automatically. In this case, data is lost permanently and cannot be recovered. To view status changes that occur after Subscription resources expire, see [#unique_6](#).

You can use the Renew for Configuration Downgrade feature to downgrade the specifications of a Subscription instance before you update its renewal to help lower costs incurred in the next billing cycle.

You can also change the billing method of your data disks from Subscription to Pay-As-You-Go.

Limits

When you use the Renew for Configuration Downgrade feature, the following limits apply:

- The feature allows you to scale down instance specifications at the time of renewal.
- Renew for Configuration Downgrade can only be used for Subscription instances.
- You can downgrade the configuration of an instance 15 days prior to its expiration date, or 15 days from its expiration date, but before the instance is released.

For example, if you have a Subscription instance that expires on April 30, you can downgrade the specifications of the instance and renew it between the dates of April 16 to April 30. If you do not renew the instance at this time, the instance enters the Recycling Upon Expiration state. If the instance enters this state, you can still renew the instance from May 1 to May 15. If you do not renew the instance, it is automatically released on May 16.

- After you renew an instance, the new package is effective from the next billing cycle. The current package continues until the end of the current billing cycle.
- If instance configurations are changed during renewal, you must [restart the instance](#) within the first seven days of the new billing cycle for the new configurations to be effective. If you restart the instance on the seventh day of the new billing cycle, the instance is considered to have used the original package for the first six days, and uses the downgraded package only after it is restarted.
- Once the renewal for configuration downgrade is complete, you cannot [upgrade configurations](#), [increase the system disk size](#), or increase the size of a Subscription data disk, which is attached to a [Linux instance](#) or a [Windows instance](#), during the rest of the current billing cycle.
- You cannot cancel the renewal orders once the payment is processed.

Procedure

To downgrade the configuration of a Subscription instance during renewal, follow these steps:

1. Find the Subscription instance. In the Actions column, click Change Configuration.
2. In the Configuration Change Guide dialog box, select Renew for Configuration Downgrade and click Continue.

3. On the Renew for Configuration Downgrade page, complete the following operations as necessary:

- Downgrade Instance Type. If you want to downgrade specifications for an instance, you must also set the Restart Time for that instance.



Note:

- Instance specifications that can be changed are displayed on the page. For more information about instance specifications, see [Instance type families](#).
 - Restarting an instance suspends your business operations on that instance. Please restart the instance during off-peak hours to reduce service impact. The restart must be completed within the first seven days of the next billing cycle.
- If a data disk was created while creating a Subscription instance, you can change the billing method of the data disk to Pay-As-You-Go. If the billing method is not changed, the data disk has the same billing cycle as the Subscription instance.
 - Set the bandwidth value.
 - Set the renewal duration.

4. Click Pay, and follow the prompts to complete the process.

5. If you have changed instance specifications, or changed the public network bandwidth of an instance in a classic network from 0 Mbit/s to a non-zero value for the first time, you must restart the instance in the console or by using the RebootInstance API within the first seven days of the next billing cycle for the new configurations to be effective.



Note:

For a VPC-connected instance, the instance does not need to be restarted.

10 Change configurations

10.1 Overview of configuration changes

You can change the configurations of an instance and its Internet bandwidth after the instance is created.

Upgrade or downgrade instance configurations

You must upgrade or downgrade the configurations of vCPUs and memory simultaneously by changing the instance type ([instance type family](#)). The following table provides the configuration change methods suitable for the billing method applied to your instance.



Note:

Before changing instance configurations, you need to check [instance type families that support instance type upgrades](#) and confirm the target configuration supported by a specific instance type family.

Billing method	Configuration upgrade method	How does the change take effect?	Configuration downgrade method	How does the change take effect?
Subscription	#unique_313	The new configurations take effect after you restart the instance in the ECS console or by calling the API action RebootInstance .	<ul style="list-style-type: none"> · #unique_302 · On-demand configuration downgrade (Whether this feature is supported is determined by your ECS instance resource usage.) 	<ul style="list-style-type: none"> · Renew for configuration downgrade: The new configurations take effect after you restart the instance in the ECS console within the first seven days of the new billing cycle. · On-demand configuration downgrade: The new configurations take effect after you restart the instance in the ECS console or by calling the API action RebootInstance.
Pay-As-You-Go	#unique_121	The new configurations take effect after you restart the instance.	#unique_121	The new configurations take effect after you restart the instance.

Adjust Internet bandwidth

You can adjust the Internet bandwidth of an instance. The methods vary according to your business needs and the billing method of the instance. The following table lists the methods.

Billing method	Supports permanent upgrade?	Does the change take effect immediately?	Feature	Description
Subscription	Yes	Yes	#unique_313	Only applicable to VPC-connected ECS instances to which no EIP addresses are attached or classic network-connected ECS instances . The public and private IP addresses remain unchanged after you increase Internet bandwidth of the instance.

Billing method	Supports permanent upgrade?	Does the change take effect immediately?	Feature	Description
Subscription	Yes	Effective from next billing cycle	#unique_302	You can adjust the Internet bandwidth when you renew the instance. When the Internet bandwidth is reduced to 0 Mbit/s, the public IP address of a VPC-connected instance is released in the new billing cycle, but the public IP address of a classic network-connected ECS instance is retained.
Pay-As-You-Go or Subscription	Yes	Yes	#unique_315	Only applicable to VPC-connected instances to which EIP addresses are attached. You can adjust the Internet bandwidth on an EIP address at any time during the current lifecycle of the instance.

Assign a public IP address

You can assign a public IP address to an ECS instance when you create the instance. If no public IP addresses are assigned at instance creation, you can assign a public IP address after an ECS instance is created (the instance must be a Subscription instance). For more information, see the following table.

Feature	Does the change take effect immediately?	Description
#unique_313	Yes	Only applicable to VPC-connected ECS instances to which no EIP addresses are attached, or classic network-connected ECS instances. Set the Internet bandwidth to a non-zero value to assign a public IP address.
#unique_302	Effective from next billing cycle	

10.2 Instance type families that support instance type upgrades

This article describes the instance type families that support instance type upgrades.

Restrictions

Upgrading instance types has the following impacts:

- Classic network instances:
 - For [phased-out instance types](#), when a non-I/O optimized instance is upgraded to an I/O optimized instance, changes are made to the private IP address, the driver name, and the software authorization code. For Linux instances, Basic Cloud Disks (`cloud`) are recognized as `xvda` or `xvdb` , while Ultra Cloud Disks (`cloud_efficiency`) and SSD Cloud Disks (`cloud_ssd`) are recognized as `vda` or `vdb` .
 - For [available instance types](#), changes are made to the private IP address of the instance.
- VPC instances:

For [phased-out instance types](#), when a non-I/O optimized instance is upgraded to an I/O optimized instance, changes are made to the driver name and the

software authorization code. For Linux instances, Basic Cloud Disks (`cloud`) are recognized as `xvda` or `xvdb` , while Ultra Cloud Disks (`cloud_efficiency`) and SSD Cloud Disks (`cloud_ssd`) are recognized as `vda` or `vdb` .

Instance type families that support upgrading instance types



Note:

Each instance type is available only in specific zones. Before upgrading an instance type, check if the target instance type (family) is available in the current zone.

In the following table, the target instance type families apply to both Subscription and Pay-As-You-Go instances.

Source instance type family	Target instance type family
g5, r5, c5, ic5	<ul style="list-style-type: none"> g5, r5, c5, ic5 sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5, hfg5, re4, t5, n4, mn4, xn4, e4
sn1ne, sn2ne, se1ne	<ul style="list-style-type: none"> sn1ne, sn2ne, se1ne c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4
se1	<ul style="list-style-type: none"> se1 sn1, sn2, n1, n2, e3, sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4
n4, mn4, xn4, e4	<ul style="list-style-type: none"> n4, mn4, xn4, e4 sn1, sn2, se1, n1, n2, e3, sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, re4, t5
re4	<ul style="list-style-type: none"> re4 sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, t5, n4, mn4, xn4, e4, ecs.se1.14xlarge
hfc5, hfg5	<ul style="list-style-type: none"> hfc5, hfg5 sn1ne, sn2ne, se1ne, c4, cm4, ce4, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4
gn4	gn4
gn5i	gn5i

Source instance type family	Target instance type family
gn6v	gn6v
t5	<ul style="list-style-type: none"> · t5 · sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, re4, n4, mn4, xn4, e4
t1, s1, s2, s3, m1, m2, c1, c2	<ul style="list-style-type: none"> · t1, s1, s2, s3, m1, m2, c1, c2 · sn1, sn2, se1, n1, n2, e3, sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4
n1, n2, e3	<ul style="list-style-type: none"> · n1, n2, e3 · sn1, sn2, se1, sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4
sn1, sn2	<ul style="list-style-type: none"> · sn1, sn2 · se1, n1, n2, e3, sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4
c4, ce4, cm4	<ul style="list-style-type: none"> · c4, ce4, cm4 · sn1ne, sn2ne, se1ne, hfc5, hfg5, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4

10.3 Upgrade configurations

10.3.1 Upgrade configurations of Subscription instances

This topic describes how to upgrade a Subscription-billed instance type and its Internet bandwidth.

You can also:

- Convert the billing method of data disks from Pay-As-You-Go to Subscription. The billing method of system disks cannot be changed.
- Change the Internet bandwidth. This applies to the instances in a classic network and instances in a VPC that are not bound with EIPs. If you do not purchase Internet bandwidth when creating an instance, no public IP address is assigned. In this case, you can use this feature to assign a public IP address to the instance when needed.

Fees

After upgrading the configuration, you must make up the difference for the rest of the current billing cycle.

Limits

This feature has the following limits:

- Only applicable to Subscription instances.
- You can upgrade an instance multiple times, but the time period between each upgrade must be at least five minutes.
- You must upgrade both the vCPU cores and memory size of an instance type. That is, you cannot upgrade one item separately.
- Not supported within or between such instance type families: d1, d1ne, i1, i2, ga1, gn5, f1, f2, f3, ebmc4, ebmg5, sccg5, and scch5. For the instance type families that support this feature and the rules for upgrading instance types, see [instance type families that support upgrading instance types](#).
- This feature can be used to change the Internet bandwidth only for VPC instances bound with no EIPs and classic network instances.
- You can change the billing method from Pay-As-You-Go to Subscription only for data disks, not for system disks.
- In the current billing cycle, if you have already performed the [renewal for configuration downgrade](#) operation, you cannot upgrade the configuration until a new billing cycle begins.
- After upgrading an instance type or changing the Internet bandwidth of a classic network instance from 0 Mbps to a non-zero value for the first time, you must restart the instance on the console or through the RebootInstance API to activate the new configuration.

Procedure

1. Select the Subscription instance to upgrade and, in the Actions column, click Change Configuration.
2. Select Upgrade Configuration and click Continue.

3. On the Upgrade Configuration page, perform any of the following operations:

- Select a new Instance Type.



Note:

The page displays all the new instance types that are available for your instance.

- If a [Pay-As-You-Go-billed data disk is attached](#) to your instance, you can convert its billing method to Subscription.
- If the instance is a classic network instance, or is VPC-Connected and not bound with an EIP, you can modify its Internet bandwidth.



Note:

If you do not purchase Internet bandwidth when creating an instance, no public IP address is assigned. In this case, you can use this feature to assign a public IP address to the instance when needed.

4. Confirm your order details, and then click Create Order. Follow additional instructions as required.
5. After upgrading an instance type or changing the Internet bandwidth of a classic network instance from 0 Mbit/s to a non-zero value for the first time, you must restart the instance through the console or through the `RebootInstance` API to activate the new configuration.



Note:

You do not have to restart a VPC-connected instance if this upgrade configuration is the first time its Internet bandwidth is increased from 0 Mbit/s to a non-zero value.

You can also use the [#unique_319](#) and [DescribeResourcesModification](#) APIs to query the instance types that can be upgraded.

10.4 Downgrade configurations

10.4.1 Renew for configuration downgrade

This topic describes how to downgrade the configuration of a Subscription instance when you renew the instance.

After a Subscription instance expires, if renewal has not been completed in the required time, the instance is released automatically. In this case, data is lost permanently and cannot be recovered. To view status changes that occur after Subscription resources expire, see [#unique_6](#).

You can use the Renew for Configuration Downgrade feature to downgrade the specifications of a Subscription instance before you update its renewal to help lower costs incurred in the next billing cycle.

You can also change the billing method of your data disks from Subscription to Pay-As-You-Go.

Limits

When you use the Renew for Configuration Downgrade feature, the following limits apply:

- The feature allows you to scale down instance specifications at the time of renewal.
- Renew for Configuration Downgrade can only be used for Subscription instances.
- You can downgrade the configuration of an instance 15 days prior to its expiration date, or 15 days from its expiration date, but before the instance is released.

For example, if you have a Subscription instance that expires on April 30, you can downgrade the specifications of the instance and renew it between the dates of April 16 to April 30. If you do not renew the instance at this time, the instance enters the Recycling Upon Expiration state. If the instance enters this state, you can still renew the instance from May 1 to May 15. If you do not renew the instance, it is automatically released on May 16.

- After you renew an instance, the new package is effective from the next billing cycle. The current package continues until the end of the current billing cycle.
- If instance configurations are changed during renewal, you must [restart the instance](#) within the first seven days of the new billing cycle for the new configurations to be effective. If you restart the instance on the seventh day of the

new billing cycle, the instance is considered to have used the original package for the first six days, and uses the downgraded package only after it is restarted.

- Once the renewal for configuration downgrade is complete, you cannot [upgrade configurations](#), [increase the system disk size](#), or increase the size of a Subscription data disk, which is attached to a [Linux instance](#) or a [Windows instance](#), during the rest of the current billing cycle.
- You cannot cancel the renewal orders once the payment is processed.

Procedure

To downgrade the configuration of a Subscription instance during renewal, follow these steps:

1. Find the Subscription instance. In the Actions column, click Change Configuration.
2. In the Configuration Change Guide dialog box, select Renew for Configuration Downgrade and click Continue.
3. On the Renew for Configuration Downgrade page, complete the following operations as necessary:

- Downgrade Instance Type. If you want to downgrade specifications for an instance, you must also set the Restart Time for that instance.



Note:

- Instance specifications that can be changed are displayed on the page. For more information about instance specifications, see [Instance type families](#).
- Restarting an instance suspends your business operations on that instance. Please restart the instance during off-peak hours to reduce service impact. The restart must be completed within the first seven days of the next billing cycle.

- If a data disk was created while creating a Subscription instance, you can change the billing method of the data disk to Pay-As-You-Go. If the billing method is not changed, the data disk has the same billing cycle as the Subscription instance.
 - Set the bandwidth value.
 - Set the renewal duration.
4. Click Pay, and follow the prompts to complete the process.
 5. If you have changed instance specifications, or changed the public network bandwidth of an instance in a classic network from 0 Mbit/s to a non-zero value

for the first time, you must restart the instance in the console or by using the `RebootInstance` API within the first seven days of the next billing cycle for the new configurations to be effective.



Note:

For a VPC-connected instance, the instance does not need to be restarted.

10.4.2 Downgrade configurations of Subscription instances

This topic describes how to downgrade configurations of Subscription instances, including the memory size and the number of vCPUs. The changes take effect immediately after an instance is restarted.

Limits

- Whether you can downgrade the configurations of a Subscription instance is determined by your ECS instance resource usage.
- You can downgrade configurations of only one instance at a time.
- You can change the instance configurations only to lower-level configurations. Changes to configurations of the same level or a higher level are not allowed.
- You can only downgrade the configurations of each instance a maximum of three times. Configuration downgrade operations include instance configuration downgrades, bandwidth configuration downgrades, and cloud disk billing method adjustments.
- The time interval between two downgrade operations must be at least 5 minutes.

Prerequisites

The configurations of an instance can be downgraded only if the instance meets the following conditions:

- The instance is in the Stopped state.
- The instance works properly. That is, the instance cannot be in an abnormal state, such as overdue, outdated, locked, or to be released.
- The instance cannot have any ongoing configuration downgrade renewal process.

Procedure

1. Find the target instance. In the Actions column, choose More > Change Configuration.
2. In the displayed dialog box, select Configuration downgrade and Instance Type.

3. Select a desired instance type, confirm the refund amount, and read and confirm that you agree with the *ECS Service Terms*.
4. Click **Downgrade Now**.

What to do next

Restart the instance for the new configurations to take effect.

10.4.3 Downgrade bandwidth configurations of Subscription instances

You can downgrade Internet bandwidth configurations of Subscription instances and change the bandwidth billing method from Pay-By-Bandwidth to Pay-By-Traffic. The configurations take effect immediately without the need to restart instances.

You can use the bandwidth configuration downgrade function to perform the following operations:

- If the current bandwidth billing method is Pay-By-Bandwidth, you can:
 - Lower the fixed bandwidth.
 - Change the billing method to Pay-By-Traffic and set the peak bandwidth.
- If the current bandwidth billing method is Pay-By-Traffic, you can:

Change the peak bandwidth. Note that you cannot change the billing method to Pay-By-Bandwidth.



Note:

If your instance uses a VPC, the process of detaching the Internet IP address will be triggered when the bandwidth is lowered to 0 Mbit/s.

Limits

- Whether the bandwidth of a Subscription instance can be downgraded is determined by your ECS instance usage.
- You can downgrade bandwidth configurations of only one instance at a time.
- You can only downgrade the bandwidth configurations of each instance a maximum of three times. Configuration downgrade operations include instance configuration downgrades, bandwidth configuration downgrades, and cloud disk billing method adjustments.
- The time interval between two downgrade operations must be at least 5 minutes.

- If the instance uses a VPC and has an elastic IP address, the bandwidth configurations of the instance cannot be downgraded.

Prerequisites

The configurations of an instance can be downgraded only if the instance meets the following conditions:

- The billing method is Subscription.
- The instance works properly. That is, the instance cannot be in an abnormal state, such as overdue, outdated, locked, or to be released.
- The instance cannot have any ongoing configuration downgrade renewal process.

Procedure

1. Find the target instance and click Change Configuration in the Actions column.
2. In the displayed dialog box, select Configuration downgrade and Bandwidth Configuration.
3. Set the bandwidth and read and confirm that you agree with the *ECS Service Terms*.
4. Click Downgrade Now.

10.5 Change configurations of Pay-As-You-Go instances

10.5.1 Change configurations of Pay-As-You-Go instances

This topic describes how to change configurations of Pay-As-You-Go instances, including the number of vCPUs and memory size.



Note:

Changing instance configurations requires stopping your instance, which disrupts services. Exercise caution when performing this action. We recommend that you perform this operation during off-peak hours.

Limits

- You can change the configurations of an instance multiple times, but the interval between two change operations must be at least five minutes.
- You cannot change the configurations of instances within or between such instance type families: d1, d1ne, i1, i2, ga1, gn5, f1, f2, f3, ebmc4, ebmg5, sccg5, and scch5. For more information, see [instance type families that support upgrading instance types](#).

Prerequisites

The instance has been stopped.

Procedure

To change configurations of an instance, follow these steps:

1. Find the target instance. In the Actions column, click Change Instance Type.
2. On the Instance Type page, select the desired instance type and click Confirm.



Note:

You can also call the [DescribeResourcesModification](#) API action to query the instance types that can be changed.

The new configuration takes effect immediately after the change is complete. You can view the instance type information in the Basic Information area of the Instance Details page, as shown in the following figure.

10.5.2 Change Internet bandwidth

This topic describes how to change Internet bandwidth of a Pay-As-You-Go instance if the current bandwidth does not meet service requirements.

If you are using a VPC-connected ECS instance to which [an Elastic IP \(EIP\) address is attached](#), you can use the Change Bandwidth feature to change the Internet bandwidth as needed, regardless of your billing method.

Limits

The Change Bandwidth feature only applies to VPC-connected instances to which an EIP address is attached.

Change bandwidth

To change the Internet bandwidth of an EIP address, follow these steps:

1. Find an instance to which an EIP address is attached, and in the Actions column, choose More > Configuration Change.
2. In the Configuration Change dialog box, select change bandwidth, and click Continue.
3. On the Confirm Order page, set the new peak bandwidth.

4. Click Activate and follow the instructions to complete the configuration.

Related operations

For Subscription instances of the classic network type or VPC-connected ECS instances to which no EIP addresses are attached, you can use the following features to change the Internet bandwidth:

- [#unique_313](#) to change the Internet bandwidth immediately.
- [#unique_302](#) to change the Internet bandwidth for the next billing cycle.

11 ECS instance FAQ

This topic provides answers to frequently asked questions (FAQ) about ECS instances.

- FAQ about purchasing instances
 - [How do I check whether instances are available for sale in a specified region or zone?](#)
 - [What can I do if no resources are available when I create an ECS instance?](#)
 - [How do I select an ECS instance that is suitable for my business?](#)
 - [How long does it take to create an ECS instance?](#)
 - [I already paid for an ECS instance but no ECS instance was created. Why?](#)
- FAQ about enterprise-level instances
 - [What are enterprise-level instances? What are entry-level instances?](#)
 - [What are the differences between enterprise-level and entry-level instances?](#)
 - [Which instance families are enterprise-level instances? Which instance families are entry-level instances?](#)
 - [In what business scenarios do I need to purchase enterprise-level instances?](#)
 - [How is the network performance of enterprise-level instances?](#)
 - [What types of disks do enterprise-level instances support?](#)
 - [What types of images do enterprise-level instances support?](#)
 - [What are the limits on upgrading or downgrading enterprise-level instances?](#)
 - [Can I upgrade an entry-level instance to an enterprise-level instance?](#)

- **FAQ about GPU instances**
 - [After the NVIDIA driver is installed, nvidia-smi does not work and an error message is returned indicating that the NVIDIA driver has not been installed. Why?](#)
 - [Why doesn't the Windows Graphics support graphics-related applications such as DirectX?](#)
 - [A g4l Windows instance uses an Alibaba Cloud Marketplace image that has an AMD driver pre-installed. However, the Windows Device Manager window shows that the GPU driver is unable to load properly. Why?](#)
 - [Why do I encounter a blue-screen error when starting an instance that uses an Alibaba Cloud Marketplace image pre-installed with an AMD driver?](#)
 - [Do GPU instances support Android emulators?](#)
 - [Can I upgrade or downgrade GPU instances?](#)
 - [Do pay-as-you-go GPU instances support the No Fees for Stopped Instances \(VPC-Connected\) feature?](#)
 - [How do I view GPU monitoring data?](#)
- **FAQ about ECS Bare Metal Instances**
 - [What are the differences between ECS Bare Metal Instances, traditional cloud hosts \(virtual machines\), and traditional physical machines?](#)
 - [How is the network performance of ECS Bare Metal Instances?](#)
 - [What types of disks do ECS Bare Metal Instances support? How many data disks can be attached to an ECS Bare Metal Instance?](#)
 - [Can I upgrade, downgrade, or migrate ECS Bare Metal Instances?](#)
- **FAQ about SCC**
 - [How do I create an SCC instance?](#)
 - [How are SCC instances billed?](#)
 - [How do I create an SCC by using E-HPC?](#)
 - [How is SCC RDMA used?](#)

- **FAQ about preemptible instances**

- **My account has no overdue payments. Why are my preemptible instances released?**
- **Will I be notified when my preemptible instance is released? And how?**
- **Can data be automatically retained after a preemptible instance is released?**
- **Can I cancel or reschedule the automatic release of my preemptible instance?**
- **Can preemptible instances be converted to subscription instances?**
- **To what resources are the preemptible instance prices applicable?**
- **How do I bid for a preemptible instance?**
- **What is the relationship between the highest bid for a preemptible instance and the current market price?**
- **Will I be charged the same fees for all my preemptible instances that start at the same time?**
- **Will I be charged at a price that fluctuates with the market price within one hour after my preemptible instance starts?**
- **Can I view the current market price of an instance type when I purchase a preemptible instance?**
- **Can I view the historical prices of a preemptible instance type? And how?**
- **Will preemptible instances continue to be billed after they are stopped?**
- **How do I view the consumption details of my preemptible instance?**
- **How many preemptible instances can be purchased under one account?**
- **How do I increase the maximum number of vCPUs?**
- **Can I change the instance type of my preemptible instance?**
- **Which instance families support preemptible instances?**
- **In which regions can I create preemptible instances?**

- **FAQ about reserved instances**

- [What is a reserved instance?](#)
- [Will a reserved instance provide a resource reservation?](#)
- [What operating systems do reserved instances support?](#)
- [Which instance families do reserved instances support?](#)
- [Can reserved instances match preemptible instances to apply a discounted rate?](#)
- [Can I change the instance families of reserved instances?](#)
- [To what scenarios are zonal reserved instances applicable?](#)
- [To what scenarios are regional reserved instances applicable?](#)
- [How is the zone flexibility of regional reserved instances applied?](#)
- [How is the instance size flexibility of regional reserved instances applied?](#)
- [Do zonal reserved instances provide instance size flexibility?](#)
- [Do zonal reserved instances provide zone flexibility?](#)
- [Can I change a zonal reserved instance to a regional one?](#)
- [Can I change the scope of a reserved instance from one region to another?](#)
- [Can I use reserved instances across accounts?](#)
- [Can I use reserved instances to cover the storage and network charges of pay-as-you-go instances?](#)
- [Can I configure a reserved instance to apply its discounted rate to a specific pay-as-you-go instance?](#)
- [How are reserved instances billed?](#)
- [When does a reserved instance take effect after I purchase it?](#)
- [When do my reserved instances take effect after I modify, split, or merge them?](#)
- [Why isn't the No Upfront payment option displayed on the purchase page?](#)
- [Can I modify the payment option of my reserved instance?](#)
- [Can I request a refund for a reserved instance?](#)
- [Can I resell a reserved instance?](#)
- [Can I use reserved instances to cover the image costs of pay-as-you-go Windows instances?](#)
- [Can I use reserved instances to cover the image costs of pay-as-you-go Linux instances?](#)

- **FAQ about connecting to instances**
 - [Can multiple users connect to the Management Terminal simultaneously?](#)
 - [What can I do if I forget the password for remote connection?](#)
 - [Why can't I connect to the Management Terminal even after I reset my password for remote connection?](#)
 - [Authentication failed on the Management Terminal. What can I do?](#)
 - [What can I do if a black screen appears when I am connected to the Management Terminal?](#)
 - [What can I do when I am unable to access the Management Terminal?](#)
 - [Why can't I use Internet Explorer 8.0 to access the Management Terminal?](#)
 - [When I use Firefox to access the Management Terminal, an error message is returned indicating that the secure connection has failed. What can I do?](#)
 - [How do I remotely log on to a Linux instance?](#)
 - [What are the default username and password for remote logon to the operating system of an ECS instance?](#)
- **FAQ about upgrading and downgrading instances**
 - [Can I upgrade my subscription instances?](#)
 - [Can I upgrade my pay-as-you-go instances?](#)
 - [How long does it take to upgrade an ECS instance?](#)
 - [How is the fee for upgrading an ECS instance calculated?](#)
 - [Does upgrading my ECS instance affect my cloud service configurations?](#)
 - [How do I upgrade my ECS resources?](#)
 - [I have upgraded an instance but no changes have taken effect yet. Why?](#)

- **FAQ about managing instances**
 - The ECS instance has been in the Starting state for an extended period of time, and AliyunService was disabled or deleted. What can I do?
 - How do I use f1 instances?
 - How do I upload files by using the FTP tool in macOS?
 - How do I apply for an ICP filing for my domain name after buying an ECS instance?
 - Why can't I start an ECS instance by loading the kernel?
 - How do I change the logon password within an instance?
 - Why can't I add sound or video cards to ECS instances?
 - Can I transfer the remaining service time of an ECS instance to another ECS instance?
 - Do ECS instances provide database capabilities by default?
 - Can I build a database on an ECS instance?
 - Do ECS instances support Oracle databases?
 - Are public IP addresses and private IP addresses independent? Can I specify or add IP addresses?
 - Can an ECS instance be used for load balancing?
 - Can I change the region of an ECS instance?
 - Can I modify the partition size of a purchased disk?
 - How do I replace the public IP address of my ECS instance with an Elastic IP Address (EIP)?
 - How do I view my subscription instances in the same region?
 - How do I view subscription instances in all regions under my account?
 - When can I forcibly stop my ECS instance? What will the consequence be?
 - Why can't I reactivate my ECS instance?
- **FAQ about instance security**
 - What is the AliVulfix process in an ECS instance?
 - How do I defend ECS instances against attacks?
 - What security services does Alibaba Cloud provide?

- **FAQ about using Linux instances**
 - [After I renew a subscription ECS Linux instance that expires, I am still unable to access the website it hosts. What can I do?](#)
 - [How do I check, partition, and format the data disks of a Linux instance?](#)
 - [How do I upload files to a Linux instance?](#)
 - [How do I change the owner and owner group of directories and files on a Linux instance?](#)
 - [How do I update the software source of Linux instances?](#)
 - [How do I install a VNC server to access a Linux system by using a GUI?](#)
- **FAQ about instance limits**
 - [What limits apply to the transfer and change of public IP addresses of ECS instances?](#)
 - [Can I access amazon.com from my ECS instance?](#)
 - [Why can't I access a website hosted outside China after I log on to my ECS instance?](#)
 - [I cannot purchase any more pay-as-you-go instances. What can I do?](#)
- **FAQ about instance billing**
 - [Are fees still incurred after a pay-as-you-go ECS instance is automatically stopped due to an overdue payment or is manually stopped?](#)
 - [The order fails to be placed when I change the billing method of my ECS instance from pay-as-you-go to subscription. What can I do?](#)
 - [How long does it take to change the billing method after you pay for the order?](#)
 - [What can I do if the change from pay-as-you-go to subscription fails?](#)
 - [Will the bandwidth billing method change after I change from pay-as-you-go to subscription?](#)
 - [I have an unpaid order for changing the billing method of an instance from pay-as-you-go to subscription. If I upgrade the instance, will the order still be valid?](#)
 - [What can I do if the change from subscription to pay-as-you-go fails?](#)
 - [When I try to change the billing method of a disk in an ECS instance, an error message appears indicating that I have already changed the billing method three times. What does this mean?](#)
 - [Why can't I change a pay-as-you-go instance to a subscription one?](#)
 - [How do I view the expiration time of my subscription instances?](#)

How do I check whether instances are available for sale in a specified region or zone?

You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.

What can I do if no resources are available when I create an ECS instance?

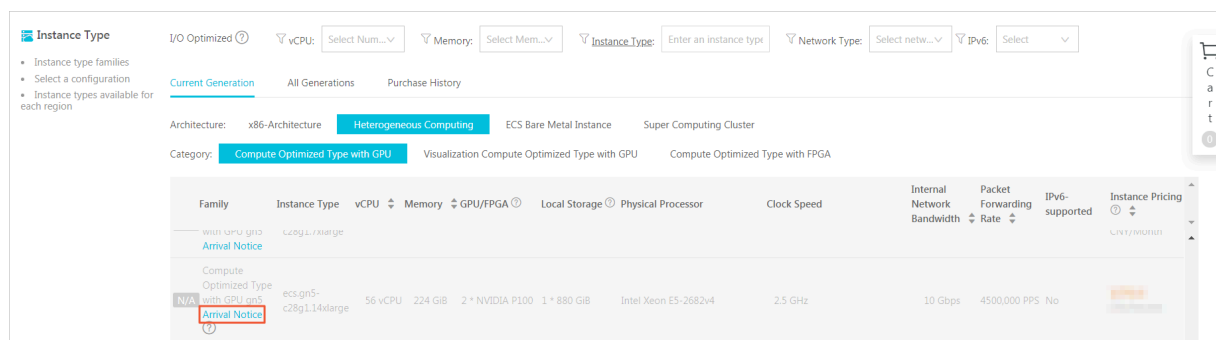
If no resources are available when you create an ECS instance, take any of the following measures:

- Select another region
- Select another zone
- Change resource configurations

If still no resources are available after you take all the preceding measures, try again after some time. Instance resources are dynamic. When resources are insufficient, Alibaba Cloud replenishes them as soon as it can.

You can also use the arrival notice feature to get notified when resources are available

.



How do I select an ECS instance that is suitable for my business?

- Select an ECS instance based on your business requirements.
- Determine your website type.
- Determine the average number of page views per day on your website.
- Determine the size of your homepage.
- Determine the data capacity.

How long does it take to create an ECS instance?

It takes one or two minutes to create an ECS instance. After the instance is created:



Note:

If an error occurs when you create an ECS instance, [submit a ticket](#).

I already paid for an ECS instance but no ECS instance was created. Why?

If the specified zone has insufficient resources to create the instance type that you specified, the ECS instance fails to be created. A refund will be automatically credited to your account by the system. If you do not receive a refund within half an hour, [submit a ticket](#).

You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.

What are enterprise-level instances? What are entry-level instances?

Enterprise-level instances are a series of instance families released by Alibaba Cloud in September 2016. Enterprise-level instances feature high performance, consistent computing power, and balanced network performance. These instances have exclusive and consistent computing, storage, and network resources, and are applicable to enterprise scenarios with high business stability requirements.

Entry-level instances are a series of instance families that are targeted at small and medium websites or individuals. Compared with enterprise-level instances that have exclusive resources, entry-level instances share resources. Therefore, entry-level instances do not provide consistent computing performance but they have a lower cost.

What are the differences between enterprise-level and entry-level instances?

Enterprise-level instances use the CPU-bound scheduling scheme. Each vCPU is bound to a CPU hyperthread. Instances do not compete for CPU resources and provide consistent computing performance as guaranteed in the SLA.

Entry-level instances use the CPU-unbound scheduling scheme. Each vCPU is randomly allocated to an idle CPU hyperthread. vCPUs of different instances compete for CPU resources. This results in fluctuations of computing performance when traffic loads are heavy. Therefore, entry-level instances guarantee only availability but not performance in the SLA.

Which instance families are enterprise-level instances? Which instance families are entry-level instances?

Among the [instance families](#) that are available for sale, n4, mn4, xn4, and e4 are entry-level instances, and the others are enterprise-level instances.

In what business scenarios do I need to purchase enterprise-level instances?

For the business scenarios to which different enterprise-level instances are applied, see [Instance families](#).

How is the network performance of enterprise-level instances?

The network performance of enterprise-level instances depends on their specifications. The larger specifications enterprise-level instances have, the stronger network performance they provide. For more information about the network performance of different instance types, see [Instance families](#).

What types of disks do enterprise-level instances support?

For the disk types that enterprise-level instances support, see [#unique_329](#).

What types of images do enterprise-level instances support?

For the public images that enterprise-level instances support, see [#unique_330](#).

You can also import custom images. For more information, see [#unique_331](#).

What are the limits on upgrading or downgrading enterprise-level instances?

For information about the limits on upgrading or downgrading enterprise-level instances, see [#unique_119](#).

Can I upgrade an entry-level instance to an enterprise-level instance?

For the instance families that support instance type upgrades, see [#unique_119](#).

After the NVIDIA driver is installed, `nvidia-smi` does not work and an error message is returned indicating that the NVIDIA driver has not been installed. Why?

Cause: The kernel version is inconsistent with the kernel-devel package version. This results in a driver compilation error when the driver is installed from the `.rpm` file.

Solution: Check the kernel version and download the correct kernel-devel package version. Then, run the `rpm -qa | grep kernel` command on the instance to check whether the kernel-devel package version matches the kernel version. Make sure that they match and then reinstall the driver.

Why doesn't the Windows Graphics support graphics-related applications such as DirectX?

Cause: On Windows instances where the installed GPU drivers have taken effect, the Windows Remote Desktop Protocol (RDP) does not support DirectX, OpenGL, or other graphics-related applications.

Solution: Install Virtual Network Computing (VNC) servers and clients or other protocols that support these applications, such as PC over IP (PCoIP) and XenDesktop HDX 3D.

A ga1 Windows instance uses an Alibaba Cloud Marketplace image that has an AMD driver pre-installed. However, the Windows Device Manager window shows that the GPU driver is unable to load properly. Why?

1. Log on to the instance.
2. Open the Device Manager window.
3. Right-click the GPU device.
4. Select Update Driver Software from the shortcut menu.
5. Then, select Search automatically for updated driver software.
6. After the device driver is updated, restart the instance.

Why do I encounter a blue-screen error when starting an instance that uses an Alibaba Cloud Marketplace image pre-installed with an AMD driver?

Alibaba Cloud will update the driver included in the Alibaba Cloud Marketplace image to solve this problem. If you are using an image of any earlier version, update the GPU driver to avoid this issue.

If you encounter a blue-screen error and cannot start an instance, you must enter Safe Mode. In Safe Mode, uninstall the existing driver, start Windows normally, and then install a GPU driver of the latest version.

Do GPU instances support Android emulators?

No, they do not support Android emulators.

Can I upgrade or downgrade GPU instances?

You cannot upgrade or downgrade GPU instances that are equipped with local disks, such as ga1 and gn5 instances. For the GPU instances that are not equipped with local disks and can be upgraded or downgraded, see [#unique_119](#).

Do pay-as-you-go GPU instances support the No Fees for Stopped Instances (VPC-Connected) feature?

GPU instances that are not equipped with local disks support the No Fees for Stopped Instances (VPC-Connected) feature. Such GPU instances include gn4 (with NVIDIA

M40 GPU processors), gn5i (with NVIDIA P4 GPU processors), and gn6v (with NVIDIA V100 GPU processors) instances. For more information, see [#unique_142](#).

How do I view GPU monitoring data?

You can log on to the [CloudMonitor console](#) or call the [DescribeMetricList](#) operation to view GPU monitoring data. For more information, see [#unique_77](#).

What are the differences between ECS Bare Metal Instances, traditional cloud hosts (virtual machines), and traditional physical machines?

For the differences between ECS Bare Metal Instances, traditional cloud hosts (virtual machines), and traditional physical machines, see [What is an ECS Bare Metal Instance?](#).

How is the network performance of ECS Bare Metal Instances?

The network performance of ECS Bare Metal Instances depends on their specifications. The larger specifications ECS Bare Metal Instances have, the stronger network performance they provide. For more information about the network performance of different ECS Bare Metal Instance types, see [Instance families](#).

What types of disks do ECS Bare Metal Instances support? How many data disks can be attached to an ECS Bare Metal Instance?

ECS Bare Metal Instances support ultra disks and standard SSDs. Up to 16 data disks can be attached to each ECS Bare Metal Instance.

Can I upgrade, downgrade, or migrate ECS Bare Metal Instances?

You cannot upgrade or downgrade ECS Bare Metal Instances. When the hardware of the physical machine where an ECS Bare Metal Instance resides fails, you can migrate the ECS Bare Metal Instance to another physical machine. Data is stored in the data disks of the instance.

How do I create an SCC instance?

You can create an SCC instance in either of the following ways:

- If you only want to use RDMA, log on to the ECS console. [#unique_111](#).
- If you want to use the HPC scheduler and cluster resizing service in addition to RDMA, log on to the [E-HPC console](#). Create an SCC and then create an SCC instance.

How are SCC instances billed?

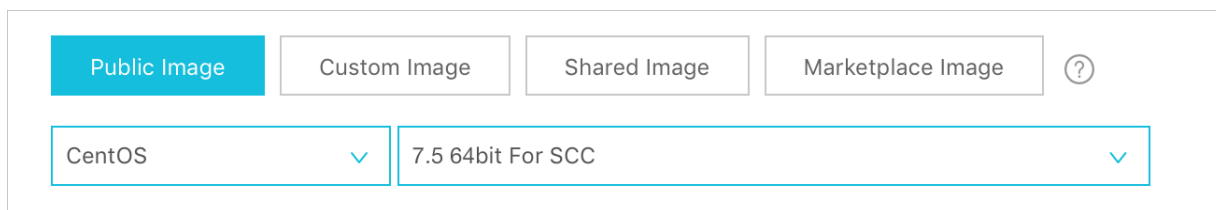
SCC instances support the weekly, monthly, and yearly subscription billing methods.

How do I create an SCC by using E-HPC?

You can log on to the [E-HPC console](#) or call the `CreateCluster` operation to create an SCC.

How is SCC RDMA used?

When you create an SCC instance, select a custom image for SCC, which supports the RDMA RoCE driver and OFED stack. You can use SCC RDMA through the IB Verbs API or manage RDMA communication through the MPI.



My account has no overdue payments. Why are my preemptible instances released?

You can reserve a preemptible instance without interruptions for at least one hour. After one hour, your bid is compared with the market price. When the market price exceeds your bid or resources are insufficient, your preemptible instance is automatically released.

Will I be notified when my preemptible instance is released? And how?

Yes, you will be notified when your preemptible instance is released. When your preemptible instance needs to be released due to a change in market price or due to insufficient resources, the instance will first enter the To Be Released state. After 5 minutes, the instance will be automatically released. You can check whether the instance in the To Be Released state based on the [instance metadata](#) or `OperationLocks` information returned by the [DescribeInstances](#) operation.

Can data be automatically retained after a preemptible instance is released?

No, data cannot be automatically retained. When you no longer need a preemptible instance, we recommend that you [create snapshots](#) to back up your data and environment and then release the instance. You can purchase new preemptible instances at any time as needed.

Can I cancel or reschedule the automatic release of my preemptible instance?

Yes, you can cancel or reschedule the automatic release of your preemptible instance at any time.

Can preemptible instances be converted to subscription instances?

No, they cannot.

To what resources are the preemptible instance prices applicable?

The preemptible instance prices apply to instance types only. Other resources of preemptible instances, such as system disks, data disks, and network bandwidth, are billed at the same rates as those of pay-as-you-go instances, without discounts.

How do I bid for a preemptible instance?

To create a preemptible instance, you must set the highest hourly price you are willing to pay for an instance type. If your bid is higher than or equal to the current market price, your instance is created and billed based on the current market price. For more information, see [#unique_336](#).

What is the relationship between the highest bid for a preemptible instance and the current market price?

The highest bid for a preemptible instance must be higher or equal to the current market price. If the highest bid is lower than the market price, no new preemptible instance is created or the running preemptible instances may be released. Your preemptible instance will be billed based on the market price regardless of your bid.

Will I be charged the same fees for all my preemptible instances that start at the same time?

Yes.

Will I be charged at a price that fluctuates with the market price within one hour after my preemptible instance starts?

No, the hourly price of your preemptible instance is set at the beginning of each hour and applies to the instance within the whole hour.

Can I view the current market price of an instance type when I purchase a preemptible instance?

Yes. When you create a preemptible instance in the ECS console, you can view the current market price and historical prices of an instance type after you select it. The total price that includes the fees for the instance type, storage, and bandwidth is

displayed at the lower-left corner of the instance creation page. The instance type price is the current market price of the specified instance type.

Can I view the historical prices of a preemptible instance type? And how?

Yes. You can view the historical prices of a preemptible instance type by selecting it in the ECS console. You can also call the [DescribeSpotPriceHistory](#) operation to view the historical prices of an instance type.

Will preemptible instances continue to be billed after they are stopped?

Yes. When you no longer need a preemptible instance, we recommend that you [create snapshots](#) to back up your data and environment and then release the instance. You can purchase new preemptible instances at any time as needed.



Note:

Preemptible instances will continue to be billed after you stop them in the ECS console or by calling the `StopInstance` operation.

How do I view the consumption details of my preemptible instance?

See [#unique_336/unique_336_Connect_42_section_pkj_3qt_qgb](#).



Note:

The total cost displayed when you create a pay-as-you-go instance is the hourly price to be charged for actual usage. However, the total cost displayed when you create a preemptible instance is the actual transaction price.

How many preemptible instances can be purchased under one account?

There is a limit on the maximum number of vCPUs that can be allocated to preemptible instances for each account. When you create a preemptible instance, you can view the number of available vCPUs after you select an instance type. For more information, see [Limits](#).

How do I increase the maximum number of vCPUs?

[Submit a ticket](#).

Can I change the instance type of my preemptible instance?

No, you cannot change the instance type of your preemptible instance.

Which instance families support preemptible instances?

The instance families that support the pay-as-you-go billing method also support preemptible instances. If you cannot create preemptible instances of an instance type due to insufficient resources, try a different instance type.

In which regions can I create preemptible instances?

You can create preemptible instances in all regions. If you cannot create preemptible instances in a region due to insufficient resources, try a different region.

What is a reserved instance?

Reserved instances automatically match pay-as-you-go instances (not including preemptible instances) in your account to provide a billing discount. Reserved instances can also be used to reserve resources for pay-as-you-go instances. Compared with subscription instances, the combination of reserved instances and pay-as-you-go instances provides similar cost-effectiveness and higher flexibility.

Will a reserved instance provide a resource reservation?

Zonal reserved instances provide resource reservations, but regional reserved instances do not.

What operating systems do reserved instances support?

Reserved instances support both Windows and Linux. For example, a Linux reserved instance can be applied to pay-as-you-go Linux instances that match its attributes, regardless of the image types (public images, custom images, shared images, and Alibaba Cloud Marketplace images) these instances use.

If your pay-as-you-go instance uses a BYOL image and you want to apply the discounted rate provided by a reserved instance to your instance, [submit a ticket](#).

Which instance families do reserved instances support?

The following instance families support reserved instances: sn1ne, sn2ne, se1ne, ic5, c5, g5, r5, c6, g6, r6, i2, i2g, hfc5, hfg5, and t5.

Reserved instances matching [t5 burstable instances](#) are only available at the zone level. These reserved instances cannot be merged or split, and their scopes cannot be modified.

Can reserved instances match preemptible instances to apply a discounted rate?

No.

Can I change the instance families of reserved instances?

No.

To what scenarios are zonal reserved instances applicable?

We recommend that you purchase zonal reserved instances when you have clear requirements to reserve resources.

To what scenarios are regional reserved instances applicable?

We recommend that you purchase regional reserved instances if you want to enjoy better zone and instance size flexibility.

How is the zone flexibility of regional reserved instances applied?

Only regional reserved instances provide zone flexibility. Example:

You have the following pay-as-you-go instance running:

One ecs.c5.xlarge Linux instance in Qingdao Zone B. The instance name is C5PAYG-b.

You have purchased the following reserved instance:

One regional ecs.c5.xlarge reserved instance in the China (Qingdao) region. The name of the reserved instance is C5RI.

C5RI matches C5PAYG-b to apply a discounted rate.

You release C5PAYG-b and start another Linux instance named C5PAYG-c, which is of the same instance type as C5PAYG-b. C5RI then matches C5PAYG-c.

How is the instance size flexibility of regional reserved instances applied?

Only regional reserved instances provide instance size flexibility. Example:

You have one regional ecs.g5.4xlarge reserved instance. It can match one ecs.g5.4xlarge pay-as-you-go instance, two ecs.g5.2xlarge pay-as-you-go instances, or four ecs.g5.xlarge pay-as-you-go instances.

You have a one-year regional ecs.g5.xlarge reserved instance. It can match one ecs.g5.xlarge instance running for one year, or one ecs.g5.2xlarge instance running for six months.

Do zonal reserved instances provide instance size flexibility?

No. A zonal reserved instance can only match pay-as-you-go instances of the same instance type as it.

Do zonal reserved instances provide zone flexibility?

No. A zonal reserved instance can only match pay-as-you-go instances in the same zone as it.

Can I change a zonal reserved instance to a regional one?

Yes. You can change the scope of a reserved instance you purchased as follows:

- From a zone to a region
- From a region to a zone
- From one zone to another within the same region for a zonal reserved instance.

Can I change the scope of a reserved instance from one region to another?

No. For example, if you have a zonal reserved instance in Hangzhou Zone B, you can change the instance scope to another zone or change the instance to a regional reserved instance in the China (Hangzhou) region. However, you cannot change it to a regional or zonal reserved instance in another region.

Can I use reserved instances across accounts?

No.

Can I use reserved instances to cover the storage and network charges of pay-as-you-go instances?

No. Reserved instances cover the compute (vCPU and memory) costs of pay-as-you-go instances. For pay-as-you-go Windows instances, reserved instances also cover the image costs.

Can I configure a reserved instance to apply its discounted rate to a specific pay-as-you-go instance?

No. When you have multiple pay-as-you-go instances that match the attributes of a reserved instance, the discounted rate is applied automatically to pay-as-you-go instances based on the optimized matching scheme.

How are reserved instances billed?

Reserved instances are billed separately and support the All Upfront, Partial Upfront, and No Upfront payment options.

The term of a reserved instance starts immediately after you purchase it. You are charged based on your selected payment option regardless of whether pay-as-you-go instances are matched. The All Upfront option is the most cost-effective.

When does a reserved instance take effect after I purchase it?

The reserved instance takes effect and starts to be billed on the hour (zero minutes and zero seconds past the hour) of your purchase. It expires at 00:00:00 the day after the term end date. For example, if you purchase a one-year reserved instance at 2019-02-26 13:45:00, the reserved instance takes effect and starts to be billed at 2019-02-26 13:00:00 and will expire at 2020-02-27 00:00:00. If you purchase a reserved instance that matches your existing pay-as-you-go instances, the discounted rate is applied to bills generated for the pay-as-you-go instances starting from 13:00 to 14:00 on February 26, 2019 until the reserved instance expires.

When do my reserved instances take effect after I modify, split, or merge them?

After reserved instances are modified, split, or merged, new reserved instances are generated, and the original ones become invalid. The new reserved instances take effect on the hour, and the original reserved instance become invalid at the same time. For example, you split an ecs.g5.2xlarge zonal reserved instance RI1 into two zonal ecs.g5.xlarge reserved instances RI2 and RI3 at 2019-02-26 13:45:00. RI1 becomes invalid and RI 2 and RI 3 take effect at 2019-02-26 13:00:00. As of 2019-02-26 13:00:00, the eligible reserved instance type also changes from ecs.g5.2xlarge to ecs.g5.xlarge. If RI2 and RI3 match pay-as-you-go instances immediately after they take effect, the discounted rates provided by RI2 and RI3 are applied to the hourly bills of ecs.g5.xlarge pay-as-you-go instances as of 2019-02-26 13:00:00.

Why isn't the No Upfront payment option displayed on the purchase page?

The availability of this option depends on your ECS usage.

Can I modify the payment option of my reserved instance?

No.

Can I request a refund for a reserved instance?

In the following cases, you can [open a ticket](#) to claim a refund:

- The instance resources are in short supply in the target region or zone after you have purchased RIs.

- The instance resources are in short supply in the target region or zone after you have split or merged RIs, or adjusted the RI scope.

Can I resell a reserved instance?

No.

Can I use reserved instances to cover the image costs of pay-as-you-go Windows instances?

Yes. Windows reserved instances already include Windows image costs. Therefore, you can use them to cover the image costs of your pay-as-you-go Windows instances.

Can I use reserved instances to cover the image costs of pay-as-you-go Linux instances?

No.

Can multiple users connect to the Management Terminal simultaneously?

No, only one user can connect to the Management Terminal at a time.

What can I do if I forget the password for remote connection?

You can reset your password for remote connection. For more information, see [#unique_72](#).

Why can't I connect to the Management Terminal even after I reset my password for remote connection?

If the ECS instance to which you are connecting is not I/O optimized, you must restart the instance in the ECS console for the new password to take effect. For more information, see [#unique_221](#).



Note:

If the ECS instance is restarted from within the instance itself, the new password will not take effect.

Authentication failed on the Management Terminal. What can I do?

If you enter an incorrect password, authentication fails on the Management Terminal. Perform the following steps to troubleshoot the problem:

1. Enter the correct password.
2. If you forget your password, [reset your password for remote connection](#) and try again.

3. If the ECS instance that you are connecting to is not I/O optimized, you must **restart the instance** in the ECS console for the new password to take effect.

What can I do if a black screen appears when I am connected to the Management Terminal?

A black screen indicates that the instance is in sleep mode. Perform the following operations based on your operating system:

- For a Linux instance, click your mouse or press any key to activate the instance.
- For a Windows instance, in the upper-left corner of the Management Terminal page, choose Send Remote Call > CTRL+ALT+DELETE to open the logon dialog box.

What can I do when I am unable to access the Management Terminal?

Use Google Chrome to connect to the Management Terminal, and press F12 to open the developer tools pane. Then click the Console tab and analyze the information displayed to find errors.

Why can't I use Internet Explorer 8.0 to access the Management Terminal?

The Management Terminal supports Internet Explorer 10 and later.

We recommend that you use Google Chrome, which is the most compatible with the Management Terminal.

When I use Firefox to access the Management Terminal, an error message is returned indicating that the secure connection has failed. What can I do?

This problem occurs if the encryption algorithm of your Firefox version is different from that of the Management Terminal.

We recommend that you use Google Chrome, which is the most compatible with the Management Terminal.

How do I remotely log on to a Linux instance?

Linux instances use SSH for remote connection. You can use any of the following methods to remotely log on to a Linux instance:

- [#unique_72](#).
- [#unique_214](#).
- [#unique_213](#).
- [#unique_215](#).

What are the default username and password for remote login to the operating system of an ECS instance?

The default username depends on the operating system.

- For a Windows instance, the default username is `administrator`.
- For a Linux instance, the default username is `root`.

The password for remote login to the operating system is set by you when you create the instance. For more information, see [#unique_69](#). If you forget your password, you can reset the password. For more information, see [#unique_219](#).



Note:

This password is used to remotely log on to the operating system of an ECS instance, not to the Management Terminal.

Can I upgrade my subscription instances?

Yes, you can upgrade your subscription instances. For more information, see [#unique_313](#).

Can I upgrade my pay-as-you-go instances?

Yes, but you must stop your pay-as-you-go instances before you can upgrade them. You can upgrade your pay-as-you-go instances by following the instructions in [#unique_121](#) or calling the [ModifyInstanceSpec](#) operation.

How long does it take to upgrade an ECS instance?

- To upgrade a subscription instance, you do not need to stop it. The upgrade process takes about 15 minutes.
- To upgrade a pay-as-you-go instance, you must stop it first. The upgrade process takes about 15 minutes.
- To upgrade the bandwidth of an instance, you do not need to stop the instance. The upgrade process takes about 5 minutes.

How is the fee for upgrading an ECS instance calculated?

The upgrade fee and its calculation are displayed when you upgrade the instance. You can also view the details of the fee on the [Account Overview](#) page.

Does upgrading my ECS instance affect my cloud service configurations?

You must stop a pay-as-you-go instance before you can upgrade it. After you upgrade a subscription instance, you must restart it for the new specifications to take effect. The upgrade operation will interrupt your services for a short period of time. We recommend that you upgrade your instance during off-peak hours. Instances can seamlessly resume services after upgrades without server environment reconfiguration.

How do I upgrade my ECS resources?

For details about how to upgrade ECS resources, see [#unique_120](#).

- Except for ECS instances that use local storage, ECS instances support online CPU scaling, online memory scaling, and online bandwidth upgrades.
- A maximum of 16 data disks can be attached to an ECS instance. You cannot reduce the size of a data disk after you expand it.
- The ECS bandwidth is measured in megabits per second (Mbit/s) and ranges from 0 Mbit/s to 200 Mbit/s. You can also temporarily upgrade the bandwidth of your ECS instance, or downgrade the instance when you renew it.

I have upgraded an instance but no changes have taken effect yet. Why?

To make the new specifications take effect, you need to [restart the instance](#).

The ECS instance has been in the Starting state for an extended period of time, and AliyunService was disabled or deleted. What can I do?

Problem description: After you start an ECS instance, it remains in the Starting state for an extended period of time before automatically shutting down. You log on to the instance and find that AliyunService was deleted or disabled in the system services.

Solution:

- If AliyunService is disabled:
 1. Change the state of AliyunService to Automatic.
 2. Restart the instance.
- If AliyunService is deleted:
 1. Run the following command to add AliyunService to the instance:

```
sc create AliyunService type = " own " start = " auto "
binPath = " C :\ Program Files \ AliyunService \ AliyunService.exe "
```

```
ice . exe - d " tag = " no " DisplayNam e = " AliyunServ ice
"
```

**Note:**

Make sure you leave a space after the equal signs (=).

2. Find the registry key `HKEY_LOCAL_MACHINE \ SYSTEM \ CurrentControlSet \ services \ AliyunServ ice` , and change `c :\ Program Files \ AliyunServ ice \ AliyunServ ice . exe - d` to `" c :\ Program Files \ AliyunServ ice \ AliyunServ ice . exe " - d` .
3. Restart the instance.

How do I use f1 instances?

After your application for an f1 instance trial is approved, Alibaba Cloud will share an FPGA development image with you. Only CentOS 7u2 images are available at this time . The image includes the complete Intel Quartus development suite and the constraint files for the f1 instance to provide a complete cloud development environment.

**Note:**

You can go to the [ECS Instance Types Available for Each Region page](#) to view the instance types available in each region.

The basic workflow for using f1 instances is as follows:

1. After development, generate an intermediate QAR file during the compilation stage, and upload the file to an OSS bucket. You can upload this file only to an OSS bucket in the China (Hangzhou) region. Then, register the QAR file information with Alibaba Cloud by calling an API operation.

We recommend that you use the free Intel Quartus development suite to perform development, compilation, and simulation operations on the cloud.

2. Alibaba Cloud verifies the registration request for the QAR file. After the verification succeeds, Alibaba Cloud sends you an email that includes an FPGA image ID.

3. When you deploy the image, call an API operation with the f1 instance ID and FPGA image ID specified to associate the instance with the image.

You can initiate the association operation in any scenario where ECS API is available.

- If no FPGA image has ever been associated with the f1 instance, directly initiate the association operation.
- If the f1 instance was previously associated with an FPGA image and had the image loaded, erase the FPGA image from the f1 instance before initiating the association operation.

4. After the association, call an API operation to load the FPGA image.

You must initiate the load operation from the f1 instance. Then the underlying service of Alibaba Cloud burns the associated FPGA image to the corresponding FPGA on the instance.

If you want to restore the f1 instance to its initial state, call an API operation to erase the burned FPGA image from the f1 instance.

For more instructions on f1 instances, see:

- [#unique_340](#)
- [#unique_341](#)
- [#unique_342](#)

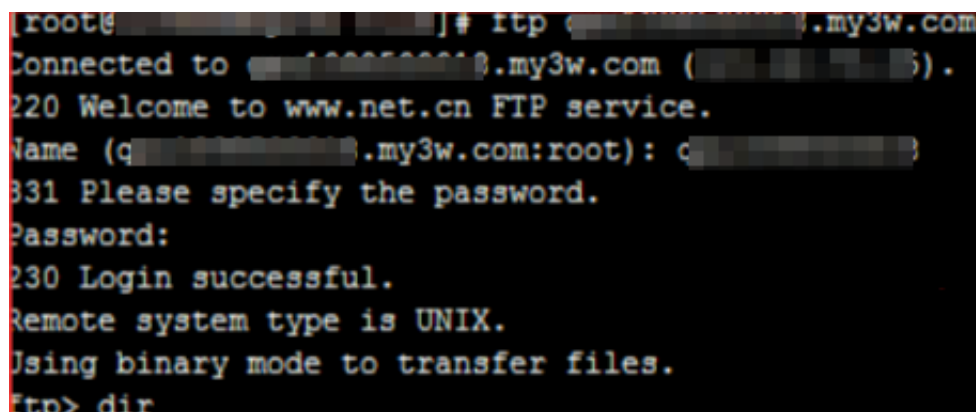
How do I upload files by using the FTP tool in macOS?

Upload files by using the Terminal in macOS

Open the Terminal in macOS or iTerm2 for Mac (click [here](#) to download iTerm2).

Make sure that you select the correct destination path.

1. Connect to the FTP server.



```
[root@ ~]# ftp 1000500001.my3w.com
Connected to 1000500001.my3w.com (1000500001.my3w.com).
220 Welcome to www.net.cn FTP service.
Name (q1000500001.my3w.com:root): c1000500001
331 Please specify the password.
Password:
230 Login successful.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> dir
```

2. Enter the destination directory. If the directory is on a Windows instance, use the directory directly. If it is on a Linux instance, change the directory to `htdocs` .

```
227 Entering Passive Mode (121,42,75,25,156,83)
150 Here comes the directory listing.
drwxr-xr-x  2 0      0      4096 Jan 13 20:30 backup
drwxr-xr-x  2 521    100    4096 Jan 13 20:30 cgi-bin
drwxr-xr-x  2 0      0      4096 Jun 03 03:40 ftplogs
drwxr-xr-x 20 521    100    4096 May 26 16:13 htdocs
drwxr-xr-x  2 521    100    4096 Jan 13 20:30 myfolder
drwxr-xr-x  2 0      0      4096 Jun 03 02:03 wwwlogs
-rw-r--r--  1 0      0      606 Feb 03 15:47 请先读我.txt
226 Directory send OK.
ftp> cd htdocs
```

3. Run the `put` command to upload files.

```
ftp> put wordpress.tar
local: wordpress.tar remote: wordpress.tar
227 Entering Passive Mode (121,42,75,25,156,121)
150 Ok to send data.
226 File receive OK.
5720564 bytes sent in 46.2 secs (145.49 Kbytes/sec)
```

Upload files by using a third-party tool

1. Click [here](#) to download Yummy FTP.
2. Install Yummy FTP.
3. Enter the server IP address, username, and password. Set Protocol to Standard (FTP) and Port to 21 or a port number you are using, and leave the SSH key field unselected.
4. Click Connect.
5. In the right section, select the destination directory. If the directory is on a Windows instance, use the directory directly. If it is on a Linux instance, change the directory to `htdocs`. In the left section, select the files and click the Upload icon to upload files.



Note:

If you are prompted with a message similar to "Your security preferences only allow the installation of applications from the Mac App Store and authorized developers" when you try to install Yummy FTP, perform these steps:

- a. Choose System Preference > Security and Privacy.

- b. Click the security lock in the lower-left corner of the window, and enter the administrator password.
 - c. Select Anywhere for Allow apps download from.
- Then, you can upload files by using Yummy FTP.

If you have further questions, [submit a ticket](#).

How do I apply for an ICP filing for my domain name after buying an ECS instance?

You can apply for a limited number of ICP filing service numbers for an ECS instance. For more information, see [#unique_343](#).

For information about how to apply for an ICP filing, see [#unique_344](#).

Why can't I start an ECS instance by loading the kernel?

Problem description: When the system starts, it does not respond when I select an option from the GRUB menu. After the LiveCD image is loaded on the ECS instance, verify that file system privileges are correct and there are no exceptions in the message logs.

Cause: The system is attacked by ransomware.

Solution: Back up your data and reinitialize the system.

For more information, see [Recover from ransomware attacks](#).

How do I change the logon password within an instance?

For details, see [#unique_264](#).

Why can't I add sound or video cards to ECS instances?

This is because the servers that Alibaba Cloud provides are not multimedia servers. These servers do not provide sound card and video card components.

Can I transfer the remaining service time of an ECS instance to another ECS instance?

No. If you want both higher flexibility and cost-effectiveness, we recommend that you use a combination of reserved instances and pay-as-you-go instances. For more information, see [#unique_127](#).

Do ECS instances provide database capabilities by default?

No, ECS instances do not provide database capabilities by default. To use database services, perform the following operations:

- Deploy your database. For more information about how to deploy a database on an ECS instance, see [#unique_345](#).
- Purchase [ApsaraDB for RDS](#) provided by Alibaba Cloud.
- Use a database image provided in Alibaba Cloud Marketplace.

Can I build a database on an ECS instance?

Yes, you can install database software and configure a database environment on an ECS instance as needed. You can also separately purchase [ApsaraDB for RDS](#) provided by Alibaba Cloud.

Do ECS instances support Oracle databases?

Yes. We recommend that before installing an Oracle database, you perform a performance stress test on an ECS instance to ensure that the instance can satisfy the read/write requirements of the database.

Are public IP addresses and private IP addresses independent? Can I specify or add IP addresses?

In classic networks, public and private IP addresses are independent of each other. Private IP addresses in classic networks are used for communication between ECS instances and communication between ECS instances and OSS or RDS instances. When the 0 Mbit/s bandwidth is selected for an ECS instance, no public IP address is allocated to the instance. Typically, public and private IP addresses in classic networks do not change. You cannot specify, select, or add IP addresses.

In VPCs, NAT gateways map public IP addresses to private IP addresses. You can specify secondary private IP addresses or add multiple IP addresses.

Can an ECS instance be used for load balancing?

ECS Linux and Windows instances can be used for load balancing. You must ensure that the configurations of ECS instances serving as Web servers meet the requirements for website code to run. Load balancing can be implemented with as few as one ECS instance under an account. However, we recommend you use two or more ECS instances for load balancing to ensure the load balancing effect.

Can I change the region of an ECS instance?

No, you cannot change the region of an ECS instance.

Can I modify the partition size of a purchased disk?

For system security and stability purposes, system disks do not support secondary partitioning on either Windows or Linux instances. If you use a third-party tool to perform secondary partitioning, unknown exceptions such as system failure and data loss might occur.

Secondary partitioning of data disks may cause data loss. We do not recommend that you perform secondary partitioning on data disks.

How do I replace the public IP address of my ECS instance with an Elastic IP Address (EIP)?

To replace the public IP address of your ECS instance with an EIP, ensure that you have selected pay-by-bandwidth as the bandwidth billing method for your ECS instance and that you have purchased a new EIP. Then, perform the following steps:

1. Change the bandwidth billing method from pay-by-bandwidth to pay-by-traffic.
2. [Convert the public IP address of the ECS instance to an EIP.](#)
3. [Disassociate the EIP obtained in the previous step from the ECS instance.](#)
4. [Associate the EIP you purchased with the ECS instance.](#)

How do I view my subscription instances in the same region?

You can use the resource overview feature to view your subscription instances in the same region. For more information, see [#unique_348](#).

How do I view subscription instances in all regions under my account?

You can go to the renewal page to view subscription instances in all regions under your account.

1. Log on to the [ECS console](#).
2. In the top navigation bar, choose Billing Management > Renew.

When can I forcibly stop my ECS instance? What will the consequence be?

If you cannot stop an instance by shutting it down properly, you can forcibly stop it. Forcible stop is equivalent to a physical shutdown, and may cause data loss if data in the instance is not yet written to disks.

Why can't I reactivate my ECS instance?

You may fail to reactivate your ECS instance due to any of the following reasons:

- Your account has overdue payments. Pay the outstanding bills and try again.

- System is busy. Try again later.
- Resources are sold out.

**Note:**

You can go to the [ECS Instance Types Available for Each Region](#) page to view the instance types available in each region.

What is the AliVulfix process in an ECS instance?

The AliVulfix process is an Alibaba Cloud Security program that detects vulnerabilities in ECS instances.

How do I defend ECS instances against attacks?

ECS instances use Alibaba Cloud Security to defend against DDoS attacks. CloudMonitor or deployed on ECS instances can automatically detect attacks and scrub suspicious traffic. For high-volume attacks, Alibaba Cloud implements black hole filtering to protect your instances. To strengthen security protection, we recommend that you install security software and disable ports that are not commonly used.

What security services does Alibaba Cloud provide?

Alibaba Cloud Security is powered by the robust data analysis capability of the Alibaba Cloud cloud computing platform to provide you with a comprehensive set of security services, such as security vulnerability detection, webpage trojan detection, host intrusion detection, and DDoS protection.

For information about more security services, see [Security](#).

After I renew a subscription ECS Linux instance that expires, I am still unable to access the website it hosts. What can I do?

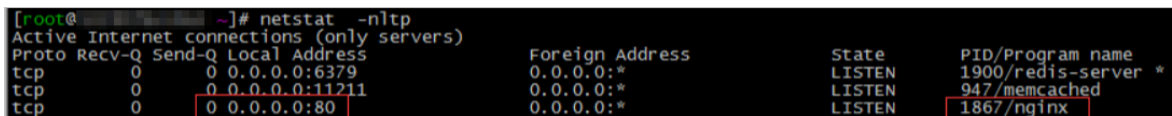
Problem description: The ECS instance is in the Stopped state after it expires. After you renew and start the instance, the website it hosts cannot be accessed.

Cause: This may be because the website service has not been started.

Solution:

1. Connect to the ECS instance remotely and run the following command to check whether the website service has been started:

```
# netstat -nltcp // Check whether port 80 on the ECS instance is being listened on .
```



```
[root@ ~]# netstat -nltcp
Active Internet connections (only servers)
Proto Recv-Q Send-Q Local Address           Foreign Address         State       PID/Program name
tcp        0      0 0.0.0.0:6379          0.0.0.0:*               LISTEN      1900/redis-server *
tcp        0      0 0.0.0.0:11211        0.0.0.0:*               LISTEN      947/memcached
tcp        0      0 0.0.0.0:80           0.0.0.0:*               LISTEN      1867/nginx
```

2. If no information about port 80 is displayed in the command output, the website service has not been started. Run a command to manually start the website service and relevant services.

In Linux, websites are typically developed based on PHP and MySQL.

- In Apache, you only need to start the website service and MySQL.

```
#!/ etc / init . d / httpd start // Start the website service . This command is applicable to Apache .
#!/ etc / init . d / mysqld start // Start MySQL .
```

- In NGINX, you need to start the website service, PHP, and MySQL.

```
#!/ etc / init . d / nginx start // Start the website service . This command is applicable to NGINX .
#!/ etc / init . d / php - fpm start // Start PHP .
#!/ etc / init . d / mysqld start // Start MySQL .
```

3. Re-check whether the website service has been started.

```
# netstat -nltcp // Check whether port 80 on the ECS instance is being listened on .
```

4. After the website service has been started, access the website again.

If the problem persists, [submit a ticket](#).

How do I check, partition, and format the data disks of a Linux instance?

You can run the `df -h` command to check the capacity and usage of disks, and run the `fdisk -l` command to view disk information. For details about how to partition and format disks in Linux instances, see [#unique_200](#).

How do I upload files to a Linux instance?

You can use the FTP service to upload files to a Linux instance.

How do I change the owner and owner group of directories and files on a Linux instance?

If the file or directory permissions are not correctly configured on the Web server, a 403 error will occur when you access websites hosted on the instance. Therefore, before you adjust a file or directory, you must identify the identity under which the file or directory process is running.

You can run the `ps` and `grep` commands to query the identities under which processes are running.

You can run the `ls -l` command to query the owners and owner groups of files and directories.

To modify the permissions, run the `chown` command. For example, you can run the `chown -R www . www / alidata / www / phpwind /` command to change the owner and owner group of all files and directories under directory `/ alidata / www / phpwind` to account `www`.

How do I update the software source of Linux instances?

You can use an automatic software source updating tool. For more information, see [Automatic source updating tool for Linux ECS](#).

How do I install a VNC server to access a Linux system by using a GUI?

For details, see [Install and run VNC Server on a Linux instance](#).

What limits apply to the transfer and change of public IP addresses of ECS instances?

The following limits apply to the transfer and change of public IP addresses of ECS instances:

- In classic networks:
 - You cannot transfer public IP addresses across accounts.
 - The public IP address of an ECS instance is mutable only within 6 hours after the instance is created, and can be changed a maximum of 3 times. For more information, see [#unique_349](#).
 - If Anti-DDoS Pro is deployed on your instance, you can change the IP address of an ECS instance a maximum of 10 times in the Anti-DDoS Pro console. For more information, see [#unique_350](#) of the Anti-DDoS Pro User Guide.

- In VPCs:

- You cannot transfer public IP addresses or EIPs across accounts.
- If no public IP address is allocated to your instance, you can bind an EIP to the instance. You can replace the public IP address of your instance with an EIP.
- If a public IP address is allocated to your instance:
 - This public IP address is mutable within 6 hours after the instance is created, and can be changed a maximum of 3 times. For more information, see [#unique_349](#).
 - You can replace this public IP address by converting it to an EIP. For more information, see [#unique_351](#).

If you have further questions, [submit a ticket](#).

Can I access amazon.com from my ECS instance?

You can access amazon.com from your ECS instance if your ECS instance can connect to the public network properly.

Why can't I access a website hosted outside China after I log on to my ECS instance?

You can access a website hosted outside China from your ECS instance only when the website complies with the local regulatory policies of the country or region where your instance is located. Make sure that your ECS instance can connect to the public network properly and that the website is compliant with the preceding local regulatory policies.

I cannot purchase any more pay-as-you-go instances. What can I do?

If you have reached the maximum number of pay-as-you-go instances that you can purchase, you cannot purchase any more pay-as-you-go instances. For more information, see [Limits](#). You can log on to the ECS console and view your resource quotas on the Privileges & Quotas page. For more information, see [#unique_352](#).

Are fees still incurred after a pay-as-you-go ECS instance is automatically stopped due to an overdue payment or is manually stopped?

When the payment for a pay-as-you-go instance is overdue, the instance is automatically stopped and taken out of service.

You can stop a running pay-as-you-go instance in the ECS console or by calling the StopInstance operation. When the instance is stopped, its status changes to Stopped.

Billing of a pay-as-you-go instance in the Stopped state depends on the instance network type.

- **VPC:** You can enable the No Fees for Stopped Instances (VPC-Connected) feature. After the feature is enabled, pay-as-you-go instances in VPCs start being billed when they are created, stop being billed when they enter the Stopped state, and resume being billed when they are started again. With the No Fees for Stopped Instances (VPC-Connected) feature enabled, the vCPUs, memory resources, and public IP addresses of pay-as-you-go instances in VPCs will not incur any additional costs after the instances are stopped, but the disks will continue to be billed. For more information, see [#unique_142](#).
- **Classic network:** Pay-as-you-go instances in classic networks will continue to be billed even when they are in the Stopped state.

The order fails to be placed when I change the billing method of my ECS instance from pay-as-you-go to subscription. What can I do?

Your order may fail to be placed due to any of the following reasons:

- The instance is in a state that does not support changing billing methods.
- The instance is scheduled for an automatic release at a specific point in time and does not support changing billing methods.
- The instance is in the Expired state.
- The instance information has changed.
- A change order is not yet paid for the instance.

If you are prompted with any of the preceding error messages, adjust the instance accordingly.

How long does it take to change the billing method after you pay for the order?

After you pay for the order, it takes 1 to 4 seconds to change the billing method of up to 20 instances at a time. After the change, the billing method of the instance is changed to Subscription in the console.

What can I do if the change from pay-as-you-go to subscription fails?

[Submit a ticket](#).

Will the bandwidth billing method change after I change from pay-as-you-go to subscription?

No. You can only change the billing method of instances and disks. For information about how to change the bandwidth billing method, see [#unique_120/unique_120_Connect_42_ChangeBilling](#).

I have an unpaid order for changing the billing method of an instance from pay-as-you-go to subscription. If I upgrade the instance, will the order still be valid?

An order is created when you change the billing method of your instance from pay-as-you-go to subscription. You must complete the payment for this order for the change to take effect. If you upgrade the instance before the order is paid for, the order payment cannot be completed because the instance components change and the original order amount does not meet the requirements of the new billing method. If you still want to change the billing method, you must cancel the unpaid order and place a new order.

What can I do if the change from subscription to pay-as-you-go fails?

The change from subscription to pay-as-you-go may fail due to any of the following reasons:

- The instance is in a state that does not support changing billing methods.
- The instance is in the Expired state.
- The instance information has changed.

If you are prompted with any of the preceding error messages, adjust the instance accordingly. If this problem persists, [submit a ticket](#).

When I try to change the billing method of a disk in an ECS instance, an error message appears indicating that I have already changed the billing method three times. What does this mean?

Each ECS instance can be downgraded a maximum of three times. Downgrade operations include instance specifications downgrades, bandwidth configuration downgrades, and the change of the disk billing method from subscription to pay-as-you-go.

Why can't I change a pay-as-you-go instance to a subscription one?

To change the billing method for an ECS instance, ensure that the instance meet the following conditions:

How do I view the expiration time of my subscription instances?

You can log on to the ECS console and go to the Instances page. View the expiration time of your subscription instances in the Billing Method column.



Note:

If the Billing Method column is not displayed, click  in the upper-right corner.

In the dialog box that appears, select Billing Method and click OK.

Instance ID/Name	Tag	Monitoring	Zone	IP Address	Status	Network Type	Specifications	Billing Method	Automatic Renewal	Stop Mode	Actions
			Shanghai Zone B			VPC	1 vCPU 512 MB (I/O Optimized) ecs.t5-ic2m1.nano 5Mbps (Peak Value)	Pay-As-You-Go 12 September 2018, 13:11 Create	No Fees for Stopped Instances (VPC-Connected)		Manage Change Instance Type More
			Shanghai Zone E			VPC	2 vCPU 8 GiB (I/O Optimized) ecs.g5.large 0Mbps (Peak Value)	Pay-As-You-Go 20 August 2018, 14:25 Create	No Fees for Stopped Instances (VPC-Connected)		Manage Change Instance Type More
			Shanghai Zone E			VPC	2 vCPU 4 GiB (I/O Optimized) ecs.c5.large 25Mbps (Peak Value)	Subscription 7 July 2020, 00:00 Expire	Renew	-	Manage Connect Change Configuration Renew More