Alibaba Cloud
Elastic Compute Service

Instances

Issue: 20190611

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 `cd / d C :/ windows` command to enter the Windows system folder.
Italics	It is used for parameters and variables.	`bae log list --instanceid` *Instance_ID*
[] or [a|b]	It indicates that it is a optional value, and only one item can be selected.	`ipconfig` *[-all|-t]*

Style	Description	Example
{} or {a|b}	It indicates that it is a required value, and only one item can be selected.	`swich {stand | slave}`

Contents

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 *Subscription* and *Pay-As-You-Go*.	Stopped	Yes
Expired and Being Recycled	Stable	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 Expired and Being Recycled state. · You can renew the instance before it enters the Expired and Being Recycled state. If the renewal is successful, all resources are retained without being affected. · After the instance enters the Expired 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 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	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. · 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 you 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 type families

This topic describes the available ECS instance type families.

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

ECS instances are categorized into specification types, which are called type families, based on the business scenarios they can be applied to. You may select multiple type families for one business scenario. Each type family contains multiple *ECS instance types* with different CPU and memory specifications, including the CPU model and clock speed. Besides the instance type, you must also define a block storage, an image, and the network service when you create an instance.

> Note:
> The availability of instance type families and their types varies from region to region. Go to the purchase page to check the available instance types.

Alibaba Cloud ECS provides two kinds of instance type families: enterprise-level instance type families and entry-level instance type families. Type families for enterprise-level computing offer stable performance and dedicated resources, while entry-level type families are ideal for small and mid-sized websites, or individual customers. For the differences, see *Enterprise-level instances and entry-level instances FAQ*.

> Note:
> - Some instance types are no longer available for purchase. For more information, see *Phased-out instance types*.
> - Upgrading instance types is supported within or between certain instance type families. For such families and corresponding upgrade rules, see *Instance type families that support upgrading instance types*.
> - Upgrading instance types is not supported within or between the following instance type families: d1, d1ne, i1, i2, i2g, vgn5i, ga1, gn5, gn6i, f1, f3, ebmc4, ebmg5, sccg5, scch5, and sccgn6.

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

- **Type families for enterprise-level computing on the x86-architecture:**
  - *g5, general-purpose type family*
  - *sn2ne, general-purpose type family with enhanced network performance*
  - *ic5, intensive compute instance type family*
  - *c5, compute instance type family*
  - *sn1ne, compute optimized type family with enhanced network performance*
  - *r5, memory instance type family*
  - *re4, memory optimized type family with enhanced performance*
  - *re4e, memory optimized type family with enhanced performance*
  - *se1ne, memory optimized type family with enhanced network performance*
  - *se1, memory optimized type family*
  - *d1ne, big data type family with enhanced network performance*
  - *d1, big data type family*
  - *i2, type family with local SSD disks*
  - *i2g, type family with local SSD disks*
  - *i1, type family with local SSD disks*
  - *hfc5, compute optimized type family with high clock speed*
  - *hfg5, general-purpose type family with high clock speed*
- **Type families for enterprise-level heterogeneous computing:**
  - *vgn5i, light-weight compute optimized type family with GPU*
  - *gn6i, compute optimized type family with GPUs*
  - *gn6v, compute optimized type family with GPU*
  - *gn5, compute optimized type family with GPU*
  - *gn5i, compute optimized type family with GPU*
  - *gn4, compute optimized type family with GPU*
  - *ga1, visualization compute optimized type family with GPU*
  - *f1, compute optimized type family with FPGA*
  - *f3, compute optimized type family with FPGA*

- ECS Bare Metal Instance type families and Super Computing Cluster (SCC) instance type families:
  - *ebmhfg5, ECS Bare Metal Instance type family with high clock speed*
  - *ebmc4, computing ECS Bare Metal Instance type family*
  - *ebmg5, general-purpose ECS Bare Metal Instance type family*
  - *scch5, Super Computing Cluster (SCC) instance type family with high clock speed*
  - *sccg5, general-purpose Super Computing Cluster (SCC) instance type family*
  - *sccgn6, compute optimized Super Computing Cluster (SCC) instance type family with GPUs*
- Type families for entry-level computing on the x86-architecture:
  - *t5, burstable instances*
  - *xn4/n4/mn4/e4, type families of previous generations for entry-level users, computing on the x86-architecture*

g5, general-purpose type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:4
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - Enterprise-level applications of various types and sizes
  - Medium and small database systems, cache, and search clusters
  - Data analysis and computing
  - Computing clusters and data processing reliant on memory

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.g5.large	2	8.0	N/A	1.0	300	2	2	6
ecs.g5.xlarge	4	16.0	N/A	1.5	500	2	3	10
ecs.g5.2xlarge	8	32.0	N/A	2.5	800	2	4	10
ecs.g5.3xlarge	12	48.0	N/A	4.0	900	4	6	10
ecs.g5.4xlarge	16	64.0	N/A	5.0	1,000	4	8	20
ecs.g5.6xlarge	24	96.0	N/A	7.5	1,500	6	8	20
ecs.g5.8xlarge	32	128.0	N/A	10.0	2,000	8	8	20
ecs.g5.16xlarge	64	256.0	N/A	20.0	4,000	16	8	20

See *other instance type families*.

sn2ne, general-purpose type family with enhanced network performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU 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
- Supports strong network performance through sufficient computing capacity

- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - Enterprise-level 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) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sn2ne. large	2	8.0	N/A	1.0	300	2	2	6
ecs. sn2ne. xlarge	4	16.0	N/A	1.5	500	2	3	10
ecs. sn2ne. 2xlarge	8	32.0	N/A	2.0	1,000	4	4	10
ecs. sn2ne. 3xlarge	12	48.0	N/A	2.5	1,300	4	6	10
ecs. sn2ne. 4xlarge	16	64.0	N/A	3.0	1,600	4	8	20
ecs. sn2ne. 6xlarge	24	96.0	N/A	4.5	2,000	6	8	20
ecs. sn2ne. 8xlarge	32	128.0	N/A	6.0	2,500	8	8	20

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sn2ne. 14xlarge	56	224.0	N/A	10.0	4,500	14	8	20

See *other instance type families*.

ic5, intensive compute instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:1
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Web front-end servers
  - Data analysis, batch compute, and video coding
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - Massively Multiplayer Online (MMO) game frontends

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.ic5.large	2	2.0	N/A	1.0	300	2	2	6
ecs.ic5.xlarge	4	4.0	N/A	1.5	500	2	3	10
ecs.ic5.2xlarge	8	8.0	N/A	2.5	800	2	4	10
ecs.ic5.3xlarge	12	12.0	N/A	4.0	900	4	6	10
ecs.ic5.4xlarge	16	16.0	N/A	5.0	1,000	4	8	20

See *other instance type families*.

c5, compute instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:2
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - Web front-end servers
  - Massively Multiplayer Online (MMO) game frontends
  - Data analysis, batch compute, and video coding
  - High-performance science and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.c5. large	2	4.0	N/A	1.0	300	2	2	6
ecs.c5. xlarge	4	8.0	N/A	1.5	500	2	3	10
ecs.c5. 2xlarge	8	16.0	N/A	2.5	800	2	4	10
ecs.c5. 3xlarge	12	24.0	N/A	4.0	900	4	6	10
ecs.c5. 4xlarge	16	32.0	N/A	5.0	1,000	4	8	20
ecs.c5. 6xlarge	24	48.0	N/A	7.5	1,500	6	8	20
ecs.c5. 8xlarge	32	64.0	N/A	10.0	2,000	8	8	20
ecs.c5. 16xlarge	64	128.0	N/A	20.0	4,000	16	8	20

See *other instance type families*.

sn1ne, compute optimized type family with enhanced network performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU 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
- Supports strong network performance through sufficient computing capacity

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

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sn1ne. large	2	4.0	N/A	1.0	300	2	2	6
ecs. sn1ne. xlarge	4	8.0	N/A	1.5	500	2	3	10
ecs. sn1ne. 2xlarge	8	16.0	N/A	2.0	1,000	4	4	10
ecs. sn1ne. 3xlarge	12	24.0	N/A	2.5	1,300	4	6	10
ecs. sn1ne. 4xlarge	16	32.0	N/A	3.0	1,600	4	8	20
ecs. sn1ne. 6xlarge	24	48.0	N/A	4.5	2,000	6	8	20
ecs. sn1ne. 8xlarge	32	64.0	N/A	6.0	2,500	8	8	20

See *other instance type families*.

r5, memory instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - High-performance databases and high memory databases
  - Data analysis and mining, and distributed memory cache
  - Hadoop, Spark, and other enterprise-level applications with large memory requirements

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.r5. large	2	16.0	N/A	1.0	300	2	2	6
ecs.r5. xlarge	4	32.0	N/A	1.5	500	2	3	10
ecs.r5. 2xlarge	8	64.0	N/A	2.5	800	2	4	10
ecs.r5. 3xlarge	12	96.0	N/A	4.0	900	4	6	10
ecs.r5. 4xlarge	16	128.0	N/A	5.0	1,000	4	8	20

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.r5. 6xlarge	24	192.0	N/A	7.5	1,500	6	8	20
ecs.r5. 8xlarge	32	256.0	N/A	10.0	2,000	8	8	20
ecs.r5. 16xlarge	64	512.0	N/A	20.0	4,000	16	8	20

See *other instance type families*.

re4, memory optimized instance type family with enhanced performance

Features

- Supports SSD Cloud Disks and Ultra Disks
- I/O-optimized
- Optimized for high-performance databases, high memory databases, and other memory-intensive enterprise applications
- 2.2 GHz Intel Xeon E7 8880 v4 (Broadwell) processors, up to 2.4 GHz Turbo Boot
- Equipped with a vCPU 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 high 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 disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.re4. 20xlarge	80	960.0	N/A	15.0	2,000	16	8	20
ecs.re4. 40xlarge	160	1920.0	N/A	30.0	4,500	16	8	20

See *other instance type families*.

re4e, memory optimized type family with enhanced performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Optimized for high-performance databases, high memory databases, and other memory-intensive enterprise applications
- 2.2 GHz Intel Xeon E7 8880 v4 (Broadwell) processors, up to 2.4 GHz Turbo Boot
- Equipped with a vCPU to memory ratio of 1:24, up to 3840.0 GiB memory
- Suitable for the following scenarios:
  - High-performance databases and high 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 disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. re4e. 40xlarge	160	3840.0	N/A	30.0	4,500	16	15	20

See *other instance type families*.

se1ne, memory optimized type family with enhanced network performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:8
- Ultra high packet receive and forwarding rate
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - High-performance databases and large memory databases
  - Data analysis and mining, and distributed memory cache
  - Hadoop, Spark, and other enterprise-level applications with large memory requirements

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. se1ne. large	2	16.0	N/A	1.0	300	2	2	6
ecs. se1ne. xlarge	4	32.0	N/A	1.5	500	2	3	10
ecs. se1ne. 2xlarge	8	64.0	N/A	2.0	1,000	4	4	10
ecs. se1ne. 3xlarge	12	96.0	N/A	2.5	1,300	4	6	10
ecs. se1ne. 4xlarge	16	128.0	N/A	3.0	1,600	4	8	20
ecs. se1ne. 6xlarge	24	192.0	N/A	4.5	2,000	6	8	20
ecs. se1ne. 8xlarge	32	256.0	N/A	6.0	2,500	8	8	20
ecs. se1ne. 14xlarge	56	480.0	N/A	10.0	4,500	14	8	20

See *other instance type families*.

se1, memory optimized type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:8

- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - High-performance databases and large memory databases
  - Data analysis and mining, and distributed memory cache
  - Hadoop, Spark, and other enterprise-level applications with large memory requirements

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.se1. large	2	16.0	N/A	0.5	100	1	2	6
ecs.se1. xlarge	4	32.0	N/A	0.8	200	1	3	10
ecs.se1. 2xlarge	8	64.0	N/A	1.5	400	1	4	10
ecs.se1. 4xlarge	16	128.0	N/A	3.0	500	2	8	20
ecs.se1. 8xlarge	32	256.0	N/A	6.0	800	3	8	20
ecs.se1. 14xlarge	56	480.0	N/A	10.0	1,200	4	8	20

See *other instance type families*.

d1ne, big data type family with enhanced network performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks

- High-volume local SATA HDD disks with high I/O throughput and up to 35 Gbit/s of bandwidth for a single instance
- Equipped with a vCPU to memory ratio of 1:4, designed for big data scenarios
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Hadoop MapReduce, HDFS, Hive, HBase, and so on
  - Spark in-memory computing, MLlib, and so on
  - Enterprises that require big data computing and storage analysis, such as those in the Internet and finance industries, to store and compute massive volumes of data
  - Elasticsearch, logs, and so on

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. d1ne. 2xlarge	8	32.0	4 * 5500	6.0	1,000	4	4	10
ecs. d1ne. 4xlarge	16	64.0	8 * 5500	12.0	1,600	4	8	20
ecs. d1ne. 6xlarge	24	96.0	12 * 5500	16.0	2,000	6	8	20
ecs. d1ne-c8d3. 8xlarge	32	128.0	12 * 5500	20.0	2,000	6	8	20
ecs. d1ne. 8xlarge	32	128.0	16 * 5500	20.0	2,500	8	8	20

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. d1ne-c14d3. 14xlarge	56	160.0	12 * 5500	35.0	4,500	14	8	20
ecs. d1ne. 14xlarge	56	224.0	28 * 5500	35.0	4,500	14	8	20

Note:

- You cannot change configurations of d1ne instances.
- For more information about d1ne type families, see *FAQ on d1 and d1ne*.

See *other instance type families*.

d1, big data type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- High-volume local SATA HDD disks with high I/O throughput and up to 17 Gbit/s of bandwidth for a single instance
- Equipped with a vCPU to memory ratio of 1:4, designed for big data scenarios
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity

- Suitable for the following scenarios:
  - Hadoop MapReduce, HDFS, Hive, and HBase
  - Spark in-memory computing and MLlib
  - Enterprises that require big data computing and storage analysis, such as those in the Internet and finance industries, to store and compute massive volumes of data
  - Elasticsearch and logs

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.d1. 2xlarge	8	32.0	4 * 5500	3.0	300	1	4	10
ecs.d1. 3xlarge	12	48.0	16 * 5500	4.0	400	1	6	10
ecs.d1. 4xlarge	16	64.0	8 * 5500	6.0	600	2	8	20
ecs.d1. 6xlarge	24	96.0	12 * 5500	8.0	800	2	8	20
ecs.d1 -c8d3. 8xlarge	32	128.0	12 * 5500	10.0	1,000	4	8	20
ecs.d1. 8xlarge	32	128.0	16 * 5500	10.0	1,000	4	8	20
ecs.d1- c14d3. 14xlarge	56	160.0	12 * 5500	17.0	1,800	6	8	20
ecs.d1. 14xlarge	56	224.0	28 * 5500	17.0	1,800	6	8	20

Note:

For more information about d1 type families, see *FAQ on d1 and d1ne*.

See *other instance type families*.

i2, type family with local SSD disks

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- High-performance local NVMe SSD disks with high IOPS, high I/O throughput, and low latency.
- Equipped with a vCPU to memory ratio of 1:8, designed for high-performance databases
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases, such as Cassandra and MongoDB
  - Search applications, such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.i2. xlarge	4	32.0	1 * 894	1.0	500	2	3	10
ecs.i2. 2xlarge	8	64.0	1 * 1788	2.0	1,000	2	4	10
ecs.i2. 4xlarge	16	128.0	2 * 1788	3.0	1,500	4	8	20
ecs.i2. 8xlarge	32	256.0	4 * 1788	6.0	2,000	8	8	20
ecs.i2. 16xlarge	64	512.0	8 * 1788	10.0	4,000	16	8	20

See *other instance type families*.

i2g, type family with local SSD disks

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- High-performance local NVMe SSD disks with high IOPS, high I/O throughput, and low latency.
- Equipped with a vCPU to memory ratio of 1:4, designed for high-performance databases
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases, such as Cassandra and MongoDB
  - Search applications, such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.i2g. 2xlarge	8	32.0	1 * 894	2.0	1,000	2	4	10
ecs.i2g. 4xlarge	16	64.0	1 * 1788	3.0	1,500	4	8	20
ecs.i2g. 8xlarge	32	128.0	2 * 1788	6.0	2,000	8	8	20
ecs.i2g. 16xlarge	64	256.0	4 * 1788	10.0	4,000	16	8	20

See *other instance type families*.

i1, type family with local SSD disks

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- High-performance local NVMe SSD disks with high IOPS, high I/O throughput, and low latency
- Equipped with a vCPU to memory ratio of 1:4, designed for big data scenarios
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases, such as Cassandra and MongoDB
  - Search applications, such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.i1. xlarge	4	16.0	2 * 104	0.8	200	1	3	10
ecs.i1. 2xlarge	8	32.0	2 * 208	1.5	400	1	4	10
ecs.i1. 3xlarge	12	48.0	2 * 312	2.0	400	1	6	10
ecs.i1. 4xlarge	16	64.0	2 * 416	3.0	500	2	8	20
ecs.i1 -c5d1. 4xlarge	16	64.0	2 * 1456	3.0	400	2	8	20
ecs.i1. 6xlarge	24	96.0	2 * 624	4.5	600	2	8	

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.i1. 8xlarge	32	128.0	2 * 832	6.0	800	3	8	20
ecs.i1- c10d1. 8xlarge	32	128.0	2 * 1456	6.0	800	3	8	20
ecs.i1. 14xlarge	56	224.0	2 * 1456	10.0	1,200	4	8	20

See *other instance type families*.

hfc5, compute optimized type family with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Stable performance
- 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors
- Equipped with a vCPU to memory ratio of 1:2
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - High-performance Web front-end servers
  - High-performance science and engineering applications
  - Massively Multiplayer Online (MMO) games and video coding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. hfc5. large	2	4.0	N/A	1.0	300	2	2	6
ecs. hfc5. xlarge	4	8.0	N/A	1.5	500	2	3	10
ecs. hfc5. 2xlarge	8	16.0	N/A	2.0	1,000	2	4	10
ecs. hfc5. 3xlarge	12	24.0	N/A	2.5	1,300	4	6	10
ecs. hfc5. 4xlarge	16	32.0	N/A	3.0	1,600	4	8	20
ecs. hfc5. 6xlarge	24	48.0	N/A	4.5	2,000	6	8	20
ecs. hfc5. 8xlarge	32	64.0	N/A	6.0	2,500	8	8	20

See *other instance type families*.

hfg5, general-purpose type family with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Stable performance
- 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors
- Equipped with a vCPU to memory ratio of 1:4, except for the 56 vCPU instance type

- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - High-performance Web front-end servers
  - High-performance science and engineering applications
  - Massively Multiplayer Online (MMO) games and video coding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. hfg5. large	2	8.0	N/A	1.0	300	2	2	6
ecs. hfg5. xlarge	4	16.0	N/A	1.5	500	2	3	10
ecs. hfg5. 2xlarge	8	32.0	N/A	2.0	1,000	2	4.	10
ecs. hfg5. 3xlarge	12	48.0	N/A	2.5	1,300	4	6	10
ecs. hfg5. 4xlarge	16	64.0	N/A	3.0	1,600	4	8	20
ecs. hfg5. 6xlarge	24	96.0	N/A	4.5	2,000	6	8	20
ecs. hfg5. 8xlarge	32	128.0	N/A	6.0	2,500	8	8	20
ecs. hfg5. 14xlarge	56	160.0	N/A	10.0	4,000	14	8	20

See *other instance type families*.

vgn5i, light-weight compute optimized type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Use an NVIDIA P4 GPU computation accelerator
- Contains a virtual GPU (which is the result of partitioned virtualization)
  - 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
- Equipped with a vCPU to memory ratio of 1:3
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Real-time online rendering required for cloud gaming and AR/VR applications
  - AI reasoning (including deep and machine learning), used in the elastic deployment of Internet services that use AI reasoning and computing
  - Educational and modeling experiment environments that use deep learning

Instance types

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/s) **	Packet forwar rate (thousa pps) ***	NIC queues ****	ENIs *****
ecs.vgn5i-m1.large	2	6	N/A	P4*1/8	1	1	300	2	2
ecs.vgn5i-m2.xlarge	4	12	N/A	P4*1/4	2	2	500	2	3

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/s) **	Packet forwar rate (thousa pps) ***	NIC queues ****	ENIs *****
ecs.vgn5i-m4.2xlarge	8	24	N/A	P4*1/2	4	3	800	2	4
ecs.vgn5i-m8.4xlarge	16	48	N/A	P4*1	8	5	1,000	4	5

Note:
For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

gn6i, compute optimized type family with GPUs

Features

- I/O-optimized
- Equipped with a vCPU to memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports ESSD Cloud Disks (million-level IOPS), SSD Cloud Disks, and Ultra Disks
- Achieves better performance with the X-Dragon new-generation compute architecture
- Uses NVIDIA T4 GPU processors:
  - Based on the new NVIDIA Turing architecture
  - Up to 320 Turing Tensor Cores
  - Up to 2,560 CUDA Cores
  - Mixed-precision Tensor Cores support 65 TFlops FP16, 130 INT8 TOPS, and 260 INT4 TOPS
  - 16 GB GPU memory capacity (320 GB/s bandwidth)

- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - AI (deep learning and machine learning) inference, computer vision, voice recognition, voice synthesization, natural language processing, machine translation, and reference systems
  - Real-time online rendering required for cloud gaming and AR/VR applications
  - Heavy-load graphic computing or graphic workstations
  - GPU-accelerated databases
  - High-performance computing

Instance types

Instance types	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****
ecs.gn6i-c4g1.xlarge	4	15	N/A	T4*1	16	4	500	2	2
ecs.gn6i-c8g1.2xlarge	8	31	N/A	T4*1	16	5	800	2	2
ecs.gn6i-c16g1.4xlarge	16	62	N/A	T4*1	16	6	1,000	4	3
ecs.gn6i-c24g1.6xlarge	24	93	N/A	T4*1	16	7.5	1,200	6	4
ecs.gn6i-c24g1.12xlarge	48	186	N/A	T4*2	32	15	2,400	12	6

Instance types	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****
ecs.gn6i-c24g1.24xlarge	96	372	N/A	T4*4	64	30	4,800	24	8
ecs.gn6i-c32g1.8xlarge	32	124	N/A	T4*1	16	10	1,600	8	6
ecs.gn6i-c48g1.12xlarge	48	186	N/A	T4*1	16	12	2,400	12	6
ecs.gn6i-c72g1.18xlarge	72	279	N/A	T4*1	16	21.5	3,600	18	8

Note:

For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

gn6v, compute optimized type family with GPUs

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Uses NVIDIA V100 GPU processors
- Equipped with a vCPU to memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors

- Uses NVIDIA V100 GPU processors (with the SXM2 module):
  - Based on the new NVIDIA Volta architecture
  - 16 GB HBM2 GPU memory capacity (900 GB/s bandwidth)
  - Up to 5,120 CUDA Cores
  - Up to 640 Tensor Cores
  - Supports up to six NVLink connections and total bandwidth of 300 GB/s (25 GB/s per connection)
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning, autonomous vehicles, voice recognition, and other AI applications
  - Scientific computing, computational finance, genomics, and environmental analysis

Instance types

Instance types	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forwarding rate (Thousand pps) ***	NIC queues ****	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	4	4	10
ecs.gn6v-c8g1.8xlarge	32	128.0	N/A	4 * NVIDIA V100	4 * 16	10.0	2,000	8	8	20
ecs.gn6v-c8g1.16xlarge	64	256.0	N/A	8 * NVIDIA V100	8 * 16	20.0	2,500	16	8	20

Note:

For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

gn5, compute optimized type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Uses NVIDIA P100 GPU processors
- No fixed ratio of vCPU to memory
- High-performance local NVMe SSD disks
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning
  - Scientific computing, such as computational fluid dynamics, computational finance, genomics, and environmental analysis
  - High-performance computing, rendering, multi-media coding and decoding, and other server-side GPU compute workloads

Instance types

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	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	1	3	10
ecs.gn5-c8g1.2xlarge	8	60.0	440	1 * NVIDIA P100	1 * 16	3.0	400	1	4	10

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.gn5-c4g1.2xlarge	8	60.0	880	2 * NVIDIA P100	2 * 16	5.0	1,000	2	4	10
ecs.gn5-c8g1.4xlarge	16	120.0	880	2 * NVIDIA P100	2 * 16	5.0	1,000	4	8	20
ecs.gn5-c28g1.7xlarge	28	112.0	440	1 * NVIDIA P100	1 * 16	5.0	1,000	8	8	20
ecs.gn5-c8g1.8xlarge	32	240.0	1760	4 * NVIDIA P100	4 * 16	10.0	2,000	8	8	20
ecs.gn5-c28g1.14xlarge	56	224.0	880	2 * NVIDIA P100	2 * 16	10.0	2,000	14	8	20
ecs.gn5-c8g1.14xlarge	54	480.0	3520	8 * NVIDIA P100	8 * 16	25.0	4,000	14	8	20

Note:

For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

gn5i, compute optimized type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Uses NVIDIA P4 GPU processors
- Equipped with a vCPU to memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning
  - Multi-media coding and decoding, and other server-side GPU compute workloads

Instance types

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/ s) **	Packet forwar rate (Thous pps) ***	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	2	2	6
ecs. gn5i- c4g1. xlarge	4	16.0	N/A	1 * NVIDIA P4	1 * 8	1.5	200	2	3	10
ecs. gn5i- c8g1. 2xlarge	8	32.0	N/A	1 * NVIDIA P4	1 * 8	2.0	400	4	4	10
ecs. gn5i- c16g1. 4xlarge	16	64.0	N/A	1 * NVIDIA P4	1 * 8	3.0	800	4	8	10

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.gn5i-c16g1.8xlarge	32	128.0	N/A	2 * NVIDIA P4	2 * 8	6.0	1,200	8	8	20
ecs.gn5i-c28g1.14xlarge	56	224.0	N/A	2 * NVIDIA P4	2 * 8	10.0	2,000	14	8	20

Note:
For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

gn4, compute optimized type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Uses NVIDIA M40 GPU processors
- No fixed ratio of CPU to memory
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning
  - Scientific computing, such as computational fluid dynamics, computational finance, genomics, and environmental analysis
  - High-performance computing, rendering, multi-media coding and decoding, and other server-side GPU compute workloads

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	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	1	3	10
ecs.gn4-c8g1.2xlarge	8	30.0	N/A	1 * NVIDIA M40	1 * 12	3.0	400	1	4	10
ecs.gn4.8xlarge	32	48.0	N/A	1 * NVIDIA M40	1 * 12	6.0	800	3	8	20
ecs.gn4-c4g1.2xlarge	8	60.0	N/A	2 * NVIDIA M40	2 * 12	5.0	500	1	4	10
ecs.gn4-c8g1.4xlarge	16	60.0	N/A	2 * NVIDIA M40	2 * 12	5.0	500	1	8	20
ecs.gn4.14xlarge	56	96.0	N/A	2 * NVIDIA M40	2 * 12	10.0	1,200	4	8	20

Note:

For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

ga1, visualization compute type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks

- AMD S7150 GPU processors
- Equipped with a vCPU to memory ratio of 1:2.5
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- High-performance local NVMe SSD disks
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Rendering, multimedia coding and decoding
  - Machine learning, high-performance computing, and high-performance databases
  - Other server-end business scenarios that require powerful concurrent floating-point compute capabilities

Instance types

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.ga1.xlarge	4	10.0	1 * 87	0.25 * AMD S7150	2	1.0	200	1	3	10
ecs.ga1.2xlarge	8	20.0	1 * 175	0.5 * AMD S7150	4	1.5	300	1	4	10
ecs.ga1.4xlarge	16	40.0	1 * 350	1 * AMD S7150	8	3.0	500	2	8	20
ecs.ga1.8xlarge	32	80.0	1 * 700	2 * AMD S7150	2 * 8	6.0	800	3	8	20
ecs.ga1.14xlarge	56	160.0	1 * 1400	4 * AMD S7150	4 * 8	10.0	1,200	4	8	20

Note:

For more information, see *Create an instance of ga1*.

See *other instance type families*.

f1, compute optimized type family with FPGA

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Intel ARRIA 10 GX 1150 FPGA
- Equipped with a vCPU to memory ratio of 1:7.5
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning and reasoning
  - Genomics research
  - Financial analysis
  - Picture transcoding
  - Computational workloads, such as real-time video processing and security

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	FPGA	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.f1-c8f1.2xlarge	8	60.0	N/A	Intel ARRIA 10 GX 1150	3.0	400	4	4	10
ecs.f1-c8f1.4xlarge	16	120.0	N/A	2 * Intel ARRIA 10 GX 1150	5.0	1,000	4	8	20

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	FPGA	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.f1-c28f1.7xlarge	28	112.0	N/A	Intel ARRIA 10 GX 1150	5.0	2,000	8	8	20
ecs.f1-c28f1.14xlarge	56	224.0	N/A	2 * Intel ARRIA 10 GX 1150	10.0	2,000	14	8	20

See *other instance type families*.

f3, compute optimized type family with FPGA

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Xilinx 16nm Virtex UltraScale + VU9P
- Equipped with a vCPU to memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning and reasoning
  - Genomics research
  - Speeding up database access
  - Picture transcoding, such as converting JPEG to WebP
  - Real-time video processing, such as H.265 video compression

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	FPGA	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.f3-c4f1.xlarge	4	16.0	N/A	1 * Xilinx VU9P	1.5	300	2	3	10
ecs.f3-c8f1.2xlarge	8	32.0	N/A	1 * Xilinx VU9P	2.5	500	4	4	10
ecs.f3-c16f1.4xlarge	16	64.0	N/A	1 * Xilinx VU9P	5.0	1,000	4	8	20
ecs.f3-c16f1.8xlarge	32	128.0	N/A	2 * Xilinx VU9P	10.0	2,000	8	8	20
ecs.f3-c16f1.16xlarge	64	256.0	N/A	4 * Xilinx VU9P	20.0	2,500	16	8	20

See *other instance type families*.

ebmhfg5, ECS Bare Metal Instance type family with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:4
- 3.7 GHz Intel Xeon E3-1240v6 (Skylake) processors, 8-core vCPU, up to 4.1 GHz Turbo Boot
- High network performance: 2 million pps packet forwarding rate
- Supports VPC network only
- Provides dedicated hardware resources and physical isolation
- Supports Intel SGX

- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources, or scenarios where binding a license to the hardware is required
  - Gaming or financial applications featuring high performance
  - High-performance Web servers
  - Enterprise-level applications, such as high-performance databases

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. ebmhfg5 .2xlarge	8	32.0	N/A	6.0	2,000	8	6	8

Note:
For more information about ECS Bare Metal Instance, see *EBM Instance overview*.

See *other instance type families*.

ebmc4, computing ECS Bare Metal Instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:2
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors, up to 2.9 GHz Turbo Boot
- High network performance: 4 million pps packet forwarding rate
- Supports VPC network only
- Provides dedicated hardware resources and physical isolation

- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - 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 Container, and Pouch)
  - Enterprise-level applications, such as medium and large databases
  - Video coding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. ebmc4. 8xlarge	32	64.0	N/A	10.0	4,000	8	12	10

Note:

For more information about ECS Bare Metal Instance, see *EBM Instance overview*.

See *other instance type families*.

ebmg5, general-purpose ECS Bare Metal Instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96-core vCPU, up to 2.7 GHz Turbo Boot
- High network performance: 4 million pps packet forwarding rate
- Supports VPC network only
- Provides dedicated hardware resources and physical isolation

- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources, or scenarios where binding a license to the hardware is required
  - 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 Container, and Pouch)
  - Enterprise-level applications, such as medium and large databases
  - Video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. ebmg5. 24xlarge	96	384.0	N/A	10.0	4,000	8	32	10

Note:

For more information about ECS Bare Metal Instance, see *EBM Instance overview*.

See *other instance type families*.

scch5, Super Computing Cluster (SCC) instance type family with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Supports both RoCE and VPC networks, of which RoCE is dedicated to RDMA communication
- With all features of ECS Bare Metal Instance
- 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors
- Equipped with a vCPU to memory ratio of 1:3

- Suitable for the following scenarios:
  - Large-scale machine learning applications
  - Large-scale high-performance scientific and engineering applications
  - Large-scale data analysis, batch computing, video encoding

Instance types

Instanc type	vCPU	Physica core	Memor (GiB)	GPU	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	RoCE (Inboun / Outbou) (Gbit /s)	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. scch5. 16xlarge	64	32	192.0	N/A	10.0	4,500	46	8	32	10

Note:
The ecs.scch5.16xlarge instance type provides 64 logical processors on 32 physical cores. For more information about SCC, see *Super Computing Clusters*.

See *other instance type families*.

sccg5, general-purpose Super Computing Cluster (SCC) instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Supports both RoCE and VPC networks, of which RoCE is dedicated to RDMA communication
- With all features of ECS Bare Metal Instance
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Equipped with a vCPU to memory ratio of 1:4

- Suitable for the following scenarios:
  - Large-scale machine learning applications
  - Large-scale high-performance scientific and engineering applications
  - Large-scale data analysis, batch computing, video encoding

Instance types

Instanc type	vCPU	Physica core	Memor (GiB)	GPU	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	RoCE (Inboun / Outbou) (Gbit /s)	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sccg5. 24xlarge	96	48	384.0	N/A	10.0	4,500	46	8	32	10

Note:
The ecs.sccg5.24xlarge instance type provides 96 logical processors on 48 physical cores. For more information about SCC, see *Super Computing Clusters*.

See *other instance type families*.

sccgn6, compute optimized Super Computing Cluster (SCC) instance type family with GPUs

Features

- I/O-optimized
- Equipped with a vCPU 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 ESSD Cloud Disks (million-level IOPS), SSD Cloud Disks, and Ultra Disks
  - Supports high-performance Cloud Parallel File System (CPFS)

- Networking:
  - Supports VPC networks equipped with two 25Gbps 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
  - Up to 640 Tensor Cores
  - Up to 5,120 CUDA Cores
  - 16 GB HBM2 GPU memory capacity (900 GB/s 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 applications
  - Large-scale high-performance scientific and emulation applications
  - Large-scale data analysis, batch computing, video encoding

Instance types

Instanc types	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	RoCE (Inboun / Outbou) (Gbit /s)	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sccgn6 . 24xlarge	96	384	N/A	V100* 8	30	4,500	25*2	8	32	10

See *other instance type families*.

t5, burstable instances

Features

- Equipped with 2.5 GHz Intel Xeon processors
- The latest DDR4 memory
- No fixed ratio of vCPU to memory

- Baseline CPU performance, burstable, but restricted by accumulated CPU credits
- Resource balance among compute, memory, and networks
- Supports VPC network only
- Suitable for the following scenarios:
  - Web application servers
  - Lightweight web servers
  - Development and testing environments

Instance types

Instan types	vCPU	Memo (GiB)	Avg baseli CPU perfor e	CPU credits /hour	Max CPU credit balanc	Local disks (GiB) *	Bandw (Gbit/ s) **	Packet forwar rate (Thous pps) ***	NIC queues ****	ENIs ****	Private IP address of a single ENI
ecs .t5- lc2m1 .nano	1	0.5	10%	6	144	N/A	0.1	40	1	1	2
ecs .t5- lc1m1 . small	1	1.0	10%	6	144	N/A	0.2	60	1	1	2
ecs .t5- lc1m2 . small	1	2.0	10%	6	144	N/A	0.2	60	1	1	2
ecs .t5- lc1m2 .large	2	4.0	10%	12	288	N/A	0.4	100	1	1	2
ecs .t5- lc1m4 .large	2	8.0	10%	12	288	N/A	0.4	100	1	1	2

Instance types	vCPU	Memory (GiB)	Avg baseline CPU performance	CPU credits/hour	Max CPU credit balance	Local disks (GiB) *	Bandwidth (Gbit/s) **	Packet forwarding rate (Thousand pps) ***	NIC queues ****	ENIs ****	Private IP address of a single ENI
ecs.t5-c1m1.large	2	2.0	15%	18	432	N/A	0.5	100	1	1	2
ecs.t5-c1m2.large	2	4.0	15%	18	432	N/A	0.5	100	1	1	2
ecs.t5-c1m4.large	2	8.0	15%	18	432	N/A	0.5	100	1	1	2
ecs.t5-c1m1.xlarge	4	4.0	15%	36	864	N/A	0.8	200	1	2	6
ecs.t5-c1m2.xlarge	4	8.0	15%	36	864	N/A	0.8	200	1	2	6
ecs.t5-c1m4.xlarge	4	16.0	15%	36	864	N/A	0.8	200	1	2	6
ecs.t5-c1m1.2xlarge	8	8.0	15%	72	1728	N/A	1.2	400	1	2	6

Instan types	vCPU	Memo (GiB)	Avg baseli CPU perfor e	CPU credits /hour	Max CPU credit balanc	Local disks (GiB) *	Bandw (Gbit/ s) **	Packet forwar rate (Thou pps) ***	NIC queue ****	ENIs ****	Private IP address of a single ENI
ecs .t5- c1m2 . 2xlarge	8	16.0	15%	72	1728	N/A	1.2	400	1	2	6
ecs .t5- c1m4 . 2xlarge	8	32.0	15%	72	1728	N/A	1.2	400	1	2	6
ecs .t5- c1m1 . 4xlarge	16	16.0	15%	144	3456	N/A	1.2	600	1	2	6
ecs .t5- c1m2 . 4xlarge	16	32.0	15%	144	3456	N/A	1.2	600	1	2	6

Note:
For more information about t5, see *Basic concepts*.

See *other instance type families*.

xn4/n4/mn4/e4, type families of previous generations for entry-level users, computing on the x86-architecture

Features

- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- The latest DDR4 memory
- No fixed ratio of CPU to memory

Type family	Features	vCPU to memory ratio	Ideal for
xn4	Compact entry-level instances	1:1	· Front ends of Web applications · Light load applications and microservices · Applications for development or testing environments
n4	General entry-level instances	1:2	· Websites and Web applications · Development environment, building servers, code repositories, microservices, and testing and staging environment · Lightweight enterprise applications
mn4	Balanced entry-level instances	1:4	· Websites and Web applications · Lightweight databases and caches · Integrated applications and lightweight enterprise services

Type family	Features	vCPU to memory ratio	Ideal for
e4	Memory entry-level instances	1:8	· Applications that require large volume of memory · Lightweight databases and cache

xn4

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.xn4. small	1	1.0	N/A	0.5	50	1	1	2

n4

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.n4. small	1	2.0	N/A	0.5	50	1	1	2
ecs.n4. large	2	4.0	N/A	0.5	100	1	1	2
ecs.n4. xlarge	4	8.0	N/A	0.8	150	1	2	6
ecs.n4. 2xlarge	8	16.0	N/A	1.2	300	1	2	6

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.n4.4xlarge	16	32.0	N/A	2.5	400	1	2	6
ecs.n4.8xlarge	32	64.0	N/A	5.0	500	1	2	6

mn4

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.mn4.small	1	4.0	N/A	0.5	50	1	1	2
ecs.mn4.large	2	8.0	N/A	0.5	100	1	1	2
ecs.mn4.xlarge	4	16.0	N/A	0.8	150	1	2	6
ecs.mn4.2xlarge	8	32.0	N/A	1.2	300	1	2	6
ecs.mn4.4xlarge	16	64.0	N/A	2.5	400	1	2	6
ecs.mn4.8xlarge	32	128.0	N/A	5	500	2	8	6

e4

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.e4. small	1	8.0	N/A	0.5	50	1	1	2
ecs.ce4. xlarge	2	16.0	N/A	0.5	100	1	1	2
ecs.ce4. xlarge	4	32.0	N/A	0.8	150	1	2	6
ecs.e4. 2xlarge	8	64.0	N/A	1.2	300	1	3	6
ecs.ce4. xlarge	16	128.0	N/A	2.5	400	1	8	6

See *other instance type families*.

* *Cache disks*, or *Local disks*, are the disks located on the physical servers (host machines) that ECS instances are hosted on. They provide temporary block level storage for instances. Block storage capacity is measured in binary units. In some cases, such as when the computing resources of an instance, including CPU and memory, are released, or an instance is inactive while migration occurs, data on the local disks is erased. For more information, see *Local disks*.

** The maximum sum of inbound and outbound bandwidth.

*** The maximum sum of inbound and outbound packet forwarding rates. For more information about packet forwarding rate testing, see *Test network performance*.

**** The maximum number of NIC queues that an instance type supports. If your instance is running CentOS 7.3, the maximum number of NIC queues is used by default.

***** An enterprise-level instance with two or more vCPU cores supports elastic network interfaces. An entry-level instance with four or more vCPU cores supports elastic network interfaces. For more information, see *ENI overview*.

4 Instance type families

4.1 Burstable performance instances

4.1.1 Basic concepts

Burstable instances (also called t5 instances) provide a baseline level of CPU performance with the ability to burst above the baseline. Each t5 instance provides a baseline CPU performance and earns CPU credits at a specified rate based on instance types. A t5 instance consumes CPU credits to meet service requirements once it is started. When the required performance is higher than the baseline, the instance consumes more CPU credits to seamlessly increase CPU performance without affecting the environment or applications on the instance.

t5 instances come with two running modes: *Standard* and *Unlimited*.

Concepts

- Baseline CPU performance

  Each t5 instance type provides a baseline level of CPU performance,

  - which means that each vCPU core has a usage limit under normal workloads. For example, when the ecs.t5-lc1m2.small standard instance runs under normal workloads, the maximum CPU usage is 10%. More credits will be consumed to burst above that baseline. After the credits are used up, the maximum CPU usage is 10%.
  - By contrast, t5 unlimited instances are not restricted by the baseline and can maintain high CPU performance for any time period. However, fees are charged for excess credits.

- CPU credits

  Each t5 instance earns CPU credits at a fixed rate based on the baseline CPU performance. One CPU credit represents the computing performance, which is related to the number of vCPU cores, CPU usage, and running time. For example:

  - One CPU credit = One vCPU core running at 100% usage for 1 minute
  - One CPU credit = One vCPU core running at 50% usage for 2 minutes
  - One CPU credit = Two vCPU cores running at 25% usage for 2 minutes

  To support a vCPU core running at 100% usage for an hour, 60 CPU credits are required.

- Initial CPU credits

  Every time you create a t5 instance, each vCPU core of the instance is immediately allocated 30 CPU credits, which are called initial CPU credits. Initial CPU credits are allocated only upon instance creation. Additionally, they are consumed first when the instance starts to spend credits.

- CPU credit acquisition rate

  t5 instances acquire CPU credits on a per-minute basis. The CPU credit acquisition rate indicates CPU credits acquired by a t5 instance per unit time (minute). It is determined by the baseline CPU performance. The calculation formula is as follows:

  ```
  CPU credit acquisitio n rate = Baseline CPU
  performanc e x number of vCPUs
  ```

  Example: Use the ecs.t5-c1m2.xlarge instance as an example. Its average baseline CPU performance is 15%, so the CPU credit acquisition rate is 0.6 CPU credit per minute, that is, 36 CPU credits per hour.

- CPU credit consumption

  A t5 instance consumes CPU credits once it is started. Initial CPU credits are consumed first. The formula for calculating the consumed CPU credits per minute is as follows:

  ```
  CPU credits consumed per minute = One CPU credit x
    actual CPU performanc e
  ```

  Example: Use the ecs.t5-lc1m2.small instance as an example. It consumes 0.2 CPU credit when it runs at 20% CPU usage for one minute.

- Accrued CPU credits

  When the CPU usage of a t5 instance is lower than the baseline CPU performanc e, the instance accrues CPU credits because the CPU credit consumption rate is lower than the CPU credit acquisition rate. Otherwise, the instance consumes CPU credits overall. The CPU credit accrual rate is determined by the difference between the actual CPU load and the baseline performance. It can be calculated using the following formula:

  ```
  CPU credits accrued per minute = One CPU credit x
  ( baseline CPU performanc e - actual CPU performanc e )
  ```

  You can *view the accrued and consumed CPU credits* on the ECS console.

- Max. CPU credit balance

  CPU credits increase when the CPU credit acquisition rate is higher than the CPU credit consumption rate. The accrued credits do not expire on a running instance. However, there is an upper limit for the credits that can be accrued by an instance , namely, the maximum CPU credit balance. The upper limit varies with instance types.

  Taking ecs.t5-lc2m1.nano as an example, the maximum CPU credit balance is 144 . When the CPU credit balance reaches 144, accrual pauses. When the balance is lower than 144, accrual restarts.

How stopping an instance impacts CPU credits

After you stop a t5 through the *view CPU utilization and CPU credits* feature or the *StopInstance API*, CPU credits change according to the billing method and network type, as shown in the following table.

Network type	Billing method	How CPU credits change after the instance is stopped
VPC	Subscription	The existing CPU credits are valid, and the credit accrual continues.
	Pay-As-You-Go (with the *no fees for stopped VPC instances* function disabled)	

Network type	Billing method	How CPU credits change after the instance is stopped
	Pay-As-You-Go (with *no fees for stopped VPC instances* function enabled)	CPU credits accrued before the stoppage become invalid. The instance acquires initial CPU credits again after it is restarted.

The instance continues to accrue CPU credits after it is restarted.

If a Pay-As-You-Go instance has overdue payment or a Subscription instance expires, its CPU credits remain valid, but no new CPU credits will be accrued. After you *reactivate* or *renew* the instance, it automatically accrues CPU credits.

Instance types

t5 instances use Intel Xeon processors. The following table lists instance types. In this table:

- CPU credits/hour indicates the total CPU credits allocated to all vCPU cores of a t5 instance per hour.
- Average baseline CPU performance indicates the average baseline CPU performance of each vCPU core of a t5 instance.

Instance type	vCPU	Average baseline CPU performance	Initial CPU credits	CPU credits/ hour	Max. CPU credit balance	Memory (GiB)
ecs.t5-lc2m1.nano	1	10%	30	6	144	0.5
ecs.t5-lc1m1.small	1	10%	30	6	144	1.0
ecs.t5-lc1m2.small	1	10%	30	6	144	2.0

Instance type	vCPU	Average baseline CPU performance	Initial CPU credits	CPU credits/ hour	Max. CPU credit balance	Memory (GiB)
ecs.t5-lc1m2.large	2	10%	60	12	288	4.0
ecs.t5-lc1m4.large	2	10%	60	12	288	8.0
ecs.t5-c1m1.large	2	15%	60	18	432	2.0
ecs.t5-c1m2.large	2	15%	60	18	432	4.0
ecs.t5-c1m4.large	2	15%	60	18	432	8.0
ecs.t5-c1m1.xlarge	4	15%	120	36	864	4.0
ecs.t5-c1m2.xlarge	4	15%	120	36	864	8.0
ecs.t5-c1m4.xlarge	4	15%	120	36	864	16.0
ecs.t5-c1m1.2xlarge	8	15%	240	72	1,728	8.0
ecs.t5-c1m2.2xlarge	8	15%	240	72	1,728	16.0
ecs.t5-c1m4.2xlarge	8	15%	240	72	1,728	32.0

Instance type	vCPU	Average baseline CPU performance	Initial CPU credits	CPU credits/ hour	Max. CPU credit balance	Memory (GiB)
ecs.t5-c1m1.4xlarge	16	15%	480	144	3,456	16.0
ecs.t5-c1m2.4xlarge	16	15%	480	144	3,456	32.0

Examples

- The following uses the ecs.t5-c1m1.xlarge instance as an example.
  - For each vCPU core, the average baseline performance is 15%. Therefore, the total baseline performance of the instance is 60% (4 vCPU x 15%). Details are as follows:
    - When the instance uses only one vCPU core, this core provides the baseline performance of 60%.
    - When the instance uses two vCPU cores, each core is allocated the baseline performance of 30%.
    - When the instance uses three vCPU cores, each core is allocated the baseline performance of 20%.
    - When the instance uses all four vCPU cores, each core is allocated the baseline performance of 15%.

    Note:
    When the business needs arise, CPU credits are consumed to improve the CPU performance. The performance of each vCPU core can increase to 100%.
  - An instance acquires 36 CPU credits per hour, which means that each vCPU core acquires nine CPU credits per hour.

- The following uses the ecs.t5-c1m2.4xlarge instance as an example.
  - For each vCPU core, the average baseline performance is 15%. Therefore, the total baseline computing performance of the instance is 240% (16 vCPU x 15%). Details are as follows:
    - When the instance uses only one vCPU core, this core provides the baseline performance of 100%.
    - When the instance uses two vCPU cores, each core is allocated the baseline performance of 100%.
    - When the instance uses three vCPU cores, each core is allocated the baseline performance of 80%.
    - When the instance uses all 16 vCPU cores, each core is allocated the baseline performance of 15%.

       Note:
      When the business needs arise, CPU credits are consumed to improve the CPU performance. The performance of each vCPU core can increase to 100%.
  - An instance acquires 144 CPU credits per hour, which means that each vCPU core acquires nine CPU credits per hour.

Billing methods

t5 instances support both the Pay-As-You-Go and Subscription billing methods. For differences between the billing methods, see *billing method comparison*.

4.1.2 t5 standard instances

t5 standard instances are ideal 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.

If the instance has accrued few credits, its performance gradually declines to the baseline level within 15 minutes, so that the instance performance does not drop dramatically when the accrued CPU credits are used up. When the accrued CPU credits are used up, the actual CPU performance of the t5 instance cannot be higher than the baseline CPU performance.

Fees

There is no additional fee except the cost of creating an instance.

Examples

Take a t5 standard instance of the ecs.t5-lc1m2.small type as an example. The following describes how its CPU credits change:

1. When the instance is created, 30 initial CPU credits are allocated to it. That is, the total CPU credits are 30 before it is started. After it is started, CPU credits accrue at the rate of 0.1 credits per minute. Meanwhile, the credits are consumed during its running.
2. During the first minute, if the CPU usage is 5%, 0.05 initial CPU credits are consumed, while 0.1 CPU credits are allocated. Therefore, 0.05 CPU credits are accrued.
3. After the instance has started for N minutes, if the CPU usage is 50%, 0.5 CPU credits are consumed while 0.1 CPU credits are allocated within one minute. Therefore, 0.4 CPU credits are consumed during this one minute.
4. When the accrued CPU credits are used up, the maximum CPU usage is 10%.

4.1.3 t5 unlimited instances

t5 unlimited instances can maintain high CPU performance for any period of time, without being limited to the baseline CPU performance.

Concepts

In addition to the *basic concepts*, you also need to understand the following concepts before using t5 unlimited instances:

- Advance credits

  The credits that are used in advance but should be obtained within the next 24 hours.
- Excess credits

  After the credits for the next 24 hours are used up, additional credits incur fees that are billed by the hour.

When a t5 unlimited instance runs out of its CPU credit balance, advance credits are used first to address the requirement of high CPU performance. When the CPU usage is lower than the baseline, the earned CPU credits are used to pay down (offset) the advance credits.

Billing rules

- Fees are not charged in the following cases:
  - The hourly t5 instance price automatically covers all interim spikes in usage if the average CPU utilization of the instance is at or below the baseline over a 24-hour period or the instance lifetime, whichever is shorter. You do not have to pay additional fees.
  - An instance earns a maximum number of credits in a 24-hour period. For example, a t5-lc1m1.small instance can earn a maximum of 144 credits in a 24-hour period. When the advance credits are less than that maximum, no additional fees are charged.
- Fees are charged in the following cases:
  - If the consumed advance credits exceed the maximum credits (that is, excess credits generated), fees are charged at the end of the time period.
  - If advance credits are used and the instance is stopped or released before the advance credits are cleared, a one-off fee is charged for the advance credits.
  - If excess credits are used after advance credits are used up, additional fees are charged.
  - If a t5 instance is changed from unlimited mode to standard mode, the fees for advance credits are charged immediately, and the accrued credits remain unchanged.

    The fees charged are shown in the following table:

Region	Windows instance (USD/credit)	Linux instance (USD/credit)
Mainland China	0.0008	0.0008
Other regions	0.0016	0.0008

Examples

Use a t5-lc1m1.small unlimited instance (purchased in US West 1 region) as an example. The following describes how its CPU credits change.

1. After a Linux instance is created, it is allocated 30 initial CPU credits. When the instance starts, it is able to spend 144 credits in advance, which are the maximum of CPU credits for the next 24 hours. Therefore, there are 174 CPU credits when the instance starts.

2. After the instance starts, assuming the CPU utilization is 50%, the instance consumes 0.5 initial CPU credits per minute, while 0.1 CPU credits are allocated in the meantime. As a result, CPU credits continue to decrease.
3. After the instance has been running for N minutes, assuming the accrued CPU credits are used up, the advance credits are spent to maintain high CPU performance.
4. After the instance has been running for N + X minutes, assuming the 144 advance credits are used up, excess credits are used to maintain high CPU performance.
5. After the instance has been running for N + X + Y minutes, assuming 50 excess credits are used and the CPU utilization drops to 5% (below the baseline), the instance begins to earn 0.1 credits per minute, which are used to pay down (offset) the consumed advance credits. When the advance credits are restored to 144, credits begin to accrue to the instance (0.1 credits per minute).

Consumption details

At the end of the N + X + Y minutes, if excess credits are no longer used, fees are charged.

During the above time period, the Linux instance uses 50 excess credits. The additional fee is: 0.0016 USD/credit x 50 credits = 0.08 USD.

4.1.4 Manage t5 instances

You can create t5 instances and change their running modes through the console or APIs.

Create an instance

You can create a t5 instance by referring to *Create an instance by using the wizard*. When creating a t5 instance, note the following:

- Network type: Only VPC is supported.
- Image: 512 MiB is the minimum memory size for a t5 instance. The OS can only be a Linux variant or Windows Server 1709. Operating systems that require a minimum memory of 1 GiB, such as Windows Server 2016, are not supported. For more information about selecting images, see *How to select a system image*.
- Running mode of t5 instances: Select the Enable Unlimited Mode for t5 Instances check box to create a t5 instance without performance constraints. If it is not

selected, a t5 standard instance (with performance constraints) is created. You can change the mode after creating a t5 instance.

Change the running mode of a t5 instance

Within the lifecycle of a t5 instance, you can change its running mode through the console or the API *#unique_73*.

Note:

A t5 instance must be in a Running state (`Running`) status before you can change its mode.

Follow the steps below to change the running mode in the console:

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. Select a region.
4. In the list of instances, find the target instance, and click the instance ID.
5. In the Basic Information part of the Instance details page, click More, and then select Enable Unlimited Mode. You can also select Disable Unlimited Mode to return to the standard mode.

The following table describes how the running mode of a t5 instance changes according to the operations or its payment status.

Operation/status	Results
Stop an instance	After restart, t5 instances that are not charged after being stopped will run in the standard mode (with performance constraints), while t5 instances that are still charged after being stopped will run in the same mode as before.
Restart an instance	After restart, the running mode remains unchanged.
Payment failed	The running mode changes to the standard mode. The original mode is restored after the payment is made successfully.

View CPU utilization and CPU credits

In the ECS console, you can view the CPU information about an instance, such as CPU utilization, consumed CPU credits, accrued CPU credits, excess CPU credits, and advance CPU credits.

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. Select a region.
4. In the instance list, find the target instance and click the instance ID. Alternatively, you can click Manage in the Actions column.
5. View the metrics in the Monitoring Information area of the Instance Details page.

You can also view the CPU utilization after connecting to an instance remotely:

- Windows: *Connect to an instance*, and then view the CPU utilization in the Task Manager.
- Linux: *Connect to an instance*, and then run the `top` command to view the CPU utilization.

Change an instance type

In the ECS console, if you find that the CPU utilization remains at the baseline performance level for a long period of time or rarely exceeds the baseline level, your instance type may not meet your needs or may exceed your needs. In either case, you may consider changing the instance type.

You can change the instance type based on the billing method:

- Subscription instances: You can change the instance type by *upgrading or downgrading instance configurations*.
- Pay-As-You-Go instances: You can change the configurations by *changing the instance type*.

For information about supported instance type families, see *Instance type families that support upgrading instance types*.

4.2 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.

xn4/n4/mn4/e4, type families of previous generations for entry-level users, computing on the x86-architecture

Features

- **Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors**
- **The latest DDR4 memory**
- **No fixed ratio of CPU to memory**

Type family	Features	vCPU to memory ratio	Ideal for
xn4	Compact entry-level instances	1:1	· Front ends of Web applications · Light load applications and microservices · Applications for development or testing environments
n4	General entry-level instances	1:2	· Websites and Web applications · Development environment, building servers, code repositories, microservices, and testing and staging environment · Lightweight enterprise applications

Type family	Features	vCPU to memory ratio	Ideal for
mn4	Balanced entry-level instances	1:4	· Websites and Web applications · Lightweight databases and caches · Integrated applications and lightweigh t enterprise services
e4	Memory entry-level instances	1:8	· Applications that require large volume of memory · Lightweight databases and cache

xn4

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.xn4. small	1	1.0	N/A	0.5	50	1	1	2

n4

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.n4. small	1	2.0	N/A	0.5	50	1	1	2
ecs.n4. large	2	4.0	N/A	0.5	100	1	1	2
ecs.n4. xlarge	4	8.0	N/A	0.8	150	1	2	6
ecs.n4. 2xlarge	8	16.0	N/A	1.2	300	1	2	6
ecs.n4. 4xlarge	16	32.0	N/A	2.5	400	1	2	6
ecs.n4. 8xlarge	32	64.0	N/A	5.0	500	1	2	6

mn4

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. mn4. small	1	4.0	N/A	0.5	50	1	1	2
ecs. mn4. large	2	8.0	N/A	0.5	100	1	1	2
ecs. mn4. xlarge	4	16.0	N/A	0.8	150	1	2	6

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. mn4. 2xlarge	8	32.0	N/A	1.2	300	1	2	6
ecs. mn4. 4xlarge	16	64.0	N/A	2.5	400	1	2	6
ecs. mn4. 8xlarge	32	128.0	N/A	5	500	2	8	6

e4

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.e4. small	1	8.0	N/A	0.5	50	1	1	2
ecs.ce4. xlarge	2	16.0	N/A	0.5	100	1	1	2
ecs.ce4. xlarge	4	32.0	N/A	0.8	150	1	2	6
ecs.e4. 2xlarge	8	64.0	N/A	1.2	300	1	3	6
ecs.ce4. xlarge	16	128.0	N/A	2.5	400	1	8	6

See *other instance type families*.

* *Cache disks*, or *Local disks*, are the disks located on the physical servers (host machines) that ECS instances are hosted on. They provide temporary block level storage for instances. Block storage capacity is measured in binary units. In some cases, such as when the computing resources of an instance, including CPU and memory, are released, or an instance is inactive while migration occurs, data on the local disks is erased. For more information, see *Local disks*.

** The maximum sum of inbound and outbound bandwidth.

*** The maximum sum of inbound and outbound packet forwarding rates. For more information about packet forwarding rate testing, see *Test network performance*.

**** The maximum number of NIC queues that an instance type supports. If your instance is running CentOS 7.3, the maximum number of NIC queues is used by default.

***** An enterprise-level instance with two or more vCPU cores supports elastic network interfaces. An entry-level instance with four or more vCPU cores supports elastic network interfaces. For more information, see *ENI overview*.

4.3 General-purpose instance type families

This topic describes the general-purpose instance type families g5 and sn2ne, and lists specific instance types.

g5, general-purpose type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:4
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity

- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - Enterprise-level applications of various types and sizes
  - Medium and small database systems, cache, and search clusters
  - Data analysis and computing
  - Computing clusters and data processing reliant on memory

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.g5. large	2	8.0	N/A	1.0	300	2	2	6
ecs.g5. xlarge	4	16.0	N/A	1.5	500	2	3	10
ecs.g5. 2xlarge	8	32.0	N/A	2.5	800	2	4	10
ecs.g5. 3xlarge	12	48.0	N/A	4.0	900	4	6	10
ecs.g5. 4xlarge	16	64.0	N/A	5.0	1,000	4	8	20
ecs.g5. 6xlarge	24	96.0	N/A	7.5	1,500	6	8	20
ecs.g5. 8xlarge	32	128.0	N/A	10.0	2,000	8	8	20
ecs.g5. 16xlarge	64	256.0	N/A	20.0	4,000	16	8	20

See *other instance type families*.

sn2ne, general-purpose type family with enhanced network performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU 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
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - Enterprise-level 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) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sn2ne. large	2	8.0	N/A	1.0	300	2	2	6
ecs. sn2ne. xlarge	4	16.0	N/A	1.5	500	2	3	10
ecs. sn2ne. 2xlarge	8	32.0	N/A	2.0	1,000	4	4	10
ecs. sn2ne. 3xlarge	12	48.0	N/A	2.5	1,300	4	6	10

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sn2ne. 4xlarge	16	64.0	N/A	3.0	1,600	4	8	20
ecs. sn2ne. 6xlarge	24	96.0	N/A	4.5	2,000	6	8	20
ecs. sn2ne. 8xlarge	32	128.0	N/A	6.0	2,500	8	8	20
ecs. sn2ne. 14xlarge	56	224.0	N/A	10.0	4,500	14	8	20

See *other instance type families*.

4.4 Compute-intensive instance type families

This topic describes specific compute instance type families, and lists the specific instance types within each of the instance type families. Specifically, the instance type families to be described are: compute-intensive instance type family ic5, compute-optimized instance type family with enhanced network performance sn1ne, and compute instance type family c5.

ic5, intensive compute instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:1
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity

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

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.ic5. large	2	2.0	N/A	1.0	300	2	2	6
ecs.ic5. xlarge	4	4.0	N/A	1.5	500	2	3	10
ecs.ic5. 2xlarge	8	8.0	N/A	2.5	800	2	4	10
ecs.ic5. 3xlarge	12	12.0	N/A	4.0	900	4	6	10
ecs.ic5. 4xlarge	16	16.0	N/A	5.0	1,000	4	8	20

See *other instance type families*.

c5, compute instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:2
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity

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

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.c5. large	2	4.0	N/A	1.0	300	2	2	6
ecs.c5. xlarge	4	8.0	N/A	1.5	500	2	3	10
ecs.c5. 2xlarge	8	16.0	N/A	2.5	800	2	4	10
ecs.c5. 3xlarge	12	24.0	N/A	4.0	900	4	6	10
ecs.c5. 4xlarge	16	32.0	N/A	5.0	1,000	4	8	20
ecs.c5. 6xlarge	24	48.0	N/A	7.5	1,500	6	8	20
ecs.c5. 8xlarge	32	64.0	N/A	10.0	2,000	8	8	20
ecs.c5. 16xlarge	64	128.0	N/A	20.0	4,000	16	8	20

See *other instance type families*.

sn1ne, compute optimized type family with enhanced network performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU 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
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - Web front-end servers
  - Massively Multiplayer Online (MMO) game frontends
  - Data analysis, batch compute, and video coding
  - High-performance science and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sn1ne. large	2	4.0	N/A	1.0	300	2	2	6
ecs. sn1ne. xlarge	4	8.0	N/A	1.5	500	2	3	10
ecs. sn1ne. 2xlarge	8	16.0	N/A	2.0	1,000	4	4	10
ecs. sn1ne. 3xlarge	12	24.0	N/A	2.5	1,300	4	6	10

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sn1ne. 4xlarge	16	32.0	N/A	3.0	1,600	4	8	20
ecs. sn1ne. 6xlarge	24	48.0	N/A	4.5	2,000	6	8	20
ecs. sn1ne. 8xlarge	32	64.0	N/A	6.0	2,500	8	8	20

See *other instance type families*.

4.5 Memory-intensive instance type families

This topic describes specific memory instance type families, and lists the specific instance types within each of the instance type families. Specifically, the instance type families to be described are: memory instance type family r5, memory-optimized instance type families with enhanced performance re4 and re4e, memory optimized instance type family with enhanced network performance se1ne, and memory-optimized instance type family se1.

r5, memory instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Ultra high packet forwarding rate
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity

- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - High-performance databases and high memory databases
  - Data analysis and mining, and distributed memory cache
  - Hadoop, Spark, and other enterprise-level applications with large memory requirements

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.r5. large	2	16.0	N/A	1.0	300	2	2	6
ecs.r5. xlarge	4	32.0	N/A	1.5	500	2	3	10
ecs.r5. 2xlarge	8	64.0	N/A	2.5	800	2	4	10
ecs.r5. 3xlarge	12	96.0	N/A	4.0	900	4	6	10
ecs.r5. 4xlarge	16	128.0	N/A	5.0	1,000	4	8	20
ecs.r5. 6xlarge	24	192.0	N/A	7.5	1,500	6	8	20
ecs.r5. 8xlarge	32	256.0	N/A	10.0	2,000	8	8	20
ecs.r5. 16xlarge	64	512.0	N/A	20.0	4,000	16	8	20

See *other instance type families*.

re4, memory optimized instance type family with enhanced performance

Features

- Supports SSD Cloud Disks and Ultra Disks
- I/O-optimized
- Optimized for high-performance databases, high memory databases, and other memory-intensive enterprise applications
- 2.2 GHz Intel Xeon E7 8880 v4 (Broadwell) processors, up to 2.4 GHz Turbo Boot
- Equipped with a vCPU 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 high 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 disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.re4. 20xlarge	80	960.0	N/A	15.0	2,000	16	8	20
ecs.re4. 40xlarge	160	1920.0	N/A	30.0	4,500	16	8	20

See *other instance type families*.

re4e, memory optimized type family with enhanced performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Optimized for high-performance databases, high memory databases, and other memory-intensive enterprise applications
- 2.2 GHz Intel Xeon E7 8880 v4 (Broadwell) processors, up to 2.4 GHz Turbo Boot
- Equipped with a vCPU to memory ratio of 1:24, up to 3840.0 GiB memory

- Suitable for the following scenarios:
  - High-performance databases and high 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 disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. re4e. 40xlarge	160	3840.0	N/A	30.0	4,500	16	15	20

See *other instance type families*.

se1ne, memory optimized type family with enhanced network performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:8
- Ultra high packet receive and forwarding rate
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - High-performance databases and large memory databases
  - Data analysis and mining, and distributed memory cache
  - Hadoop, Spark, and other enterprise-level applications with large memory requirements

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. se1ne. large	2	16.0	N/A	1.0	300	2	2	6
ecs. se1ne. xlarge	4	32.0	N/A	1.5	500	2	3	10
ecs. se1ne. 2xlarge	8	64.0	N/A	2.0	1,000	4	4	10
ecs. se1ne. 3xlarge	12	96.0	N/A	2.5	1,300	4	6	10
ecs. se1ne. 4xlarge	16	128.0	N/A	3.0	1,600	4	8	20
ecs. se1ne. 6xlarge	24	192.0	N/A	4.5	2,000	6	8	20
ecs. se1ne. 8xlarge	32	256.0	N/A	6.0	2,500	8	8	20
ecs. se1ne. 14xlarge	56	480.0	N/A	10.0	4,500	14	8	20

See *other instance type families*.

se1, memory optimized type family

Features

- I/O-optimized

- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:8
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - High-performance databases and large memory databases
  - Data analysis and mining, and distributed memory cache
  - Hadoop, Spark, and other enterprise-level applications with large memory requirements

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.se1. large	2	16.0	N/A	0.5	100	1	2	6
ecs.se1. xlarge	4	32.0	N/A	0.8	200	1	3	10
ecs.se1. 2xlarge	8	64.0	N/A	1.5	400	1	4	10
ecs.se1. 4xlarge	16	128.0	N/A	3.0	500	2	8	20
ecs.se1. 8xlarge	32	256.0	N/A	6.0	800	3	8	20
ecs.se1. 14xlarge	56	480.0	N/A	10.0	1,200	4	8	20

See *other instance type families*.

4.6 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.

d1ne, big data type family with enhanced network performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- High-volume local SATA HDD disks with high I/O throughput and up to 35 Gbit/s of bandwidth for a single instance
- Equipped with a vCPU to memory ratio of 1:4, designed for big data scenarios
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity

- Suitable for the following scenarios:
  - Hadoop MapReduce, HDFS, Hive, HBase, and so on
  - Spark in-memory computing, MLlib, and so on
  - Enterprises that require big data computing and storage analysis, such as those in the Internet and finance industries, to store and compute massive volumes of data
  - Elasticsearch, logs, and so on

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. d1ne. 2xlarge	8	32.0	4 * 5500	6.0	1,000	4	4	10
ecs. d1ne. 4xlarge	16	64.0	8 * 5500	12.0	1,600	4	8	20
ecs. d1ne. 6xlarge	24	96.0	12 * 5500	16.0	2,000	6	8	20
ecs. d1ne-c8d3. 8xlarge	32	128.0	12 * 5500	20.0	2,000	6	8	20
ecs. d1ne. 8xlarge	32	128.0	16 * 5500	20.0	2,500	8	8	20
ecs. d1ne-c14d3. 14xlarge	56	160.0	12 * 5500	35.0	4,500	14	8	20

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. d1ne. 14xlarge	56	224.0	28 * 5500	35.0	4,500	14	8	20

Note:

- You cannot change configurations of d1ne instances.
- For more information about d1ne type families, see *FAQ on d1 and d1ne*.

See *other instance type families*.

d1, big data type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- High-volume local SATA HDD disks with high I/O throughput and up to 17 Gbit/s of bandwidth for a single instance
- Equipped with a vCPU to memory ratio of 1:4, designed for big data scenarios
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Hadoop MapReduce, HDFS, Hive, and HBase
  - Spark in-memory computing and MLlib
  - Enterprises that require big data computing and storage analysis, such as those in the Internet and finance industries, to store and compute massive volumes of data
  - Elasticsearch and logs

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.d1.2xlarge	8	32.0	4 * 5500	3.0	300	1	4	10
ecs.d1.3xlarge	12	48.0	16 * 5500	4.0	400	1	6	10
ecs.d1.4xlarge	16	64.0	8 * 5500	6.0	600	2	8	20
ecs.d1.6xlarge	24	96.0	12 * 5500	8.0	800	2	8	20
ecs.d1-c8d3.8xlarge	32	128.0	12 * 5500	10.0	1,000	4	8	20
ecs.d1.8xlarge	32	128.0	16 * 5500	10.0	1,000	4	8	20
ecs.d1-c14d3.14xlarge	56	160.0	12 * 5500	17.0	1,800	6	8	20
ecs.d1.14xlarge	56	224.0	28 * 5500	17.0	1,800	6	8	20

Note:
For more information about d1 type families, see *FAQ on d1 and d1ne*.

See *other instance type families*.

4.7 Local SSD-equipped instance type families

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

i2, type family with local SSD disks

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- High-performance local NVMe SSD disks with high IOPS, high I/O throughput, and low latency.
- Equipped with a vCPU to memory ratio of 1:8, designed for high-performance databases
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases, such as Cassandra and MongoDB
  - Search applications, such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.i2. xlarge	4	32.0	1 * 894	1.0	500	2	3	10
ecs.i2. 2xlarge	8	64.0	1 * 1788	2.0	1,000	2	4	10
ecs.i2. 4xlarge	16	128.0	2 * 1788	3.0	1,500	4	8	20
ecs.i2. 8xlarge	32	256.0	4 * 1788	6.0	2,000	8	8	20
ecs.i2. 16xlarge	64	512.0	8 * 1788	10.0	4,000	16	8	20

See *other instance type families*.

i2g, type family with local SSD disks

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- High-performance local NVMe SSD disks with high IOPS, high I/O throughput, and low latency.
- Equipped with a vCPU to memory ratio of 1:4, designed for high-performance databases
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases, such as Cassandra and MongoDB
  - Search applications, such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.i2g. 2xlarge	8	32.0	1 * 894	2.0	1,000	2	4	10
ecs.i2g. 4xlarge	16	64.0	1 * 1788	3.0	1,500	4	8	20
ecs.i2g. 8xlarge	32	128.0	2 * 1788	6.0	2,000	8	8	20
ecs.i2g. 16xlarge	64	256.0	4 * 1788	10.0	4,000	16	8	20

See *other instance type families*.

i1, type family with local SSD disks

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks

- High-performance local NVMe SSD disks with high IOPS, high I/O throughput, and low latency
- Equipped with a vCPU to memory ratio of 1:4, designed for big data scenarios
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases, such as Cassandra and MongoDB
  - Search applications, such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.i1. xlarge	4	16.0	2 * 104	0.8	200	1	3	10
ecs.i1. 2xlarge	8	32.0	2 * 208	1.5	400	1	4	10
ecs.i1. 3xlarge	12	48.0	2 * 312	2.0	400	1	6	10
ecs.i1. 4xlarge	16	64.0	2 * 416	3.0	500	2	8	20
ecs.i1 -c5d1. 4xlarge	16	64.0	2 * 1456	3.0	400	2	8	20
ecs.i1. 6xlarge	24	96.0	2 * 624	4.5	600	2	8	
ecs.i1. 8xlarge	32	128.0	2 * 832	6.0	800	3	8	20
ecs.i1- c10d1. 8xlarge	32	128.0	2 * 1456	6.0	800	3	8	20

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.i1. 14xlarge	56	224.0	2 * 1456	10.0	1,200	4	8	20

See *other 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 µs.

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 performanc e. 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.
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.

4.8 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.

hfc5, compute optimized type family with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Stable performance
- 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors
- Equipped with a vCPU to memory ratio of 1:2
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - High-performance Web front-end servers
  - High-performance science and engineering applications
  - Massively Multiplayer Online (MMO) games and video coding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. hfc5. large	2	4.0	N/A	1.0	300	2	2	6
ecs. hfc5. xlarge	4	8.0	N/A	1.5	500	2	3	10
ecs. hfc5. 2xlarge	8	16.0	N/A	2.0	1,000	2	4	10
ecs. hfc5. 3xlarge	12	24.0	N/A	2.5	1,300	4	6	10
ecs. hfc5. 4xlarge	16	32.0	N/A	3.0	1,600	4	8	20
ecs. hfc5. 6xlarge	24	48.0	N/A	4.5	2,000	6	8	20

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. hfc5. 8xlarge	32	64.0	N/A	6.0	2,500	8	8	20

See *other instance type families*.

hfg5, general-purpose type family with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Stable performance
- 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors
- Equipped with a vCPU to memory ratio of 1:4, except for the 56 vCPU instance type
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - High-performance Web front-end servers
  - High-performance science and engineering applications
  - Massively Multiplayer Online (MMO) games and video coding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. hfg5. large	2	8.0	N/A	1.0	300	2	2	6

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. hfg5. xlarge	4	16.0	N/A	1.5	500	2	3	10
ecs. hfg5. 2xlarge	8	32.0	N/A	2.0	1,000	2	4.	10
ecs. hfg5. 3xlarge	12	48.0	N/A	2.5	1,300	4	6	10
ecs. hfg5. 4xlarge	16	64.0	N/A	3.0	1,600	4	8	20
ecs. hfg5. 6xlarge	24	96.0	N/A	4.5	2,000	6	8	20
ecs. hfg5. 8xlarge	32	128.0	N/A	6.0	2,500	8	8	20
ecs. hfg5. 14xlarge	56	160.0	N/A	10.0	4,000	14	8	20

See *other instance type families*.

4.9 Compute optimized type family with GPU

4.9.1 Compute optimized and GPU-equipped instance type families

This topic describes specific compute optimized instance type families with GPUs, and lists the specific instance types within each of the instance type families.

Specifically, the instance type families to be described are as follows: vgn5i, gn6i, gn6v, gn5, gn5i, and gn4.

vgn5i, light-weight compute optimized type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Use an NVIDIA P4 GPU computation accelerator
- Contains a virtual GPU (which is the result of partitioned virtualization)
  - 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
- Equipped with a vCPU to memory ratio of 1:3
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Real-time online rendering required for cloud gaming and AR/VR applications
  - AI reasoning (including deep and machine learning), used in the elastic deployment of Internet services that use AI reasoning and computing
  - Educational and modeling experiment environments that use deep learning

Instance types

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/ s) **	Packet forwar rate (thousa pps) ***	NIC queues ****	ENIs *****
ecs. vgn5i -m1. large	2	6	N/A	P4*1/ 8	1	1	300	2	2
ecs. vgn5i -m2. xlarge	4	12	N/A	P4*1/ 4	2	2	500	2	3

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/s) **	Packet forwar rate (thousa pps) ***	NIC queues ****	ENIs *****
ecs.vgn5i-m4.2xlarge	8	24	N/A	P4*1/2	4	3	800	2	4
ecs.vgn5i-m8.4xlarge	16	48	N/A	P4*1	8	5	1,000	4	5

Note:

For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

gn6i, compute optimized type family with GPUs

Features

- I/O-optimized
- Equipped with a vCPU to memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports ESSD Cloud Disks (million-level IOPS), SSD Cloud Disks, and Ultra Disks
- Achieves better performance with the X-Dragon new-generation compute architecture
- Uses NVIDIA T4 GPU processors:
  - Based on the new NVIDIA Turing architecture
  - Up to 320 Turing Tensor Cores
  - Up to 2,560 CUDA Cores
  - Mixed-precision Tensor Cores support 65 TFlops FP16, 130 INT8 TOPS, and 260 INT4 TOPS
  - 16 GB GPU memory capacity (320 GB/s bandwidth)

- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - AI (deep learning and machine learning) inference, computer vision, voice recognition, voice synthesization, natural language processing, machine translation, and reference systems
  - Real-time online rendering required for cloud gaming and AR/VR applications
  - Heavy-load graphic computing or graphic workstations
  - GPU-accelerated databases
  - High-performance computing

Instance types

Instance types	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****
ecs.gn6i-c4g1.xlarge	4	15	N/A	T4*1	16	4	500	2	2
ecs.gn6i-c8g1.2xlarge	8	31	N/A	T4*1	16	5	800	2	2
ecs.gn6i-c16g1.4xlarge	16	62	N/A	T4*1	16	6	1,000	4	3
ecs.gn6i-c24g1.6xlarge	24	93	N/A	T4*1	16	7.5	1,200	6	4
ecs.gn6i-c24g1.12xlarge	48	186	N/A	T4*2	32	15	2,400	12	6

Instance types	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****
ecs.gn6i-c24g1.24xlarge	96	372	N/A	T4*4	64	30	4,800	24	8
ecs.gn6i-c32g1.8xlarge	32	124	N/A	T4*1	16	10	1,600	8	6
ecs.gn6i-c48g1.12xlarge	48	186	N/A	T4*1	16	12	2,400	12	6
ecs.gn6i-c72g1.18xlarge	72	279	N/A	T4*1	16	21.5	3,600	18	8

Note:
For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

gn6v, compute optimized type family with GPUs

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Uses NVIDIA V100 GPU processors
- Equipped with a vCPU to memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors

- Uses NVIDIA V100 GPU processors (with the SXM2 module):
  - Based on the new NVIDIA Volta architecture
  - 16 GB HBM2 GPU memory capacity (900 GB/s bandwidth)
  - Up to 5,120 CUDA Cores
  - Up to 640 Tensor Cores
  - Supports up to six NVLink connections and total bandwidth of 300 GB/s (25 GB/s per connection)
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning, autonomous vehicles, voice recognition, and other AI applications
  - Scientific computing, computational finance, genomics, and environmental analysis

Instance types

Instance types	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forwarding rate (Thousand pps) ***	NIC queues ****	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	4	4	10
ecs.gn6v-c8g1.8xlarge	32	128.0	N/A	4 * NVIDIA V100	4 * 16	10.0	2,000	8	8	20
ecs.gn6v-c8g1.16xlarge	64	256.0	N/A	8 * NVIDIA V100	8 * 16	20.0	2,500	16	8	20

Note:

For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

gn5, compute optimized type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Uses NVIDIA P100 GPU processors
- No fixed ratio of vCPU to memory
- High-performance local NVMe SSD disks
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning
  - Scientific computing, such as computational fluid dynamics, computational finance, genomics, and environmental analysis
  - High-performance computing, rendering, multi-media coding and decoding, and other server-side GPU compute workloads

Instance types

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	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	1	3	10
ecs.gn5-c8g1.2xlarge	8	60.0	440	1 * NVIDIA P100	1 * 16	3.0	400	1	4	10

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.gn5-c4g1.2xlarge	8	60.0	880	2 * NVIDIA P100	2 * 16	5.0	1,000	2	4	10
ecs.gn5-c8g1.4xlarge	16	120.0	880	2 * NVIDIA P100	2 * 16	5.0	1,000	4	8	20
ecs.gn5-c28g1.7xlarge	28	112.0	440	1 * NVIDIA P100	1 * 16	5.0	1,000	8	8	20
ecs.gn5-c8g1.8xlarge	32	240.0	1760	4 * NVIDIA P100	4 * 16	10.0	2,000	8	8	20
ecs.gn5-c28g1.14xlarge	56	224.0	880	2 * NVIDIA P100	2 * 16	10.0	2,000	14	8	20
ecs.gn5-c8g1.14xlarge	54	480.0	3520	8 * NVIDIA P100	8 * 16	25.0	4,000	14	8	20

Note:

For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

gn5i, compute optimized type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Uses NVIDIA P4 GPU processors
- Equipped with a vCPU to memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning
  - Multi-media coding and decoding, and other server-side GPU compute workloads

Instance types

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	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	2	2	6
ecs.gn5i-c4g1.xlarge	4	16.0	N/A	1 * NVIDIA P4	1 * 8	1.5	200	2	3	10
ecs.gn5i-c8g1.2xlarge	8	32.0	N/A	1 * NVIDIA P4	1 * 8	2.0	400	4	4	10
ecs.gn5i-c16g1.4xlarge	16	64.0	N/A	1 * NVIDIA P4	1 * 8	3.0	800	4	8	10

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forwarding rate (Thousand pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.gn5i-c16g1.8xlarge	32	128.0	N/A	2 * NVIDIA P4	2 * 8	6.0	1,200	8	8	20
ecs.gn5i-c28g1.14xlarge	56	224.0	N/A	2 * NVIDIA P4	2 * 8	10.0	2,000	14	8	20

Note:

For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

gn4, compute optimized type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Uses NVIDIA M40 GPU processors
- No fixed ratio of CPU to memory
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning
  - Scientific computing, such as computational fluid dynamics, computational finance, genomics, and environmental analysis
  - High-performance computing, rendering, multi-media coding and decoding, and other server-side GPU compute workloads

Instance types

Instanc type	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	GPU memor (GB)	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	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	1	3	10
ecs.gn4-c8g1.2xlarge	8	30.0	N/A	1 * NVIDIA M40	1 * 12	3.0	400	1	4	10
ecs.gn4.8xlarge	32	48.0	N/A	1 * NVIDIA M40	1 * 12	6.0	800	3	8	20
ecs.gn4-c4g1.2xlarge	8	60.0	N/A	2 * NVIDIA M40	2 * 12	5.0	500	1	4	10
ecs.gn4-c8g1.4xlarge	16	60.0	N/A	2 * NVIDIA M40	2 * 12	5.0	500	1	8	20
ecs.gn4.14xlarge	56	96.0	N/A	2 * NVIDIA M40	2 * 12	10.0	1,200	4	8	20

Note:

For more information, see *Create a compute optimized instance with GPUs*.

See *other instance type families*.

4.9.2 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 *#unique_95* 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_install.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_install.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,

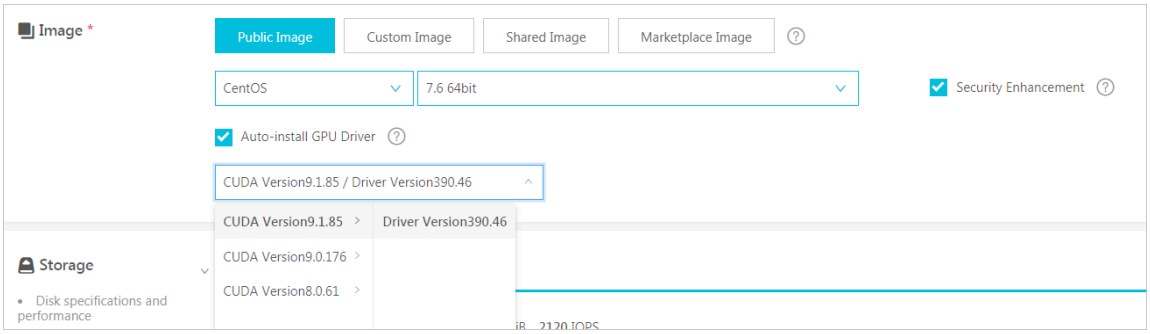
the regions and zones displayed on the page are the actual regions and zones available.

- Instance Type: Choose Heterogeneous Computing > GPU Compute, 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 *Install the GPU driver* 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.

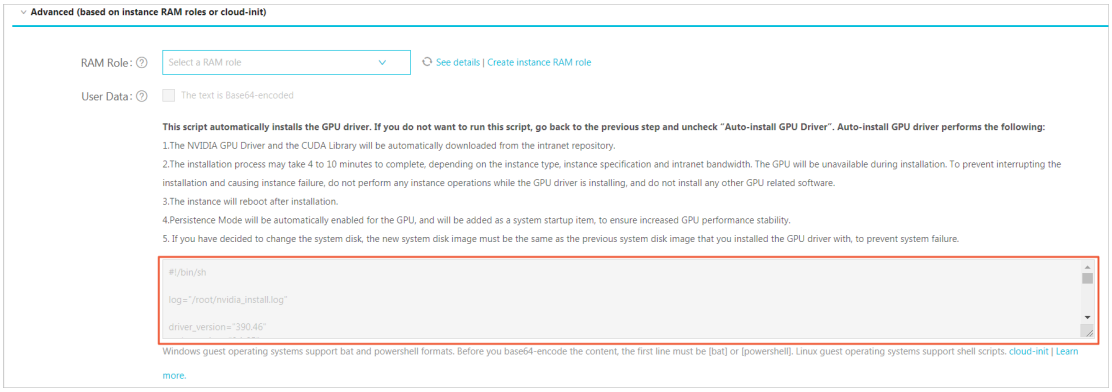


3. Complete the Networking configurations as follows:
   - Network: Select VPC.
   - Network Billing Method: Select a bandwidth as needed.

     > **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 XenDeskop 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.



     - 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	· China North 2 (zone A), China East 2 (zone B) · China South 1 (zone C)
gn5	· China North 2 (zones C and E), China North 5 (zone A) · China East 1 (zones G and F), China East 2 (zones D, B, and E) · China South 1 (zone D) · Hong Kong (zones C and B) · Asia Pacific SE 1 (zones B and A), Asia Pacific SE 2 (zone A), Asia Pacific SE 3 (zone A), Asia Pacific SE 5 (zone A) · US West 1 (zones B and A), US East 1 (zones B and A) · EU Central 1 (zone A)
gn5 (NGC environment)	The instance type family gn5 does not fully support the NVIDIA GPU CLOUD (NGC) environment. For more information, see *Deploy an NGC on gn5 instances*.
gn5i	· China North 2 (zones C, E, and A) · China East 1 (zone B), China East 2 (zones D and B) · China South 1 (zone A)
gn6v	China East 2 (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	The following versions are supported: · CentOS 64-bit (All available versions are supported.) · Ubuntu 16.04 64-bit · SUSE Linux Enterprise Server 12 SP2 64-bit
Alibaba Cloud Marketplace	Depending on your image requirements, obtain the required image by using either of the following methods: · 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_installer_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	- gn5 - gn5i - gn6v - gn4
9.0.176	- 390.46 - 384.125 - 384.111	- gn5 - gn5i - gn6v - gn4
8.0.61	- 390.46 - 384.125 - 384.111	- gn5 - gn5i - gn4

- If you configure the installation script in the User Data area, we recommend that you use `nvidia_ins tall_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_ver  sion =" 410 . 79 "
cuda_versi  on =" 9 . 0 . 176 "
cudnn_vers  ion =" 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	- 7.4.2 - 7.3.1
9.2.148	- 410.79 - 396.44	- 7.4.2 - 7.3.1 - 7.1.4
9.0.176	- 410.79 - 396.44 - 390.46	- 7.4.2 - 7.3.1 - 7.1.4 - 7.0.5
8.0.61	- 410.79 - 396.44 - 390.46	- 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_ver  sion =$ 1
 cuda_versi  on =$ 2
 cudnn_vers  ion =$ 3
```

```
NVIDIA_DIR ="/ root / nvidia "
log =${ NVIDIA_DIR }"/ nvidia_ins  tall . log "
PROCESS_NA  ME ="/ var / lib / cloud / instance / scripts / part -
001 "
DRIVER_PRO  CESS_NAME =${ NVIDIA_DIR }"/ NVIDIA - Linux - x86_64 "
CUDA_PROCE  SS_NAME =${ NVIDIA_DIR }"/ cuda "
CUDNN_PROC  ESS_NAME =${ NVIDIA_DIR }"/ cudnn "
DOWNLOAD_P  ROCESS_NAM  E =" wget "
SUCCESS_ST  R =" NVIDIA   INSTALL   OK "
DOWNLOAD_S  UCCESS_STR =" Download   OK "
DRIVER_FAI  L_STR =" Driver   INSTALL   FAIL "
CUDA_FAIL_  STR =" CUDA   INSTALL   FAIL "
CUDNN_FAIL  _STR =" CUDNN   INSTALL   FAIL "
DOWNLOAD_F  AIL_STR =" Download   FAIL "
install_no  tes =" The   script   automatica  lly   downloads   and
installs   a   NVIDIA   GPU   driver   and   CUDA / CUDNN   Toolkit .
1 .  The   installati  on   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   successful  ly .
2 .  After   the   GPU   is   installed   successful  ly ,  the
instance   will   restarts   automatica  lly ."
check_inst  all ()
{
    b =''
    if  [ "$ 1 " = " NVIDIA " ];  then
        ProcessNam  e =$ DRIVER_PRO  CESS_NAME
        t = 2
    elif  [ "$ 1 " = " cuda " ];  then
        ProcessNam  e =$ CUDA_PROCE  SS_NAME
        t = 2 . 5
    elif  [ "$ 1 " = " cudnn " ];  then
        ProcessNam  e =$ CUDNN_PROC  ESS_NAME
        t = 0 . 5
    fi
    i = 0
    while   true
    do
        pid_num =$( ps  - ef  |  grep  $ ProcessNam  e  | 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_down  load ()
{
    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_proces  s ()
{
    echo  " CHECKING   NVIDIA   INSTALL ,  PLEASE   WAIT  ......"
    echo  "$ install_no  tes "
    echo
    while   true
    do
        pid_num =$( ps  - ef  |  grep  $ PROCESS_NA  ME  | 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_P  ROCESS_NAM
E  | grep   driver  | grep  - v   grep  |  wc  - l )
            if  [ $ pid_num  - gt   0  ]; then
                echo  " Driver -${ 1 }  downloadin  g ,  need   10
seconds .  Remaining   installati  on   time   360  -  600   seconds
!"
                check_down  load  " NVIDIA "
            fi
            pid_num =$( ps  - ef  |  grep  $ DOWNLOAD_P  ROCESS_NAM
E  | grep   cuda  | grep  - v   grep  |  wc  - l )
            if  [ $ pid_num  - gt   0  ]; then
                echo  " CUDA -${ 2 }  downloadin  g ,  need   150
or   more   seconds .  Remaining   installati  on   time   350  -
590   seconds !"
                while   true
                do
                    check_down  load  " cuda "
                    sleep   1
                    pid_num =$( ps  - ef  |  grep  $ DOWNLOAD_P
ROCESS_NAM  E  | grep   cuda  | grep  - v   grep  |  wc  - l )
                    if  [ $ pid_num  - eq   0  ]; then
                        break
                    fi
                done
            fi
            pid_num =$( ps  - ef  |  grep  $ DOWNLOAD_P  ROCESS_NAM
E  | grep   cudnn  | grep  - v   grep  |  wc  - l )
            if  [ $ pid_num  - gt   0  ]; then
                echo  " cuDNN -${ 3 }  downloadin  g ,  need   about
  30   seconds .  Remaining   installati  on   time   200  -  430
seconds !"
                check_down  load  " cudnn "
            fi
            pid_num =$( ps  - ef  |  grep  $ DRIVER_PRO  CESS_NAME  |
grep  - v   grep  |  wc  - l )
            if  [ $ pid_num  - gt   0  ]; then
                echo
                echo  " Driver -${ 1 }  installing ,  need   30  -
160   seconds .  Remaining   installati  on   time   160  -  400
seconds !"
                check_inst  all  " NVIDIA "
                check_inst  all_log  " DRIVER "
            fi
            pid_num =$( 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 -${ 3 }  installing ,  need   10
 seconds .  Installati  on   will   be   successful   soon ,  please
 wait ......"
                check_inst  all  " cudnn "
                check_inst  all_log  " CUDNN "
            fi
        fi
        sleep   1
    done
}
 create_nvi  dia_repo_c  entos ()
{
    baseurl_ce  ntos =$( 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_ke rnel_cento s ()
{
    kernel_ver sion =$( uname - r )
    kernel_dev el_num =$( rpm - qa | grep kernel - devel |
 grep $ kernel_ver sion | wc - l )
    if [ $ kernel_dev el_num - eq 0 ]; then
        echo "****** exec \" yum install - y kernel - devel -$
 kernel_ver sion \""
        yum install - y kernel - devel -$ kernel_ver sion
        if [ $? - ne 0 ]; then
            echo " INSTALL_ER ROR : install kernel - devel
 fail !!!"
            return 1
        fi
    fi
    return 0
}
 install_ke rnel_suse ()
{
    kernel_ver sion =$( uname - r | awk - F '-' '{ print $ 1 "-"$
 2 }')
    kernel_dev el_num =$( rpm - qa | grep kernel - default -
 devel | wc - l )
    if [ $ kernel_dev el_num - eq 0 ]; then
        echo "*** exec \" zypper install - y kernel - default
 - devel =$ kernel_ver sion \""
        zypper install - y kernel - default - devel =$
 kernel_ver sion
        if [ $? - ne 0 ]; then
            echo " error : install kernel - default - devel
 fail !!!"
            return 1
        fi
    fi
}
 install_ke rnel_ubunt u ()
{
    kernel_ver sion =$( uname - r )
    linux_head ers_num =$( dpkg -- list | grep linux - headers
 | grep $ kernel_ver sion | wc - l )
    if [ $ linux_head ers_num - eq 0 ]; then
        echo "*** exec \" apt - get install - y -- allow -
 unauthenti cated linux - headers -$ kernel_ver sion \""
        apt - get install - y -- allow - unauthenti cated
 linux - headers -$ kernel_ver sion
        if [ $? - ne 0 ]; then
            echo " error : install linux - headers fail !!!"
            return 1
        fi
    fi
}
 download ()
{
    download_u rl ="${ baseurl }/ opsx / ecs / linux / binary /
 nvidia "
    wget ${ download_u rl }/ driver /${ driver_fil e }
    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 UCCESS_STR !"
    return 0
}
 install_dr iver ()
{
   / root / nvidia /$ driver_fil e -- silent
    if [ $? - ne 0 ]; then
        echo " INSTALL_ER ROR : driver install fail !!!"
        return 1
    fi
    echo "$ DRIVER_SUC CESS_STR !"
    return 0
}
 install_cu da ()
{
    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_fi le ="/ 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_fi le ="/ root /. bash_profi le "
 fi
 if [ "$ 1 " = " check " ]; then
```

```
    check_inst  all_proces  s  $ driver_ver  sion  $ cuda_versi  on
$ cudnn_vers  ion
    sed  - i  '/ part - 001  / d ' $ profile_fi  le
    exit  0
else
    mkdir  $ NVIDIA_DIR
    echo  " begin  to  install ,  driver : $ driver_ver  sion ,
cuda : $ cuda_versi  on ,  cudnn : $ cudnn_vers  ion  " >> $ log  2
>& 1
    driver_fil  e =" NVIDIA - Linux - x86_64 -"${ driver_ver  sion
}". run "
    cuda_big_v  ersion =$( echo  $ cuda_versi  on  |  awk  - F '.' '{
print  $ 1 "."$ 2 }')
    cudnn_file =" cudnn -"${ cuda_big_v  ersion }"- linux - x64 - v
"${ cudnn_vers  ion }". tgz "
    echo  " sh  / var / lib / cloud / instance / scripts / part - 001
  check " |  tee  - a  $ profile_fi  le
fi
echo  " os :$ os " >> $ log  2 >& 1
if  [ "$ os " = " ubuntu " ];  then
    disable_no  uveau_ubun  tu  >> $ 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_releas  e " ];  then
       pkgname =$( yum  provides  / usr / bin / lsb_releas  e  |
grep  centos | grep  x86_64  | head  - 1  | awk  - F : '{ print  $
1 }')
       if  [ - z  "$ pkgname " ];  then
           echo  " INSTALL_ER  ROR : / usr / bin / lsb_releas  e
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  uveau_cent  os  >> $ log  2 >& 1
    str =$( lsb_releas  e  - r  |  awk  - F '[:.]' '{ print  $ 2 }')
```

```
    version =$( echo  $ str  |  sed  ' s / // g ')
    create_nvi  dia_repo_c  entos
fi
install_ke  rnel_ ${ os } >> $ log   2 >& 1
if  [ $? - ne   0  ];  then
    echo  " INSTALL_ER  ROR :  kernel - devel   install   fail !!!"
>> $ log   2 >& 1
    exit   1
fi
cd  / root / nvidia
begin_down  load =$( date  '+% s ')
download  >> $ log   2 >& 1
if  [ $? - ne   0  ];  then
    exit   1
fi
end_downlo  ad =$( date  '+% s ')
time_downl  oad =$(( end_downlo  ad - begin_down  load ))
echo  " NVIDIA   download   OK !  Using   time  $ time_downl  oad   s
 !!" >> $ log   2 >& 1
begin =$( date  '+% s ')
install_dr  iver  >> $ log   2 >& 1
if  [ $? - ne   0  ];  then
    exit   1
fi
end =$( date  '+% s ')
time_insta  ll =$(( end - begin ))
echo  " NVIDIA   install   driver   OK !  Using   time  $ time_insta
ll   s  !!" >> $ log   2 >& 1
begin =$( date  '+% s ')
install_cu  da  >> $ log   2 >& 1
if  [ $? - ne   0  ];  then
    exit   1
fi
end =$( date  '+% s ')
time_insta  ll =$(( end - begin ))
echo  " NVIDIA   install   cuda   OK !  Using   time  $ time_insta
ll   s  !!"  >> $ log   2 >& 1
begin =$( date  '+% s ')
install_cu  dnn  >> $ log   2 >& 1
if  [ $? - ne   0  ];  then
    exit   1
fi
end =$( date  '+% s ')
time_insta  ll =$(( end - begin ))
echo  " NVIDIA   install   cudnn   OK !  Using   time  $ time_insta
ll   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*.

4.9.3 Install the GPU driver

This topic describes how to download and install the GPU driver. If you do not configure automatic installation of the GPU driver when you create a GPU instance, you must install it manually after the creation.

Download the GPU driver

1. Go to *NVIDIA website*.
2. Find the drivers for your NVIDIA products, and click Search. The following table shows relevant metrics.

Item	gn4	gn5	gn5i	gn6v	vgn5i	gn6i
Product type:	Tesla	Tesla	Tesla	Tesla	Tesla	Tesla
Product series	M-Class	P-Series	P-Series	V-Series	P-Series	T-Series
Product	M40	Tesla P100	Tesla P4	Tesla V100	Tesla P4	Tesla T4
Operating system (OS)	Select a release according to the image of your instance. Note: · 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.					

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

Install the GPU driver for a Linux instance

1. Download and install the kernel-devel and kernel-header packages based on your kernel version.

   > Note:
   > If version conflict exists between kernel-devel and your kernel, 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.

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 and the following information appears, this means that the installation is successfully completed.

   ```
   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.

   Take Linux 64-bit Ubuntu 14.04 for example:

Install the GPU driver for a Windows instance

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

> Note:
> 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 by yourself (for example, PCOIP, and XenDeskop HDX 3D).

4.9.4 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
```

4.9.5 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 :// nvdia - 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.

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

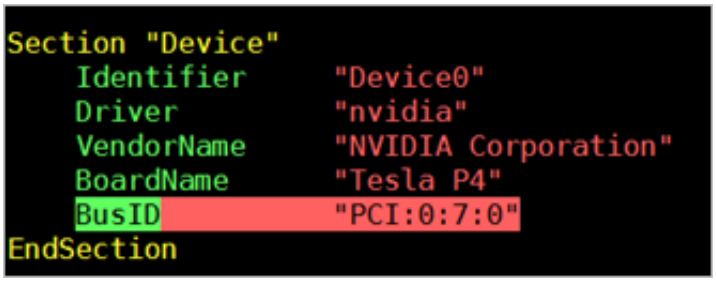
8. Run the following command to install x11vnc.

```
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`.



    c. Run `reboot` to restart the system.

Verify installation

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

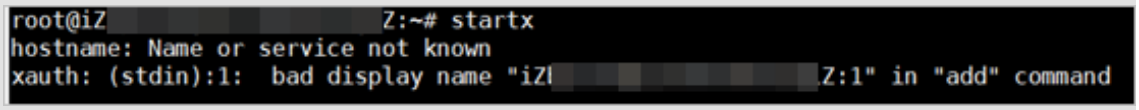
1. Run the following command to install the GLX application.

```
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.



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.

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   groupinsta  ll  " 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 :// nvdia - 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:

   a. Run `cd /etc/nvidia` to switch to the /etc/nvidia directory.

   b. Run `cp gridd.conf.template gridd.conf` to create a file named 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
```

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 taret 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.

4.10 Visualization compute type family with GPU

4.10.1 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.

ga1, visualization compute type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- AMD S7150 GPU processors
- Equipped with a vCPU to memory ratio of 1:2.5
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- High-performance local NVMe SSD disks
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Rendering, multimedia coding and decoding
  - Machine learning, high-performance computing, and high-performance databases
  - Other server-end business scenarios that require powerful concurrent floating-point compute capabilities

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	GPU	GPU memory (GB)	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.ga1.xlarge	4	10.0	1 * 87	0.25 * AMD S7150	2	1.0	200	1	3	10
ecs.ga1.2xlarge	8	20.0	1 * 175	0.5 * AMD S7150	4	1.5	300	1	4	10
ecs.ga1.4xlarge	16	40.0	1 * 350	1 * AMD S7150	8	3.0	500	2	8	20
ecs.ga1.8xlarge	32	80.0	1 * 700	2 * AMD S7150	2 * 8	6.0	800	3	8	20
ecs.ga1.14xlarge	56	160.0	1 * 1400	4 * AMD S7150	4 * 8	10.0	1,200	4	8	20

Note:
For more information, see *Create an instance of ga1*.

See *other instance type families*.

4.10.2 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: Select VPC. The current GPU-rendered ga1 instances only support VPCs.
- Instance: Select Heterogeneous Computing > GPU Visualization Compute > 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 XenDeskop HDX 3D and PCOIP.

4.11 Compute optimized type family with FPGA

4.11.1 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.

f1, compute optimized type family with FPGA

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Intel ARRIA 10 GX 1150 FPGA
- Equipped with a vCPU to memory ratio of 1:7.5
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Supports strong network performance through sufficient computing capacity
- Suitable for the following scenarios:
  - Deep learning and reasoning
  - Genomics research
  - Financial analysis
  - Picture transcoding
  - Computational workloads, such as real-time video processing and security

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	FPGA	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.f1-c8f1.2xlarge	8	60.0	N/A	Intel ARRIA 10 GX 1150	3.0	400	4	4	10
ecs.f1-c8f1.4xlarge	16	120.0	N/A	2 * Intel ARRIA 10 GX 1150	5.0	1,000	4	8	20
ecs.f1-c28f1.7xlarge	28	112.0	N/A	Intel ARRIA 10 GX 1150	5.0	2,000	8	8	20
ecs.f1-c28f1.14xlarge	56	224.0	N/A	2 * Intel ARRIA 10 GX 1150	10.0	2,000	14	8	20

See *other instance type families*.

f3, compute optimized type family with FPGA

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Xilinx 16nm Virtex UltraScale + VU9P
- Equipped with a vCPU to memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Supports strong network performance through sufficient computing capacity

- Suitable for the following scenarios:
  - Deep learning and reasoning
  - Genomics research
  - Speeding up database access
  - Picture transcoding, such as converting JPEG to WebP
  - Real-time video processing, such as H.265 video compression

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	FPGA	Bandwidth (Gbit/s) **	Packet forward rate (Thousand pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs.f3-c4f1.xlarge	4	16.0	N/A	1 * Xilinx VU9P	1.5	300	2	3	10
ecs.f3-c8f1.2xlarge	8	32.0	N/A	1 * Xilinx VU9P	2.5	500	4	4	10
ecs.f3-c16f1.4xlarge	16	64.0	N/A	1 * Xilinx VU9P	5.0	1,000	4	8	20
ecs.f3-c16f1.8xlarge	32	128.0	N/A	2 * Xilinx VU9P	10.0	2,000	8	8	20
ecs.f3-c16f1.16xlarge	64	256.0	N/A	4 * Xilinx VU9P	20.0	2,500	16	8	20

See *other instance type families*.

4.11.2 Create an f1 instance

This article 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 East 1 (Hangzhou) > China East 1 Zone F.
- Instance Type: Select Heterogeneous Computing > FPGA > Compute, 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: 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)*

4.11.3 Create an f3 instance

This article 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 East 2 (Shanghai).
- Instance Type: Select Heterogeneous Computing > FPGA Compute, 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: Select VPC.

Best practices

Best practices for OpenCL on an f3 instance

Best practices for RTL design on an f3 instance

4.12 ECS bare metal instance type family

4.12.1 EBM Instance overview

This topic describes ECS Bare Metal (EBM) Instances and the instance type families that provide EBM Instance types (specifically, high-clock speed type family ebmhfg5, compute type family ebmc4, and general-purpose type family ebmg5).

Background information

Alibaba Cloud EBM Instances use next-generation virtualization technology that supports multiple virtual cloud servers and nested virtualization while also maintaining excellent resource elasticity and a seamless user experience.

Compared with other available instance types, EBM Instances offer the following advantages:

- Exclusive computing resources

  EBM Instances provide better performance and isolation than standard physical servers and enable exclusive computing resources without virtualization performance overheads or feature loss. EBM Instances support 8, 32, and 96 CPU cores and ultra high frequencies. For example, an EBM Instance with 8 cores supports a frequency of up to 3.7 to 4.1 GHz.
- Chip-level security

  EBM Instances implement Intel® SGX, which allows EBM Instances to compute only encrypted data in a safe and trusted environment, providing advanced security for your data in the cloud. For more information, see *Install SGX*.
- Compatible with multiple private clouds

  EBM Instances supports re-virtualization. As a result, private cloud users can seamlessly migrate to Alibaba Cloud without experiencing performance overhead issues that typically occur in standard nested virtualization methods.
- Support for heterogeneous instruction set processors

  EBM Instances are based on proprietary virtualization 2.0 technology developed by Alibaba Cloud, which support ARM and other instruction set processors without incurring additional costs.

When you use EBM Instances, note the following:

- The specifications of EBM Instances cannot be upgraded or downgraded.
- If a hardware fault occurs to an EBM Instance, a failover occurs and data is stored in the cloud disks of the EBM Instance.

Comparison of EBM Instances, physical machines, and virtual machines

The following table compares EBM Instance, physical servers, and virtual servers. In this table, Y means supported, N means not supported, and N/A means not applicable .

Feature type	Feature	EBM Instance	Physical server	Virtual server
Automated O&M	Delivery in minutes	Y	N	Y

Feature type	Feature	EBM Instance	Physical server	Virtual server
Computing	Zero performance loss	Y	Y	N
	Zero feature loss	Y	Y	N
	Zero resource preemption	Y	Y	N
Storage	Compatible with ECS cloud disks	Y	N	Y
	Started from cloud disks (system disks)	Y	N	Y
	Quick resetting of the system disk	Y	N	Y
	Compatible with ECS images	Y	N	Y
	Cold migration between physical and virtual servers	Y	N	Y
	Free of OS installation	Y	N	Y
	Free of local RAID, and stronger protection of data in cloud disks	Y	N	Y
Network	Compatible with the ECS VPC networks	Y	N	Y
	Compatible with the ECS classic networks	Y	N	Y

Feature type	Feature	EBM Instance	Physical server	Virtual server
	Free of communication bottlenecks between physical and virtual server clusters in the VPC	Y	N	Y
Management	Compatible with the existing ECS management system	Y	N	Y
	Consistent user experience on VNC and other features with that of virtual servers	Y	N	Y
	Out-of-band (OOB) network security	Y	N	N/A

ebmhfg5, ECS Bare Metal Instance type family with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:4
- 3.7 GHz Intel Xeon E3-1240v6 (Skylake) processors, 8-core vCPU, up to 4.1 GHz Turbo Boot
- High network performance: 2 million pps packet forwarding rate
- Supports VPC network only
- Provides dedicated hardware resources and physical isolation
- Supports Intel SGX

- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources, or scenarios where binding a license to the hardware is required
  - Gaming or financial applications featuring high performance
  - High-performance Web servers
  - Enterprise-level applications, such as high-performance databases

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. ebmhfg5 .2xlarge	8	32.0	N/A	6.0	2,000	8	6	8

Note:
For more information about ECS Bare Metal Instance, see *EBM Instance overview*.

See *other instance type families*.

ebmc4, computing ECS Bare Metal Instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:2
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors, up to 2.9 GHz Turbo Boot
- High network performance: 4 million pps packet forwarding rate
- Supports VPC network only
- Provides dedicated hardware resources and physical isolation

- Suitable for the following scenarios:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication information
  - 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 Container, and Pouch)
  - Enterprise-level applications, such as medium and large databases
  - Video coding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. ebmc4. 8xlarge	32	64.0	N/A	10.0	4,000	8	12	10

Note:

For more information about ECS Bare Metal Instance, see *EBM Instance overview*.

See *other instance type families*.

ebmg5, general-purpose ECS Bare Metal Instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Equipped with a vCPU to memory ratio of 1:4
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96-core vCPU, up to 2.7 GHz Turbo Boot
- High network performance: 4 million pps packet forwarding rate
- Supports VPC network only
- Provides dedicated hardware resources and physical isolation

- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources, or scenarios where binding a license to the hardware is required
  - 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 Container, and Pouch)
  - Enterprise-level applications, such as medium and large databases
  - Video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB) *	Bandwid (Gbit/s) **	Packet forwardi rate (Thousar pps) ***	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. ebmg5. 24xlarge	96	384.0	N/A	10.0	4,000	8	32	10

Note:
For more information about ECS Bare Metal Instance, see *EBM Instance overview*.

See *other instance type families*.

Billing methods

EBM Instances support Subscription and Pay-As-You-Go billing methods. For more information, see *Billing method comparison*.

Related document

For more information, see *EBM Instance FAQ*.

4.12.2 Create an EBM instance

Follow the steps in *creating an instance by using the wizard* to create an EBM instance. However, the following configurations are recommended:

- Region: Currently, EBM instances are available in the following regions and zones: China East 2 (Shanghai), Zone D, China North 2 (Beijing), Zone C, China East 1 (Hangzhou), Zone G, and China South 1 (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	- 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 - OpenSUE 42.3 64-bit - SUSE Linux Enterprise Server 12 SP2 64-bit - Aliyun Linux 17.1 64-bit
Windows	- 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: EBM 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: Only VPC is supported.

4.12.3 Install SGX

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

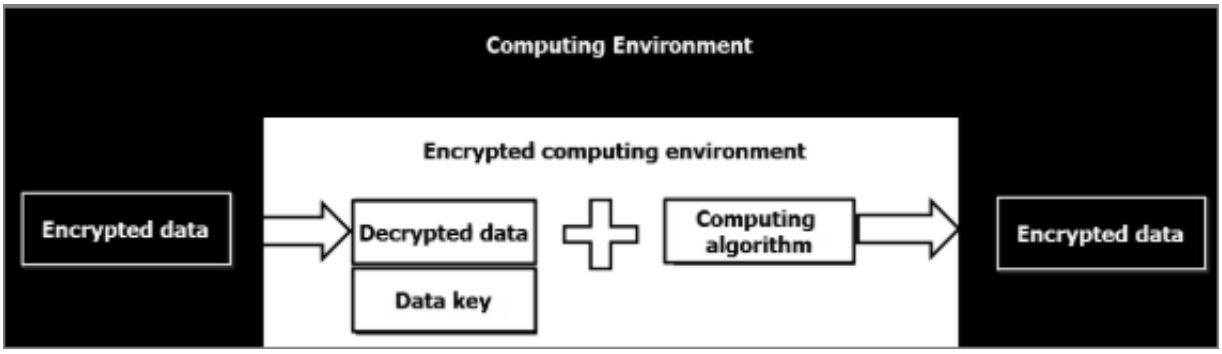
What is SGX?

SGX is an Intel architecture extension designed to increase the security of applicatio n 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.

4.13 Super Computing Cluster instance type family

4.13.1 Super Computing Clusters

This topic describes the features of three Super Computing Cluster (SCC) instance type families: scch5, sccg5, and sccgn6. The instance type family scch5 features high

clock speed, sccg5 is for general use scenarios, and sccgn6 is compute optimized. Their specific instance types are also provided.

Overview

SCCs are based on ECS Bare Metal (EBM) Instances. Given 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 EBM 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 or engineering computing, data analysis, and audio and video processing. In the clusters, nodes are connected by 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 the Alibaba Cloud elastic high-performance computing (E-HPC) platform with the ultimate high performance parallel computing resources, making supercomputing on the cloud a reality.

Comparison of SCCs to physical and virtual servers

The following table compares SCC, physical servers, and virtual servers.

Metric	Feature	SCC	Physical server	Virtual server
Automated O&M	Delivery in minutes	Yes	No	Yes
Computing	Zero performance loss	Yes	Yes	No
	Zero feature loss	Yes	Yes	No
	Zero resource competition	Yes	Yes	No

Metric	Feature	SCC	Physical server	Virtual server
Storage	Fully compatible with ECS cloud disks	Yes	No	Yes
	Start from cloud disks (system disks)	Yes	No	Yes
	System disk can be quickly reset	Yes	No	Yes
	Uses ECS images	Yes	No	Yes
	Supports cold migration between physical and virtual servers	Yes	No	Yes
	Requires no installation of operating system	Yes	No	Yes
	Discards local RAID, and provides stronger protection of data on cloud disks	Yes	No	Yes
Network	Fully compatible with ECS VPCs	Yes	No	Yes
	Fully compatible with the ECS classic network	Yes	No	Yes

Metric	Feature	SCC	Physical server	Virtual server
	Free of bottlenecks for communications between physical and virtual server clusters in a VPC	Yes	No	Yes
Management	Fully compatible with the existing ECS management system	Yes	No	Yes
	Consistent user experience with virtual servers regarding VNC and other features	Yes	No	Yes
	Guaranteed out of band network security	Yes	No	No

scch5, Super Computing Cluster (SCC) instance type family with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Supports both RoCE and VPC networks, of which RoCE is dedicated to RDMA communication
- With all features of ECS Bare Metal Instance
- 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors
- Equipped with a vCPU to memory ratio of 1:3

- Suitable for the following scenarios:
  - Large-scale machine learning applications
  - Large-scale high-performance scientific and engineering applications
  - Large-scale data analysis, batch computing, video encoding

Instance types

Instanc type	vCPU	Physica core	Memor (GiB)	GPU	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	RoCE (Inboun / Outbou) (Gbit /s)	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. scch5. 16xlarge	64	32	192.0	N/A	10.0	4,500	46	8	32	10

Note:
The ecs.scch5.16xlarge instance type provides 64 logical processors on 32 physical cores. For more information about SCC, see *Super Computing Clusters*.

See *other instance type families*.

sccg5, general-purpose Super Computing Cluster (SCC) instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Disks
- Supports both RoCE and VPC networks, of which RoCE is dedicated to RDMA communication
- With all features of ECS Bare Metal Instance
- Equipped with 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Equipped with a vCPU to memory ratio of 1:4

- Suitable for the following scenarios:
  - Large-scale machine learning applications
  - Large-scale high-performance scientific and engineering applications
  - Large-scale data analysis, batch computing, video encoding

Instance types

Instanc type	vCPU	Physica core	Memor (GiB)	GPU	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	RoCE (Inboun / Outbou) (Gbit /s)	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sccg5. 24xlarge	96	48	384.0	N/A	10.0	4,500	46	8	32	10

Note:
The ecs.sccg5.24xlarge instance type provides 96 logical processors on 48 physical cores. For more information about SCC, see *Super Computing Clusters*.

See *other instance type families*.

sccgn6, compute optimized Super Computing Cluster (SCC) instance type family with GPUs

Features

- I/O-optimized
- Equipped with a vCPU 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 ESSD Cloud Disks (million-level IOPS), SSD Cloud Disks, and Ultra Disks
  - Supports high-performance Cloud Parallel File System (CPFS)

- Networking:
  - Supports VPC networks equipped with two 25Gbps 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
  - Up to 640 Tensor Cores
  - Up to 5,120 CUDA Cores
  - 16 GB HBM2 GPU memory capacity (900 GB/s 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 applications
  - Large-scale high-performance scientific and emulation applications
  - Large-scale data analysis, batch computing, video encoding

Instance types

Instanc types	vCPU	Memor (GiB)	Local disks (GiB) *	GPU	Bandw (Gbit/s) **	Packet forwar rate (Thous pps) ***	RoCE (Inboun / Outbou) (Gbit /s)	NIC queues ****	ENIs *****	Private IP address of a single ENI
ecs. sccgn6 . 24xlarge	96	384	N/A	V100* 8	30	4,500	25*2	8	32	10

See *other instance type families*.

Billing methods

SCC instances support the Subscription billing method. For more information about billing methods, see *Billing method comparison.*

Related documents

For more information, see *SCC FAQ*.

4.13.2 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 (EBM) 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 EBM 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

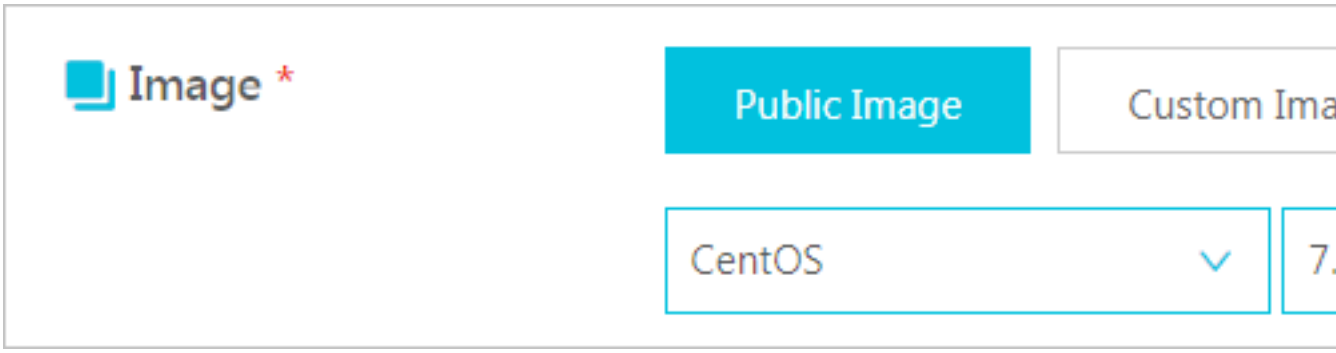
You can create an SCC instance by following the instructions described in *Create an ECS Instance*.

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: 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 East 2 (Zones D and B)
sccg5	China East 2 (Zones D and B)
sccgn6	· China East 2 · China North 2 · China North 3

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.

4.14 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 *ENI overview*.

Instance configuration upgrade

Some phased-out instance type families support upgrading to a newer instance type. For more information, see *Instance type families that support instance type upgrades*.

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	· 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	· Medium-sized Web servers · Batch processing · Distributed analysis · Advertisement services · Hadoop clusters
e3	Memory entry-level instances	1:8	· 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

For the Subscription billing method, you can use resources only after you have paid for them.

Note:
All the charging rules described in this article is for reference purpose only. For more information, please contact your service provider.

Applicable resources

Currently, Subscription billing is used for the following ECS resources:

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

If you create an ECS instance that uses Subscription billing, the Instance Cost displayed in the lower right of the instance creation page is the total cost of the three resources listed.

Payment methods

Credits are used to pay for the resources of the Subscription billing method.

Billing period

Resources in Subscription mode are billed on a monthly basis. The billing period is based on the UTC+8 time zone. It starts from the resource activation time and ends at 00:00:00 on the second day after one month or one year.

For example, if you activate an ECS instance in Monthly Subscription mode at 13:23 :56 March 12, 2017, the first billing cycle for the instance ends at 00:00:00 April 13, 2017.

The price unit varies depending on the resource. The following table lists the price units for various resources.

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

To continue using a resource after the billing period has ended, you can renew your ECS instance at the end of a billing period. For more information about the renewal procedure, see *Manual renewal* or *Auto-renewal*.

Status changes after expiration

The status of a Subscription instance after it expires will change depending on whether the auto-renewal feature is enabled.

- Auto-renewal disabled

  If auto-renewal is disabled, and a Subscription instance is not renewed at the end of the billing period, it goes out of service within the 24 hours from 00:00:00 on the

expiration day to 00:00:00 the next day. The status changes of related resources are shown in the following table.

Period	ECS instances and images	System disks	Data disks	Internet IP addresses	Snapshots
On the expiration day	Out of service*, and images are disabled.	Out of service, but data is retained.	Out of service, but data on cloud disks, local disks, or shared block storage devices is retained.	For an instance of the Classic network type: If an Internet IP address is assigned, it is retained. For a VPC-Connected ECS instance: If an Internet IP address is assigned, it is retained. If an EIP address is bound, it is retained.	Retained.

Period	ECS instances and images	System disks	Data disks	Internet IP addresses	Snapshots
15 days after expiration	Released automatically	Released along with the instance, and data cannot be recovered.	By default, cloud disks are released along with the instance and data on them cannot be recovered. If you have set the cloud disks not to be released along with the instance, the disks stop working. Local disks are released along with the instance and data on them cannot be recovered. Shared block storage is detached automatically from the instance.	For an instance of the Classic network type: If an Internet IP address is assigned, it is released. For a VPC-Connected ECS instance: If an Internet IP address is assigned, it is retained. If an EIP address is bound, it is unbound from the instance.	Automatic snapshots are automatically deleted, but those manually created are retained. Manual snapshots are not affected.

* When an instance is Out of service, you cannot connect to it, the website deployed on the instance cannot be accessed, and your business operations may be affected.

 Note:
You cannot enable the auto-renewal feature for an instance after it expires.

- Auto-renewal enabled

  If auto-renewal is enabled for an instance, but it fails to be renewed in the specified period, the status changes of the related resources are shown in the following table.

Period	ECS instances and images	System disks	Data disks	Internet IP addresses	Snapshots
Within 15 days of expiration	Running properly*.	Running properly.	Running properly.	Retained.	Retained.

Period	ECS instances and images	System disks	Data disks	Internet IP addresses	Snapshots
15 days after expiration	The instance goes out of service** at any time within the 24 hours from 00:00:00 on the 15th day after expiration to 00:00:00 the next day.	Out of service, but data is retained.	Out of service, but data is retained.	For an instance of the Classic network type: If an Internet IP address is assigned, it is released. For a VPC-Connected ECS instance: If an Internet IP address is assigned, it is retained. If an EIP address is bound, it is retained.	Retained.

Period	ECS instances and images	System disks	Data disks	Internet IP addresses	Snapshots
30 days after expiration	Automatically released.	Released along with the instance, and data cannot be recovered.	By default, the cloud disks are released along with the instance. If you set them not to be released along with the instance, they stop working. Local disks are released along with the instance. Shared block storage is automatically detached.	For an instance of the Classic network type: If an Internet IP address is assigned, it is released. For a VPC-Connected ECS instance: If an Internet IP address is assigned, it is released. If an EIP address is bound, it is unbound from the instance.	Automatic snapshots are automatically deleted, but those that are manually created are retained.

* Running properly means you can start and stop the instance properly and connect to the instance by using the Management Terminal in the console or other remote connection methods.

** When an instance is Out of service, you cannot connect to it, the website deployed on the instance cannot be accessed, and your business operations may be affected.

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. Costscan be reduced by30% to 80% compared toa traditional host investment, with which many resources may be wasted at times.

 Note:
All the charging rules described in this article is for reference purpose only. For more information, please contact your service provider.

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/or data disks

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

You can make 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 *Change configurations of Pay-As-You-Go instances*.

- Billing method change:

  Instances, system disks, and data disks support switching from Pay-As-You-Go billing to Subscription billing. For more information, see *Switch from Pay-As-You-Go to Subscription billing*.

Payment methods

Credits are used to pay for the resources of the Pay-As-You-Go billing method.

Billing period

A Pay-As-You-Go resource is billed by the second after is it created, and billing stops after it is released.

For a VPC instance, you can enable the *no fees for stopped instances feature*.When the feature is enabled, a VPC instance is not billed when it is in aStopped status. This feature is only available for instances, and not for other ECS resources.

The billing cycle varies depending on the resource types. The minimum charge for the lifecycle of an ECS instance (from creation to release) is USD 0.01.

Item	Instances + Images	System disks	Data disks
Billing cycle	One second	One second	One second
Price unit	USD/hour	USD/(GiB * hours)	USD/(GiB * hours)

Settlement cycle

Pay-As-You-Go resources are billed by the second, but settled by the 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.

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/or 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.

- 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 the 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 the Pay-As-You-Go resources 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 + data disks	Internet IP address	Snapshots
Within 15 days of the instance going out of service (T+15 to T+30)	Both stop working.	When the instance is out of service*, the capability of the cloud disks and the local disks is limited. But the data on them is retained.	· 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 + data disks	Internet IP address	Snapshots
15 days after the instance goes out of service (T+30)	Released automatically. You are notified in advance by emails when the resources will be released.	All cloud disks, including system disks and data disks that are created separately or together with the instance, or that are attached to the instance or not, are released automatically. The data cannot be recovered. The local disks are released automatically, and the data on them cannot be recovered. If shared block storage is attached to the instance, it is detached automatically, and the data on it is retained.	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.	The automatic snapshots are deleted automatically. The snapshots that are manually created are retained.

* When a Pay-As-You-Go instance is out of service, the instance is in anExpiredstatus. During the period it is out of service, no fees are incurred.

FAQs

If a Pay-As-You-Go ECS instanceis 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 an Expired status, and no fees are incurred.

A stopped instance is in a Stopped status and has been stopped*in the ECS console* or by using the *StopInstance* interface. Billing of a stopped instance varies according to the network type of the instance:

- VPC: You can enable the *No fees for stopped instances (VPC-Connected)* feature. When this feature is enabled, an instance is not billed when it is in aStopped status. This feature is only available for instances, and not for other resources.
- Classic: An instance continues to be billed even after it is in aStopped 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.

Life cycle

The following figure shows the life cycle 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_99* interface to create an instance, set the `SpotStrate gy` parameter to `SpotAsPric eGo`.

- 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_99* 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 *Billing of Internet bandwidth*.

- Billing method

  Preemptible instances are billed by the 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_66* 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

For more information about the preemptible instance quota, see *Limits*.

Create a preemptible instance

You can purchase a preemptible instance by using the *#unique_99* 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_66* 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: · 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: · 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	During the guaranteed duration, the instance can be restarted successfully only if the resource stock is sufficient. After the guaranteed duration: · 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	During the guaranteed duration, the instance can be restarted successfully only if the resource stock is sufficient. After the guaranteed duration: · 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 *User data* 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

You can create a preemptible instance in the ECS console. This document describes the steps and relevant operations.

Procedure

1. Log on to the *ECS console*.
2. In the left-side navigation pane, select Instances.
3. Click Create Instance.
4. Select Preemptible Instance for Billing Method.
5. Set the Single instance price limit.

   You need to enter the maximum price you are willing to pay. When your bid is higher than the current market transaction price, the instance starts to run. Eventually, the market transaction price is charged for the instance.
6. Select or enter the number of instances to purchase.
7. Complete other settings. For more information, see *create an instance by using the wizard*.
8. 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. After opening the instance details page, you can view the bidding policy set during instance creation in the Payment Information area.

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 cases.

Procedure

1. Log on to the *ECS console*.
2. In the left-side navigation pane, select Instances.
3. On the Instances page, click More > Instance Status > Stop in the actions column to the right of the target instance.
4. In both the Notes and Stop dialog boxes that appear in order, click OK.

Alternatively, you can call *StopInstance* to stop a preemptible instance by using developer tools such as Alibaba Cloud CLI, OpenAPI Explorer, and Alibaba Cloud SDK.

Can preemptible instances be restarted after they are stopped?

All preemptible instances can be restarted after they are stopped during and after their guaranteed duration. However, the network type, bidding mode, and stop mode of an instance affect the conditions of the instance restart. The following table describes these differences.

 Note:
Only preemptible instances in a VPC support the *No fees for stopped VPC instances* feature.

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.

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.
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.
	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.

5.4 Reserved Instances

5.4.1 RI overview

A Reserved Instance (RI) is a discount coupon with specific attributes. It can automatically match one or more Pay-As-You-Go instances (excluding preemptible instances) in your account to provide a billing discount. Compared with Subscription instances, the combination of RIs and Pay-As-You-Go instances provides higher flexibility and cost-effectiveness.

Feature release

Comparison of RIs against Subscription and Pay-As-You-Go instances

The following table compares RIs with Subscription and Pay-As-You-Go instances.

Item	RI	Pay-As-You-Go instance	Subscription instance
Definition	A type of discount coupon.	An instance with the *Pay-As-You-Go* billing method, equivalent to a virtual machine.	An instance with the *Subscription* billing method, equivalent to a virtual machine.
Usage	RIs cannot be used alone. Instead, they must match instances with specific attributes to generate a billing discount.	Pay-As-You-Go instances can be managed independently. They can be used as simple Web servers, or be combined with other Alibaba Cloud products to deliver powerful solutions.	Subscription instances can be managed independently. They can be used as simple Web servers, or be combined with other Alibaba Cloud products to deliver powerful solutions.

Payment methods, term, and instance count

When you purchase an RI, you can specify the payment method, term, and instance count based on your budget.

- Three payment methods are available:

  All Upfront, Partial Upfront, and No Upfront. For more information, see *Reserved Instance billing*.

  > Note:
  > Your Alibaba Cloud account needs to reach a certain level before you can select No Upfront.

- Two terms are available:

  1 year and 3 years.

  > Note:
  > After an RI expires, the matched Pay-As-You-Go instances will still operate normally, but they will be billed without a discount.

- Instance count:

  The instance count refers to the number of Pay-As-You-Go instances that can be matched by an RI at the same time.

Attributes

An RI has specific attributes that automatically match the corresponding Pay-As-You-Go instances. You can also split an RI, merge multiple RIs, or change the scope of an RI. In this way, your RIs can flexibly match your Pay-As-You-Go instances. The attributes include:

- Operating system: Currently, RIs can only match Pay-As-You-Go Linux instances.
- Instance type: The type of an RI, which indicates the instance type family and the instance size. This attribute is used to match the corresponding Pay-As-You-Go instances.
- Scope: This attribute indicates the matching scope of an RI. Depending on the matching scope, RIs can be classified into regional RIs and zonal RIs.

  > Note:
  > We recommend that you purchase both zonal RIs and regional RIs to meet all of your requirements. In detail, zonal RIs apply when you are certain of the zone in which you want to use it. If you are uncertain about the specific use of your RI, we recommend that you use a regional RI to meet your wider needs.

- Computational power: This attribute indicates the upper limit of computational resources that an RI matches. The computational power is determined by the instance type and the instance count.

Limits

Currently, RIs have the following limits:

- Number of RIs
  - Number of regional RIs: Each account can have up to 20 regional RIs in all regions.
  - Number of zonal RIs: Each account can have up to 20 zonal RIs in each zone.

  For example, in China East 1 (Hangzhou) and China North 1 (Qingdao), you can purchase 10 regional RIs respectively because the upper limit of regional RIs is 20 per account. In zones B and H of China East 1 (Hangzhou), you can purchase up to 20 zonal RIs respectively. If you need more RIs, you can open a ticket.
- Matchable instances: RIs only match Pay-As-You-Go instances (excluding *preemptible instances*).
- Instance type family: Currently, RIs support the following instance type families: sn1ne, sn2ne, se1ne, ic5, c5, g5, r5, hfc5, hfg5, and t5. For more information, see *Instance type families*.

   Note:
  The RIs of *t5 burstable instances* (t5) are only available at the zonal level.
  Additionally, they do not support merging, splitting, or scope changing.

Fees

For more information about RI billing, see *Reserved Instance billing*.

References

For the matching rules, see *Matching rules of Reserved Instances*.

For purchasing operations, see *Purchase a Reserved Instance*.

For managing operations, see *Manage Reserved Instances flexibly*.

For information on how to use an API to purchase an RI, see *PurchaseReservedInstancesOffering*.

For information on how to use an API to query RIs, see *DescribeReservedInstances*.

For information on how to use an API to manage RIs, see *ModifyReservedInstances*.

Contact us

If you have any questions when you use RIs, you can scan the following QR code to join the Reserved Instances Service Group for technical support (*Download DingTalk*).

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.	You have the following running Pay-As-You-Go instances: Two ecs.c5.xlarge Linux instances in China North 1 (Qingdao). The instance names are C5PAYG-1 and C5PAYG-2 respectively. You purchase the following RI: One regional ecs.c5.2xlarge RI in China North 1 (Qingdao). The RI name is C5RI. After the purchase, C5RI matches C5PAYG1 and C5PAYG2 simultaneously to provide a billing discount.

Feature	Regional RI	Zonal RI	Example
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.	You have the following running Pay-As-You-Go instance: One ecs.c5.xlarge Linux instance in zone B of China North 1 (Qingdao). The instance name is C5PAYG-b. You purchase the following RI: One regional ecs.c5.xlarge RI in China North 1 (Qingdao). The RI name is C5RI. After the purchase, C5RI matches C5PAYG-b to provide a billing discount. 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.
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 North 1 (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 North 1 (Qingdao) for one year.

5.4.3 Purchase a Reserved Instance

This topic describes how to purchase a Reserved Instance (RI).

Before you begin:

- Before purchasing an RI, make sure that the Pay-As-You-Go instances to be matched do not exceed the *Limitations*.

- The matching status between an RI and a Pay-As-You-Go instance cannot be manually managed. We recommend that you read the *Matching rules of RIs* to fully understand the matching rule requirements.

Procedure

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Reserved Instances. If you have not tried this feature, click Apply to apply for testing.
3. On the Reserved Instances page, click Purchase Reserved Instance.
4. Set the region-related parameters.
   a. Select a region.
   b. Set Resource Reservation.

   > Note:
   > Only zonal RIs support resource reservation. However, regional RIs can apply to Pay-As-You-Go instances of different sizes or in different zones.
5. Configure the RI.
   a. Select an Instance Type.

   > Note:
   > You must select an instance size when you purchase a regional RI, but you do not need to specify the instance size for the RI to match Pay-As-You-Go instances.

   b. Select Payment Option.

      The options are All Upfront, Partial Upfront, and No Upfront. For more information, see *Reserved Instance billing*.
6. Set the purchase parameters.
   a. (Optional) Enter a Name.
   b. Enter a Term.

      The options are 1 Year(s) and 3 Year(s).
   c. Enter an Instance Count.

      The Reserved Instance can match the specified count of Pay-As-You-Go instances with the specified instance type. For example, if the instance type is

ecs.g5.large and instance count is 3, the Reserved Instance can match 3 Pay-As-You-Go instances with instance type as ecs.g5.large.

7. Read and confirm that you agree to the *ECS Service Level Agreement* and *Product Terms of Service*, and then click Purchase.
8. In the confirmation dialog box, confirm the parameters, and then click Create Order.
9. Confirm the payment information, and then click Pay.

What to do next

After you have successfully purchased an RI, you can immediately benefit from the billing discount when the RI matches one or more Pay-As-You-Go instances. You can also *Manage Reserved Instances* to quickly adjust to any changes made to your Pay-As-You-Go instances.

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.

Before you begin:

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*.

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. Log on to the *ECS console*.
2. In the left-side navigation pane, click Reserved Instances.
3. On the Reserved Instances page, click Split in the Actions column of the original RI.
4. 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.

5. 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. Log on to the *ECS console*.
2. In the left-side navigation pane, click Reserved Instances.
3. On the Reserved Instances page, click Merge in the Actions column of the original RI.
4. 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.
5. 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 East 1 (Hangzhou), you can modify it as a zonal RI in another zone of China East 1 (Hangzhou), or as a regional RI in China East 1 (Hangzhou). However, you cannot modify it as a regional or zonal RI in another region.

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Reserved Instances.

3. On the Reserved Instances page, click Modify in the Actions column of the original RI.
4. On the Modify Reserved Instance Page, modify the parameters as needed.
5. 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 stoppedstatus, 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

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

- **Alibaba Cloud users who have reached a certain membership level can change the billing method from Subscription to Pay-As-You-Go.**
- **Each account is limited by a maximum monthly refund amount. The maximum refund amount varies according to the membership level. 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 vCPU × 1 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 ( vCPUs ) × 60 ( remaining days ) × 24 ( hours / day ) = 5760 ( refund amount )`.

- Note that only the actual payment amount is refunded to the original payment channel. Vouchers or coupons that have been redeemed cannot be returned.
- 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 status.

Procedure

1. Log on to the *ECS console*.
2. In the left-side navigation pane, select Instances.
3. Select a region.
4. Select the relevant instance. In the Actions column, select More > Configuration Change > Switch to Pay-As-You-Go.
5. On the conversion page, read the notes and confirm you agree to the *ECS Service Terms*.
6. Click Confirm Conversion.

   After the billing method is converted successfully, you can see that the Billing Method of the instance has 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. If you want to create a custom image from a snapshot of your system disk, and then use the custom image to create an ECS instance, see how to create an instance from a custom image.

Prerequisites

- Before creating an ECS instance, you must complete 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 the *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, click instances.
3. On the Instances list page, click Create Instance.
4. Complete the Basic Configurations as follows:
   a) Select a Billing Method: Subscription, Pay-As-You-Go or *Preemptible Instance*.

      Note:
      For how to create preemptible instances, see *Create preemptible instances*.

   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.
      - Note that some instance type families are not supported in all regions. For more information, see *Create a compute optimized instance with GPUs*, *Create an f1 instance*, *Create an SCC server instance*, and *Create an EBM instance*.

   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 vCPU cores or an entry-level instance type with at least four vCPU cores. 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 Data, select an image as instructed in *User data*.
- Public images only include the initial system environment, and more images are available in the image Marketplace.

e) Select storage devices:

- System Disk: Required. 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, with a maximum size of 500 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 shown in the following table.

Image	Available size range
Linux (excluding CoreOS) FreeBSD	[max{20, ImageSize}, 500] GiB. Where, the public image size is 40 GiB for Ubuntu 14.04 32-bit, Ubuntu 16.04 32-bit, and CentOS 6.8 32-bit.
CoreOS	[max{30, ImageSize}, 500] GiB
Windows	[max{40, ImageSize}, 500] GiB

- Data Disk: Optional. If you create a cloud disk as a data disk at this time, you must select the disk type, size, and quantity, and set whether to *encrypt* it. You can 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 type 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. 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 network billing, see *Billing of Internet bandwidth*.
      - 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 *Default security group rules*.
   d) Add an Elastic Network Interface (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 *#unique_195* interface.

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. At the left side, API Workflow explains the related APIs and request parameter values for the current operation. At 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 shown 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 *Format a data disk for Windows instances* or *Format a data disk of a Linux instance*.

6.3 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_204*
  - *#unique_205*
- If the custom image and the instance are in different regions, copy the custom image to the target region. For more information, see *#unique_206*.
- If the image to be used is owned by another account, it must be shared with you. For more information, see *share images*.

Procedure

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.

   Alternatively, you can click Images to find the target image, and then click Create Instance in the Actions column.
3. In the upper-right corner of the Instances page, click Create Instance.
4. Follow the stepswhen 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. You cannot decrease it.

5. Confirm the order.

6.4 Create an instance of the same configuration

To duplicate ECS instances of the same configurations, use the Buy Same Type feature.

Procedure

1. Log on to the *ECS console*.
2. Select the target region.
3. In the left-side navigation pane, click Instances.
4. Find your ECS instance and, in the Actions column, select 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 Use a launch template

Prerequisites

You have *created a template* or *created a version*.

Procedure

1. Log on to the *ECS console*.
2. In the left-side navigation pane, select Launch Templates.

3. Locate the template or version that you want to use, and then click Create Instance in the Actions column.

4. On the ECS buy 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 to 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 to 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 Launch template

7.1 Launch templates

A launch template helps you quickly create an ECS instance. A template contains configurations that you can use to create instances for various scenarios with specific requirements.

A template can include any configurations except passwords. It can include key pairs, RAM roles, instance type, and network configurations.

You can create multiple versions of each template. Each version can contain different configurations. You can then create an instance using any version of the template.

Console operations

- *Create a template*
- *Create multiple versions in one template*
- *Change the default version*
- *Use a launch template*
- *Delete a template or version*

API operations

- *CreateLaunchTemplate*
- *CreateLaunchTemplateVersion*
- *DescribeLaunchTemplates*
- *DescribeLaunchTemplateVersions*
- *ModifyLaunchTemplateDefaultVersion*
- *DeleteLaunchTemplate*
- *DeleteLaunchTemplateVersion*

7.2 Create a template

You can create a launch template using the following methods:

- *Create a launch template in the ECS console* if you want to create launch templates first, and then create instances using a specific launch template in one click.

- *Create a launch template on the ECS buy page* to create an instance and save its configuration information as a launch template.

> Note:
>
> - Each account can create a maximum of 30 launch templates per region.
> - All parameters are optional when you create a template using the ECS console. However, if the template that you want to use to create an instance does not have all required parameters (such as an image), then you must specify the required parameters at instance creation.
> - A template cannot be modified after it is created.

Create a template in the ECS console

1. Log on to the *ECS console*.
2. In the left-side navigation pane, select Launch Templates, and then click Create Template.

3. Go to the Launch Template page and complete the basic configurations and advanced configurations.

> Note:
> During your first template creation, the Clone Template area is unavailable. If you have already created templates, you can select an existing template, and version, and then modify its configurations.

4. On the Confirm Configuration page, enter a template name and description, and then click Create Launch Template.

> Note:
> All parameters are optional when you create a template. However, on the Confirm Configuration page, we recommend that you configure the required parameters so that you can create instances in one click as needed.

Click View Template in the Activated dialog box to view the template you have created.

Create a template on the ECS buy page

1. Go to the ECS product details page, and then click Buy Now.
2. Configure the required parameters, and then click Save as launch template..
3. In the dialog box that appears, select Create Template, enter a template name and description, and then click Save.

Click View Template in the Activated dialog box to view the template you have created.

7.3 Create a template version

One template can have multiple versions. The default version number of a newly created template is 1, and you can create additional versions based on this template . The version number increments automatically as you create a new version. You cannot customize the version number, but you can set any of the template versions as the default version.

Note:

- Each template can have a maximum of 30 versions.
- All parameters are optional when you create a template version.
- A template version cannot be modified once you have created it.

You can create a template version using the following methods:

- *Create an instance using the ECS console* to create versions of a template for future use.
- *Create an instance on the ECS buy page* to create an instance, save its configurations, and create versions of a template.

Prerequisite

You have already *created a template*.

Create an instance using the ECS console

1. Log on to the *ECS console*.
2. In the left-side navigation pane, select Launch Template.
3. Select a template ID to view its configurations, and then click New Version. You can also click New Version in the Actions column.

4. On the Launch Template page, set the parameters.

   Note:
   You can also go to the Clone Template area, select an existing template and version, and then set the parameters.

5. On the Confirm Configuration page, select Create New Version, and then select a template to save the version.
6. Click Create Launch Template.
7. In the dialog box that appears, click View New Version to view the version you have created.

Create an instance on the ECS buy page

1. Go to the ECS product details page, and then click Buy Now.
2. On the ECS buy page, configure the parameters.
3. On the Preview page, click Save as launch template.
4. In the dialog box that appears, click Create New Version, and then select a template to save the version.
5. In the Activated dialog box, click View New Version to view the version you have created.

Change the default version

1. In the ECS console, select a template ID that has multiple versions.
2. Locate the version you want to set as default, and then click Set as Default in the Actions column.

7.4 Delete a template or version

You can delete templates and versions through the ECS console. Once you delete a template, all associated versions of that template are also deleted.

Delete a version

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Launch Templates.
3. Select the target template ID.

4. In the Version Information area, locate the version you want to delete and, in Actions column, click Delete.

   Note:
   You cannot delete the default template version. If the version you want to delete is the default version, change it to a non-default version, and then delete it. If you no longer need any versions of a single template, delete the template.

5. Click OK.

Delete a template

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Launch Templates.
3. Locate the version you want to delete, and click Delete in the Actions column.
4. Click OK.

   Note:
   When you delete a template, all versions of the template are also deleted.

8 Connect to instances

8.1 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: · 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: · 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*.

8.2 Connect to Linux instances

8.2.1 Connect to an instance by using the Management Terminal

You can use the Management Terminal, also known as VNC, to connect to an ECS instance. This method is suitable for when other remote access software programs such as PuTTy, Xshell, or SecureCRT, do not work.

Scenarios

The Management Terminal can be used to:

- Check the status of an ECS instance.
- Reconfigure the firewall if a remote connection fails due to software error within the ECS instance.
- End abnormal processes that consume excessive CPU usage 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

- You have an ECS instance. For more information, see *create an ECS instance*.
- You have set the logon password of the ECS instance. If not, you can use the *reset password* function.

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. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. Select the target region.
4. In the instance list, find your instance and then, in the Actions column, click Connect.
5. 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.

6. 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:

  - If you do not know the logon password of your instance, *reset the password*.
  - 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, select 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 see a black screen, the Linux instance may be in sleep mode. To exit sleep mode, click your mouse or press any key.

- For a Windows instance: In the upper-left corner of the Management Terminal page, select Send Remote Command > CTRL+ALT+DELETE. The Windows logon interface is displayed. Enter the user name (`Administra tor`) and the logon password.

  Note:

  If you do not know the logon password of your instance, *reset the password*.

Other Operations

Change the VNC connection password

If you do not know your VNC connection password, follow these steps to change the password.

> Note:
> If the instance that you are connecting to is not I/O optimized, you must restart your instance in the ECS console to apply the new VNC password. The restart operation stops your instance and interrupts your business operations. Therefore, proceed with caution.

1. Open the Management Terminal page.
2. Close any dialog box that displays.
3. In the upper-right corner of the Management Terminal page, click Modify Management Terminal Password.
4. Enter a new password. The password must be at least six characters in length and can contain letters and numbers only.
5. Depending on the instance, the new password takes effect as follows:
   - 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.

     > Note:
     > Restarting the operating system does not apply the new password.

Input commands

If you are connecting to a Linux instance, use the Input Commands feature to type long text, such as a complex command or a URL.

To use Input Commandes, follow these steps:

1. Open the Management Terminal page.
2. In the upper-right corner of the Management Terminal page, click Input Commands.
3. Enter the commands and click OK.
4. Press the Enter key to run the commands.

FAQ

- Can multiple users simultaneously connect to the Management Terminal?

  No. Only one user can connect to the Management Terminal at a time.

- Why am I unable to connect to an instance by using the Management Terminal even after changing the password?

  Make sure that you enter the correct VNC password. If the instance that you are connecting to is not I/O optimized, you must restart the instance in the ECS console . This action helps the new VNC password to take effect.
- Why do I see a black screen after logging on to my instance?

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

  For a Linux instance, click your mouse or press any key to activate the screen.

  For a Windows instance, click Send remote command > CTRL+ALT+DELETE to view the logon interface.
- Why am I unable to access the Management Terminal?

  Open your browser, connect to the Management Terminal, and press F12 to open the developer tool. You can then go to the Console tab to analyze the Management Terminal information and locate errors under.
- Can I use IE or Firefox to access the Management Terminal?

  IE version 10 and later support the Management Terminal. Only certain versions of Firefox are supported.

  > Note:
  > We recommend that you use Google Chrome as it offers the best support for Management Terminal.

8.2.2 Connect to a Linux instance by using an SSH key pair

This document describes how to use an SSH key pair to log on to a Linux instance in the following OSs:

- *Local Windows OS*
- *Local Linux OS or other OSs supporting SSH commands*

> Note:
> You can also use your account and password to connect to a Linux instance. For detailed operations, see *connect to a Linux instance by using a password* and *connect to an instance by using the Management Terminal*.

Local Windows OS

The following uses PuTTY and PuTTYgen as an example to describe how to use a key pair generated by Alibaba Cloud to log on to a Linux instance through the SSH remote access tool on a Windows OS.

Prerequisites

- You have downloaded and installed PuTTY and PuTTYgen. The download links are as follows:
  - PuTTY: *https://the.earth.li/~sgtatham/putty/latest/w64/putty.exe*
  - PuTTYgen: *https://the.earth.li/~sgtatham/putty/latest/w64/puttygen.exe*
- You have a Linux instance allocated with a key pair. You can allocate a key pair when creating an instance or *bind a key pair* for the instance.
- The following security group rules must be added to the security group where the instance resides. For detailed operations, see *add security group rules*.

Network type	Network card 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							

Procedure

1. (Optional) If you are using a .pem private key file generated by Alibaba Cloud, you must do the following to convert it to a .ppk key file: If you are using a .ppk private key file, skip this step.

   Note:
   You can download the .pem private key file when you *create an SSH key pair*.

   a. Start PuTTYgen. PuTTYgen 0.68 is used in this example.
   b. In the Parameters area, select `RSA` for Type of key to generate.

      Note:

You do not need to set Number of bits in a generated key. PuTTYgen will automatically updates the parameter value according to the information about the imported private key.

c. Click Load, select All Files (*.*) from the drop-down list after the file name field, and then locate your `.pem` file.

Note:
By default, only `.ppk` files are displayed.

d. Select the .pem private key file you have downloaded from Alibaba Cloud and click Open.
e. Click OK to close the confirmation dialog box.
f. Click Save private key. In the PuTTYgen Warning dialog box indicating saving the key without a passphrase to protect it, click Yes.
g. Set the private key name to the key pair name and save the name. PuTTY will automatically add the .ppk extension to the file.

2. Start PuTTY.
3. Choose Connection > SSH > Auth from the left navigation pane, click Browse··· in the right pane, and then select the generated .ppk file.
4. In the left navigation pane, click Session.
   - 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 in `root @ IP address` format.
   - In the `Port` text box, enter the port number `22` .
   - Select `SSH` for `Connection type` .
5. Click Open to start connecting to your Linux instance.

When `Connection establishe d .` is displayed, you have successfully logged on to the instance by using the key pair.

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.

Prerequisites

You have a Linux instance with a key pair allocated. You can *allocate a key pair when creating an instance* or *bind a key pair* for the instance.

The following security group rules must be added to the security group where the instance resides. For detailed operations, see *add security group rules*.

Network type	Network card 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							

Procedure

- Method 1

1. Locate the directory for saving the .pem private key file on your local PC, for example, `/ root / xxx . pem .`

   Note:
   You can download the .pem private key file when you *create an SSH key pair*. `xxx . pem` is your private key file.

2. Run `chmod 400 [ Directory for saving the . pem private key file on your local PC ]` to modify the attributes of the private key file, for example, `chmod 400 / root / xxx . pem .`
3. Run `ssh - i [ Directory for saving the . pem private key file on your local PC ] root @[ Internet IP address ]` to connect to the instance, for example, `ssh - i / root / xxx . pem root @ 10 . 10 . 10 . 100 .`

- Method 2

  You can simplify the connection commands through SSH configurations.

  1. Enter 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 .
IdentityFi le ~/. ssh / ecs . pem // Enter the directory
for saving the . pem private key file on your
local PC .
```

  2. Save the *config* file.
  3. Restart SSH.
  4. Run `ssh [ ECS name ]` to connect to your ECS instance, for example, `ssh ecs`.

8.2.3 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*.

8.2.4 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

Network type	NIC	Rule direction	Authorization policy	Protocol type	Port range	Authorization type	Authorization object	Priority
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 `I 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.

8.3 Connect to Windows instances

8.3.1 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

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. On the Instances page, select the target region.
4. In the instance list, find the target instance. Then, click Connect in the Actions column.

5. 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.
6. Enter the user name and password to log on to the ECS instance.

   > Note:
   > In the upper left corner of the Management Terminal page, select 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 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.

Related documents

- *Connect to a Windows instance*.
- *Connect to an instance on a mobile device*.

8.3.2 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 `Administra tor`

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

8.3.3 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.

9 Manage instances

9.1 Start or stop an instance

This article describes how to start or stop an ECS instance.

Start an instance

You can start an instance in the ECS console. When an instance starts successfully, it is in the Running status.

Prerequisite

The instance must be in the Stopped status.

Procedure

To start an instance, follow these steps:

1. Log on to the *ECS Management console*.
2. In the left-side navigation pane, click Instances.
3. Select the target region.
4. Find the instance to be started and, in the Actions column, select More > Instance Status > Start. 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.

The instance is in the Running status after it is started.

Stop an instance

To stop an instance is to shut it down. You can stop an ECS instance in the ECS console. When an instance stops successfully, it is in the Stopped status.

> Note:
> Stopping an instance disrupts services. Exercise caution when performing this action.

If you stop a (Subscription) instance before its billing cycle is completed, the bill for that cycle is not affected. If the auto-renewal service is activated, you are still billed for the stopped instance at the start of each new billing period.

For a Pay-As-You-Go instance, its network type and the No Fees for Stopped Instances (VPC-Connected) feature determine billing:

- VPC: If the No Fees for Stopped Instances (VPC-Connected) feature is enabled, you can decide whether to continue being billed for the instance. However, you are still billed for other ECS-related resources. For more information, see *no fees for stopped instances (VPC-Connected)*. If this feature is not enabled, billing continues after the instance is stopped.
- Classic network: A stopped instance still incurs fees. Billing stops only after you *release the instance*.

Prerequisite

The instance is in the Running status.

Procedure

To stop an instance, follow these steps:

1. Log on to the *ECS Management console*.
2. In the left-side navigation pane, click Instances.
3. Select the target region.
4. Find the instance to be stopped and, in the Actions column, select More > Instance Status > Stop. If you want to stop multiple Running instances, select the required instances and then, under the instance list, click Stop.
5. According to the billing method and network type of the instance, complete the required actions:
   - Subscription instance or classic network pay per volume instance: In the Stop Instance dialog box, select Stop or Force Stop, and then click OK.
   - A VPC-Connected Subscription instance:
     - If the No Fees for Stopped Instances (VPC-Connected) feature is enabled, read the Notice, select Stop or Force Stop in the Stop Instance dialog box, select a

mode (whether to keep the instance after stopping and continue charging), and then click OK.

- If the No Fees for Stopped Instances (VPC-Connected) feature is disabled, in the Stop Instance dialog box, select Stop or Force Stop.

  Note:
  To disable the No Fees for Stopped Instances (VPC-Connected) feature, see *no fees for stopped instances*.

Once the instance is successfully stopped, the instance enters the Stopped status. For a VPC-Connected Pay-As-You-Go instance, if you select not to keep the instance, Stop Instance, No Fees is shown in the instance list. Otherwise, Keep Instance, Fees Apply is shown. For other ECS instances, no information is shown.

Related APIs

Start instance: *#unique_266*

Stop instance: *#unique_66*

9.2 Restart an instance

This topic describes how to restart an instance through the ECS console. You can also choose to call the RebootInstance API action.

Limits

- Only instances in the Running state can be restarted.
- When you restart an instance, the instance is stopped. As a result, services provided by the instance are disrupted.

Procedure

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. Select the target region.

4. 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.
5. In the displayed Restart Instance dialog box, select a Restart Mode, and then click OK.

Related API: *RebootInstance*

9.3 Release an instance

You can release a Pay-As-You-Go instance when you no longer need it to avoid excess charges.

For a Pay-As-You-Go instance, if the *No fees for stopped VPC instances* 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.

To release a Pay-As-You-Go instance, you can choose either of the following options:

- Release immediately, which releases the pay-per-order instance at once.
- Scheduled Release, which customizes the releasetime of your Pay-As-You-Go instance. The time to release the instance must be at least 30 minutes from the current time. Applying new schedules overwrites the previous ones.

 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.

Release an instance immediately

To release an instance immediately, 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. Set release:
   - If you want to release only one instance, find the instance that you want to release and then, in the Actions column, select 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 Setting at the bottom of the list.
5. In the dialog box, select Release Now.
6. Click Next, and then click OK.

Enable automatic release

To enable automatic release, 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. Set the release:
   - If you want to release only one instance, find the instance that you want to release, and in the Actions column, select 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 Setting at the bottom of the list.
5. In the dialog box, select Scheduled Release.
6. 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.
7. Click Next, and then click OK.

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. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.

3. Select the target region.
4. Set the release:
   - If you want to disable automatic release for an instance, find the instance and then, in the Actions column, select 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 Setting at the bottom of the list.
5. In the dialog box, select Scheduled Release.
6. Turn off the Automatic Release switch.
7. Click Next, and then click OK.

Related API

#unique_271

9.4 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.

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 status.

You have settled the payment by opening a ticket.

Procedure

To reactivate an instance in the ECS console, 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 instance to be reactivated, and then select More > Reactivate at the bottom of the instance list.
5. 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 status after about 10 minutes.

Alternatively, you can also complete the task by calling the ECS API `ReactivateInstances`.

9.5 Check instance information

Through the console, you can:

- *View all ECS instances under your account on the Overview page*.
- *View details of an ECS instance on the Instance Details 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 ECS Overview page, including:

- Total number of ECS instance, and numbers of instances under each status.
- Number of resources in different regions and numbers of ECS instances under each status.

The homepage of the ECS console is the Overview page by default.

View the information of ECS instances on the Instance List page

To navigate to the Instance List page, follow these steps:

1. Log on to the *ECS console*.
2. On the left-side navigation pane, click Instances.
3. Select a region.

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.

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. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. Select a region.
4. 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.

9.6 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.

Prerequisite

The instance must be in a stable status, such as Stopped and Running. For more information, see *ECS instance life cycle*.

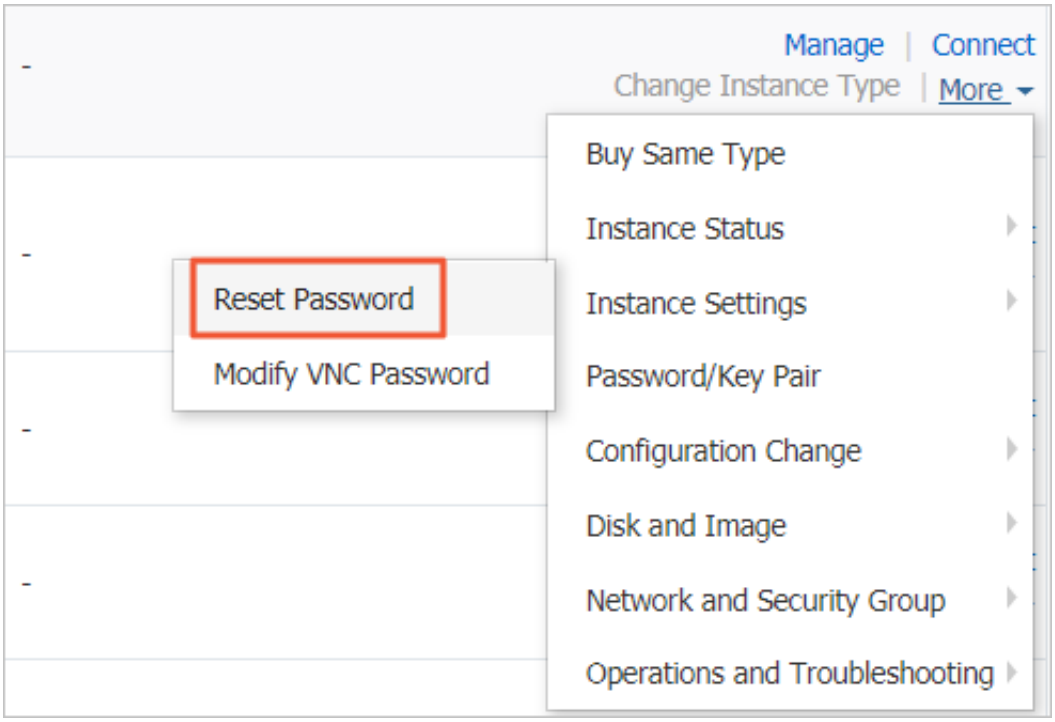
Procedure

To reset the password of one or multiple ECS instances, follow these steps:

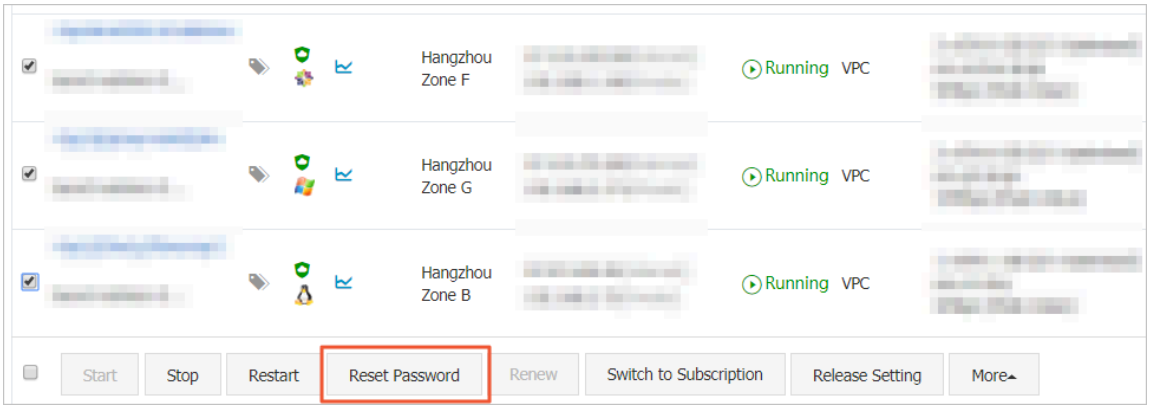
1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. Select the target region.

4. Select one or more target instances as follows:

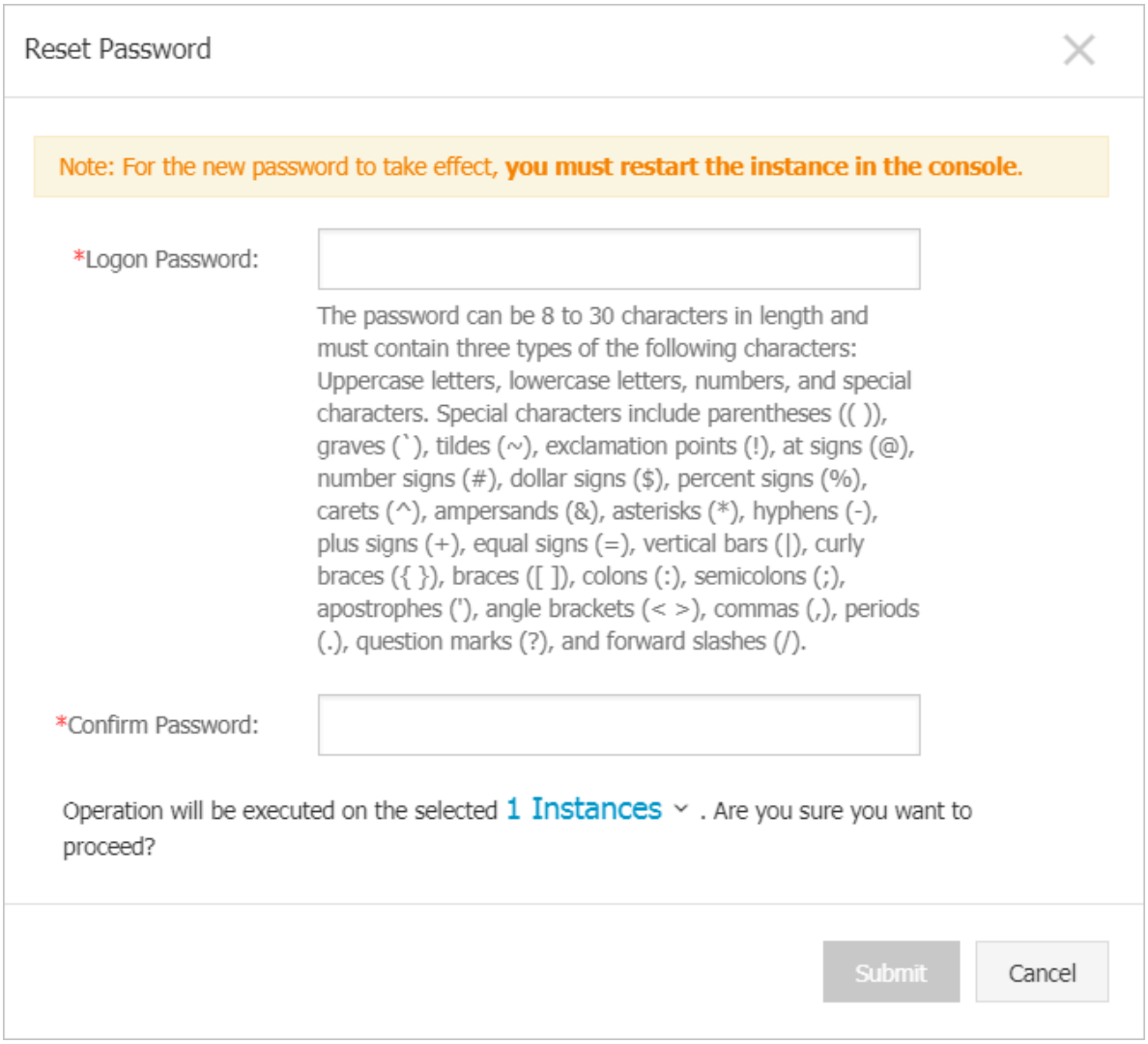
   · To reset the password of one instance, locate the target instance, and then choose More > Password/Key Pair > Reset Password in the Actions column.



   · To reset the password of multiple instances, select the target instances, and then click Reset Password under the instance list.

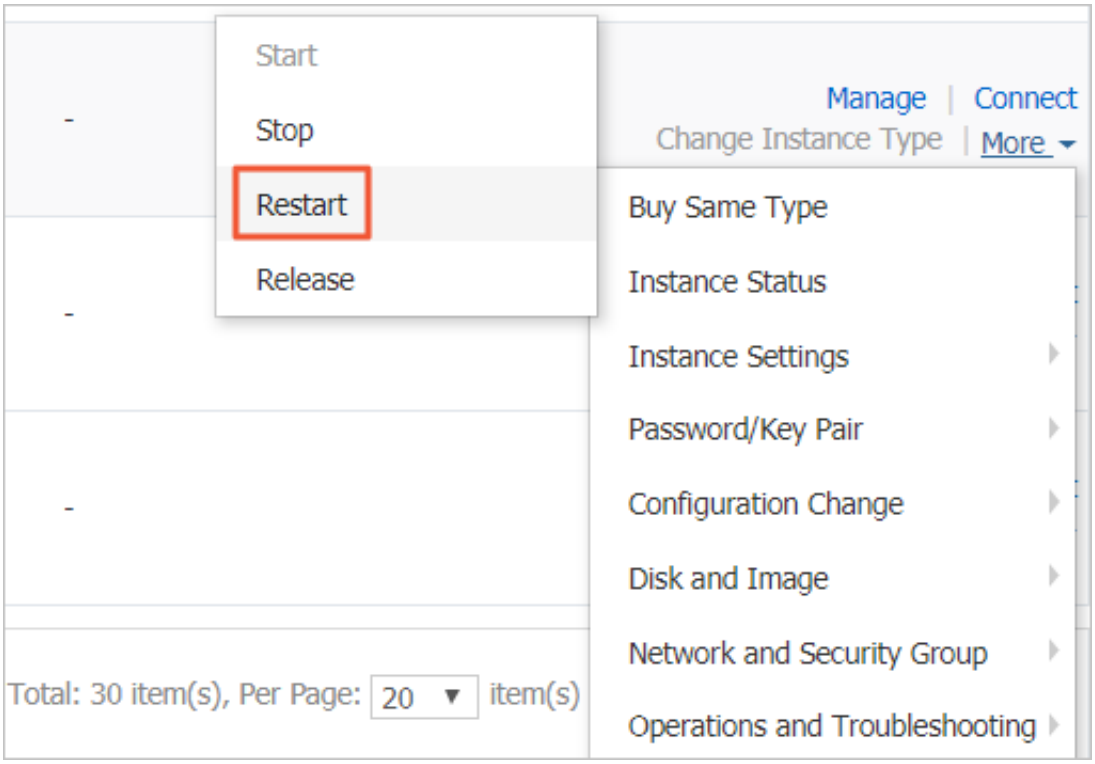


5. Specify a new valid password, and then click Submit.

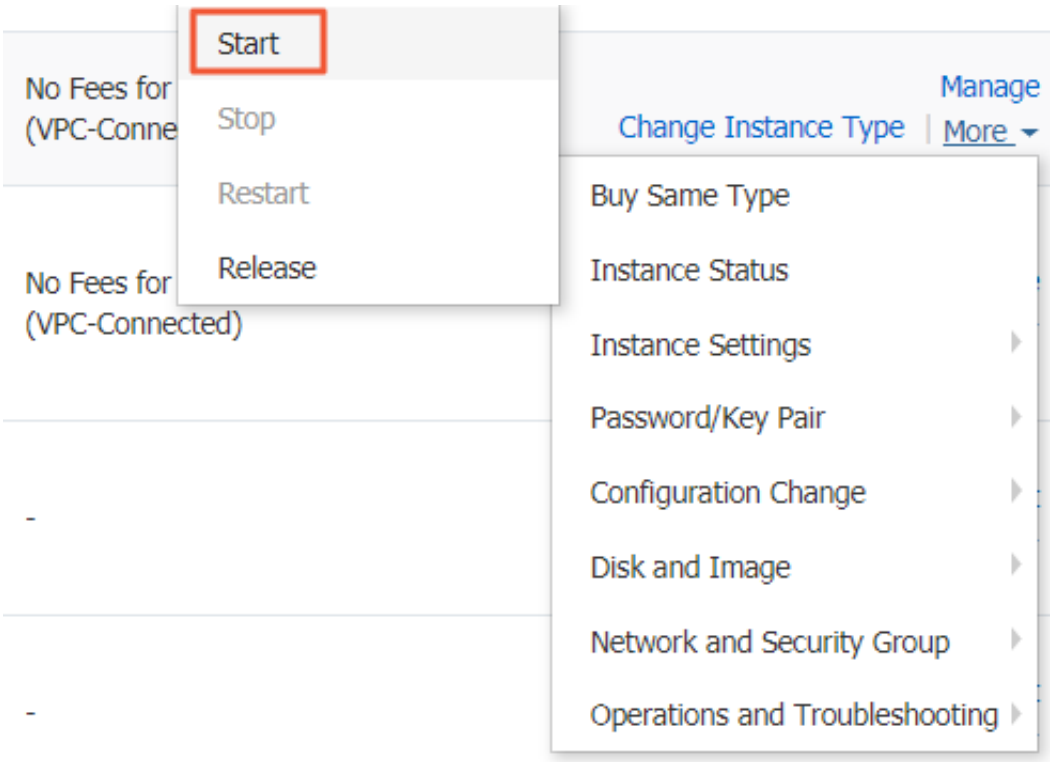


6. Perform the following operations as needed:

   · 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.

9.7 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 *Overview*.
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 *Overview*.
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.

9.8 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

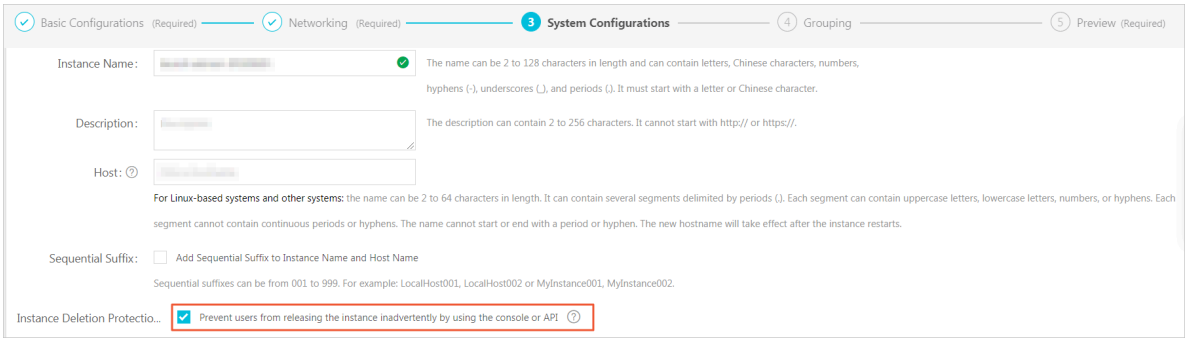
Note:

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. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. On the Instances page, click Create Instance.
4. On the Basic Configurations page, set the Billing Method to Pay-As-You-Go, set other parameters as needed, and then click Next: Networking.
5. On the Networking page, set the parameters as needed, and then click Next: System Configurations.

6. 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.



7. 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 *#unique_284*, you can use the `DeletionPr otection` parameter to enable or disable instance release protection.

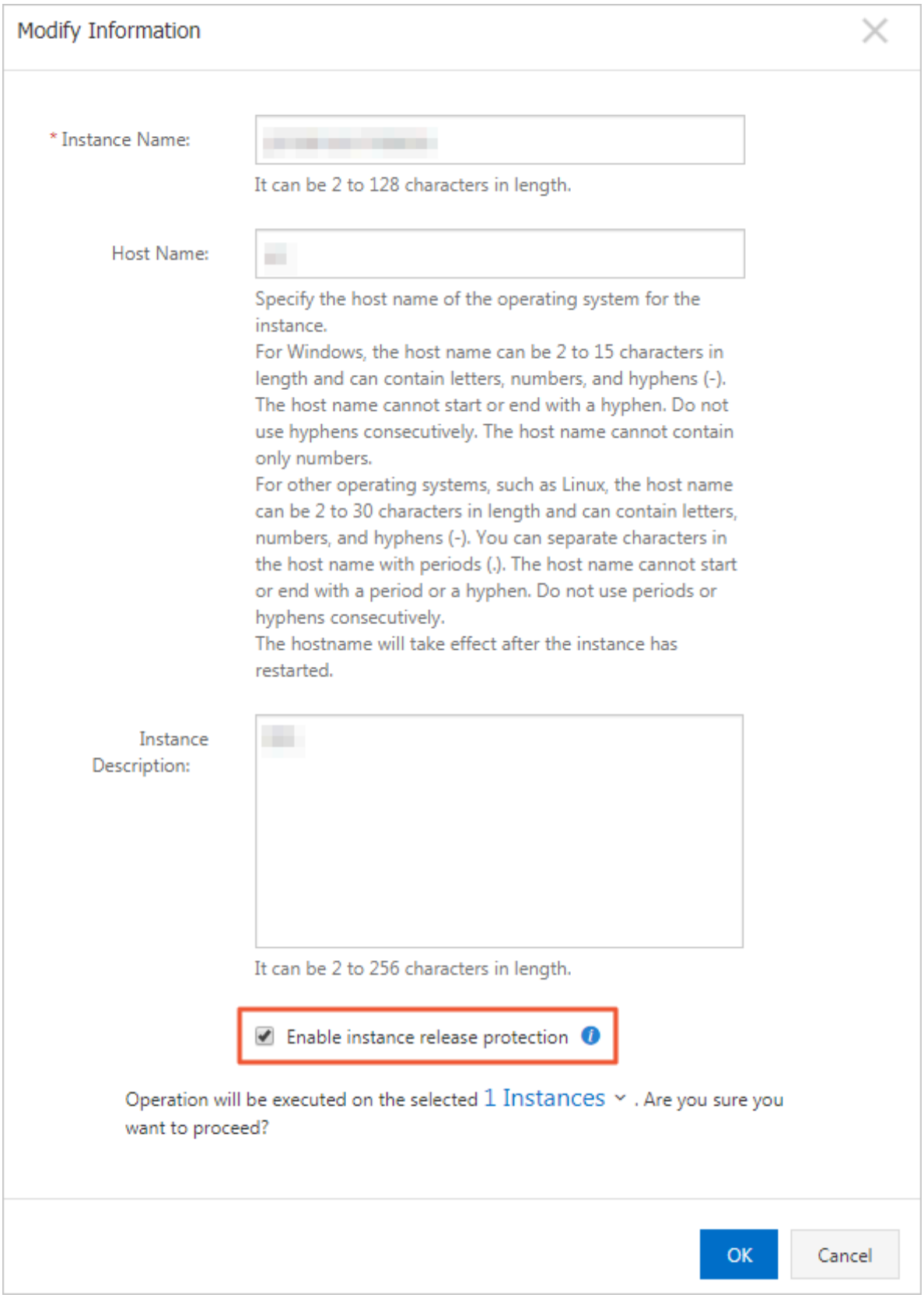
Enable or disable instance release protection by modifying instance information

To enable or disable instance release protection by modifying the information of an instance, follow these steps:

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. On the Instances page, select More > Instance Settings > Modify Information in the Actions column of the instance to be modified.
4. In the Modify Information dialog box, select Enable instance release protection.

 **Note:**

To disable the instance release protection, clear Enable instance release protection.



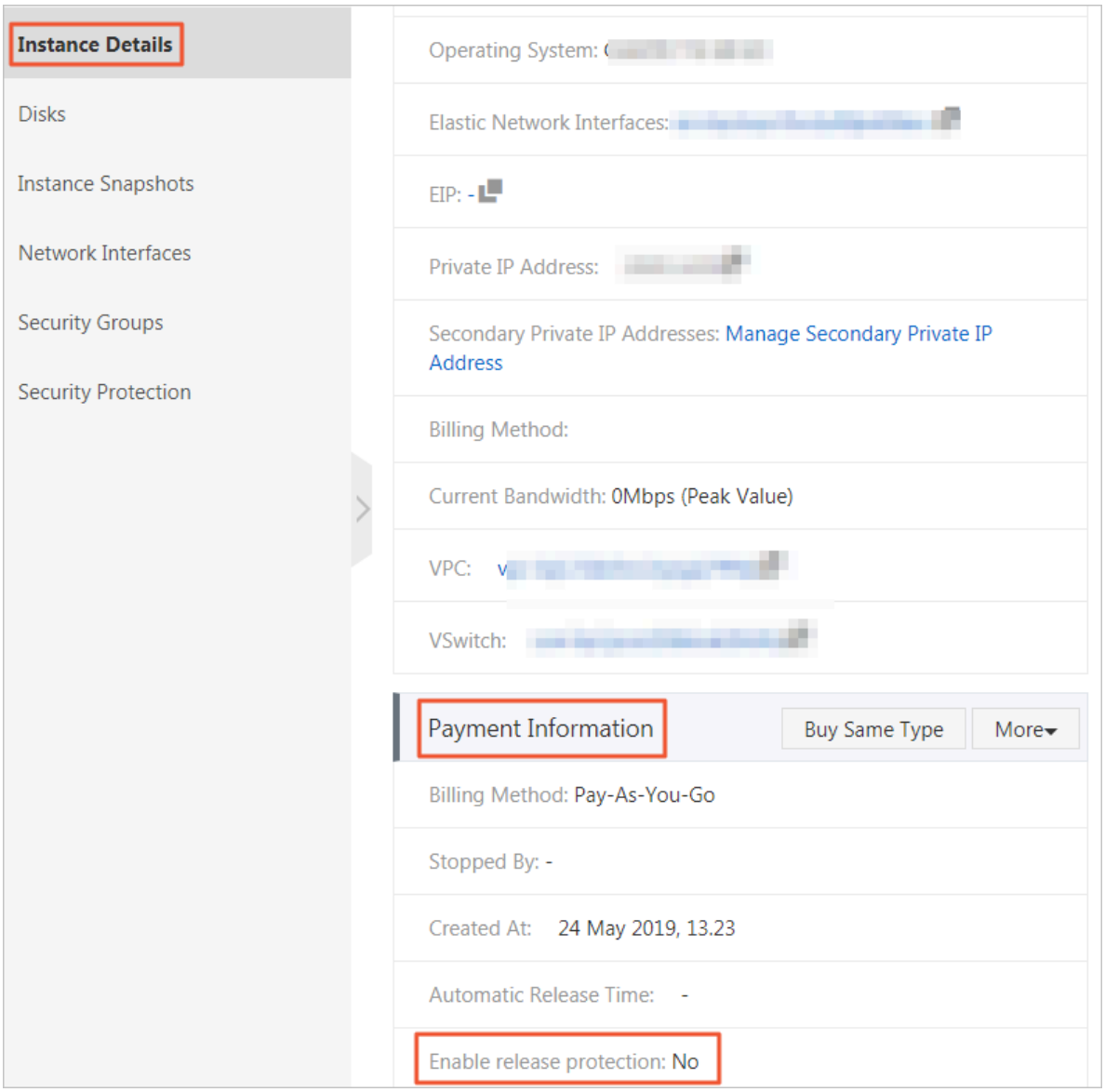
5. Click OK.

When you modify instance information through *ModifyInstanceAttribute*, you can use the `DeletionPr otection` parameter to enable or disable instance release protection.

View the instance release protection status

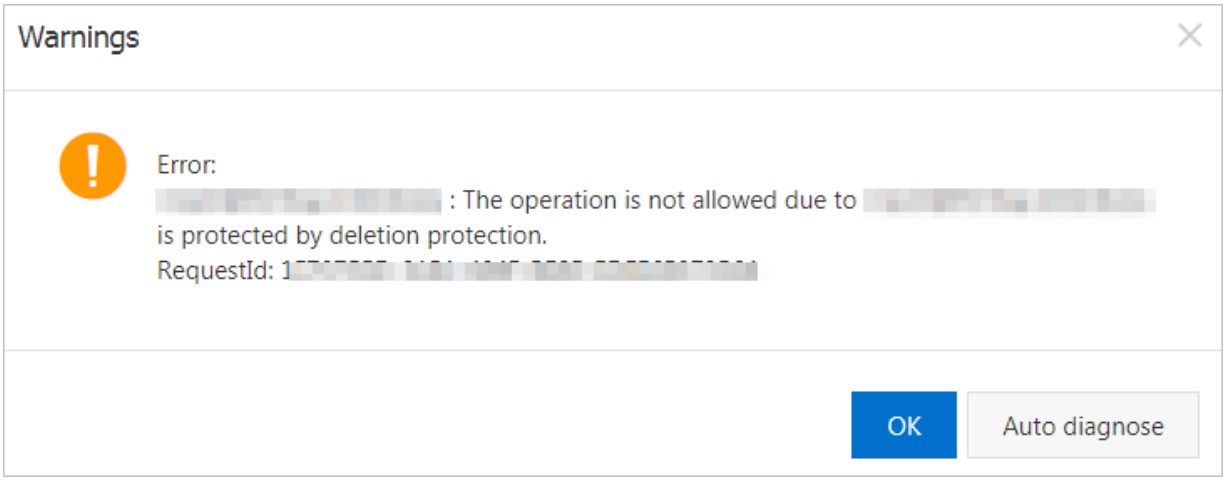
To view the release protection status of an instance, follow these steps:

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. On the Instances page, click the target instance in the Instance ID/Name column, or click Manage in the Actions column of the instance.
4. On the Instance Details page, view the status of Enable release protection in the Payment Information area.



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.

9.9 User-defined data and metadata

9.9.1 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. *Connect to a Linux instance by using a password*.

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. *Connect to a Windows instance*.
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 Marketpalce.	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]	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. Note: A new STS credential is available 30 minutes prior to the expiration of the old one.	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/macs	The MAC address list of Network Interface Controllers (NICs).	2016-01-01
network/interfaces/macs/[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 / maintenanc  e /
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
```

9.9.2 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 / maintenanc  e /
     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
```

9.10 User-defined data

9.10.1 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_99*) 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	- 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_99* + Parameters `UserData`

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_296*

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_73* + Parameters `Userdata`

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 *Metadata* 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, the 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
 descriptio  n  " 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*.

9.10.2 Configure 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 newer
Linux	- CentOS - Ubuntu - SUSE Linux Enterprise - OpenSUSE - Debian - Aliyun Linux

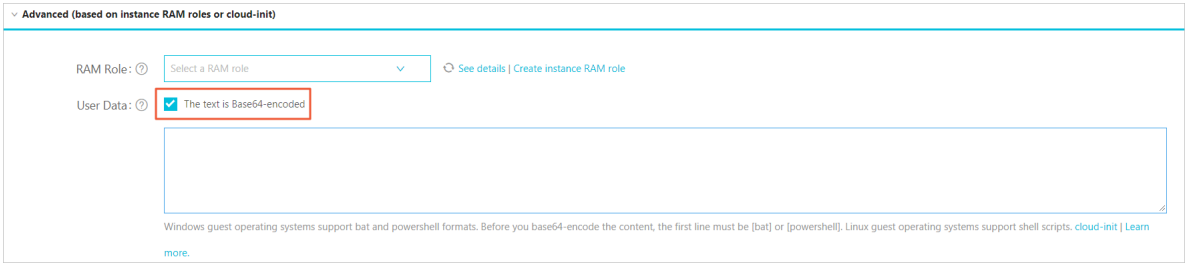
Configure user data for an instance

Note:
The following procedure takes a Linux instance running CentOS as an example.

1. Log on to the *ECS console*.

2. Create a Linux instance by following the instructions in *Step 2. Create an ECS instance*.

> **Note:**
> During instance creation, enter user data in the user data box below Advanced (based on instance RAM roles or cloud-init). If your user data is Base64-encoded, select the The text is Base64-encoded check box.



3. After the instance starts, connect to it by following the instructions in *Overview of connecting to an ECS instance*.
4. 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-RestMethod 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.

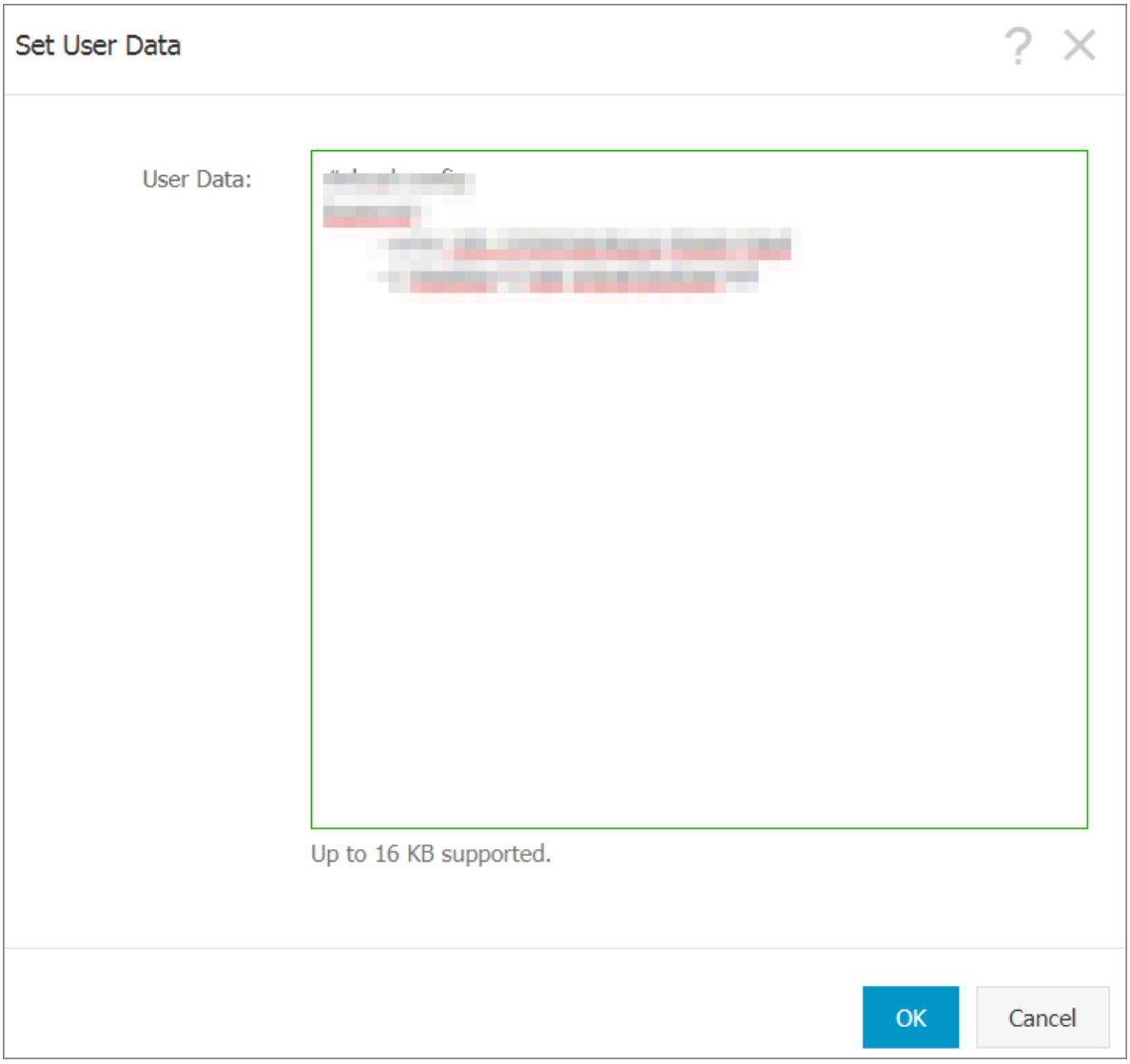
Modify user data

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. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. Select the target region.
4. Find the target instance, and in the Actions column, select Instance Settings > Set User Data.

5. In the displayed dialog box, enter the relevant information and click OK.



You can also modify user data by calling *ModifyInstanceAttribute* through Alibaba Cloud CLI, OpenAPI Explorer, or Alibaba Cloud SDK.

9.10.3 User-defined yum sources, NTP services and DNS services

User-defined scripts are a type of script provided by Alibaba Cloud for users to customize the startup behaviors of ECS instances. For more information, see *User-defined data*.

This example uses a Linux instance to demonstrate 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. Log on to the *ECS console* and create an instance. Configure the instance as follows:
   - Network Type: Select VPC.
   - Instance Type: Select an I/O-optimized instance.
   - Operating System: Select CentOS 7.2 in Public Image tab.
2. Enter the following script in the User Data box on the instance creation page.

```
#! / 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.

3. Complete the security settings as needed.
4. After you complete the configuration, click Buy Now and activate the instance following the instructions on the page.

   After the instance is created, you can connect to the instance to view the implementation details, as shown in the following figure.

   The preceding figure shows that you have successfully customized the DNS service , the NTP service, and the yum repository.

9.10.4 Create a new account with the root user privilege

User-defined scripts are a type of script provided by Alibaba Cloud to enable users to customize the startup behavior of ECS instances. For details, see *User-defined data*.

This example uses a Linux instance to demonstrate 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 a Linux instance*. Configure the instance as follows:
   - Network Type: Select VPC.
   - Instance Type: Select an I/O-optimized instance.
   - Operating System: Select CentOS 7.2 in Public Image tab.
2. Enter the following script in the User Data box on the instance creation page:

```
#! / 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   AAAAB3NzaC  1yc2EAAAAB  JQAAAQEAhG  qhEh
 / 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  YuSwhpXzt4  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.

3. Select post-creation settings in security settings.
4. After you finish the configuration, click Buy Now and activate the instance by following instructions on the page.

After the instance is created, you can use the new test user to connect to the instance using an SSH private key. You can also escalate the permission level using the `sudo` command and run operations that require the root user privilege, as shown in the following figure.

9.11 Instance identity

Instance identity is a part of *Metadata* 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

Properties	Description	Can it be changed?
create-time	Instance creation time	No
instance-id	Instance ID.	No
mac	MAC address of the instance primary network interface	No
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.

4. Verify the instance identity by using OpenSSL.

```
openssl smime -verify -in $signature -inform PEM -
content $DOCUMENT -certfile AliyunPubkey -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 `AliyunPubkey` with the *Alibaba Cloud public certificate*.

The public certificate of Alibaba Cloud in all regions is as follows.

```
-----BEGIN CERTIFICATE-----
MIIDdzCCAl+gAwIBAgIEZmbRhzANBgkqhkiG9w0BAQsFADBsMRA
wDgYDVQQGE wdV
bmtub3duMR AwDgYDVQQI EwdVbmtub3 duMRAwDgYD VQQHEwdVbm
tub3duMRAw DgYD
```

```
 VQQKEwdVbm  tub3duMRAw  DgYDVQQLEw  dVbmtub3du  MRAwDgYDVQ
 QDEwdVbmtu  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  3OZgJpYIoC
 Gik + 9lG
 5gVAIr0 +/ 3rZ61IbeVE + vDenDd8g / m / YIdYBfC2Ib  zgS9EVGAf /
 gJdtDODXrD  fQj
 Fk2rQsvpft  VOUs3Vpl9O + jeCQLoRbZY  m0c5v7jP / L2lK0Mjhiy
 wPF2kpDeis  MtnD
/ ArkSPIlg1q  VYm3F19v3p  a6ZioM2hnw  Xg5DibYlgV  vsIBGhvYqd
 Q1KosNVcVG  GQa
 HCUuVGdS7v  HJYp3byH0v  QYYygzxUJT  2TqvK7pD57  eYMN5drc7e
 19oyRQvbPQ  3kkC
 AwEAAaMhMB  8wHQYDVR0O  BBYEFAwwrn  HlRgFvPGo + UD5zS1xAkC
 91MA0GCSqG  SIb3
 DQEBCwUAA4  IBAQBBLhDR  gezd / OOppuYEVNB  9 + XiJ9dNmcuH
 UhjNTnjiKQ  WVk / YDA
 v + T2V3t9yl8L  8o61tRIVKQ ++ lDhjlVmur / mbBN25 / UNRpJllfpU
 H6oOaqvQAz  e4a
 nRgyTnBwVB  ZkdJ0d1siv  L9NZ4pKelJ  F3Ylw6rp0Y  MqV + cwkt /
 vRtzRJ31ZE  eBhs7
 vKh7F6BiGC  HL5ZAwEUYe  8O3akQwjgr  MUcfuiFs4 / sAeDMnmgN6
 Uq8DFEBXDp  AxVN
 sV / 6Hockdfinx  85RV2AUwJG  fClcVcu4hM  hOvKROpcH2  7xu9bBIeMu
 Y0vvzP2VyO  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 *Metadata*.

```
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   CERTIFICAT  E -----" >  signature
 curl   100 . 100 . 100 . 200 / latest / dynamic / instance -
 identity / pkcs7  >>  signature
 echo  "" >>  signature
 echo  "----- END   CERTIFICAT  E -----" >>  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 `Verificati on 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 =$( curl   100 . 100 . 100 . 200 / latest / dynamic /
 instance - identity / document )
 audience_j  son =',"  audience ":'"'"${ audience }'"'}'
```

```
 echo  - n  ${ document %?} ${ audience_j  son } >  document
 echo  "----- BEGIN   CERTIFICAT  E -----" >  signature
 curl   100 . 100 . 100 . 200 / latest / dynamic / instance -
 identity / pkcs7 ?  audience =${ audience } >>  signature
 echo  "" >>  signature
 echo  "----- END   CERTIFICAT  E -----" >>  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 `Verificati on successful`, remove the restriction and run the image in the instance.

9.12 Configure time

9.12.1 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: `http :// www . alidns . com`
Open source images: `http :// mirrors . aliyun . com`	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.

9.12.2 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.

9.12.3 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/XXXXXXX /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 `ntpstat` 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.

10 Renew Subscription instances

10.1 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.

10.2 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 inanExpired 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 Subscriptioninstances, you can choose manual renewal.

- Your instance will still work normally when the instance is in anExpiredstatus. 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 enterthenew billing cycle from the day of renewal, and your data will be retained for the new cycle.

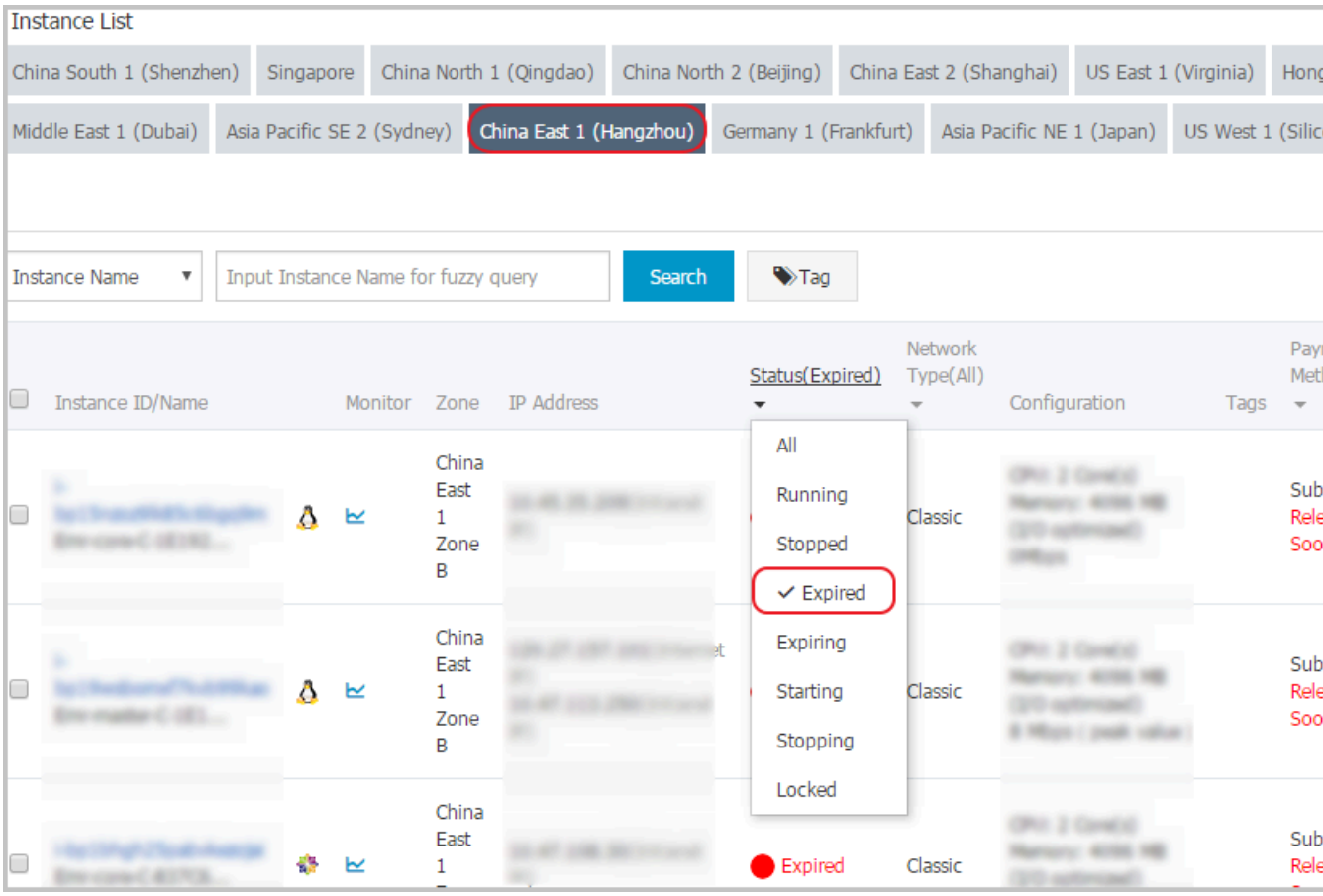
    For example, if your instance was shut down at 00:00:00 on May 10, 2016, but you successfully renewed it for one month at 08:09:35 on May 23, 2016, the billing cycle for this renewal is from 08:09:35 on May 23, 2016 to 00:00:00 on June 24, 2016.
  - If the instance is not renewed within 15 days of the 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 be1 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.

10.3 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 subscripti on 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 cankeep 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 theActions 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 theAuto-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 theManually Renew tab.

10.4 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 *Subscription*.

Youcan 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 Expired and Being Recycled 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 period, 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. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. Select a region.
4. Find the Subscription instance. In the Actions column, click Change Configuration.
5. In the Configuration Change Guide dialog box, select Renew for Configuration Downgrade and click Continue.
6. 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 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.
   - Set the bandwidth value.
   - Set the renewal duration.

7. Click Pay, and follow the prompts to complete the process.
8. (Optional) 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 *#unique_268* API within the first seven days of the next billing period. This step is necessary for the new configurations to be effective.

 **Note:**
For a VPC instance, the instance does not need to be restarted.

11 Change configurations

11.1 Overview of configuration changes

You can change the configurations of an instance and its Internet bandwidth after it is created.

Upgrade or downgrade instance configurations

You can only upgrade or downgrade the configurations of vCPU and memory (that is, *instance type family*) simultaneously by changing the instance type. Depending on the method of billing applied to your instance, you can change an instance type as follows:

- Subscription:
  - Upgrade: See *upgrade configurations*. The new configurations take effect after you *restart the instance* in the console or by using the *#unique_268* interface.
  - Downgrade: See *renewsal for configuration downgrade* . You can downgrade the configuration of an instance when you renew the instance. The new configuration takes effect after you *restart the instance* in the ECS console within the first seven days of the new billing cycle.
- Pay-As-You-Go: See *change configurations of Pay-As-You-Go instances*. You must stop the instance to use this feature.

> Note:
> Stopping an instance disrupts services. Exercise caution when performing this action.

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?	Is it effective immediately?	Available feature	Description
Subscription	Yes	Yes	*Upgrade configurations of Subscription instances*	Only applicable to VPC-Connected ECS instances to which no EIP addresses are attached, or classic network-connected ECS instances. The Internet and intranet IP addresses remain unchanged after you upgrade your configurations.
Subscription	Yes	Effective from next billing cycle	*Renew for configuration downgrade*	Adjust bandwidth in the new billing cycle. When the Internet bandwidth is set to 0 Mbit/s, the Internet IP address of a VPC-Connected instance is released in the new billing cycle, but that of a classic network-connected ECS instance is retained.

Billing method	Supports permanent upgrade?	Is it effective immediately?	Available feature	Description
Pay-As-You-Go or Subscription	Yes	Yes	*Change EIP Internet bandwidth*	Only applicable to VPC-Connected instances to which *EIP addresses* are bound. You can adjust the Internet bandwidth on an EIP address at any time.

Assign a public IP address

Assign a public IP address to an ECS instance while *creating it*. If you skip it, you can even assign after an ECS instance is created. However, the feature is only available for Subscription instances. For more information, see the following table.

Feature	Is it effective immediately?	Description
Upgrade configurations of Subscription instances	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.
Renew for configuration downgrade	Effective from next billing cycle	

11.2 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_effi ciency`) 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_effi ciency`) 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	· g5, r5, c5, ic5 · sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5 , hfg5, re4, t5, n4, mn4, xn4, e4
sn1ne, sn2ne, se1ne	· sn1ne, sn2ne, se1ne · c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4
se1	· se1 · sn1, sn2, n1, n2, e3, sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4

Source instance type family	Target instance type family
n4, mn4, xn4, e4	· 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	· re4 · sn1ne, sn2ne, se1ne, c4, cm4, ce4, hfc5, hfg5, g5, r5, c5, ic5, t5, n4, mn4, xn4, e4, ecs.se1.14xlarge
hfc5, hfg5	· hfc5, hfg5 · sn1ne, sn2ne, se1ne, c4, cm4, ce4, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4
gn4	gn4
gn5i	gn5i
gn6v	gn6v
t5	· 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	· 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	· 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	· 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	· c4, ce4, cm4 · sn1ne, sn2ne, se1ne, hfc5, hfg5, g5, r5, c5, ic5, re4, t5, n4, mn4, xn4, e4

11.3 Upgrade configurations

11.3.1 Upgrade configurations of Subscription instances

You can upgrade a Subscription-billed instance type.

you can also:

- Convert the billing method of data disksfrom 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 *#unique_268* API to activate the new configuration.

Procedure

1. Log on to the *ECS Console*.
2. In the left-side navigation pane, click Instances.
3. Select a region.
4. Select the Subscription instance to upgrade and, in the Actions column, click Change Configuration.
5. Select Upgrade Configuration and click Continue.
6. 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.
7. Confirm your order details, and then click Create Order. Follow additional instructions as required.
8. 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 through the console or through the *#unique_268* API to activate the new configuration.

> Note:
> You do not have to restart a VPC instance if this upgrade configuration is the first time its Internet bandwidth is increased from 0 Mbps to a non-zero value.

You can also use the *DescribeResourcesModification* API to query the instance types that can be upgraded.

11.4 Downgrade configurations

11.4.1 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 *Subscription*.

Youcan 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 Expired and Being Recycled 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 period, 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. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. Select a region.
4. Find the Subscription instance. In the Actions column, click Change Configuration.
5. In the Configuration Change Guide dialog box, select Renew for Configuration Downgrade and click Continue.
6. 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 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.
- Set the bandwidth value.
- Set the renewal duration.

7. Click Pay, and follow the prompts to complete the process.
8. (Optional) 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 *#unique_268* API within the first seven days of the next billing period. This step is necessary for the new configurations to be effective.

 Note:
For a VPC instance, the instance does not need to be restarted.

11.4.2 Downgrade configurations of Subscription instances

You can downgrade configurations (including the memory size and the number of vCPU cores) of Subscription instances. The configurations immediately take effect after instance restart.

Limits

- Only members who reach certain membership levels can downgrade instance configurations.
- You can downgrade configurations of only one instance at a time.
- You can change the instance configurations only to lower-level configurations . That is, 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 billing method is Subscription or weekly payment.
- The instance is in 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.

Fees

A configuration downgrade may result in a refund. The refund amount is the result of the following formula: Refund amount = Remaining amount of the configuration fee before the downgrade - Price of the new configurations. (The refund will be returned only when the result is a positive value.)

Procedure

1. Log on to the ECS console.
2. Find the target instance and click Change Configuration in the Action column.
3. In the displayed dialog box, select Configuration downgrade and Instance Type.
4. Select a desired instance type, confirm the refund amount, and read and confirm that you agree with the *ECS Service Terms*.
5. Click Downgrade Now.

What to do next

Restart the instance for the new configurations to take effect.

11.4.3 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

- Some accounts support this feature (based on your ECS 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 configurat ions 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. Log on to the ECS console.
2. Find the target instance and click Change Configuration in the Action column.
3. In the displayed dialog box, select Configuration downgrade and Bandwidth Configuration.

4. Set the bandwidth and read and confirm that you agree with the *ECS Service Terms* .
5. Click Downgrade Now.

11.5 Change configurations of Pay-As-You-Go instances

11.5.1 Change configurations of Pay-As-You-Go instances

This article describes how to change configurations of Pay-As-You-Go instances. For information about how to change configurations of (Subscription) instances, see *overview of configuration changes*.

 Note:
Changing instance configurations requires stopping your instance, which disrupts services. Exercise caution when performing this action.

Limits

- You can upgrade an instance multiple times, but the time period between each upgrade must be at least five minutes.
- Not supported 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*.

Prerequisite

The instance has been stopped.

Procedure

To change instance type configurations of the 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. In the Actions column, click Change Instance Type.
5. On the Instance Type page, select the desired instance type and click Confirm.

 Note:
You can also enter the instance type information in the search box to filter instance types.

Once the change is complete, it takes effect immediately. You can view the instance type information in the Basic Information area of the Instance Details page, as shown in the following figure.

Then, restart the instance to restore your services.

You can also use the *DescribeResourcesModification* API to query the instance types that can be changed.

11.5.2 Change EIP Internet bandwidth

If you are using a VPC-Connected ECS instance, and *an Elastic IP (EIP) address is bound to it*, you can use the Change Bandwidth feature to change the Internet bandwidth as needed, regardless of your billing method.

Restrictions

The Change Bandwidth feature only applies to VPC-Connected instances, that are bound with an Elastic IP (EIP) address.

Change bandwidth

To change the Internet bandwidth of an EIP address, follow these steps:

1. Log on to the *ECS console*.
2. In the left-side navigation pane, click Instances.
3. Select a region.
4. Find an instance to which an EIP address is bound, and in the Actions column, click Change Configuration.
5. In the Change Configurations dialog box, select change bandwidth, and click Continue.
6. On the Confirm Order page, set the new peak bandwidth.
7. Click Activate and follow the instructions to complete the configuraiton.

Related operations

For Subscription instances of the classic network type or VPC-Connected ECS instances to which no EIP addresses are bound, you can use the following features to change the Internet bandwidth:

- Use the *Upgrade Configuration* feature to change the Internet bandwidth immediately.
- Use the *Renew for Configuration Downgrade* feature to change the Internet bandwidth for the next billing cycle.